Data Mining the Electronic Medical Record with Intelligent Agents to Inform Decision Support Systems

Karl Jablonowski

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Linda Shapiro, Chair

Adam Wilcox

Meliha Yetisgen

Program Authorized to Offer Degree:
Biomedical Informatics and Medical Education

University of Washington

Abstract

Data Mining the Electronic Medical Record with Intelligent Agents to Inform Decision Support Systems

Karl Jablonowski

Chair of the Supervisory Committee:
Dr. Linda Shapiro
Department of Biomedical Informatics and Medical Education

An intelligent agent framework is used on an ICU EMR to create prediction models for disease onset. Eleven models are created to inspect 5 diseases: acute respiratory distress syndrome (ARDS); severe acute hypoxemic respiratory failure (SAHRF); acute kidney injury (AKI); sepsis; and disseminated intravascular coagulation (DIC).

Four of the models (ARDS, AKI Stage 1, AKI Stage 2, and sepsis) are competitive or superior to the best comparable peer-reviewed models. The other seven are novel, including: SAHRF (AUC=0.952); DIC from ARDS positive patients (AUC=0.722); ARDS from DIC positive patients (AUC=0.675); AKI Stage 3 (AUC=0.983); the progression from AKI Stage 1 to Stage 2 (AUC=0.930); the progression from AKI Stage 2 to Stage 3 (AUC=0.951); and DIC (AUC=0.838).

In derivative work: a correlation between pre-DIC patients and metabolic acidosis is shown, a meta-analysis on misclassified patients is given, a disease pathway that demonstrates how ARDS and DIC can interact in a positive feedback loop is presented. DIC is shown to be implicated in 78% of all in-hospital mortality of ARDS patients.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AIDS: Acquired Immune Deficiency Syndrome

AKI: Acute Kidney Injury

ALI: Acute Lung Injury

ALT: Alanine Aminotransferase

APACHE: Acute Physiology and Chronic Health Evaluation

ARDS: Acute Respiratory Distress Syndrome

AST: Asparate Aminotransferase

AUC: Area Under Curve

BIDMC: Beth Israel Deaconess Medical Center

BSA: Body Surface Area

BUN: Blood Urea Nitrogen

CCU: Coronary Care Unit

CI: Confidence Interval

CPAP: Continuous Positive Airway Pressure

CPK: Creatine Phosphokinase

CPU: Central Processing Unit

CSRU: Cardiac Surgery Recovery Unit

DIC: Disseminated Intravascular Coagulation

EMR: Electronic Medical Record

FN: False Negative

FP: False Positive

FPR: False Positive Rate

GCS: Glasgow Coma Score

GFR: glomerular filtration rate

GT: Gastronomy Tube

HDL: High Density Lipids

HIV: Human immunodeficiency virus

HR: Heart Rate

ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification

ICU: Intensive Care Unit

INR: Internation Normalization Ratio

ISTH: International Society on Thrombosis and Haemostasis

IV: Intravenous

JAAM: Japanese Association for Acute Medicine

KB: Knowledge Base

KDIGO: Kidney Disease Improving Global Guidelines

LD: Lactate Dehydrogenase

LDL: Low Density Lipids

LIPS: Lung Injury Prediction Score

LMWH: Low Molecular Weight Heparin

LOS: Length Of Stay

LVSW: Left Ventricle Stroke Work

MAP: Mean Arterial Pressure

MEWS: Modified Early Warning Score

MIMIC-III: Medical Information Mart for Intensive Care III

N: Number of samples

NA: Not Available

NAGMA: Non-Anion Gap Metabolic Acidois

NBP: Non-invasive Blood Pressure

NPV: Negative Predictive Value

PEEP: Positive End-Expiratory Pressure

PO: Per Os, by mouth

PPV: Positive Predictive Value

PT: Prothrombin Time

Q1: First Quartile

Q3: Third Quartile

QSOFA: Quick Sequential Organ Failure Assessment

RAM: Random Access Memory

RIFLE: Risk, Injury, Failure, Loss of kidney function, and End-stage kidney disease

ROC: Receiver Operating Characteristic

RRT: renal replacement therapy

SAHRF: Severe Acute Hypoxemic Respiratory Failure

SAPS: Simplified Acute Physiology Score

SDB: Support Database

SICU: Surgical ICU

SIRS: Systemic Inflamatory Response Syndrome

SQL: Structured Query Language

SOFA: Sequential Organ Failure Assessment

TN: True Negative

TP: True Positive

TPR: True Positive Rate

TSICU: Trauma Surgical ICU

WBC: White Blood Cell

WEKA: Waikato Environment for Knowledge Analysis

ACKNOWLEDGMENTS

- My parents, Katherine and Alexander Jablonowski, my partner, Abigail Cromwell, and my children, Madison and Gwendolyn, for their infinite patience and boundless support
- Dr. Linda Shapiro for her continued guidance
- The late Dr. Fredric Wolf for his direction
- Dr. Eshana Shah for her clinical collaboration
- PhysioNet for diligently producing one of the greatest resources for predictive bioinformatics, MIMIC-III
- The University of Washington Department of Emergency Medicine for their support
- The Open Science Grid and the Globus Alliance (from the University of Chicago and Argonne National Laboratories) for computing resources [71] [79]
- Funded, in part, by a fellowship from the National Library of Medicine
- The open source community: Linux; Ubuntu; VirtualBox; Python (numpy, scipy, and sklearn); Perl; Gnu Compiler Collection; Java; Java-Bridge; WEKA; PostgreSQL; SQLite; and MySQL

DEDICATION

From Katherine and Alexander and all that they gave
with help from Abigail
to Madison and Gwendolyn
may you have a better world

Chapter 1

INTRODUCTION

1.1 Motivation

The motivation for the research contained in this dissertation is very much an effort to predict the onset of high-mortality diseases in the intensive care unit. The earlier the prediction, the more time a clinician has to alter the patient's trajectory, avoid mortality, and mitigate adverse outcomes.

The research contained in this dissertation makes several scientific discoveries and novel characterizations of disease. More importantly, the intelligent agent framework described has a proven ability to drive scientific discovery. The sophisticated tool may be generalized and applied to other problems. Indeed, the one tool is applied to five different diseases. The tool may predict the onset of a disease, and may be applied to any disease where onset can be defined. In this research the intelligent agent framework is applied to ARDS, SAHRF, AKI, sepsis and DIC. Generally, the intelligent agent framework can data mine a database and predict an event.

Carried forward, this research and derivative works will aide in patient care and curb adverse outcomes. The ICU setting is chosen, because it is there where the health data is rich and it is there where interventions can have a huge impact. ARDS, SAHRF, AKI, sepsis, and DIC were chosen because these diseases have: a clearly defined disease onset; a large impact on the patient (an in-hospital mortality rate of 28%, 23%, 18% (as high as 47% for Stage 3 AKI), 44%, and 27% (38% when also afflicted with ARDS) respectively); and actionable early therapeutics to change the patients trajectory. Sophisticated predictive models may enable clinical tools to identify at-risk patients early, allowing for the early application of therapies and mitigating adverse outcomes.

1.2 Research Objectives

The goals of this research are to 1) create an intelligent agent tool capable of: navigating an EMR; generating and testing hypotheses within the confines of the EMR; and produce good performing models of disease; 2) apply the intelligent agent framework to the diseases of ARDS, AKI and sepsis; and 3) apply the intelligent agent framework to derivative investigations of diseases, which are SAHRF, Stage 2 AKI, Stage 3 AKI, Progression from Stage 1 to Stage 2 AKI, Progression from Stage 2 to Stage 3 AKI, DIC from ARDS patients, ARDS from DIC patients, and DIC.

1.3 Thesis Overview

Chapter 2 examines the peer-reviewed literature of: the machine learning algorithms used; ARDS; AKI; sepsis; and DIC. Chapters 4, 5, and 6 examine the predictive capability of the intelligent agent framework on ARDS, AKI, and sepsis respectively.

During the course of researching ARDS and AKI opportunities arose to study the depth of disease states and disease interactions. Table 1.1 is the calculated P-value of diseases co-occurring at random. The table is not symmetric along the diagonal, because exclusion criteria for diseases are unique, and therefore do not have a commutative property. That is, the probability of disease A given disease B is not equivalent to the probability of disease B given disease A in this framework. Derivative investigations are bolded. SAHRF (Chapter 4) is a necessary (but not sufficient) condition for ARDS. The intelligent agent derived model predicts the onset of SAHRF very well. This dissertation conclusively shows a correlation between ARDS and DIC in Chapter 4. The large number of AKI positive patients allowed for deeper investigation, and in Chapter 5 AKIs are predicted and characterized at all three Stages and between progressively severe Stages. Chapter 7 is dedicated to the investigation of predicting the onset of DIC.

	Hypoxemia	ARDS	AKI1	AKI2	AKI3	Sepsis	jaamDIC	isthDIC
Hypoxemia		**0.036**	0.294	0.333	0.341	0.363	0.188	0.322
ARDS	0.036		0.273	0.249	0.243	0.278	**0.051**	0.304
AKI1	0.199	0.273		**0.155**	0.162	0.338	0.253	0.277
AKI2	0.201	0.106	0.108		**0.013**	0.162	0.212	0.201
AKI3	0.205	0.093	0.108	0.011		0.171	0.184	0.181
Sepsis	0.222	0.113	0.281	0.139	0.156		0.164	0.142
jaamDIC	0.177	**0.044**	0.328	0.332	0.317	0.313		0.423
isthDIC	0.315	0.296	0.349	0.325	0.316	0.297	0.424	

Table 1.1: P-value of co-occurring diseases: The probability that a patient afflicted with a disease in the vertical also is afflicted with a disease in the horizontal at random. It is the P-value of the two diseases co-occurring. Bold faced represent disease relationships inspected in this dissertation, in addition to the diseases themselves.

Chapter 2
RELATED LITERATURE

2.1 Introduction

This chapter is an examination of peer-reviewed literature relevant to this dissertation. It covers the 17 machine learning algorithms utilized via the Java based WEKA libraries. ARDS, AKI, sepsis, and DIC are each described with current knowledge of epidemiology, pathogenesis, and capabilities of predictive models.

2.2 Machine Learning Algorithms

This dissertation makes use of 17 machine learning algorithms. The non-linear Random Forest algorithm based on work by Breiman [4] dominates this dissertation and is the primary algorithm used for all final models. The other 16 algorithms include DecisionStump based on work by Iba and Langly [36]; Decision Tables based on work by Kohavi [42]; the J48 and REP Tree algorithms based on work by Quinlan [73]; the JRip algorithm based on work by Cohen [9]; the KStar algorithm based on work by Cleary and Trigg [8]; Logistic Model Trees and Simple Logistic Regression based on work by Landwehr et al. [45]; Logistic Regression based on work by Le Cessie and van Houwelingen [48]; Multilayer Perceptrons based on work by Hastie et al. [32]; the Naive Bayes algorithm based on work by John and Langley [37]; the One R algorithm based on work by Holte [35]; the PART algorithm based on work by Frank and Witten [23]; Random Trees based on work by La Valle and Kuffner [46]; the Sequential Minimum Optimization algorithm for training a support vector classifier based on work by Platt [70]; and Voted Perceptrons based on work by Freund and Schapire [24].

The aforementioned machine learning algorithms are made accessible to the intelligent agents of this dissertation via Java libraries that are part of the Waikato Environment for

Knowledge Analysis (WEKA) by Frank et al. [14]. WEKA is made available in the Python programming language, the native language of the intelligent agent framework, by Java-Bridge.

2.3 Acute Respiratory Distress Syndrome

ARDS existed as a concept since 1967, when Ashbaug et al. [1] characterized the clinical, radiographic, and pathologic features. ARDS presently afflicts an estimated 200,000 patients per year in the US, and is associated with significant morbidity and a mortality rate of 37.5% [20].

Koh [41] describes the pathogenesis of the disease taking two pathways that lead to ARDS. Either the capillary-membrane or the surfactant producing cells may be damaged, which propogates a sequence of events leading to alvelolar damage and impaired oxygenation and ventilation at the capillary-alveolar junction. With ongoing alveolar damage and in the setting of the inflammatory milieu, normal lung parenchyma is replaced by stiff fibrotic tissue that lacks normal compliance and is incapable of CO2 and O2 gas exchange. These structural and functional changes lead to an increase in dead space (inspired air cannot participate in gas exchange by the alveoli) as described by Nuckton et al. [65].

There are several known risk factors associated with ARDS. Nosocomial, or in-hospital, causes of ARDS include mechanical ventilation. Gajic et al. [26] found that 24% of ARDS patients studied had onset after ventilation. The insult to lung tissue may be caused by a traumatic intubation or delivery of an excessive amount (tidal volume) of air into the lung leading to alveolar overdistension. Gajic later co-authored a paper [25] in 2011 which described risk factors that predispose a patient to acquiring ARDS. These include: admission source; APACHE II score; shock; aspiration; cardiac surgery; aortic vascular complications; smoke inhalation; lung contusion; tachypnea (high respiratory rate); low SpO_2; high FiO_2; low albumin; and low pH.

Kallet et al. [38] describes early therapeutic interventions to be taken as part of the ARDSnet Mechanical Ventilation Protocol, which is now standardized management for these pa-

tients. It includes setting a relatively low tidal volume (6-8 mg/kg of predicted body weight) in order to minimize alveolar hyperinflation and patient pronation (lay patient face down) which would recruit additional dependent alveoli. Gong et al. [30] identified an additional action, which is to avoid packed red blood cell transfusions if possible.

ARDS, as a disease, has had several different diagnostic criteria. Previously, the distinguishing factor was the degree of hypoxemia (denoted as the P/F ratio) and acute lung injury (ALI) wa characterized to be a milder form of ARDS. At the 2011 Berlin conference, the European Society of Intensive Care Medicine, the American Thoracic Society, and the Society of Critical Care Medicine developed a definition for the diagnostic criteria [74] (known as the Berlin definition). This removes ALI as a separate concept. As ARDS is diagnosed from a constellation of findings rather than a single test or finding, the Berlin definition defines these as: 1) worsening respiratory status or onset within one week after a known insult; 2) radiographic bilateral opacities not fully explained by an alternative etiology; 3) pulmonary edema not due to cardiogenic origin; 4) and delineates the stages of ARDS by the PaO_2/FiO_2 ratio starting at hypoxemic (or <300 mmHg).

Gajic et al. [25] derived a Lung Injury Prediction Score (LIPS) to describe patients at risk of developing ARDS. The scoring system includes the following risk factors: shock; aspiration; sepsis; pneumonia; orthopedic spine surgery; acute abdominal surgery; cardiac surgery; aortic vascular surgery; traumatic brain injury; smoke inhalation; near drowning; lung contusion; multiple fractures; alcohol abuse; obesity; hypoalbuminemia; chemotherapy; fraction of inspired oxygen; tachypnea; oxyhemoglobin saturation <95%; acidosis; and diabetes mellitus. Though the scoring system has an impressive AUC of 0.79, the positive predictive value is merely 0.18 while the negative predictive value is 0.97. This makes the scoring system much more valuable in predicting patients who will not become afflicted with ARDS than it is for predicting patients who will become afflicted. LIPS is validated in a population based cohort by Trillo-Alvarez et al. [89], five academic medical centers by Soto et al. [84], and a surgical critical care cohort by Bauman et al. [18]. All studies have similar results: AUC=0.79-0.84; positive predictive value = 0.17-0.24; and negative predictive

value=0.95-0.97.

This dissertation describes a correlation, and probable interaction, between ARDS and Disseminated Intravascular Coagulation (DIC). Two definitions of DIC are considered. In 2001, the "Scientific Subcommittee on Disseminated Intravascular Coagulation (DIC) of the International Society on Thrombosis and Haemostasis (ISTH)" put forward a diagnostic definition described by Taylor et al. [86]. The more liberal defining criteria for DIC was put forward by the Japanese Association for Acute Medicine (JAAM) described by Gando et al. [27]. The representation of ARDS and DIC correlating in the peer-reviewed literature is scarce. In 1977 Ogawa et al. [67] studied 9 ARDS patients, 6 of whom also had DIC. In 1986 El-Kassimi et al. [15] studied 52 heat stroke patients, 12 of whom had ARDS and DIC. In 2014 Miyoshi et al. [58] performed a retrospective study of 142 patients with both ARDS and DIC and found that administering a neutrophil elastase inhibitor (to minimize tissue damage in the ARDS pathway) and a thrombin coagulation inhibitor (to mitigate the consumption of clotting agents by DIC) combined was more effective than either drug alone, though did not go on to describe a causal or correlative process.

This dissertation goes on to create a predictive model for ARDS patients to develop DIC at the time of ARDS onset, and a predictive model for DIC patients to develop ARDS at the time of DIC onset. There are no such predictive models in the peer-reviewed literature. Additionally it addresses a predictive model for Severe Acute Hypoxemic Respiratory Failure (SAHRF) which is not found in the peer-reviewed literature.

2.4 Acute Kidney Injury

Acute Kidney Injury (AKI) is the sudden decrease of renal function, characterized by elevated serum creatinine and oftentimes decreased urine output. The kidneys are amongst the most sensitive organs to changes in blood perfusion or other abnormalities, therefore a dysfunction elsewhere often leads to renal injury. The sensitivity of the kidneys is evident by the large number of comorbid diseases that are associated with an increase in incidence of AKIs, including: inflammation and sepsis; endoplasmic reticulum stress (caused by nephrotoxic

drugs or iodine-based contrast agents); endothelium and vascular damage; diabetes; cancer; cardiac surgery; and human immunodeficiency virus (HIV) as described by Farooqi et al. [21].

Epidemiologically AKIs are estimated to cause 2 million deaths per year, afflicting 7-18% of hospitalized patients and 50% of ICU patients [6]. Negi et al. [62] describe the long term risks associated with AKIs which include: progression to chronic kidney disease; progression to end-stage renal disease requiring dialysis; cardiovascular events such as myocardial infarction (heart attack), congestive heart failure, stroke and increased mortality.

There are many variations of disease definition for AKIs. The Acute Kidney Injury Network (AKIN) definition put forth by Mehta et al. [57] is used here. The most popular alternative is RIFLE described by Bellomo et al [3]. Bellomo is less desirable for a retrospective study as it includes as a variable GFR (glomerular filtration rate). GFR is measured or calculated for 29.3% of ICU patients in this dissertation, as opposed to the more commonly captured urine output (99.9%) and serum creatinine (93.1%). To include GFR would be to introduce a bias in the sampled patients by those whom are already suspected of renal complications.

In the AKIN definition of AKI, establishing a serum creatinine baseline is critical. One sufficient condition of the AKIN definition is an elevation above baseline. The De Rosa et al. [12] definition of baseline serum creatinine is used here. The definition takes into account age, gender, and race.

Gaudry et al. [29] highlight the benefits of early renal replacement therapy (RRT) for AKI patients in a randomized trial. For the early strategy cohort, RRT was applied within six hours of an AKI event. The lag strategy cohort started RRT per standard indications (acidosis, critical electrolyte disturbances, volume overload, or symptomatic uremia, generally >112 mg/dL). Though negative effects of the early strategy include catheter related infection and hypophosphatemia (low levels of phosphate in the blood), the positive effects include a decrease in the number of medical interventions related to hyperkalemia (elevated potassium in blood serum), metabolic acidosis, and diuretics. Other studies by Zarbock et

al. [96] and Barbar et al. [2] sought to examine the benefit of early RRT but either lacked statistical power or were stopped by a data and safety monitoring board.

There exist several predictive algorithms for AKI that will be used as a basis of comparison for the models that are created in this study. Kate et al. [39] use logistic regression of patient data 60 years or older to predict AKI hospital-wide. Cheng et al. [7] use a Random Forest algorithm to predict AKI hospital-wide. He et al. [33], with overlapping authorship of Cheng et al., use a Random Forest algorithm to predict AKI hospital-wide. Mohamadlou et al. [59] use a XGBoost algorithm to predict AKI Stages 2 and 3 in an ICU and hospital-wide.

Topics are addressed in this dissertation that are not present in the peer-reviewed literature. Specifically prediction models for exclusively AKI Stage 3, the progression from AKI Stage 1 to AKI Stage 2, and the progression from AKI Stage 2 to AKI Stage 3 are included.

2.5 Sepsis

Sepsis syndromes can develop from an infection and/or bacteremia (bacteria in the bloodstream). There is a spectrum of clinical states ranging from asymptomatic or mildly symptomatic infection requiring minimal interventions to organ dysfunction, septic shock, and death. Sutton and Friedman [85] found that sepsis affects 651 per 100,000 persons in the US. Mayr et al. [55] found severe sepsis to affect 300 per 100,000 persons and carries a mortality of 25% in the US. Torio and Andrews [88] estimate the associated health care cost of sepsis at $24 billion per year.

Various patient-related factors (increasing age, immuncompromised host system) and disease-related factors (particularly virulent organisms such as E. coli or gram-positive organisms such as S. aureus) can impact the severity. One of the most important factors that affects severity and mortality is the time to recognition and treatment. Rivers et al. [75] found that early goal-directed therapy, including fluid resuscitation and timely antibiotic administration, compared with standard-of-care of the day reduced mortality by one-half (30% vs 46%, respectively). In 2004, the Surviving Sepsis Campaign published its first evidence-based guidelines focused on multiple aspects of clinical care. This has been further refined in

2008 and in 2016, with each iteration emphasizing prompt recognition and treatment. Once the Surviving Sepsis Campaign published therapy guidelines, it was validated to reduce mortality by: Kumar et al. [44]; Gao, et al. [28]; Kortgen, et al. [43]; Nguyen, et al. [64]; Ferrer, et al. [22]; Teles, et al. [87]; Zambon, et al. [95]; and Shiramizo, et al. [82]).

Though there are well-informed practices and creation of "septic bundles" to guide care once sepsis is recognized, the detection of sepsis remains critical to providing such care. Unfortunately, sub-optimal definitions of sepsis have long existed. Previously one of the leading tools involved the use of Systemic Inflammatory Response Syndrome (SIRS) criteria which primarily focused on abnormal vital signs (temperature, heart rate, respiratory rate). Its greatest drawback was that numerous non-infectious etiologies can cause a patient to meet SIRS criteria and therefore lead to inappropriate and unnecessary care.

An improved definition was proposed in 2016 with the publication of the Sepsis-3 definition criteria by Singer et al. [83]. The Sepsis-3 definition describes sepsis as "life-threatening organ dysfunction caused by a dysregulated host response to infection." Organ dysfunction is characterized by abnormalities in mentation (Glasgow Coma Score), respiration, cardiac (hypotension or need for vasopressor support), hepatic (bilirubin), renal (serum creatinine, urine output) and coagulation systems.

There are many predictive models in the peer-reviewed literature, narrowed down 19 to use in this dissertation as a foundation of comparison. Desautels et al [13] has an impressive performance, but suffers from a bias due to an artificial reduction in false positives from the inclusion criteria. Van Wyk et al [91] requires continuous physiological data as an inclusion criteria. Saqib et al. [78] uses the Angus criteria and excludes patients without measured data for all variables. Shankar-Hari et al. [80] us based on neutrophil CD279 tests. McCoy et al. [56] employs an interventional SIRS screen twice a day. Mao et al. [52] excludes patients without measured data for all variables. Layios et al. [47] examines platelet activation markers. For Rothman et al. [76], the difference between development performance and validation performance implies over-fitting of the development model. Lindner et al. [49] only applies to polytrauma patients. Mardi et al. [53] examines the IL-6 blood test results.

Lukaszewski et al. [50] includes only patients already at risk for sepsis. Other comparison models included are: Vassiliou et al. [92] in an ICU; Gouel-Cheron et al. [31] after major trauma; Faisal et al. [19] in emergency medicine; Nemati et al. [63] in an ICU; Shashikumar et al. [81] using continuous physiological data in an ICU; Danner et al. [11] in a whole hospital; Olenick et al. [68] in a 12 hospital cohort; and Wang et al. [94] in a national population-based study for community acquired sepsis.

2.6 Disseminated Intravascular Coagulation

Matsuda [54] found DIC present in roughly 1% of tertiary care patients, though there are mild and sub-clinical manifestations that may not be accounted. Treatment of DIC is focused on treating the underlying condition [93] that provides the procoagulants which starts the DIC cascade by causing excess thrombin formation. Though this dissertation describes ISTH's definition of DIC, provided by Taylor et al. [86], it primarily makes use of JAAM's DIC definition provided by Gando et al. [27].

The prediction of the onset of DIC, addressed in this research, is not present in the peer-reviewed literature.

2.7 Summary

This chapter summarized the peer-reviewed knowledge relevant to this dissertation. It included the machine learning algorithm implementation. Also reviewed were the diseases of ARDS, AKI, sepsis, and DIC with current knowledge of epidemiology, pathogenesis, and capabilities of predictive models.

Chapter 3
METHODOLOGY

3.1 Introduction

The chapter describes the methodology of creating the intelligent agent framework and the tools the intelligent agents use to data mine the EMR. The data source of this dissertation is an EMR, MIMIC-III. The knowledge necessary for efficiently mining the EMR (variable names, variable location, statistics associated with variables, aggregate definitions, logic definitions, metric definitions, operator definitions) and running machine learning algorithms, with adjustable parameters, are stored in the knowledge base. The agents update the support database with progress made in modeling. The disease condition definition sets the time of disease onset that the agents use as reference when constructing the feature space. Missing data is handled by a nearest neighbor approach. The process overview of an intelligent agent is included, along with feature space learning and sub-population concept hypothesis learning and testing. Included are all metrics of analysis, methods of comparison, and the reporting of variables, demographics, and comorbidities. Surrogate variables, variables that are artifacts and most likely infer a deeper phenomenon, are described. This chapter concludes with biases and limitations of the methodology. further details are given in the appendices.

3.2 Data Source

MIMIC-III (Medical Information Mart for Intensive Care III) is a de-identified electronic medical record derived from intensive care units at the Beth Israel Deaconess Medical Center between 2001 and 2012. The database covers 46,520 patients during 58,976 individual hospital admissions and culminates in 61,532 individual ICU stays.

The majority of the features selected by agents in these experiments are derived from seven high temporal resolution tables: chartevents (n=330,712,483 including attributes like heart rate, fingerstick glucose, ventilator setting, etc.); inputevents_cv and inputevents_mv (derived from two different vendors, CareVue and Metavision) (n=21,147,885 including attributes like albumin 20%, saline 0.9%, dextrose 5%, etc.); labevents (n=27,854,055 including attributes like free calcium, lactate, glucose, etc.); microbiologyevents (n=631,726 including attributes like blood culture, bone marrow cytogenetics, rapid respiratory viral screen and culture, etc.); outputevents (n=4,349,218 including attributes like urine flush, hemodialysis, thoracentesis, etc.); and, procedureevents_mv (n=258,066 including attributes like CT scan, bronchoscopy, tunneled (Hickman) line, etc.).

The demographics of the patient population of the MIMIC-III database are contained in Table 3.1. The generation of this demographic signature is necessary to highlight demographic differences in the sub-populations of studies derived from this database. Of the 61,532 ICU patient stays in MIMIC, the median age is 62 years, the population is predominantly male (56.0%), 13.1% of patients have a 30 day mortality, and the median length of stay is 2.09 days (Q1=1.11, Q3=4.48). The breakdown of services within the ICU are as follows: the Medical ICU (MICU) has 34.3% of the patients; the Cardiac Surgery Recovery Unit (CSRU) has 15.5% of patients; the Surgical ICU (SICU) has 14.4% of patients; the Coronary Care Unit (CCU) has 12.6% of patients; and, the Trauma Surgical ICU (TSICU) has 10.4% of patients. Of the patient population, the ethnic/racial makeup is as follows: 70.1% of patients identify as White; 9.8% of patients identify as Black; 3.6% of patients identify as Hispanic; and 3.3% of patients identify as Asian. The insurance coverage makeup is as follows: 48.4% of patients are covered by Medicare; 37.8% of patients are covered by private insurance; 9.8% of patients are covered by Medicaid; 3.0% of patients are covered by government insurance; leaving 1.0% of patients as uninsured/self-pay.

The Entity Relationship Diagram for the subset of MIMIC-III used in this study is presented in Figure 3.1. There are three basic types of tables: patient; hospital admission; and ICU stay. One patient can have multiple hospital admissions, and one hospital admission

can have multiple ICU stays. The ICU stays table is the central table of the MIMIC database. MIMIC contains three definition tables: items; laboratory items; and ICD diagnoses. MIMIC contains 8 other tables that record the granular data, each of them links to a definition table and either the ICU stay table or the hospital admission table. Output events, chart events, input events, and procedure events all link to the ICU stay table and the item definition table. The ICD diagnosis table links to the ICD diagnosis definition table and the hospital admissions table. Both the laboratory events and microbiology events tables link to the laboratory item definition table as well as the hospital admissions table. This is because laboratory and microbiological findings outside the ICU remain very relevant to the care of the patient in the ICU.

Variable	Population	(%) or (Q1,Q3)
N	61532	
Age, median	62.0	
Male	34469	(56.0%)
30 Day Mortality	8089	(13.1%)
ICU LOS, median(Q1,Q3)	2.09	(1.11, 4.48)
ICU		
CCU	7726	(12.6%)
CSRU	9312	(15.1%)
MICU	21088	(34.3%)
SICU	8891	(14.4%)
TSICU	6415	(10.4%)
ethnicity		
Asian	2032	(3.3%)
Black	6003	(9.8%)
Hispanic	2197	(3.6%)
White	43154	(70.1%)
insurance		
Government	1837	(3.0%)
Medicaid	6017	(9.8%)
Medicare	29812	(48.4%)
Private	23244	(37.8%)
Self Pay	622	(1.0%)

Table 3.1: MIMIC Demographics.

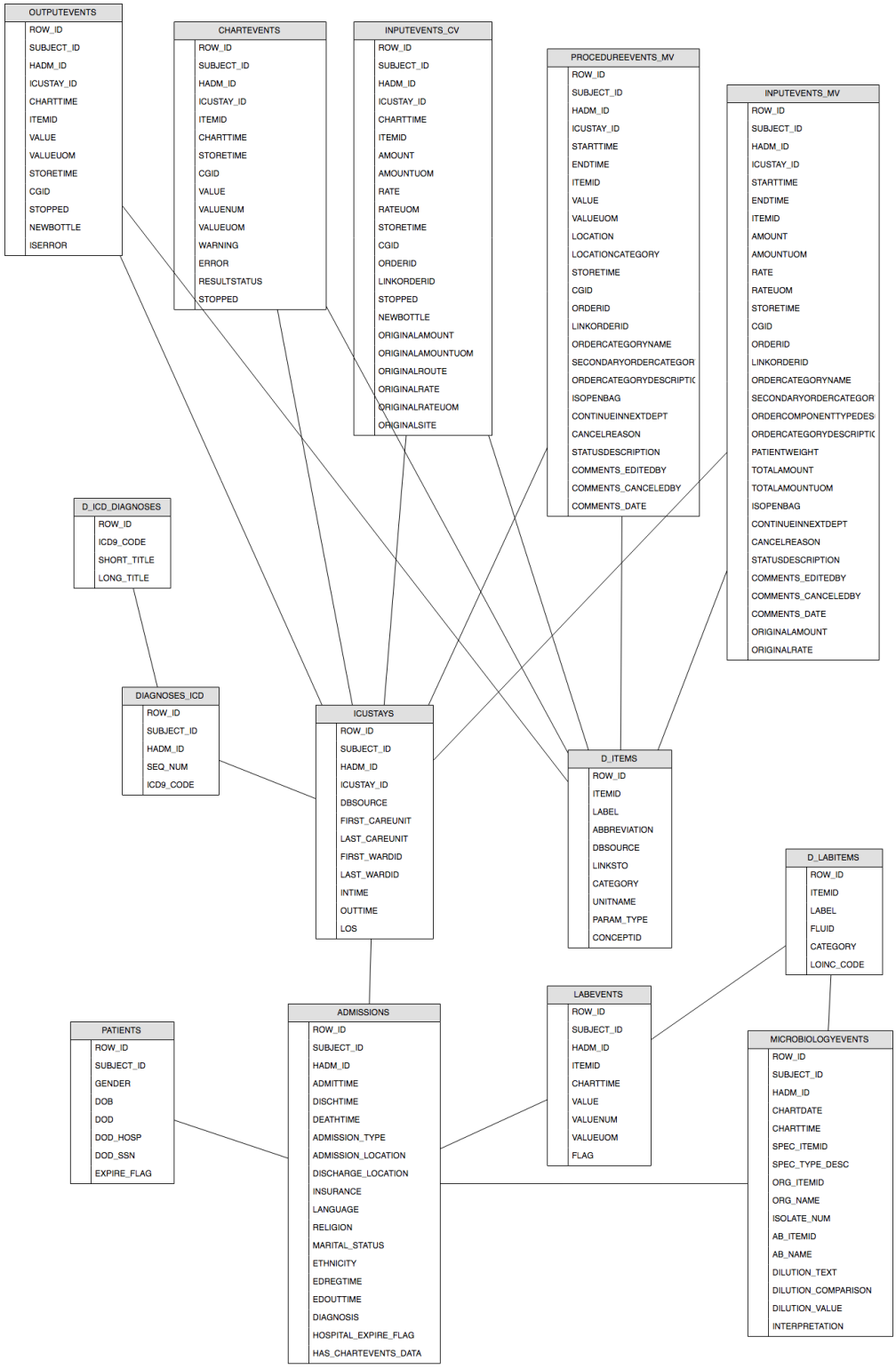


Figure 3.1: MIMIC Entity Relationship Diagram.

3.3 Knowledge Base

The knowledge base in Figure 3.2 was developed to organize information to support intelligent agents operation and implementation of complex SQL queries and machine learning algorithms using the MIMIC database. The knowledge base consists of three sections of tables.

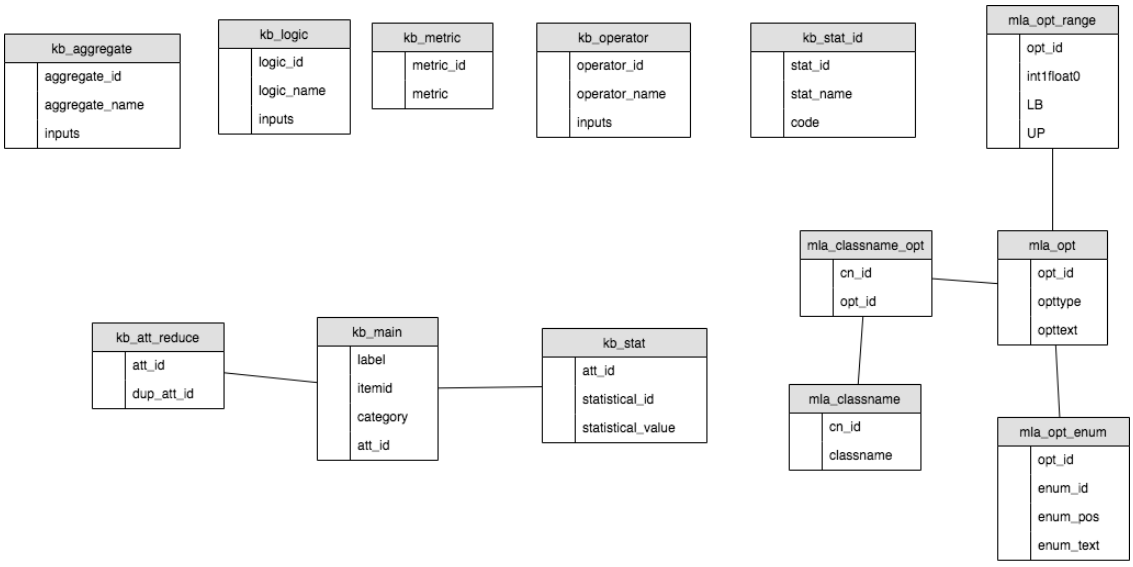


Figure 3.2: Knowledge Base Diagram.

Three constructed tables are dedicated to describing the items in the MIMIC database. These items describe the granular data of MIMIC in output events (subsection A.0.12), chart events (subsection A.0.2), input events (subsections A.0.8 and A.0.9), procedure events (subsection A.0.14), laboratory events (subsection A.0.10), and microbiological laboratory events (subsection A.0.11). The main table (subsection B.0.4) describes the items used in this dissertation. There are 575 attributes populated in this table which are present in a minimum of 5% of patients. There are many other attributes in MIMIC; to pare down the computational load the decision was made to limit the attributes to those that appear in a minimum of 5% of patients. A lower percentage would include more attributes and dramatically increase

the feature space, and the added attributes would not prove statistically significant given the size of the patient population. The statistics table (subsection B.0.7) contains statistical distributions of the values of the items in the MIMIC database. Previously computed statistical distributions are important for intelligent agents, which will be selecting and adjusting parameters of models based on these items. The third table related to MIMIC items is an attribute reduction table (subsection B.0.2) which merges multiple identical attributes of the main table under a single attribute identifier. For example, over the 12-year range of MIMIC there have been 6 different non-overlapping identifiers for white blood cell count. The attribute reduction table amalgamates duplicate items under a single attribute identifier.

Five tables are dedicated to executing machine learning algorithms on a data set. The table mla_classname (subsection B.0.9) sets a unique identifier to each one of the 17 algorithms used by the intelligent agents. Each algorithm has a set of definable options that direct the performance of the algorithm and are amassed in the mla_classname_opt table (subsection B.0.10). Each option has a type, distinguishable in the mla_opt table (subsection B.0.11). The allowable types are either a range with an upper bound and lower bound set in the table mla_opt_range (subsection B.0.13), or an enumeration of text, such as selecting a filter by name, set in the table mla_opt_enum (subsection B.0.12). Possible options include estimators (Bayes Net, Bayesian Model Averaging, or simple); search algorithms (K2, genetic, hill climber, look ahead hill climber, repeated hill climber, simulated annealing, tabu search, or global score); SVM kernels (poly kernel, normalized poly kernel, cached kernel, or radial basis function); and neighbor distance calculators (Euclidean, Chebyshev, edit distance, or Manhattan distance).

The third class of tables found in the knowledge base contain basic definitions for writing complex SQL queries. Though these tables could be completely contained in the execution programs themselves, for stability and ease of access they were placed in stand-alone tables. Aggregate (subsection B.0.1) defines and indexes the SQL aggregate functions the intelligent agents can use (min, max, avg, etc.). Logic (subsection B.0.3) defines and indexes the logical operators the intelligent agents can use in either SQL queries or in constructing

complex concepts (AND, OR, etc.). The metric table (subsection B.0.5) defines and indexes some basic metrics intelligent agents use to evaluate a model's performance (TP, TN, etc.). The operator table (subsection B.0.6) defines and indexes basic operators used in setting parameters (less than, greater than, between two numbers, etc.). The statistical identifier table (subsection B.0.8) defines and indexes statistical metrics used in describing data, and is most notably used in describing statistical distributions such as in the statistics table (subsection B.0.7) of this knowledge base.

3.4 Support Database

A supporting database (SDB) was created to support the functions of intelligent agents as the data mining of the MIMIC database commences. The supporting database is broken out into several key components: problem declaration; feature space; concept formation; and evaluation of models.

The problem declaration table, sdb_target_class, holds data surrounding the onset of disease, breaking out positive patients and control patients. For disease positive patients the timestamp records the earliest conditions necessary for the disease to be diagnosed during their ICU stay, or a randomly generated timestamp within the bounds of their ICU stay for patients who do not contract the disease.

The feature space per disease is calculated from MIMIC and populated in the fdb table (subsection C.0.2). For every disease inspected it contains an attribute, value, and the timestamp for that value for all attributes in the main table of the knowledge base (see subsection B.0.4). The table also contains a t-minus calculation which indicates the time before the onset of disease as defined in the sdb_target_class table (subsection C.0.15). Using this feature space table in conjunction with the knowledge base main table, the intelligent agents create an instance (subsection C.0.3). An instance is associated with a set of attributes found in the feature space - attribute table (subsection C.0.4) and constitutes an iteration the intelligent agents are inspecting. A feature space - patient table (subsection C.0.5) is also created to declare the patients for whom the attributes are valid. The feature space - patient table

saves computation time when amassing data files from the fdb table data, and also allows for inclusion/exclusion concepts to be applied. The selected attributes and patients (with or without inclusion/exclusion concepts) constitute a model. The statistics and description of the machine learning algorithm that ran the model to predict the onset of the disease are stored in the gfs_stat table (subsection C.0.6), and the higher resolution incremental threshold curve data is stored in gfs_stat_thresh (subsection C.0.7).

I included seven supporting database tables to support the formation and evaluation of concepts. A "concept", as the term is used in this dissertation, forms an inclusion criterion for a sub-population of patients to be included in a model. The sdb_main table (subsection C.0.12) defines an instance of a hypothesis. The intelligent agents evolve a hypothesis into one or more derivative hypotheses. The provenance of every hypothesis is tracked through a parent-child schema. sdb_main also contains a target_id that links to the definition of disease (subsection C.0.15). The hypothesis table, sdb_hyp (subsection C.0.11), contains the concept_id associated with the instance of the hypothesis. A concept under the hypothesis framework may be made of one or more other concepts, assembled with logic operators found in the sdb_concept_logic table (subsection C.0.10). The root concept, found in the sdb_concept table (subsection C.0.9), contains the necessary parameters to write a SQL query and include a sub-population that meets the criteria of the concept. The cdb table (subsection C.0.1) is a measure to save on on-the-fly computation time by previously assembling the patient, attribute, value of attribute, and timestamp for every concept. Patients who qualify in a hypothesis' sub-population may be incorporated into the feature space - patient table (subsection C.0.5) or evaluated against the disease as a stand-alone model. Evaluations are stored in the sdb_stat table (subsection C.0.14), and the machine learning algorithms that evaluated those models may be found in the sdb_mla table (subsection C.0.13).

Evaluating a model against an accepted standard is necessary to identify the impact of any findings the intelligent agents discover. To evaluate the intelligent agent models, the efforts of this research compare their performance with the performance metrics of the publication. Additionally, whenever possible, the published model is re-implemented on MIMIC

and comparisons are made by the performance metrics using the same patient population. Disease specific database tables were designed to house other published models that are re-implemented on the MIMIC database. An example may be found in the LIPS (Lung Injury Prediction Score) table (Section 4.5) where every parameter of the scoring method is itemized the by authors.

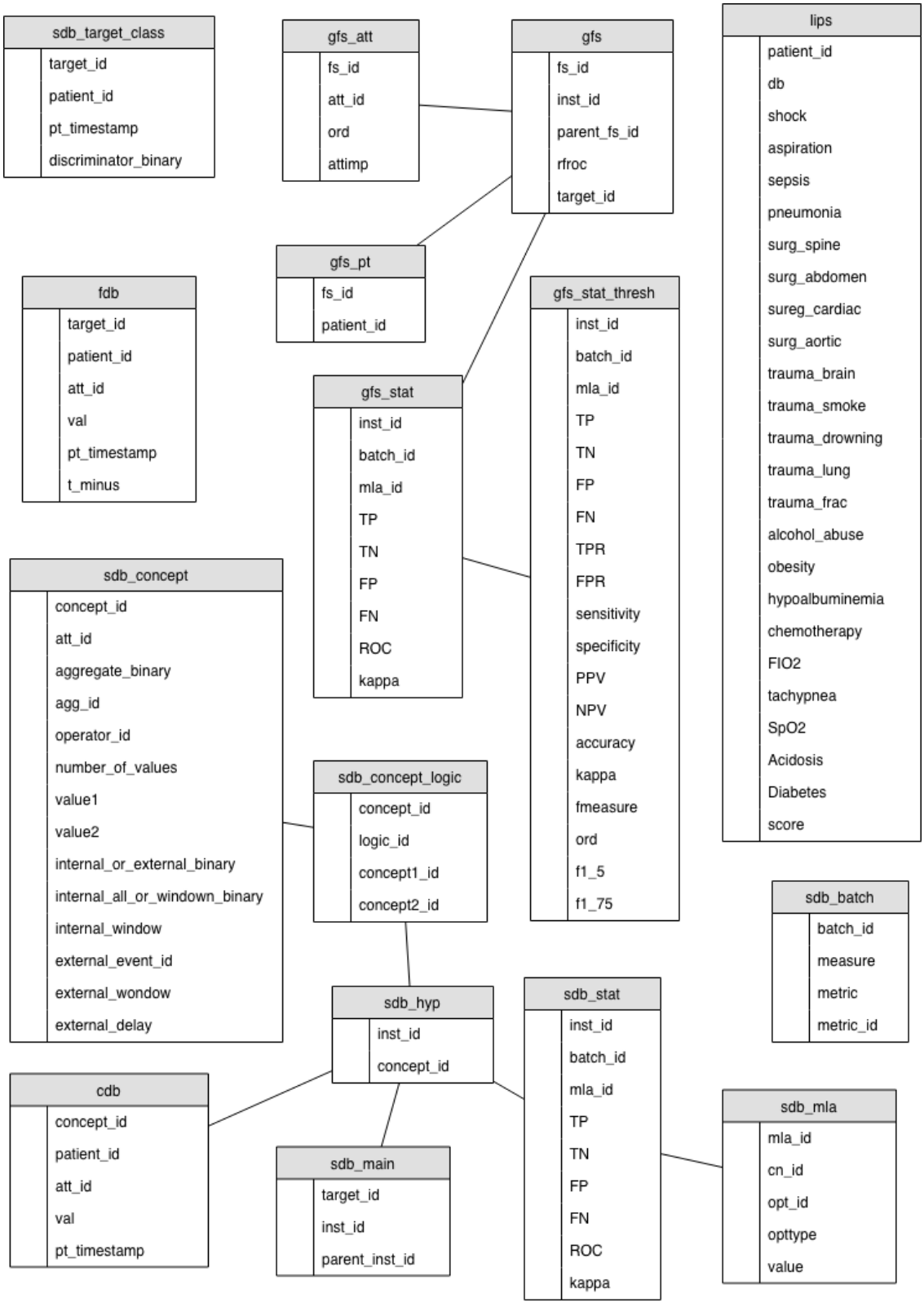


Figure 3.3: Support Base Diagram.

3.5 Seeding the Databases

3.5.1 Knowledge Base

"Seeding the databases" means populating the tables with information the intelligent agents will need to function. A table is "stand-alone" if it may be populated independently of other tables. A table is "dependent" if populating the table requires information from another table.

The stand-alone tables reside in the knowledge base, and are responsible for defining and organizing basic functions the intelligent agents access while constructing complex queries to data mine the MIMIC database or construct concepts. These knowledge base tables are the Aggregate (subsection B.0.1), Logic (subsection B.0.3, Metric (subsection B.0.5), Operator (subsection B.0.6), and Statistics Identification (subsection B.0.8) tables. These stand-alone tables are populated directly from written SQL inserts.

The five knowledge base tables dedicated to executing machine learning algorithms were also populated directly from written SQL inserts. The design and values of the tables were derived from WEKA's classifiers and affiliated options. For each of the 17 classifiers selected from the WEKA library (subsection B.0.9), an internal identifier was created to link to an option (subsection B.0.10). The options required another layer (subsection B.0.11) to disambiguate options which included ranges (subsection B.0.13) from options that were an enumeration of text (subsection B.0.12).

Seeding the knowledge base main table (subsection B.0.4) required collating patient information from seven different tables in the MIMIC database. To be included in this table, an attribute must appear in a minimum of 5% of patients in the MIMIC database. This measures was taken to reduce the feature space and associated computation load for attributes that would not likely be statistically significant. The category field identifies the MIMIC table and the itemid identifies the MIMIC identifier. They are assigned an attribute identifier native to the knowledge base and support database. The statistics table (subsection B.0.7) uses the statistic metrics identified in the previously seeded statistics identification table,

and computes every metric for every attribute. Once the main table is complete, a thorough screening of attributes is performed to identify duplication events. The attribute reduce table (subsection B.0.2) was seeded by written SQL inserts to compensate for duplicate attributes.

3.6 Disease Condition Definition

This dissertation focuses on predicting the temporal onset of disease. For a disease to be considered, the onset of the disease must be able to be calculated from data collected in the EMR. The onset of ARDS, AKI, sepsis, SAHRF, and DIC can all be reliably calculated or inferred from the data contained in MIMIC.

The problem declaration table (subsection C.0.15) is populated through SQL queries which identifies disease onset in MIMIC. Every disease has its own uniquely specific diagnostic criteria. Throughout the Methods chapter ARDS examples will accompany descriptions to demonstrate the application of the methods. ARDS has a set of diagnosis codes (518.5, 518.51, 518.52, 518.53, and 518.52). By the Berlin definition, ARDS also has an oxygen insufficiency requirement of PaO_2/FiO_2 less than or equal to 300 (hypoxemic). PaO_2 is the partial pressure of dissolved oxygen in the blood (80 mmHg is considered normal). FiO_2 is the fraction of inspired oxygen (room air is 0.21, since Earth's atmosphere is 21% oxygen). For this study, it is considered sufficient to identify the onset of ARDS for patients having an ARDS diagnosis code as the time the PaO_2/FiO_2 ratio is less than or equal to 300. All patients are excluded who enter the ICU already hypoxemic, as their disease onset precedes the data represented in MIMIC. The PaO_2/FiO_2 ratio and timestamp was calculated for every patient in the MIMIC database. The timestamp for the first occurrence of the ratio less than or equal to 300 is inserted into the problem declaration table as a condition positive patient. For all other patients (condition negative or control) the timestamp used is randomly selected between their arrival and departure from the ICU.

3.7 Feature Space Database

Predicated upon the problem declaration table being populated, the feature selection database (subsection C.0.2) can be populated using the timestamps previously calculated. The feature selection database consists of all values of attributes considered and timestamps within the considered range, for all patients in MIMIC. The considered temporal range considered for this dissertation is between just before onset and six hour prior. Values are sampled at fifteen minute intervals. The development feature database considers all attributes in the knowledge base main table (subsection B.0.4).

3.8 Missing Data Handling

The feature space database, derived from MIMIC's vast archive of patient's attribute values, has missing values. In the initial attribute list selection (subsection B.0.4), an attribute was viable for this study if it is present for just 5% of patients. Most attributes are present in most patients, but there are some attributes sparsely populated.

There are several methods for the handling of missing data in data sets to be used with machine learning algorithms. A main function of this study is to enable intelligent agents with multiple tools for knowledge discovery. It would be consistent to allow intelligent agents to access multiple methods for the handling of missing data. Pragmatically, this is not feasible due to the enormous computational and storage resources required. The intelligent agents are therefor limited to a singular method for missing data handling.

Some machine learning algorithms function with missing data inherently. These are powerful algorithms, but there are only a few of them. Of the few, the options available are also limited. Allowing the algorithms to handle the missing data themselves would restrict the number of allowable algorithms. To bestow intelligent agents with multiple and varied options, this research effort opted to handle missing data prior to machine learning.

The method of deleting entries with any missing data is appealing, because it retains the maximum integrity of the data set. There are two problems with this approach. First,

though selection bias is not a typical problematic concept for the deleting method in many disciplines, the data is not missing at random. The data in the EMR is constructed by the actions and measurements of the care providers, and inherent bias cannot be ignored. The second problem with the deletion method for handling missing data is attrition of the data set. The prevalence of ARDS is between 2% and 3% of ICU patients. Removing patients with missing data in one of the many attributes selected for each model would reduce the disease positive patients to an insufficient number for statistical consideration. The deleting method is inappropriate for this study.

The most common method for handling missing data is to replace the missing value with the mean, median, or mode of the entire data set for that attribute. Though this is a highly useful technique, the prevalence for creating statistical artifacts makes it unattractive as a solution. The median heart rate in MIMIC is 82 beats per minute. The median respiration rate is 20 breaths per minute. Though the two vital signs are not strictly tied to each other, their operation and adaptation are both regulated by similar stimuli. For example, one would not expect that a patient with a 90th percentile breath rate (28 breaths per minute) would have a 50th percentile heart rate (82 beats per minute). A high breath rate would imply a stress on the pulmonary system, which would typically be accompanied by a higher heart rate. If the heart rate were a missing data point for a pulmonary-system-stressed patient, replacing the missing data with a mean, median, or mode metric would create a statistical artifact of a patient presenting with both stressed and normal vital signs in two variables known to be linked. The method of replacing missing data with the mean, median, or mode of the data set is not chosen for this study.

To handle missing data, this study employs a nearest neighbor approach. The nearest neighbor approach, using the ball-tree algorithm (utilizing the euclidean distance metric, Minkowski, with p=2), is used to find each patient's nearest neighbors based on the values of attributes in the feature selection database. A patient inherits the value for any missing data from their nearest neighbor with a value. As this approach does not discard patients with missing data, it is less biased towards patients with more data (typically more severe

conditions or patients who instigate clinician's suspicion). The nearest neighbor approach does still introduce a source of bias to favor existing values. The data replication occurs at a distributed local level (the nearest neighbor). A single patient may inherit data from multiple near neighbors depending on which values are missing. As a result no one patient is over-represented in the data set, mitigating some biases. This method allows intelligent agents to retain many machine learning algorithm options, maintains a significant patient population size, and is less likely to create statistical artifacts.

3.9 Intelligent Agent Process Overview

Intelligent agents of this study are charged with learning predictors for the onset of diseases. The overall process is depicted in Figure 3.6.

Intelligent agents, described in Figure 3.4, are a program or tool that make an observation of the environment, select an action based on a rule set trying to achieve some goal, and then act on the environment. The environment has therefor changed for the next iteration of the intelligent agent or other agents. A simple example is how a thermostat continuously monitors the temperature of a room and based on a rule set engages the heater to change the environment with the goal of maintaining a pre-set minimum temperature. In the case of this dissertation the agents are complex, but maintain the same fundamental structure, as seen in Figure 3.5. The environment that is observed and acted upon is the EMR and the support database. The intelligent agents are also equipped with a knowledge base that serves to provide: SQL functions, so that the agents may query and update databases; an EMR map, so the agents know where to go to retrieve specific data; and machine learning libraries, so the agents may test a hypothesis it generates.

An intelligent agent is assigned a disease upon initialization. If it is the first such agent to be assigned the disease, it generates a hypothesis at random. Otherwise each intelligent agent has the following agenda:

1. Query support database to determine previous progress made by other intelligent

agents.

2. Select a subset of the features to try.

3. Select a classifier type to use.

4. Select a set of parameters for that classifier.

5. Generate a hypothesis consisting of the feature set, the classifier, and the parameters.

6. Run the algorithm for this hypothesis, get the results, and update the support database with the results.

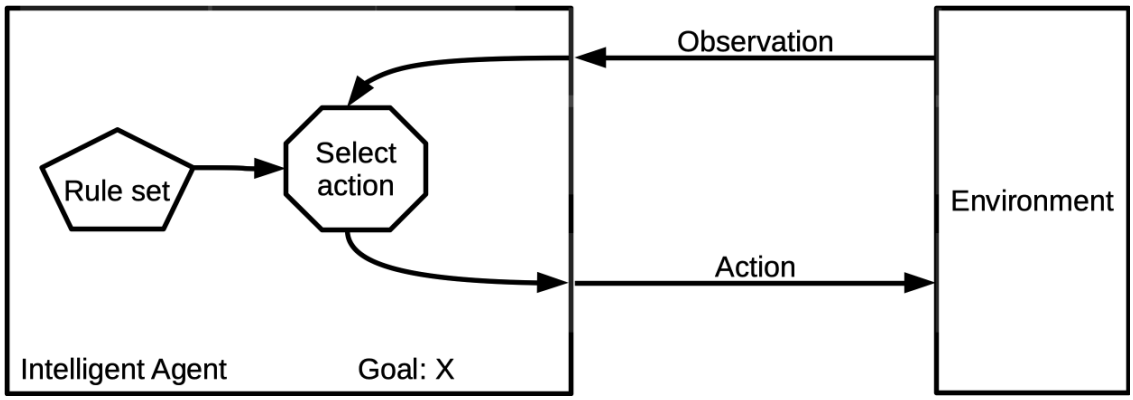


Figure 3.4: Simple Intelligent Agent.

The agents have two methods of exploring the "search space" (which is 10% of the MIMIC database set aside for model development). An agent can perform feature selection (visualized in Figure 3.7) where it may take three different approaches to evaluate novel selection criteria: creating *de novo*; editing an existing criterion; or, a genetic algorithm approach of splicing two existing criteria. An agent may alternatively perform hypothesis generation to learn on sub-populations. Hypothesis generation is accomplished by one of three methods: generating a hypothesis *de novo* (the process of which is depicted in Figure

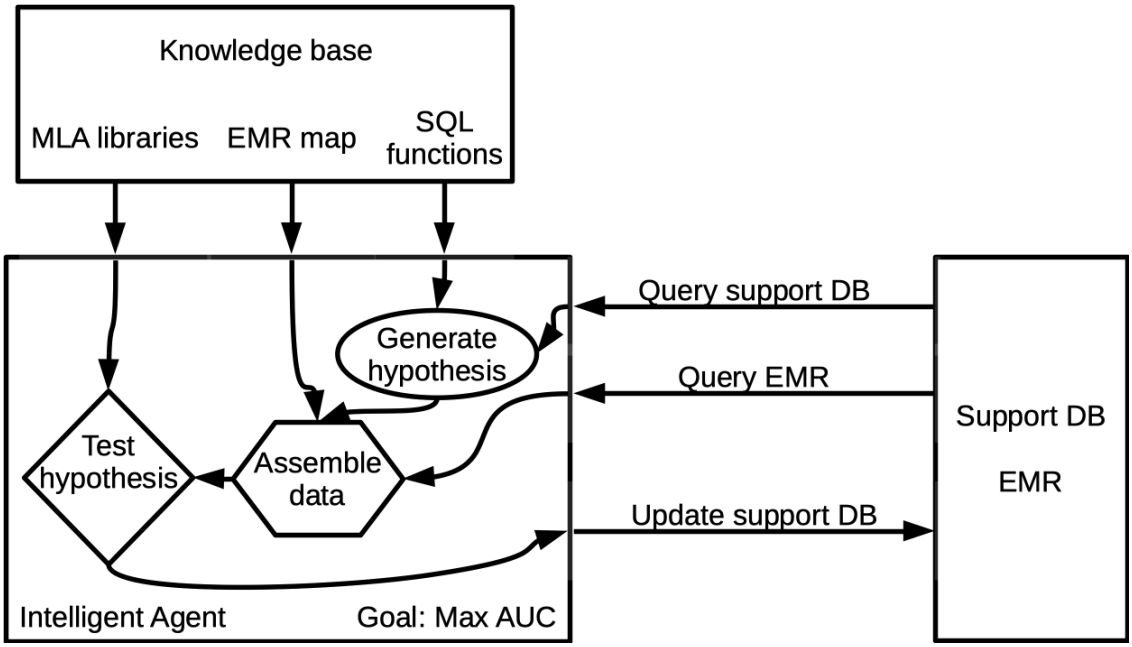


Figure 3.5: Intelligent Agent Framework Overview.

3.10); mutating an existing hypothesis (the attribute remains the same, a parameter is selected and altered at random); or constructing a new hypothesis using logical operators on existing hypotheses.

The agents evaluate performance of a criterion by implementing a machine learning algorithm experiment (10-fold cross validation). The performance metrics are stored in the support database (SDB) where future instances of agents will access the results to select attributes. The agent is then re-initialized.

The intelligent agents were designed in this dissertation to be ideal for massively parallel computing and data mining. It deviates from more traditional concepts of intelligent agents in two ways. First, an agent does not share data with another agent in real time. The only data sharing is through a retrospective record in the support database. Second, an agent additionally does not have a persistent and unique behavior. This study relies on the cumulative work done by all agents rather than creating a competitive environment that

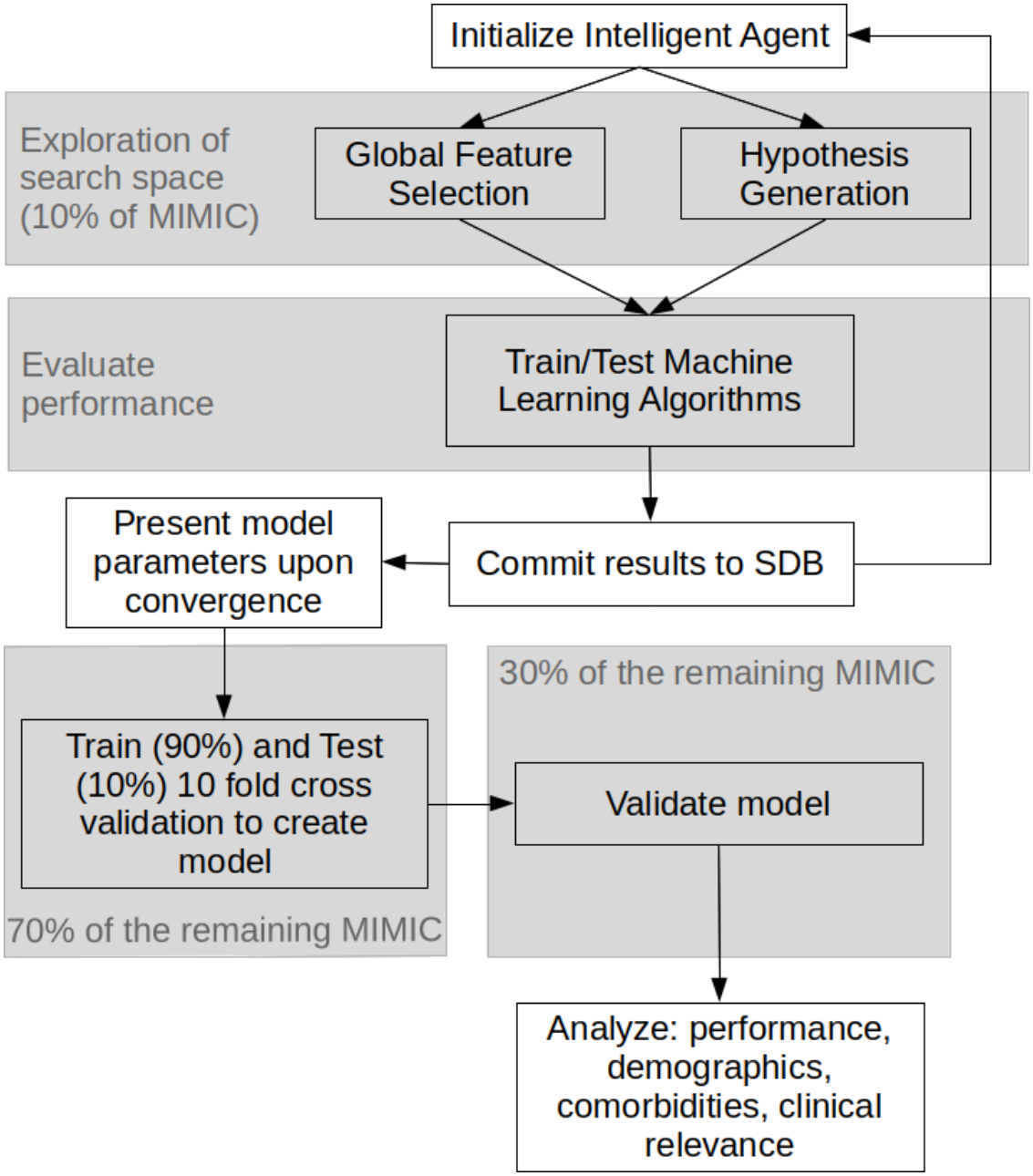


Figure 3.6: Intelligent Agent Process Overview.

rewards agents based on a predetermined disposition as many intelligent agent studies do. The only interactions agents do have is in consuming limited computational resources such

as: CPU, RAM, database access, and network bandwidth.

When the performance of a criterion does not significantly improve over a long duration of time, the agents have achieved convergence. The exploration of the search space and evaluation of performance is discontinued, and the best performing model parameters are locked. Of the remaining MIMIC patients not used in development, 70% are used to train and test to create a model. On this 70%, a ten-fold cross validation approach is used, with a 90% training data set and a 10% testing data set for each instance. The threshold cutoff of the model is set equal to the maximum F-measure where β=1.75 for studies drawn directly from the ICU and β=1.00 for studies drawn from a subset of the ICU population (such as predicting ARDS from DIC patients or predicting AKI Stage 2 from AKI Stage 1 patients). A β of 1.75 is chosen because it is deemed more clinically relevant to reduce false negatives at the expense of increasing false positives in the context of the ICU and in relation to the diseases of this dissertation. The model with the aforementioned threshold cutoff is then validated on the remaining 30% of the non-development MIMIC data set. The results are analyzed for performance, demographics, comorbidities, and clinical relevance.

3.10 Feature Space Learning

The intelligent agent's main tool is based on feature space learning. Agents are able to add or remove attributes from selection criteria one at a time. The framework takes advantage of a multiple-gradient descent approach. Gradient descent algorithms are greedy algorithms and subject to the local optimization trap. Multiple-gradient descent algorithms mitigate the risk of local optimization traps by employing the algorithm distributed over a vast search space. A measure of randomization and non-optimal criteria selection are built into this framework to foster criteria diversity and to escape potential local optimization traps.

When an intelligent agent creates a new criterion, it has three options. It can create a *de novo* criterion (initiate a criterion not derived from a previous instance), edit an existing criterion, or perform a genetic algorithm by randomly splicing together two existing criteria. Creating a *de novo* criterion requires selecting a new set of attributes. When editing an

existing criterion, the intelligent agents have two options. It can either add a new attribute to an existing criteria, or it can mutate an existing criterion. When adding to an existing criterion, the agent inserts identifiers, include linking the predecessor, and select a new attribute. When mutating an existing criterion, the agent needs to insert identifiers, include linking the predecessor, remove one attribute and add a new attribute. A genetic algorithm is the third option, where two high performing criteria are selected as the "parents". A random sampling of criterion from each are spliced together to create a new criterion. This process is depicted in Figure 3.7.

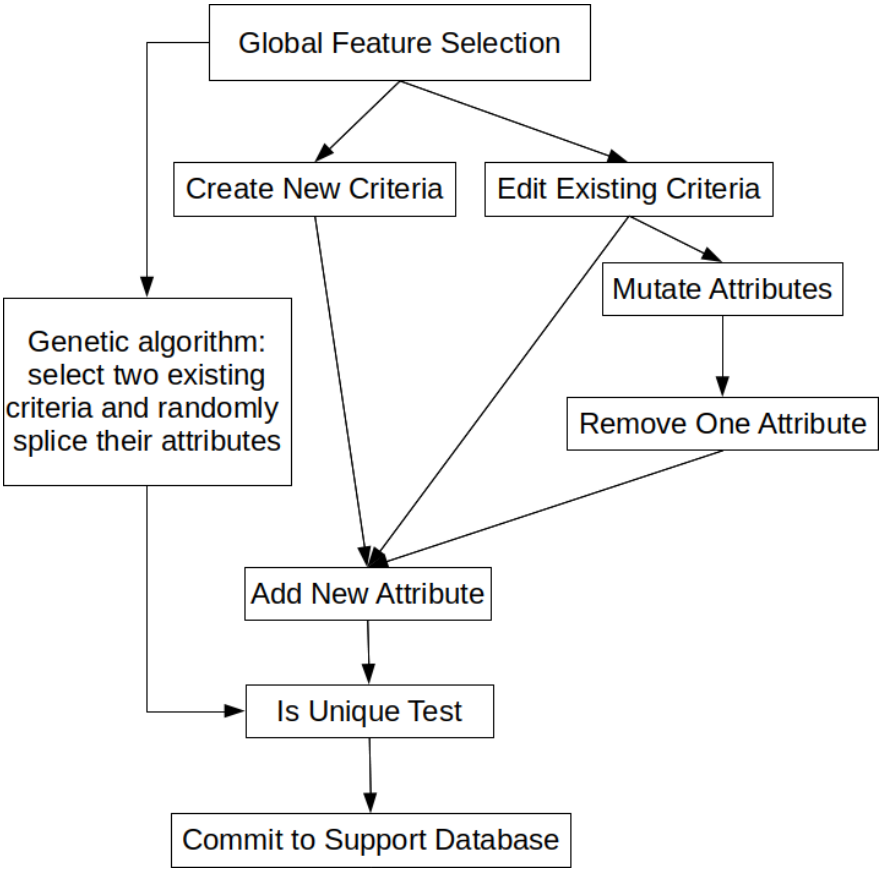


Figure 3.7: Feature Selection Diagram.

3.10.1 New vs. Edit

Many options have parameters, and those parameters are adaptable over time. The option to create a new criterion as opposed to editing an existing criteria is regulated by parameter α. The very first learning iteration is defaulted to create a new criteria, as there would be no existing criterion to edit. The second learning iteration is defaulted to edit one of the existing criteria. Every learning iteration is evaluated by an area under the curve of a receiver operating characteristic curve (denoted ROC). The probability that an intelligent agent will select the option to create a new criteria is equal to α. The parameter α is equal to the normalized average ROC of all learning iterations that were derived directly from a new criteria. That is, if the ROC of new criteria learning iterations was new_ROC, and the average ROC of existing criteria learning iterations was $existing\_ROC$, then $\alpha = new\_ROC/(new\_ROC+existing\_ROC)$. The probability that an agent will select the option to edit an existing criteria is then 1-α.

3.10.2 Edit Selection

If an intelligent agent selects the option to edit an existing criterion, the criterion is selected using a 90-10 10-90 rule. That is, the agents have a 90% probability of selecting the top 10% performers, and a 10% probability of selecting the remaining 90% performers. Performance is measured by the ROC. The 90-10 10-90 rule is in place to have preferential attraction towards high performing criteria while maintaining the diversity of criteria. This diversity is a compensation of the greedy algorithm drawbacks of gradient algorithms, mainly the trap of local optimization where there exists a higher performing global optimization.

3.10.3 Edit: New vs. Mutate vs. Splicing

Editing an existing criterion has three options. This framework maintains a 45% probability of adding a new attribute to an existing criteria, a 45% probability of mutating an attribute, and a 10% probability of randomly splicing with another criterion.

3.10.4 Adding and Removing Attribute Selection

This framework takes advantage of three methods to select an attribute to add (see Figure 3.8). A new attribute may be selected at random from the list of attributes in the knowledge base main table (subsection B.0.4) less the list of attributes already in the criteria (if editing an existing criteria). This option is available at all times, and maintains a constant diversity of attributes to escape the local optimization trap of greedy algorithms. The probability that an agent will select an attribute at random is 100% when there are no other methods available, 50% when there is only one other method available, and 33% when there are two other methods available.

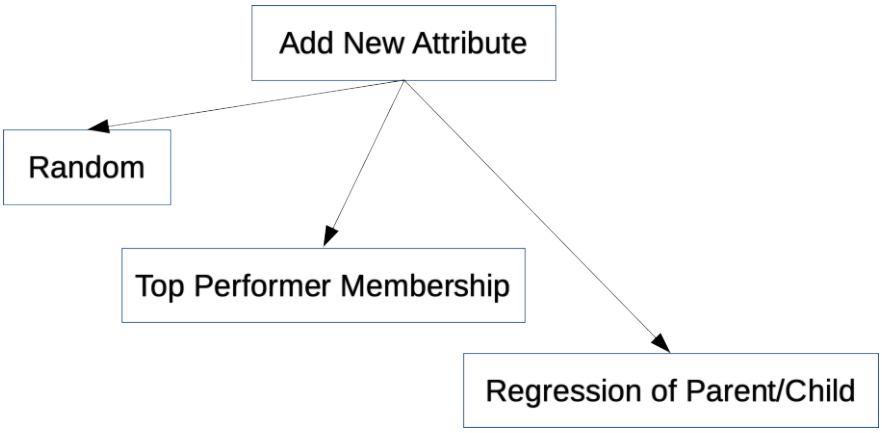


Figure 3.8: New Attribute Selection Diagram.

Top performer membership is a secondary method to add an attribute. Top performers are defined as a models with an ROC in the top 10% of all models. Membership is defined as an attribute belonging to a model. This option is available only after 10 learning iterations.

Regression of a parent's and child's ROC is a tertiary method to add an attribute. For every instance an attribute is added to an existing criteria, there exists a parent-child relationship. The difference between the parent's criteria and the child's criteria is the existence of the added attribute. It is therefor possible to trend the performance of adding a particular attribute to models across all the models the intelligent agents create. For every attribute

that has at least two parent-child relationship examples. The algorithm creates a set of two dimensional points where the x-axis represents the parent's ROC and the y-axis represents the respective child's ROC. The algorithm then perform a linear regression on the set of points. The slope of the regression line is a measure of performance for that attribute across multiple models. The top performers are defined as attributes with 10% of the greatest slopes. This option is available when there are 10 qualifying attributes, that is 10 attributes that belong to at least two parent-child relationships.

The method of mutating a criteria is a two-step. First the removal of an existing attribute, then the addition of an attribute. Adding an attribute has been described above. The method for removing an attribute (see Figure 3.9) has much the same framework. An attribute may be removed at random from the list of attributes associated with the criteria. This option is available at all times. The probability that an agent will select an attribute at random is 100% when there are no other methods available, 50% when there is only one other method available, and 33% when there are two other methods available.

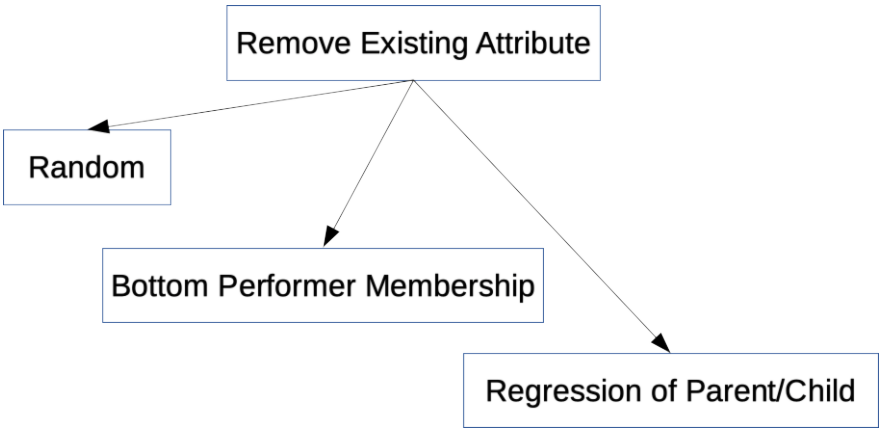


Figure 3.9: Remove Existing Selection Diagram.

Bottom performer membership is a secondary method to add an attribute. Bottom performers are defined as a model with an ROC in the bottom 10% of all models. Membership is defined as an attribute belonging to a model. This option is available only after 10 learning

iterations.

Regression of a parent's and child's ROC is a tertiary method to remove an existing attribute. Similar to the adding attribute method using an evaluation of regression of points comprising parent-child relationships, the algorithm performs a linear regression on the set of points that are the ROC metrics for parent and child. The slope of the regression line is a measure of performance for that attribute across multiple models. The bottom performers are defined as attributes with 10% of the least slopes. This option is available when there are 10 qualifying attributes, that is 10 attributes that belong to at least two parent-child relationships.

3.10.5 Test for Uniqueness

There are two main reasons for ensuring the same model is not evaluated repeatedly. The first reason is that each model run is a strain on computational resources. A sufficiently complicated model may take minutes or even hours to compile and evaluate on even high end computers. The second reason comes in the form of a repetition trap of optimization algorithms. An optimization algorithm may thrash between two equally performing points. For many applications of optimization, it is sufficient to write the algorithm to ensure it does not visit the same point it evaluated in the step preceding. There is always the possibility of being trapped between three points. Since this framework creates many complex queries, there is always the possibility of convergent evolution, where two unrelated models through a few steps align with the same set of attributes. The agents were designed to avoid such traps. The intelligent agents communicate their findings via the support database. To avoid traps and resource expenditure, every model that is proposed by an intelligent agent is checked against all previous models to ensure it is unique. If a model proves to be unique, it is committed to the database and written into the appropriate tables. If it is otherwise found to be a duplication of an intelligent agents' previous efforts, the model is discarded and the efforts to add an attribute begin again.

3.11 Sub-population Concept Hypothesis Learning and Testing

Machine learning algorithms applied to a narrow list of attributes in the feature space can be powerful for detecting patterns in patients before the onset of diseases. Built into the framework is the ability for the intelligent agents to perform sub-population data mining and analysis to enhance the power of these applications. Sub-population formation in this framework is centered around the instantiating of hypotheses through concepts. A concept is any collection of characteristics that separates a group of patients from the whole of the patient population of MIMIC. *A hypothesis is a collection of concepts.*

3.11.1 Anatomy of a Hypothesis

A hypothesis links to concepts, and the provenance is tracked by a parent-child relationship schema. The structure of a concept is such that it is sufficient to facilitate the writing of simple or complex SQL queries to identify a population sub-group (depicted in Figure 3.10). The concepts are uniquely identified. Each concept can answer a host of questions necessary for writing the SQL query. For example, *what is the attribute to evaluate? Does the predicate have an aggregate function or not? If the predicate has an aggregate function, what is it (Mean, Max, Min, etc.)? What operator is used in the predicate (greater than, less than, greater than or equal to, between, etc.)? How many values does the operator take (one for use in greater than the argument, or two for use in the argument is between two numbers, etc)? What is the value or values? Is the predicate relating exclusively to the internal concept surrounding the attribute, or are there external concepts to constrain this concept? If it an internal concept, does the aggregate function relate to part of the patient's stay or all of the patient's stay? If the internal concept relates to part of the patient's stay, what is the time window? If this is a concept with external concepts constraining it, what is the concept identifier of that external concept? What is the time window for that external concept? What is the delay between the end of the external constraining concept's time window and the current concept?*

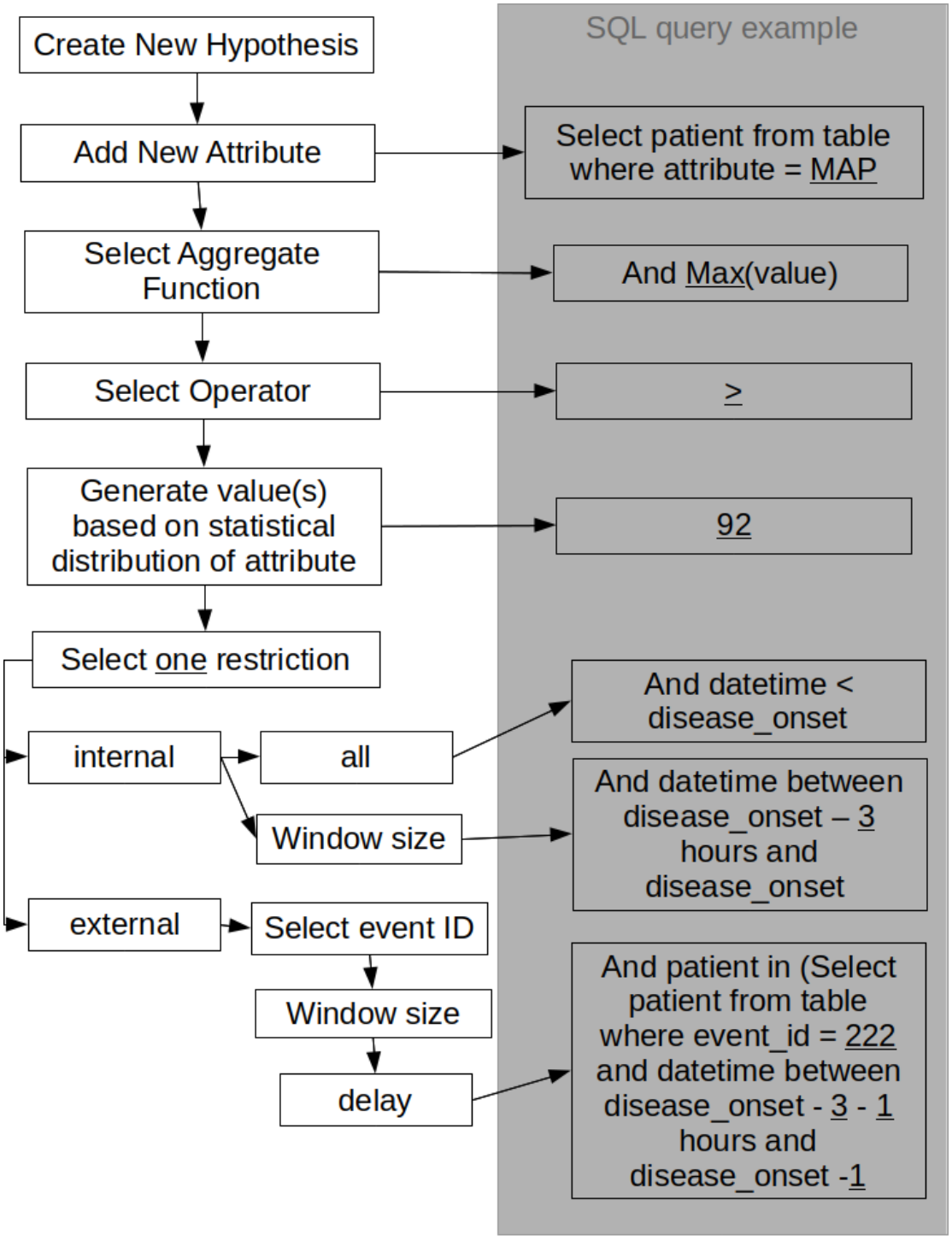


Figure 3.10: Creation of an hypothesis with an SQL query example.

The structure above outlines sufficient information for pursuing a vast array of medically relevant hypotheses and sub-population evaluations. By using the structure of a concept, the intelligent agents have written SQL queries as simple as: *create a list of all patients whose average mean arterial pressure is greater than 65 mmHg using their entire stay.* By using the combination of concepts technique built into the structure, the intelligent agents have written a complex concept SQL query: *create a list of all patients whose average mean arterial pressure was greater than 65 mmHg within two hours after a patient's heart rate was recorded at 90 beats per minute.*

A complex concept may also be created using logical operators on concepts. Each concept can create a list of patients. Logical operators would effectively perform logical operations on those lists. Group A, logical operator AND, and group B would produce the intersection of the two groups. Group A, logical operator OR, and group B would produce the union of the two groups. The framework is flexible enough that intelligent agents can create compounding unions and intersections to create a highly specified patient sub-population.

All hypotheses that result in a non-zero patient list are stored and may be used by all subsequent models created by intelligent agents. Patients who qualify in a hypotheses' sub-population may be incorporated into a feature space or evaluated against the disease as a stand-alone model.

3.12 Analysis

The focus of the intelligent agents is to find patterns that predict the onset of disease. There exists many different measurements of success for the purposes of predicting an outcome. The main metric used in this framework is the area under the curve of an ROC curve (for an example see Figure 4.3). ROC curves are published in the main body of this thesis, while the figures for other metrics are published in the appendices.

3.12.1 Receiver Operating Characteristic

The intelligent agents construct a model by using training data. When a model is complete, it takes as input testing data and outputs a resulting metric on how closely aligned the input testing data compares with the constructed model. This constitutes an order of confidence of the prediction. A table of test data entries is created in order of best-to-worst performers based on that produced resultant metric. Each test data entry also has a positive or negative data label that the model is trying to predict. The last step to finishing a model is to set the threshold, so that above the threshold the model considers entries to be positive and below the threshold the model considers entries to be negative. This presents a dichotomy where the model and the labelled data may not agree. A true positive (TP) means the entry is above the metric threshold for the model and the data label is positive. A true negative (TN) means the entry is below the metric threshold for the model and the data label is negative. A false positive (FP) occurs when the entry is above the metric threshold for the model but the data label is negative, also known as a type I error. A false negative (FN) occurs when the entry is below the metric threshold for the model but the data label is positive, also known as a type II error. Every row in the table represents a different confidence threshold, and for every confidence threshold a TP, TN, FP, and FN is calculated.

3.12.2 Receiver Operating Characteristic Curve

The true positive rate (TPR) is calculated as the number of true positives divided by the number of absolute positives (TP+FN). Likewise the false positive rate (FPR) is calculated as the number of false positives divided by the number of negatives (FP+TN). When these two metrics are plotted against each other, ordered by the ordered list, they constitute the ROC curve. The area under the curve (AUC) is derived by the numeric integration of the curve. That is:

$$AUC = \sum_{n=1}^{m} \frac{f(x_n) + f(x_{n+1})}{2} * (x_{n+1} - x_n) \tag{3.1}$$

Where ever possible, and every time this study generates the AUC, the AUC will be

accompanied by a Confidence Interval (CI) of 95% represented in parentheses immediately following the AUC as (CI lower bound, CI upper bound). For plotting AUCs over time the CI upper bound will be indicated as ROC_ub, and CI lower bound will be indicated as ROC_lb.

The CI is calculated using the following equations:

$$q_0 = AUC * (1 - AUC) \tag{3.2}$$

$$q_1 = \frac{AUC}{2 - AUC} - AUC^2 \tag{3.3}$$

$$q_2 = \frac{2 * AUC^2}{(1 + AUC)} - AUC^2 \tag{3.4}$$

$$se = \sqrt{\frac{q_0 + (N_1 - 1)q_2 + (N_2 - 1)q_2}{N_1 N_2}} \tag{3.5}$$

$$CI\_lower\_bound = AUC - z\_crit * se \tag{3.6}$$

$$CI\_upper\_bound = AUC + z\_crit * se \tag{3.7}$$

Variable definitions:

N_1 = number of positives

N_2 = number of negatives

AUC = area under the curve

z_crit = the two-tailed critical value of the standard normal distribution where alpha = 0.05 (CI = 1 - alpha) and has a value of 1.96

3.12.3 True and False Positive Rates

The true and false positive rates may be plotted in the context of the confidence threshold (for an example see Figure D.1). The confidence threshold axis is created by normalizing the ordered list outputted by the model.

3.12.4 Sensitivity and Specificity

Sensitivity and specificity are common metrics to evaluate models. Sensitivity is the true positive rate, TP/(TP+FN). Specificity is the true negative rate, TN/(TN+FP) (for an example see Figure D.2).

3.12.5 Positive and Negative Predictive Value

Positive predictive value and negative predictive value are measures of the respective predictive capability of the model. A positive predictive value (precision) is the TP/(TP+FP). It is a measure of how often a model is correct when it classifies an entry as positive. Likewise the negative predictive value, TN/(TN+FN), measures how often a model is correct when it classifies an entry as negative (for an example see Figure D.3).

3.12.6 Accuracy

Accuracy is a quantification of model performance in the presence of type I and type II errors. Accuracy is defined as (TP+TN)/(TP+TN+FP+FN) (for an example see Figure D.4).

3.12.7 Cohen's kappa coefficient

Cohen's kappa coefficient (κ), in the context of binary classification, is a measure of agreement between the model and the label of the data. It takes into account the observed agreement (p_o) and the probability of random agreement Random agreement is defined as both random positive agreement (p_p) and random negative agreement (p_n) (for an example see Figure D.5).

$$p_o = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.8}$$

$$p_p = \frac{TP + FP}{TP + TN + FP + FN} * \frac{TP + FN}{TP + TN + FP + FN} \tag{3.9}$$

$$p_n = \frac{TN + FN}{TP + TN + FP + FN} * \frac{TN + FP}{TP + TN + FP + FN} \tag{3.10}$$

$$p_r = p_p + p_n \tag{3.11}$$

$$\kappa = \frac{p_o - p_r}{1 - p_r} \tag{3.12}$$

3.12.8 F-Measures

F-measure is the harmonic mean of precision (positive predictive value) and recall (sensitivity). Harmonic means are appropriate when averaging ratios, and are the reciprocal of arithmetic means. Computational models when applied to a patient population will have false positives (patients without the disease that are classified as having the disease) and false negatives (patients with a disease that are classified as not having the disease). The practical implications of these two misclassifications are different and depend highly on the clinical context where the model is employed.

If a model is used that has an endemic high false positive rate (type I errors), there are two prominent negative consequences. Harm may be done to a patient for the treatment of a disease the patient doesn't have by medication or medical procedures that exposes the patient to risks. Harm may also be done to the patient while the care providers are confirming the model's findings by exposing the patient to unnecessary imaging or other diagnostic tests, which would otherwise not have been preformed. In addition to patient harm, the care providers may suffer alert fatigue for a system that too often misidentify patients.

If a model is used that has an endemic high false negative rate (type II errors), the greatest harm is that of the patient with a disease that goes untreated or delayed treatment. Early detection of many emergent diseases can lead to vastly improved patient outcomes.

F-measure is a metric that balances the trade-offs of type I and type II errors. The most common F-measure parameter is where β=1.00. A $\beta = 1.00$ evenly balances the precision and recall trade-off. Type I errors and type II errors incur different costs in the medical setting (as well as many other settings), and it is therefor prudent to evaluate different F-measure settings. The confidence threshold for models in this framework are set at the

maximum value of the F-measure where β=1.75. The β was chosen so that the clinically relevant emphasis was put on type I and type II errors. In the context of an ICU with the diseases modeled in this study, a greater emphasis is put on minimizing the number of patients misclassified as false negative for the diseases. This is a subjective decision as, at the time of this publication, there is insufficient literature supporting an evidence-based approach for balancing type I and type II errors in computer clinical decision systems in an ICU setting.

$$F_\beta = \frac{(1+\beta^2) * TP}{(1+\beta^2) * TP + \beta^2 * FN + FP} \tag{3.13}$$

As evident in Figure D.6, as the β increases, the curve skews towards a lower confidence threshold. This moves the maximum value for each curve also to a lower threshold. The result of a greater β is a lower type II error (condition positive patients predicted, falsely, as negative for the condition). Expectantly, the true positives increase and the false negatives decrease at the expense of true negatives decreasing and false positives increasing. This translates into raising the sensitivity at the expense of lowering the specificity and lowering the accuracy.

3.12.9 As a function of Time

The machine learning algorithms are evaluated over a course of time. T-minus is defined as the number of minutes before disease onset. The onset of disease occurs at t-minus=0. The data collected per patient at t-minus=0 represents the patient data just prior to the disease onset. Data is collected at fifteen minute intervals for the six hours (360 minutes) prior to the onset of disease. For the diseases in this study six hours represents a significant early warning to be clinically actionable. When evaluating a criteria selected by intelligent agents, the machine learning algorithms perform a classification at every time point between onset of disease and six hours prior at the fifteen minute intervals. The AUC for the ROC is calculated at each time point. The resulting data (as seen in Figure D.13) is used to evaluate each

machine learning algorithm on the selected criteria. Random Forests outperform all other machine learning algorithms evaluated both near the onset of disease and six hours prior. The various machine learning algorithm performance on the ARDS model at t-minus=0 may be found in Table 3.2. Random Forests continued to outperform other machine learning algorithms throughout this research.

Machine Learning Algorithm	Area Under Curve
NaiveBayes	0.82
Logistic	0.86
MultilayerPerceptron	0.82
SimpleLogistic	0.86
SMO	0.79
VotedPerceptron	0.70
Kstar	0.70
DecisionTable	0.85
JRip	0.82
OneR	0.73
PART	0.77
DecisionStump	0.70
J48	0.76
LMT	0.87
RandomForest	0.88
RandomTree	0.68
REPTree	0.82

Table 3.2: Algorithm's performance on the ARDS selection criteria at t-minus=0.

3.12.10 Selection Criteria

The selection criteria is outputted to human readable form (see Table 4.2). The criteria is accompanied by a variable importance metric, which is a natural derivative from the Random Forest algorithm, and an AUC calculation for the model excluding that variable.

3.12.11 Competing Models

Competing models are selected from the available literature on the specific disease. This study follows two approaches for model comparison: The intelligent agent derived model is compared to published metrics of similar models; additionally, wherever possible, published models are implemented on the MIMIC database for comparison.

In the particular case of ARDS, the Lung Injury Prediction Score (LIPS) is widely used for research purposes and is fully computable in MIMIC. The LIPS criteria are found in Table 4.1. A published account of LIPS applied to an ICU setting predicts ARDS with an AUC of 0.79. For a comparison study, LIPS was implemented on the MIMIC population predicting ARDS (AUC=0.80). Though LIPS is a valid contender for a competing model, for many diseases there exists no qualifying published model for predicting the onset of the disease. LIPS itself is most commonly applied retrospectively, for research purposes, after the patient has been diagnosed and treated for the disease.

3.12.12 Model Comparison

An example of model comparisons may be found in Table 4.8. The best model constructed by the intelligent agents in this research is compared to various published metrics and re-implemented using the MIMIC data. The emphasis of this dissertation is to minimize the false negative rate (by choosing an F-measure β=1.75 for most models). For compound models or models for which the ICU population is not the base, the model's cutoff threshold is set by choosing an F-measure β=1.00. Compound models are constructed in both the ARDS and AKI chapters.

The LIPS publication uses a point system and has prescribed a cutoff value, which means the authors have already chosen an appropriate threshold. The model the intelligent agents present maintain a confidence threshold consistent with the maximum F-measure where β=1.75. The AUC of a model does not alter with the choice of a threshold, and LIPS AUCs are consistent with both publications and the re-implementation. Kappa is also consistent between the models, but the sensitivity, specificity, and positive predictive value vary greatly from published to re-implemented. The context of an ICU may be the driving force behind the discrepancy observed.

3.12.13 Variable, Demographics, and Comorbidity Analysis

This dissertation analyzes the diseases in the context of each cohort's variable values, demographic distributions, and comorbidities.

For the variables the model uses, a table of disease positive and disease negative cohorts is constructed. The first, second (median), and third quartile of variable values is computed accompanied by a P-value for statistical significance (Kruskal-Wallis test). The intelligent agents select the variable because it aides in the prediction of disease in the context of other variables. The computed variable distributions for each cohort allow for the analysis of whole cohort significance of each variable (for an example see Table D.1).

A demographics table is constructed to place the disease population in the context of the remaining ICU population. For each demographic item either a distribution is generated or a percentile is calculated accompanied by a P-value (either Kruskal-Wallis or Pearson's Chi-Squared test). For the diseases inspected by this dissertation, the in-hospital mortality and 30 day mortality rates are the most prominent features of the demographics table (for an example see Table D.5).

A comorbidity is a co-occurring disease. Elixhauser et al. [16] clustered ICD-9-CM codes for diseases under classifications that had an impact on patient outcome. The Elixhauser comorbidities are calculated for both disease positive and disease negative cohorts. This study utilizes the Enhanced ICD-9-CM of Elixhauser comorbidities as described by Quan, et al.

[72]. This study augments the Elixhauser comorbidity list with a trauma designation. ICD-9 coded trauma is a significant insult to the human body and has a wide range of both subtle and pronounced outcomes. This study uses the Washington State Department of Health's "Washington State Trauma Registry Users Guide" [66]. This study augmented the trauma registry to include the trauma of frostbite, as it was excluded from the publication. The P-value (Pearson's Chi-Squared test) accompanies the patient percentiles. Examining the diseased patients' comorbidities in context of other ICU patients may help explain patterns in variable selection (for an example see Table D.5). Most significantly, the Elixhauser comorbidity calculations allowed this dissertation to identify the significant link between ARDS and DIC.

For each table described above (variables, demographics, and comorbidities) an accompanying table is generated for true positive, true negative, false positive, and false negative predicted patients. This measure is taken to identify the subsets of patients the model misclassifies and further understand the limitations and biases of the model. Misclassification analysis is important for clinicians who use predictive methods. Using the misclassification analysis, in a personalized medicine approach, a clinician may assess the predictive performance on individual patients. Model performance described in other sections of this dissertation are the result of applying the model to all valid patients. A clinician who has a one-on-one responsibility to a patient may use the misclassification tables to assess model performance on sub-populations relevant to the individual patient under their care.

3.13 Surrogate Variables

The intelligent agent models are a result of a data mining exploration of the EMR. Variables the intelligent agents select are grounded in evidence. The clinical interpretation however may be different than the literal variable. It will be consider it to be a surrogate variable when a selected variable has a different clinical interpretation.

Within this dissertation two prominent surrogate variables are found. GT flush, or gastronomy tube flush, likely does not mean the flushing of the gastronomy tube is significant

to the onset of disease. Rather the existence of a gastronomy tube, indicative of intubated patients, is a significant discriminator between those patients who will succumb to the disease and those who will not. Gastronomy tube flush is a significant variable in Tables 5.4, E.7, E.8, 5.7, E.13, E.14, 6.2, F.1, and F.2. Arterial line zero/calibrate does not mean the calibration of an arterial line is significant to the onset of sepsis (Tables 6.2, F.1, and F.2), but that the clinician's concern of of the patient has led to high resolution hemodynamic monitoring. The clinician's concern is correlated with the onset of sepsis.

An EMR has many variables that represent truth in differing degrees of accuracy. Some of the variables are close to a truthful interpretation, like a heart rate of 68 beats per minute. Some variables are a surrogate variable for the truthful interpretation, like arterial line calibration implies the placement of the arterial line and the clinician's considerations thereof.

3.14 Biases and Limitations

3.14.1 Design Bias

The intelligent agents are designed to data mine the EMR based on previous work of other agents. The experiments that had early initially good performance are weighted more significantly than lesser performance or randomly generated. Though this is by design to evolve a model to a better performance, it has similar pitfalls of a greedy algorithm. A greedy algorithm may find a local optimization and be unable to find a global optimization. Several measures were built in to avoid local optimization (introduction of random models, random variations of good performing models, and accepting a breadth of good performing models as opposed to just the best performer). Because the intelligent agents have preferential attachment towards good performers, and they operate in a resource-restricted competitive environment, it is unavoidable that they have a manifestation of anchoring bias (heavily preferring initial information).

As aforementioned, the intelligent agents operate in a resource-restricted competitive

environment. A single agent instance does not compete, as it would have full access to the database, CPU, and RAM. The support database would only update experimental results when the single agents completes one. A single agent approach would render the vast data mining required for this research an intractable problem. The agents are designed to operate in parallel with many instances running simultaneously. Plurality necessitates competition as resources such as RAM, CPU, bandwidth, and database access become diminished. The faster an agent updates progress on an experiment, the faster other agents may improve upon it. The resource-restriction sets up the competitive environment for agents to process experiments quickly. The unavoidable competition for resources sets a condition where faster models are iterated upon more than slower models. The manifestation in this study is by the preferential attachment to Random Forest machine learning algorithms (as they are among the faster of the non-linear ensemble algorithms), and would likely have preferential attachment towards variables that produced quick good performing models.

3.14.2 Coder Bias

The diagnostic criteria for both ARDS and sepsis depend on an ICD-9-CM code being assigned to the patient. During the catchment time frame for MIMIC the coding practices changed (as they change often). Coders are under a fair degree of time pressure, have varying degrees of experience, and have differing biases based on the patient records they receive. These features cumulatively make up a coder bias, where documentation is non-homogeneous throughout the institution and over time.

3.14.3 Caregivers

Caregivers have an enormous impact on data collected, diagnoses, procedures performed, and patient outcomes. It is impossible to observe and account for all biases in a de-identified retrospective database. There are 1,392 attending physicians and 6,175 other caregivers responsible for treatment and care of the 10+ year MIMIC database. Saposnik et al. [77] characterizes 19 biases related to medical decision making: multiple alternative/decoy bias;

outcome bias; information bias; risk aversion; ambiguity tolerance; overconfidence; omission bias; naturalness bias; reflexive reasoning; deliberation without attention; framing effect; anchoring; availability bias; premature closure; feedback bias; confirmation bias; blind obedience; Gambler's and Conjunction fallacy [61]. With 200+ medical schools in the United States, education and experience is not uniform and necessarily affects medical decision making. Caregivers also employ sampling bias, where sicker patients will be more likely to receive an ICD-9-CM diagnosis than marginal patients, even for the same disease.

3.14.4 Data Abstraction

This study utilizes an EMR that constitutes a data abstraction of ICU patients. Biological life is an extremely complicated system or, more accurately, an ensemble of extremely complicated systems interacting. Diseases manifest at the same level of complexity. The measurements and observations that become part of the patient's electronic medical record cannot capture the whole of the state of the patient's biology. Based on education and experience clinicians may create mental models of a patient's disorder and apply therapeutics to regulate the disorder. Though MIMIC is an impressive collection of data, there are some nuanced and significant events that do not become a part of the database. This study is therefor limited by the design and implementation of the collection of data thought to be a significant part of an EMR.

3.14.5 Data

EMRs are typically plagued by measurement bias where, not at random, incorrectly: measured; read; or entered observations become part of the record. It is unlikely that a patient is transferred to the ICU with a temperature of 9 degrees Fahrenheit for example. Erroneous data is so prevalent in MIMIC that this dissertation does not include minimum or maximum values for variables in tables, as those values are most often data errors.

Most data in MIMIC is directly from the ICU, with the notable exception of laboratory results and ICD-9-CM coding which represent whole-hospital stays. All predictive models

created for this dissertation should be considered only in the context of ICU patients. ICU patients are sicker than most hospitalized patients as they warrant intensive care. When comparing the intelligent agent's models to other published works, this research goes to great lengths ensure they are set in the ICU or to represent those models in an ICU context by re-implementing them in MIMIC. The predictive models disambiguate the signal from the noise, where the origin of the signal are patients who will become afflicted with a disease of interest and the origin of the noise is all other sick patients in the ICU. It is therefor important to interpret the intelligent agent's predictive models in the context of an ICU and not generally hospitalized patients. The work in this dissertation may be re-implemented on a hospital-wide EMR to target the hospitalized patient population.

MIMIC is derived from the ICUs of Beth Israel Deaconess Medical Center (BIDMC), which is a non-profit teaching hospital of Harvard University in Boston Massachusetts with a Level I Trauma Center enabled emergency department. Not only are all predictive models in this dissertation set in the context of the ICU, they are only in the context of BIDMC's ICUs. To assert the intelligent agents' predictive models may be generalized to other ICUs, one would first have to validate the models in ICUs different from the ones available at BIDMC.

The design of an EMR introduces a discontinuity in reporting. Diagnoses of patients may be interpreted from the ICD-9-CM codes for the entire hospital stay, but not at the resolution to disambiguate the ICU from the rest of the hospital. A diagnosis of a disease or suspicion of a disease may be extracted from free text notes. Free text notes are inaccessible to the intelligent agents of this study. The intelligent agent framework does not include the complex and computationally expensive natural language processing. Radiographic and other imaging findings in this EMR are only found in free text form. The discontinuity blinds the intelligent agents to some data (like medical imaging analyses) and requires inference with the possibility of error for other data elements (like the onset of ARDS in the ICU based on hypoxemia and an ICD-9-CM code). The EMR was designed for tracking and compiling a patient's record. Research is a secondary use, and is subject to the limitations introduced

by the EMR design.

The process of handling missing data using nearest neighbors is unique among all peer-reviewed published predictive models cited in this dissertation. All other cited predictive models exclude patient's entirely if the patient does not have a value for a considered variable. By using a nearest neighbor approach, a patient with missing data inherits the value of the next nearest patient. The nearness of patients is computed from the entire six hour history of all variables considered. Excluding patients entirely based on missing data introduces a selection bias that over weighs patients with more data. Sick patients often have more data, and by the exclusion of missing data patients a model is ignoring the more marginally sick and decreasing diversity of the training, testing, and validation sets. Though a nearest neighbor approach to missing data preserves some diversity of patients and greatly mitigates the associated bias, the approach inevitably introduces over representation of some values.

A common bias that cannot exist in models devised by the intelligent agents is confirmation bias. The intelligent agents have no external influences of the diseases they are inspecting other than identification of the subset of patients who will become afflicted with the disease. The intelligent agents do not have access to the results of models for comparison, and strictly devise *de novo* models free of confirmation bias.

3.15 Comparable approach

Lee et al. [10] examines and predicts the risk of acute disease onset using the MIMIC database along with a University of Washington equivalent data set. The author uses ICD-9 discharge diagnoses to identify disease positive patients for acute heart failure, acute lung injury, acute kidney injury, and acute liver failure. Utilizing outcome-related interventions, they are able to predict the need for such interventions among the ICU patients.

Lee's study and this dissertation have similar goals. Where this study aims to predict the onset of disease, Lee's study aims to predict the need for an intervention. Both goals require a time-sensitive prediction of ICU patients afflicted with a particular disease.

The Lee study and this dissertation differ in significant implementation. Lee uses ICD-9

discharge diagnosis codes to identify disease positive patients. The diseases examined in this dissertation are often missed or are sub-symptomatic; as a consequence ICD-9 codes are used sparingly in favor of computable definitions. Sepsis and ARDS are the exceptions because select criteria of those definitions (specifically, the suspicion of infection requirement for sepsis, and the interpretation of radiologists for ARDS) are difficult to (and therefor unreliable to) compute. The time Lee seeks tp predict is the intervention, while this dissertation predicts the time when the patient achieves the criteria for the computable definition. Where Lee optimize the algorithm on the AUPRC (area under the precision-recall curve), this dissertation uses the AUROC (area under the receiver operating characteristic curve) metric. AUPRC is more suited for alerts, where the precision can be kept high and minimize alert fatigue through false positives. The ROC curve is suited for a screening algorithm and can be used to emphasize minimizing false negatives. The Lee study uses the Gradient Boosted Tree as the model building algorithm, where this dissertation allows agents a multitude of algorithms for which to build models (though Random Forests were the preferred algorithm of choice made by the agents).

Lee's approach performs remarkably well, especially for acute lung injury (precision of 0.62 and recall of 0.74) and acute kidney injury (precision of 0.41 and recall of 0.51). The closest comparable results from this dissertation are for a lesser performance of ARDS (precision of 0.11 and recall of 0.56) and a greater performance for AKI (precision of 0.48 and recall of 0.87). However, it is important to note that the Lee study and this dissertation use different definitions and different methods, and they seek different goals.

3.16 Summary

The chapter described in detail databases, methods, tools, assumptions, biases and limitations of implementing an intelligent agent framework to data mine an EMR. It also described how findings are reported and interpreted.

Chapter 4

ACUTE RESPIRATORY DISTRESS SYNDROME

4.1 Introduction

First described in 1967 by Ashbaug et al. [1] as a respiratory distress that did not respond to usual and ordinary methods of respiratory therapy. ARDS currently affects approximately 200,000 patients per year in the United States, resulting in 75,000 patient deaths [20]. Globally ARDS represents 10% of all ICU admits and is estimated to afflict 3 million patients per year.

This chapter predict the onset of ARDS as illustrated in Figure 4.1. The model predicts the onset of ARDS from the general ICU patient population (with an AUC = 0.861 (0.838, 0.884)). This chapter also predicts the onset of SAHRF, a necessary condition for ARDS. The onset of SAHRF is predicted from the general ICU population (with an AUC = 0.952 (0.947, 0.957)).

This chapter makes three significant contributions to disease prediction and disease modeling: ARDS prediction, severe acute hypoxemic respiratory failure (SAHRF); a link between ARDS and Disseminated Intravascular Coagulation (DIC).

First, the ARDS model outperforms the existing standard model with higher AUC, specificity, and accuracy. Second, there does not exist a peer-reviewed published prediction model for SAHRF. This chapter introduces one based on common clinical variables. Third, by modeling two different components of ARDS this study was able to isolate a comorbidity, DIC, that is significant in the ARDS model, not significant in the SAHRF model, and significant in the ARDS patients predicted by the SAHRF model. The pattern of significance implies a link in the evolution and progression of the two diseases. The ARDS-DIC link is not well represented in the peer-reviewed literature. This chapter proposes a theoretical disease path-

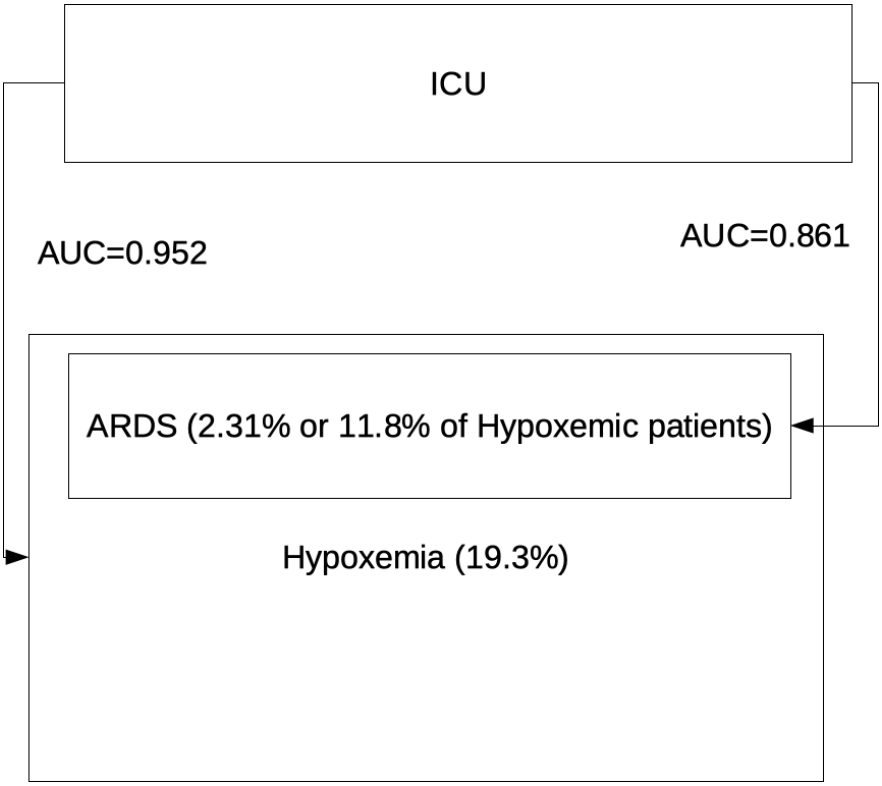


Figure 4.1: Two paths to predict the onset of ARDS.

way where the ARDS pathway merges with the DIC pathway to create a positive feedback loop, adding to the progression of both diseases. The ramifications of the pathway loop may have a broad impact in disease modeling and therapeutic strategies.

4.2 Pathogenesis

Healthy lungs are compliant (easily expandable) able to keep aveoli sacs open and relatively dry to allow gas exchange. There are two paths to ARDS from the initial insult to lung tissue. The damage can be of the alveolar-capillary membrane damage which may lead to a decrease in oxygen permeability in the tissue and can precipitate a collapse of the alveolar (possibly decreasing surfactant production needed to keep alveolar open). The initial insult may otherwise cause cellular injury which may inhibit surfactant production (causing alveolar

collapse). Both paths may culminate in partial or whole lung collapse (atelectasis), decrease in lung volume, or non-compliant tissue as described by Koh [41].

Generally, injured lung tissue initiates pro-inflammatory milieu and can cause excess fluid in the aveoli sacs. Rigidity in the tissue is caused by diffuse aveolar damage, which naturally progresses to irreversible fibrosis (lung tissue scaring). The excess fluid can cause the aveoli sac to collapse or otherwise impair gas exchange. The non-compliance of the tissue can cause the lung's tidal volume (the inhalation and exhalation volume) to decrease. Pulmonary dead-space, as described by Nuckton et al. [65], is then increased and is clinically observed as a ventilation and oxygenation mismatch.

4.3 Value in Early Detection

Early detection of the onset of ARDS would allow for preventative treatment or management treatment to avoid a worsening condition arising for that patient's outlook, as described by Gajic et al. [25]. The primary focus for early action is for the ARDSnet Mechanical Ventilation Protocol, as described by Kallet et al. [38], which prioritizes lung protective measures by reducing the tidal volume and thereby preventing lung hyperventilation, which may cause aveaolar rupture. Early pronation of the patient (turn patient to lay on his or her front, reduces the weight of other organs in the abdomen on the lungs) with the aim to decrease the dead-space, if possible, avoiding a packed red blood cell transfusion, which has been linked to development of ARDS and a high mortality, as described by Gong et al. [30].

The value in the early detection of SAHRF is in both treatment and diagnostic benefit. A patient who is afflicted with SAHRF is treated by the increase of oxygen support, either changing the ventilator settings or putting the patient on Bilevel Positive Airway Pressure (BiPAP). The causal source of SAHRF is typically: an infection; pulmonary edema (both cardiogenic and non-cardiogenic); transfusion complications (transfusion-associated circulatory overload (TACO) or transfusion-related acute lung injury (TRALI)); chronic obstructive pulmonary disease (COPD) (both normal and exacerbated); drug complications (which lead to pneumonitis - inflammation of lung tissue). A predicted positive patient may be asymp-

tomatic of these underlying diseases, and an early detection may allow a clinician to diagnose and treat the underlying cause before the onset of further complications.

4.4 Diagnostic Criteria and Implementation

ARDS has had multiple iterations of definition; the international community has most recently settled on the Berlin definition, as described in the Journal of American Medical Association article [74] in 2012. The Berlin definition requires ARDS to be acute (symptoms appearing within 7 days of a clinical insult or worsening respiration), have respiratory failure not explained by fluid overload or cardiac failure, and have imaging evidence (in the form of bilateral opacities not fully explained by lobar collapse, lung collapse, pleural effusions, or pulmonary nodules). The oxygenation requirement for respiratory failure is having a PaO_2/FiO_2 less than or equal to 300 mmHg on a ventilator (PEEP or CPAP) greater than or equal to 5 cm H_2O.

This study considers a hypoxemia patient one whose measurement of the PaO_2/FiO_2 ratio falls below 300 mmHg. The time point for the onset of hypoxemia is the time point of the measured ratio. An ARDS positive patients is considered by having an ARDS ICD9 code (518.5, 518.51, 518.52, 518.53, or 518.52). The time point for the onset of ARDS is the first measurement of the PaO_2/FiO_2 ratio to fall below 300 mmHg for ARDS positive patients. Patients with an ARDS ICD9 code whose PaO_2/FiO_2 ratio did not fall below 300 mmHg in the ICU did not experience the onset of ARDS in the ICU and are considered ARDS negative patients. Patients whose PaO_2/FiO_2 ratio was below 300 mmHg within the first six hours of their ICU stay were excluded, as the purpose of this study is to identify the onset of disease and not differentiate patients who have already contracted it.

4.5 Competing Model for ARDS

The Lung Injury Prediction Score (LIPS) [25] is the most widely used lung injury prediction method. The authors compiled known risk factors for lung injury 5,584 patients (377 positive for acute lung injury or acute respiratory distress syndrome, and 5,207 negative for lung

injury). The authors proceeded to preform a logistic regression of the risk factors, which included predisposing conditions and risk modifiers. From the logistic regression the authors created a scoring system, measured the performance of that scoring system, and assigned a cutoff "based on the AUC analysis." The LIPS scoring method has a 0.80 AUC in the seminal publication. Applying their established cut point their model produces a positive predictive value of 0.18 and a negative predictive value of 0.97. LIPS was evaluated in an ICU setting and had a performance of 0.79 AUC. This study applied the LIPS scoring method to MIMIC patients in the validation set and found an AUC of 0.847 (0.823, 0.871) (see Figure 4.2).

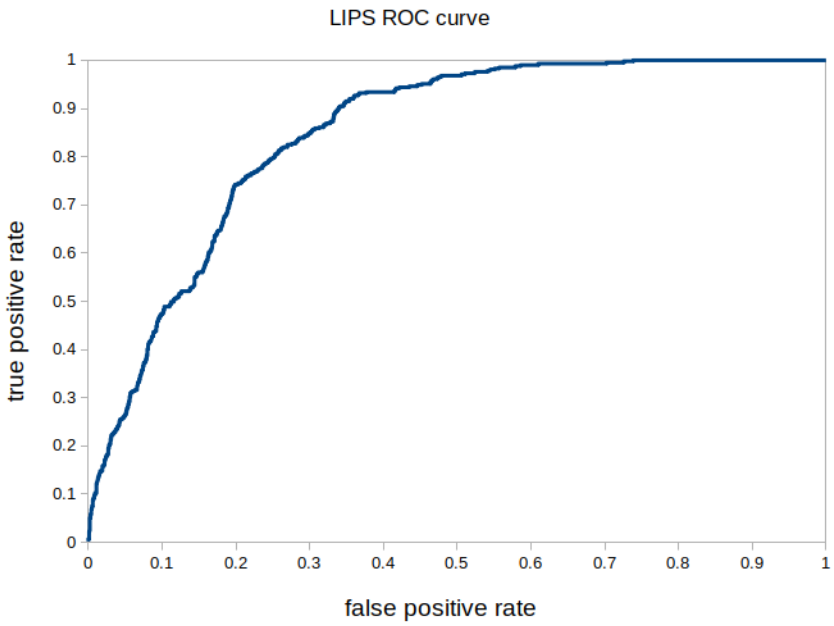


Figure 4.2: LIPS applied to the validation set, ROC curve AUC=0.847 (0.823, 0.871).

Variable	Points
Shock	2
Aspiration	2
Sepsis	1
Pneumonia	1.5
Orthopedic spine surgery	1
Acute abdomen surgery	2
Cardiac surgery	2.5
Aortic vascular surgery	3.5
Traumatic brain injury	2
Smoke inhalation trauma	2
Near drowning trauma	2
Lung contusion	1.5
Multiple fractures	1.5
Alcohol abuse	1
Obesity (BMI>30)	1
Hypoabluminemia	1
Chemotherapy	1
FiO_2 >0.35 (>4 L/min)	2
Tachypnea (RR>30)	1.5
SpO_2 <95%	1
Acidosis (pH <7.35)	1.5
Diabetes mellitus (with sepsis)	-1
Emergency surgery	1.5

Table 4.1: LIPS Criteria.

4.6 ARDS

4.6.1 Variable Selection

Using the 10% data set, the intelligent agents performed 596 viable models. The variables and corresponding variable importance metric for the best performing model is described by Table 4.2. The variables selected for the model by the intelligent agents fall under categories: oxygen and carbon dioxide exchange; blood chemistry; functional; and procedural.

Some of the higher ranked variables by importance have to do with O_2 and CO_2 exchange and generally the mechanism of breathing. The variable pCO_2 is the partial pressure of CO_2 in the venous blood, tidal volume is the amount of air pushed into the lungs by a mechanical ventilator, compliance is a measure of the distensibility of elastic tissue in the lungs, mean airway pressure is the pressure applied during mechanical ventilation, and SpO_2 is the peripheral oxygen saturation measured by a pulse oximeter. O_2 flow liters per minute and the addition of a cannula concern oxygen delivery to the patient.

Another category of variables is blood chemistry. Creatine kinase is a blood assay that may indicate a breakdown of tissues rich in creatine kinase such as inflammation in the heart muscle (myocarditis) and other muscle (myositis), heart attack (myocardial infarction), or a breakdown of skeletal tissue (rhabdomyolysis). Creatine kinase levels may also be influenced by non-clinically relevant conditions that are asymptomatic, exercise, and some medications. Chloride levels are carefully regulated by the kidneys and play a functional part in CO_2 respiration via the Chloride Shift whereby the red blood cells export bicarbonate into the plasma and intakes chloride ions from the plasma. High pCO_2 levels causes a buildup of bicarbonate in red blood cells which is then exchanged for chloride in the plasma. Asparate aminotransferase (AST) is a blood assay that is primarily a marker for liver damage, but may also be indicative of damage to other organs that produce the enzyme like muscle, heart or lung. Hematocrit is the percentage of red blood cells in blood by volume. It may be used as a means to differentiate male from female blood (47% +/- 5% and 42% +/- 5% respectively). Elevated hematocrit may be indicative of dehydration or hypoxia. Albumin

is the primary protein found in blood plasma, and an elevated level may be indicative of dehydration whereas a low level may indicate abnormal liver function, where it is produced. Lactic acid is a blood assay to determine the status of pH regulation in extracellular fluid. High lactic acid levels may be a result of the body breaking down carbohydrates for energy without ample oxygen present.

The intelligent agents additionally selected functional and procedural variables. "Urine out foley" indicates the volume of urine produced. "Metered dose inhaler" is used to deliver aerosolized medication to the lungs. Magnesium sulfate is a medication administered for various reasons including organ and neural protection, as described by Panahi et al. [69]. "20 gauge placed in outside facility" refers to the placement of a 20-gauge needle for an IV. "IV/Saline lock" also indicated the patient is receiving IV fluids or medication. Eye opening, which was the only cognitive based variable, was selected. Other variables include age, gender ("male") and weight.

The intelligent agents selected a myriad of comorbidity variables concerning heart function and blood related conditions (cardiac arrhythmias, congestive heart failure, hypertension, chronic pulmonary disease, valvular disease, peripheral vascular disorder, pulmonary circulation disorders, coagulopathy). Diseases related to organs and the regulation of the body's chemistry were also selected (diabetes, fluid and electrolyte disorders, liver disease, renal failure, hypothyroidism). Various disparate diseases were included (depression, solid tumor without metastasis, neurological disorders, drug abuse, rheumatoid arthritis, and alcohol abuse). Truama was also found to be relevant in predicting the onset of ARDS.

4.6.2 Model training and testing

Using the variables from Table 4.2, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 36,348 patients. At t-minus zero, the time immediately before the onset of ARDS, the model has an area under the ROC curve of 0.847 (0.832, 0.862) (see Figure 4.3. The model performance on the training/testing data set is outlined in Table 4.3. With the cutoff set at the maximum value for the F-

Label	Variable Importance	AUC excluding variable
Tidal Volume	0.29	0.839 (0.815, 0.863)
O_2 Flow (additional cannula)	0.25	0.843 (0.819, 0.866)
pCO_2 (venous)	0.3	0.841 (0.817, 0.865)
20 Gauge placed in outside facility	0.24	0.845 (0.821, 0.868)
Compliance	0.29	0.837 (0.812, 0.861)
Metered Dose Inhaler	0.27	0.845 (0.821, 0.869)
Mean Airway Pressure	0.27	0.83 (0.806, 0.855)
Creatine Kinase, MB Isoenzyme	0.28	0.844 (0.82, 0.868)
Urine Out Foley	0.28	0.846 (0.822, 0.87)
Chloride, Whole Blood	0.27	0.844 (0.82, 0.868)
IV/Saline lock	0.17	0.843 (0.819, 0.867)
SpO_2 (arterial)	0.27	0.84 (0.816, 0.864)
Eye Opening	0.23	0.842 (0.818, 0.866)
Magnesium Sulfate	0.24	0.853 (0.829, 0.876)
Lactic Acid	0.24	0.841 (0.817, 0.865)
Asparate Aminotransferase (AST)	0.27	0.85 (0.826, 0.873)
O_2 Flow (lpm)	0.23	0.83 (0.805, 0.854)
Albumin	0.26	0.852 (0.829, 0.875)
Hematocrit	0.25	0.849 (0.826, 0.873)
trauma	0.2	0.838 (0.814, 0.863)
weight	0.18	0.839 (0.814, 0.863)
male	0.22	0.841 (0.817, 0.865)
age	0.23	0.837 (0.812, 0.861)

Table 4.2: Variables and variable importance of best performing ARDS model generated by intelligent agents.

measure ($beta$=1.75) the model produces: 560 true positives; 31,189 true negatives; 4,154 false positives; 445 false negatives; a true positive rate (sensitivity) of 0.557 (see Figure D.1); a false positive rate of 0.118; a specificity of 0.882 (see Figure D.2); a positive predictive value of 0.119 (see Figure D.3; a negative predictive value of 0.986; an accuracy of 0.873 (see Figure D.4); with a kappa of 0.935 (see Figure D.5).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	0	35342	1	1005	0	0.000	0	1.000	0	0.972	0.972	0.986
1.50	500	31917	3426	505	0.498	0.097	0.498	0.903	0.127	0.984	0.892	0.945
1.75	560	31189	4154	445	0.557	0.118	0.557	0.882	0.119	0.986	0.873	0.935

Table 4.3: ARDS model performance on the training/testing data set.

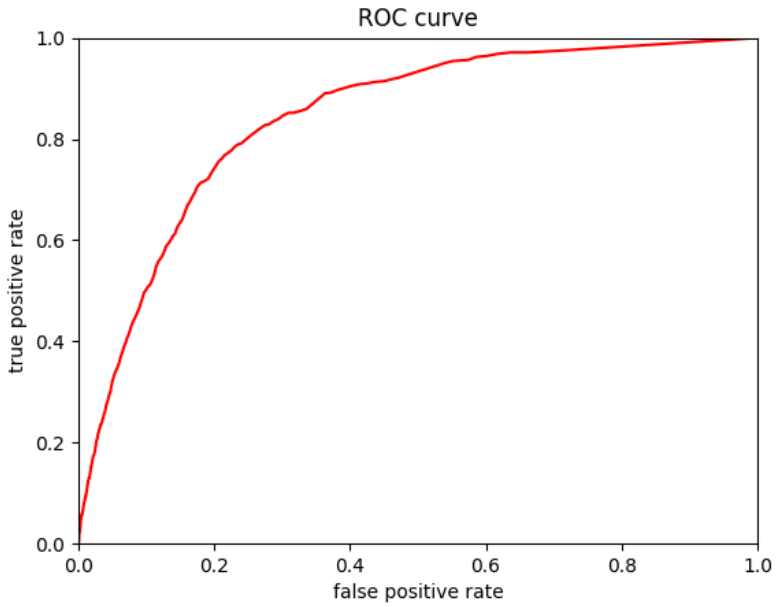


Figure 4.3: ARDS training/testing ROC curve, AUC=0.847 (0.832, 0.862).

4.6.3 Model validation

The threshold resulting for the maximum F-measure (*beta*=1.75) is used to set the threshold for the validation set containing 15,638 patients. The performance of the model against the validation set is largely consistent with the performance on the training/testing data set which indicated a well balanced model with no evidence of over fitting. The ROC curve in Figure 4.4 has an area under the curve of 0.861 (0.838, 0.884) compared to the training/testing are of 0.847 (0.832, 0.862). The performance of the model on the validation set may be found in Table 4.4 with 230 true positives, 13,359 true negatives, 1,868 false positives, and 181 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.560 (see Figure D.7) as opposed to 0.557; a false positive rate of 0.123 as opposed to 0.118; a specificity of 0.877 (see Figure D.8) as opposed to 0.882; a positive predictive value of 0.110 (see Figure D.9) as opposed to 0.119; a negative predictive value of

0.987 as opposed to 0.986; an accuracy of 0.869 (see Figure D.10) as opposed to 0.873; with a kappa of 0.933 (see Figure D.11) as opposed to 0.935.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	0	15226	1	411	0	6.57E-05	0	1.000	0	0.974	0.974	0.987
1.50	204	13671	1556	207	0.496	0.102	0.496	0.898	0.116	0.985	0.887	0.943
1.75	230	13359	1868	181	0.560	0.123	0.560	0.877	0.110	0.987	0.869	0.933

Table 4.4: ARDS model performance on the validation data set.

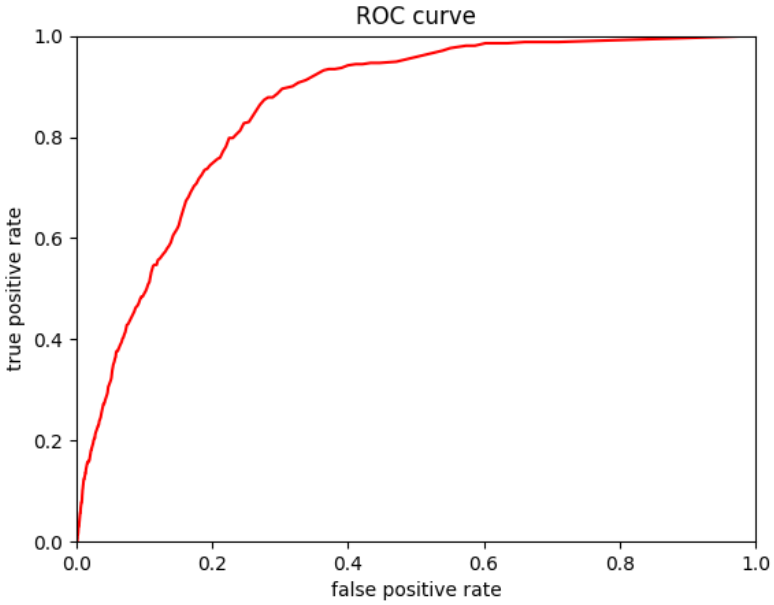


Figure 4.4: ARDS validation ROC curve, AUC=0.861 (0.838, 0.884).

4.6.4 As a function of time

The AUC metric is fairly stable in Figure 4.5, losing 8.25% with a temporal distance of six hours. At t-minus zero the AUC is 0.861 (0.838, 0.884), and at t-minus six hours the AUC is 0.790 (0.764, 0.816). This implies the ARDS model is a stable predictive algorithm relying on mostly static variables with respect to data collection during a patient's ICU stay.

Additional lower performing machine learning algorithms over time may be found in Figure D.13.

4.6.5 Variable composition, demographics and comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table D.1. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate ARDS positive patients from ARDS negative patients. Tidal volume, pCO_2 (venous), mean airway pressure, creatine kinase, urine

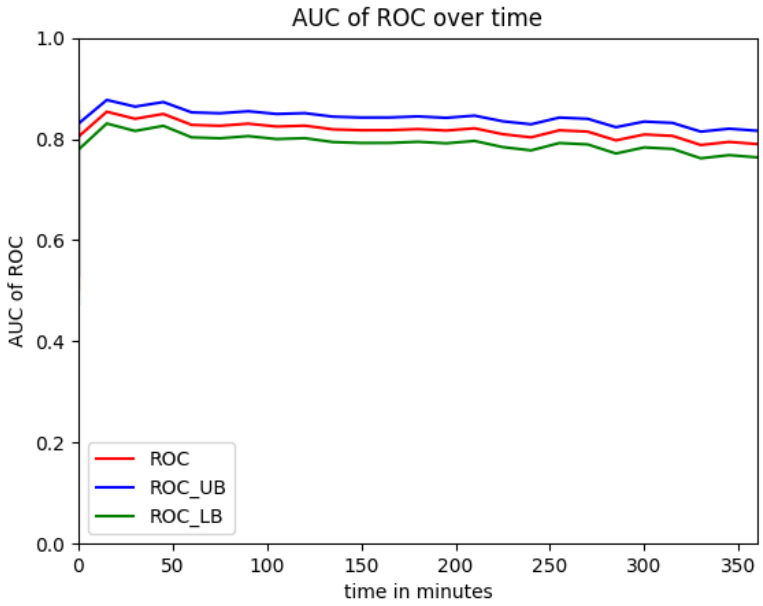


Figure 4.5: ARDS validation ROC curve over time. AUC at t-minus=0 is 0.861 (0.838, 0.884). AUC at t-minus=360 minutes is 0.790 (0.764, 0.816).

output, whole blood chloride, and O_2 flow (lpm) are all of a higher value in ARDS positive patients. Eye opening had a lower value in ARDS positive patients.

The elevated tidal volume, mean airway pressure, and O_2 flow are consistent with aggressive oxygenation ventilation. As pCO_2 (venous) increases, so does the chloride. Elevated chloride may also indicate renal dysfunction, but is most likely an iatrogenic anomaly caused by a non-anion gap metabolic acidois (NAGMA) from the administration of NaCl. Elevated creatine kinase may indicate a breakdown of muscle tissue. A larger urine output may indicate a more aggressive fluid therapy for an underlying condition. Elevated asparate aminotransferase may indicate liver (or other organ) damage. Depressed eye opening implies diminished cognition in condition positive patients. Depressed albumin indicated abnormal liver function, and in the context of elevated asparate aminotransferase liver damage is likely.

Variable composition by classification was assessed in Table D.2 to greater understand the

model's false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they present with: lower AST; and lower hematocrit.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table D.3. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate ARDS positive patients from ARDS negative patients. ARDS positive compared with ARDS negative patients had a higher in-hospital mortality (28% as opposed to 17%), a higher 30-day mortality (30% as opposed to 20%), and a greater ICU length of stay (median 9.22 days as opposed to 2.21 days). The breakdown of ARDS positive and ARDS negative patients by type of ICU is also significant. ARDS patients are underrepresented in: the Cardiac Care Unit (9% as opposed to 14%) and the Medical ICU (29% as opposed to 38%). ARDS patients are overrepresented in: the Surgical ICU (21% as opposed to 15%) and the Trauma Surgery ICU (22% as opposed to 11%).

Demographic and key variables were assessed by classification in Table D.4. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their demographics include: over-representation in the CSRU; and under-representation in the MICU.

Comorbidities of condition positive and condition negative patients were assessed in Table D.5. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate ARDS positive patients from ARDS negative patients. ARDS positive patients present disproportionately greater for: peripheral vascular disorders (18% as opposed to 11%); coagulopathy (15% as opposed to 10%); sustained trauma (29% as opposed to 9%). ARDS positive patients present disproportionately fewer for:hypothyroidism (5% as opposed to 11%) and depression (3% as opposed to 9%).

Comorbidities by classification is assessed in Table D.6. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their comorbidities include: congestive heart failure; hypertension; under-represented in

hypothyroidism; renal failure; and under-represented in coagulopathy.

4.7 SAHRF

Severe Acute hypoxemic respiratory failure can be completely defined by the oxygenation requirement of ARDS. This makes ARDS a special case of acute hypoxemic respiratory failure. This section of study aims at predicting the onset of severe acute hypoxemic respiratory failure. The process follows the plans outlined in the methodology chapter. Variables are selected using the intelligent agent framework; a model is built, trained, and tested; the model is then validated and evaluated.

4.7.1 Variable selection

Using the 10% development data set, the intelligent agents performed 824 iterations. Of these, 415 created viable models. The variables and corresponding variable importance metric for the best performing model are described by Table 4.5. The variables selected for the model by the intelligent agents fall under categories: scoring systems, one-time measures, vital signs, blood laboratory, blood-gas makeup, and management. Scoring systems include the Braden score (a metric that predicts the occurrence of pressure ulcers in patients) and the Glasgow Coma Scale (which includes motor response, verbal response, and eye movement). One-time measures include weight, height, BSA (body surface area, usually calculated from height, weight, and gender). Vital signs include the systolic blood pressure, diastolic blood pressure, mean blood pressure, heart rate, and respiratory rate. Blood laboratory includes WBC (white blood cell count), creatinine, hemoglobin, platelet count, hematocrit, chloride, PTT, magnesium, BUN, INR(PT), Potassium, red blood cell count, and sodium. Blood-gas makeup includes SpO_2 and carbon dioxide. Management includes O_2 flow.

Label	Variable Importance	AUC excluding variable
Motor Response	0.47	0.950 (0.955, 0.944)
SpO_2	0.43	0.949 (0.954, 0.943)
Admit Wt	0.43	0.948 (0.954, 0.943)
Admit Ht	0.43	0.949 (0.954, 0.944)
BSA	0.4	0.949 (0.954, 0.944)
Carbon Dioxide	0.41	0.949 (0.954, 0.944)
WBC (4-11,000)	0.38	0.949 (0.954, 0.944)
Braden Score	0.38	0.949 (0.954, 0.943)
GCS Total	0.34	0.950 (0.955, 0.944)
NBP Mean	0.34	0.950 (0.955, 0.945)
NBP [Diastolic]	0.36	0.950 (0.955, 0.945)
Eye Opening	0.32	0.950 (0.955, 0.945)
NBP [Systolic]	0.32	0.949 (0.955, 0.944)
HR Alarm [Low]	0.33	0.950 (0.955, 0.945)
Creatinine	0.38	0.950 (0.955, 0.945)
Hemoglobin	0.37	0.950 (0.956, 0.945)
Verbal Response	0.24	0.948 (0.954, 0.943)
Platelet Count	0.35	0.949 (0.955, 0.944)
Hematocrit	0.36	0.951 (0.956, 0.946)
HR Alarm [High]	0.25	0.950 (0.955, 0.945)
Chloride	0.36	0.950 (0.955, 0.945)
PTT	0.32	0.951 (0.956, 0.946)
Magnesium	0.34	0.950 (0.955, 0.944)
O_2 Flow (lpm)	0.27	0.942 (0.947, 0.936)
BUN (6-20)	0.34	0.950 (0.955, 0.945)
SpO_2 Alarm [High]	0.24	0.950 (0.955, 0.945)
INR(PT)	0.31	0.950 (0.955, 0.945)
Potassium	0.33	0.950 (0.955, 0.945)
Respiratory Rate	0.27	0.949 (0.955, 0.944)
SpO_2 Alarm [Low]	0.24	0.949 (0.955, 0.944)
Red Blood Cells	0.31	0.950 (0.955, 0.945)
Sodium	0.32	0.950 (0.955, 0.945)
Heart Rate	0.29	0.950 (0.955, 0.945)

Table 4.5: Variables and variable importance of best performing hypoxemic model generated by intelligent agents.

4.7.2 Model training and testing

Using the variables from Table 4.5, a 10-fold cross validation model for predicting ARDS (with 90% for training and 10% testing) was created on a training/testing set of 42,162 patients. At t-minus zero, the time immediately before the onset of severe hypoxemic respiratory syndrome, the model has an area under the ROC curve of 0.950 (0.947, 0.953) (see Figure 4.6). Other measures important for evaluating the performance of the model are: true positive and false positive rate (see Figure D.14); sensitivity and specificity (see Figure D.15); positive predictive value and negative predictive value (see Figure D.16); accuracy (see Figure D.17); kappa (see Figure D.18); and F-measure evaluations (see Figure D.19).

The model performance on the training/testing data set is outlined in Table 4.6. The algorithm defaults to a threshold where the F-measure (β=1.00) is at a maximum, 0.51 and greater indicate a predicted positive result, and 0.50 and below indicate a predicted negative result. At this threshold the model produces 5448 true positives and 32486 true negatives, accompanied by 1589 false positives and 2639 false negatives. This yields a true positive rate (or sensitivity) of 0.674, a false positive rate of 0.047, a specificity of 0.953, a positive predictive value of 0.774, a negative predictive value of 0.925, an accuracy of 0.900, and a kappa of 0.978. This table also evaluates the threshold where there is a maximum F-measure having a β=1.50. The results may be found in the table with a threshold of 0.27.

This study uses an F-measure with a β of 1.75 because of the diseases in clinical context it is deemed important to more heavily weigh sensitivity over specificity. The maximum F-measure value produces a threshold of 0.22. This cutoff yields significantly more true positives (7333 as opposed to the default algorithm of 5448), fewer true negatives (29431 as opposed to 32486), many more false positives (4644 as opposed to 1589), much fewer false negatives (754 as opposed to 2639). The F-measure β of 1.75 setting therefor yields a true positive rate (sensitivity) significantly greater than the default (0.907 as opposed to 0.674), a false positive rate increase (0.136 as opposed to 0.047), a lower specificity (0.864 as opposed to 0.953), a lower positive predictive value (0.612 as opposed to 0.774), a higher negative

predictive value (0.975 as opposed to 0.925), a decrease in accuracy (0.872 as opposed to 0.900) and a similar kappa (0.987 as opposed to 0.978).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	5448	32486	1589	2639	0.674	0.047	0.674	0.953	0.774	0.925	0.900	0.978
1.50	7090	30356	3719	997	0.877	0.109	0.877	0.891	0.656	0.968	0.888	0.991
1.75	7333	29431	4644	754	0.907	0.136	0.907	0.864	0.612	0.975	0.872	0.987

Table 4.6: Hypoxemic model performance on the training/testing data set.

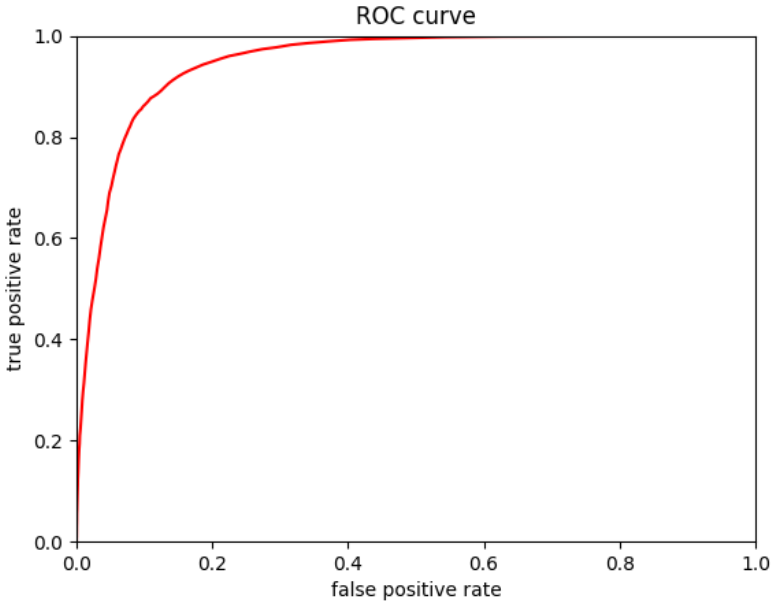


Figure 4.6: Hypoxemic training/testing ROC curve, AUC=0.950 (0.947, 0.953).

4.7.3 Model validation

The threshold (0.22) resulting from the maximum F-measure (β=1.75) is to be used to evaluate the model performance against a validation data set containing 18,074 patients. The performance of the model against the validation set is largely consistent with the performance on the training/testing data set, which indicates a well-balanced model with no evidence of overfitting. The ROC curve (see Figure 4.7) has an area under the curve of 0.952 (0.947, 0.957) compared to the training/testing area of 0.950 (0.947, 0.953). The performance of the model on the validation set may be found in Table 4.7. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has a true positive rate (sensitivity) of 0.914 (see Figure D.20) as opposed to 0.907, a false positive rate of 0.138 as opposed to 0.136, a specificity of 0.862 (see Figure D.21) as opposed to 0.864, a positive predictive value of 0.612 (see Figure D.22) which is exactly the same as the training/testing set, a negative predictive value of 0.977 as opposed to 0.975,

an accuracy of 0.872 (see Figure D.23) which is exactly the same as the training/testing set, and a kappa of 0.989 (see Figure D.24) as opposed to 0.987.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	2365	13911	679	1119	0.679	0.047	0.679	0.953	0.777	0.926	0.901	0.979
1.50	3086	12941	1649	398	0.886	0.113	0.886	0.887	0.652	0.97	0.887	0.992
1.75	3184	12574	2016	300	0.914	0.138	0.914	0.862	0.612	0.977	0.872	0.989

Table 4.7: Hypoxemic model performance on the validation data set.

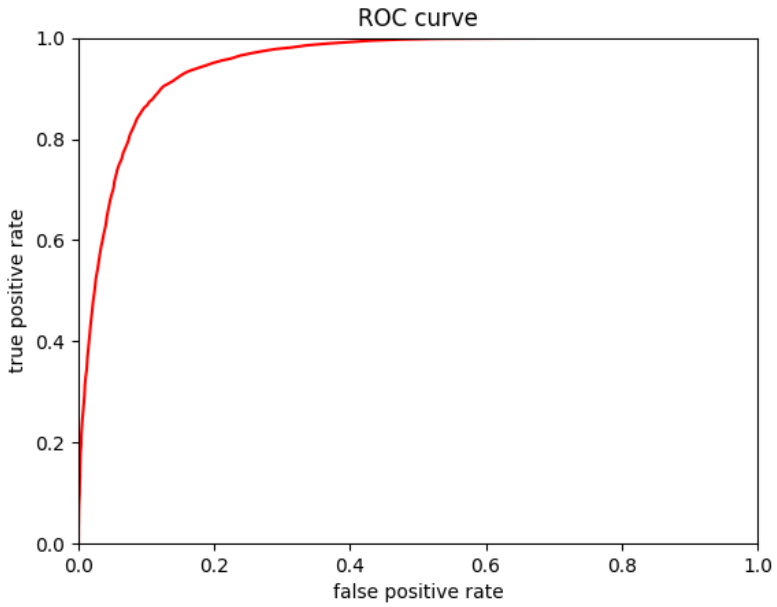


Figure 4.7: Hypoxemic validation ROC curve, AUC = 0.952 (0.947, 0.957).

4.7.4 As a function of time

The model maintains excellent performance over time previous to the onset of severe acute hypoxemic respiratory failure. At t-minus zero, just prior to the onset of respiratory failure, the model's area under the ROC curve is 0.952 (0.947, 0.957). At t-minus six hours, six hours prior to respiratory failure, the model's area under the ROC curve is 0.878 (0.870, 0.886), an excellent predictive tool even losing 7.77% AUC over six hours. See Figure 4.8 to see how the AUC of the ROC changes over time.

Additional lower performing machine learning algorithms over time may be found in Figure D.26.

4.7.5 Variable composition, demographics, and comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table D.7. There are several statistically significant (Kruskal-Wallis

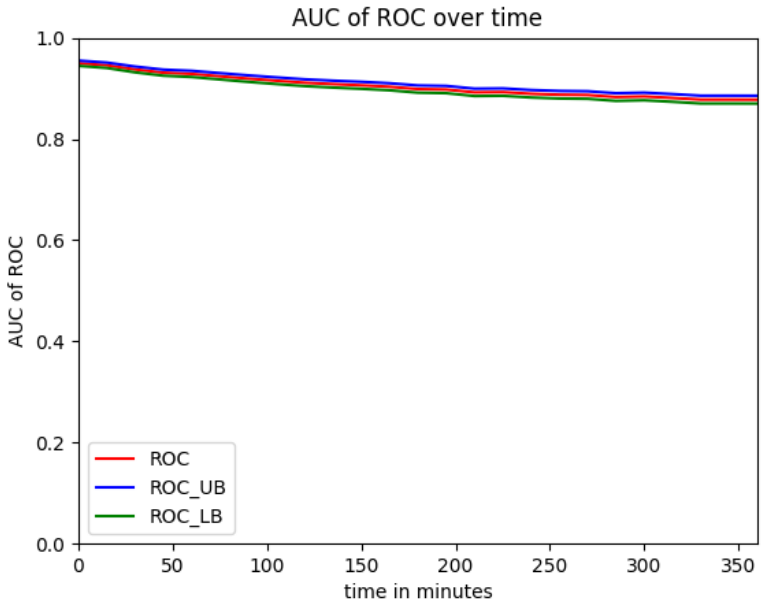


Figure 4.8: Hypoxemic validation AUC of the ROC over time. AUC at t-minus=0 is 0.952 (0.947, 0.957). At t-minus six hours, the model's area under the ROC curve is 0.878 (0.870, 0.886).

test P-value <0.0001) variables that differentiate Hypoxemic positive patients from Hypoxemic negative patients. Hypoxemic positive patients present with statistically significant higher values than negative patients in: SpO_2 (arterial); weight; CO_2 (venous); white blood cell count; diastolic blood pressure; heart rate; creatinine; chloride; PTT; magnesium; O_2 flow; BUN; INR(PT). Hypoxemic positive patients present with statistically significant lower values than negative patients in: Braden score; systolic blood pressure; hemoglobin; verbal response; platelet count.

Variable composition by classification is assessed in Table D.8. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they present with: higher motor response; a higher body surface area; increased eye opening; lower respiratory rate; lower red blood cell count; and higher sodium.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table D.9. There are several statistically significant (Kruskal-Wallis test P-value<0.0001) variables that differentiate Hypoxemic positive patients from Hypoxemic negative patients. Hypoxemic positive compared with Hypoxemic negative patients had a higher weight (80.65 kg as opposed to 75.00 kg), a correspondingly higher body mass index (28.93 as opposed to 27.11), a greater male population (61% as opposed to 55%), a higher in-hospital mortality (23% as opposed to 14%), a higher 30-day mortality (25% as opposed to 16%), and a greater ICU length of stay (median 4.13 days as opposed to 1.90 days). The breakdown of Hypoxemic positive and Hypoxemic negative patients by type of ICU is also significant. Hypoxemic patients are under represented in the Medical ICU (28% as opposed to 36%). Hypoxemic patients are over represented in: the Cardiac Surgery Recovery Unit (33% as opposed to 11%); Trauma Surgery ICU (13% as opposed to 10%). Non-white ethnicities were underrepresented in Hypoxemic patients (Asian 2% as opposed to 4%, Black 6% as opposed to 10%, and Hispanic 3% as opposed to 4%).

Demographic and key variables were assessed by classification in Table D.10. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their demographics include: lower representation of Medicare; and higher representation of private insurance. This phenomenon is likely due to the age gap, where the median false positives age is 74 years old while the median false negative age is 64 years old.

Comorbidities of condition positive and condition negative patients were assessed in Table D.11. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate Hypoxemic positive patients from Hypoxemic negative patients. Hypoxemic positive patients present disproportionately greater for: congestive heart failure (33% as opposed to 24%); cardiac arrhythmias (38% as opposed to 29%); valvular disease (21% as opposed to 12%); peripheral vascular disorder (13% as opposed to 10%); hypertension, uncomplicated (43% as opposed to 35%); diabetes, uncomplicated (22% as opposed to 18%). Hypoxemic positive patients present disproportionately fewer for: hyper-

tension, complicated (8% as opposed to 13%); hypothyroidism (6% as opposed to 10%); fluid and electrolyte disorders (21% as opposed to 27%); depression (4% as opposed to 9%).

Comorbidities by classification are assessed in Table D.12. Of the statistically significant variables not included above, there exist no significant comorbidity related to misclassifications.

4.8 Model Comparisons

There exists, in the peer-reviewed body of literature, no comparison model that predicts the onset of ARDS. The models of this study are designed to focus only on data points previous to the onset of disease. This strategy allows for adoption into an EMR as an early warning system. LIPS is the closest comparable model, but operates on retrospective patient data. Some risk factors may be computed in real-time, the primary core of the model is predisposing conditions which are immutable duing a patient's stay. The analytics of model comparisons are found in Table 4.8.

Model analytics for LIPS were extracted from the LIPS author group [25]. For the purposes of this study, Gajic, et al. is considered the foundational model, as others are derived from it. The paper did not include some important metrics nor did it include a confusion matrix. To adequately compare the foundational model's performance, the relative ratios of the confusion matrix was calculated, setting true positive equal to one.

Computing confusion matrix ratio from Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value based on True Positive = 1, and solving in terms of known variables:

$$TPR = \frac{TP}{TP + FN} \tag{4.1}$$

$$FN = \frac{TP}{TPR} - TP \tag{4.2}$$

$$PPV = \frac{TP}{TP + FP} \tag{4.3}$$

$$FP = \frac{TP}{PPV} - TP \tag{4.4}$$

$$NPV = \frac{TN}{TN + FN} \tag{4.5}$$

$$TN = \frac{FN * NPV}{1 - NPV} \tag{4.6}$$

Though the LIPS reported specificity is 0.78, the calculated specificity is 0.761, denoted by the double dagger (‡).

Bauman et al. [18] published a LIPS model performance in an ICU setting, which had a similar AUC (0.79) as the foundational model performance (0.80). Bauman, et al. provide no other analytic metrics, but are included to validate the application of LIPS on an ICU patient population. This study implemented LIPS method on the MIMIC database validation data set for an in-kind comparison. This study's ARDS model is extracted from Table 4.4, and SAHRF model is extracted from 4.7.

The overall performance of the three LIPS implementations (LIPS foundational model, LIPS in an ICU, LIPS applied to MIMIC) are comparable in AUC (0.80, 0.79, 0.84 (0.823, 0.871)), kappa (0.69, NA, 0.722), specificity (0.78, NA, 0.773), negative predictive value (0.97, NA, 0.993), accuracy (0.756, NA, 0.773), and false positive rate (0.239, NA, 0.227). The LIPS foundational model differs from LIPS applied to MIMIC in sensitivity or true positive rate (0.69 as opposed to 0.769), and positive predictive value (0.18 as opposed to 0.073). The discrepancy likely lies in the mix of patients. LIPS was derived from modeling an inclusive set of in-patients with a wide distribution of prognoses, diagnoses, and overall severity of illness. When applied to a focused patient population of ICU patients, all of whom have a heightened severity of illness relative to the total in-patient population, the model contributes the high severity of illness to an ARDS origin, thus resulting in a high false positive rate, and consequently a lower positive predictive value. Compared to the LIPS applied to MIMIC model, this study's ARDS model has a higher AUC (0.861 (0.838, 0.884) as opposed to 0.847 (0.823, 0.871)), kappa (0.852 as opposed to 0.722), sensitivity and true positive rate(0.560 as opposed to 0.769), specificity (0.877 as opposed to 0.773), lower negative predictive value (0.987 as opposed to 0.993), and accuracy (0.869 as opposed to 0.773), a higher positive predictive value (0.110 as opposed to 0.073) and a lower false

positive rate (0.123 as opposed to 0.227).

	AUC	Kappa	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN	ACC	TPR	FPR
LIPS [25]	0.8	0.694*	0.69	0.78‡	0.18	0.97	1†	14.527†	4.556†	0.449†	0.756*	0.690*	0.239*
LIPS in ICU [18]	0.79												
LIPS to MIMIC	0.847 (0.823, 0.871)	0.722	0.769	0.773	0.073	0.993	316	13659	4004	95	0.773	0.769	0.227
Intelligent Agent ARDS, this study	0.861 (0.838, 0.884)	0.852	0.560	0.877	0.110	0.987	544	10651	4062	381	0.869	0.560	0.123

Table 4.8: ARDS model comparison. Values denoted with an asterisk (*) are derived from imputing the confusion matrix ratio. Values denoted with a dagger (†) are ratios relative to true positive value of one, and derived from the equations in the Model Comparison section.

4.9 Derivative Investigation - Disseminated Intravascular Coagulation

The information derived from the intelligent agents experimenting with ARDS and SAHRF patients has presented a mosaic of variables, demographics and comorbidities. There is an interesting pattern that developed during the production of this study. In the single-model ARDS, patients with a comorbidity of coagulopathy was found to be statistically significant (P-value 0.0023 in Table D.1). In the model which predicts Severe Acute Hypoxemic Respiratory Disorder, a necessary condition for ARDS, patients with a comorbidity of coagulopathy were not found to be statistically significant (P-value <0.0605 in Table D.7). This implies a context specific ARDS comorbidity that is not significant in Hypoxemic patients.

This study has shown a correlation between ARDS and Coagulopathy. To advanced the investigation this study inspects a special case of Coagulopathy, Consumptive Coagulopathy, also known as Disseminated Intravascular Coagulation (DIC) which afflicts between 9% and 19% of ICU patients (from a study by Van der Linden et al. [90]). This study implemented the Japanese Association of Acute Medicine (JAAM) diagnostic algorithm and the International Society on Haemostasis (ISTH) non-overt diagnostic criteria to identify the onset of DIC. Some clinicians are critical of the JAAM definition for depending on SIRS criteria (making it sensitive to sepsis related DIC, as described by Wada et al. [93]) leading to a relatively larger inclusion of patients defined by the disease compared to other diagnostic criteria. This study incorporates ISTH's non-overt as the more conservative diagnostic criteria, as described by Gando et al. [27], recruiting fewer patients into the disease definition. This study found 20.2% of ICU patients met the diagnostic criteria for JAAM's DIC, and 15.2% met diagnostic criteria for ISTH's DIC. Of ARDS patients, JAAM's DIC negative patients have an in-hospital mortality of 19%, while JAAM's DIC positive patients have an in-hospital mortality of 36% (P-value <0.0001). Of ARDS patients, ISTH's DIC negative patients have an in-hospital mortality of 25%, while ISTH's DIC positive patients have an in-hospital mortality of 47% (P-value <0.0001).

JAAM's DIC and ARDS positive patients in MIMIC were analyzed to determine temporal

placement of the two diseases in the patient's ICU stay. Of the 1,419 ARDS positive patients, 915 of them (64.5%) are also afflicted with JAAM's DIC. Of all patients with both ARDS and JAAM's DIC during their ICU stay, 498 (or 54.4%) contracted JAAM's DIC before contracting ARDS. The median time or ARDS preceding JAAM's DIC is 34.1 hours (Q1 = 11.3 hours, Q3 = 3 days). The median time of JAAM's DIC preceding ARDS is 6.2 hours (Q1 = 1.5 hours, Q3 = 31.2 hours).

ISTH's DIC and ARDS positive patients were also analyzed. Of the 1,419 ARDS patients 305 of them (or 21.5%) are also afflicted with ISTH's DIC. Of all patients with both ARDS and ISTH's DIC during their ICU stay, 234 (or 76.7%) contracted ISTH's DIC before contracting ARDS. The median time or ARDS preceding ISTH's DIC is 15.1 hours (Q1 = 6.2 hours, Q3 = 3 days). The median time of ISTH's DIC preceding ARDS is 2.5 days (Q1 = 10 hours, Q3 = 9.4 days).

Strongly correlating conditions (ARDS and DIC) where neither one overwhelmingly precedes the other does not lend to an independent causal argument, though leaves open the possibility of either a mutually advancing interaction or an unobserved causal event for both diseases.

There are three studies to present a baseline of research implicating DIC causing or correlated with ARDS. In 1977 Ogawa et al. [67] ran a study of nine ARDS positive individuals, six of whom proved positive for DIC as well. In 1986 El-Kassimi et al. [15] conducted a study of 52 heat stroke patients. Twelve had both ARDS and DIC, and only one had exclusively DIC. In 2014 Miyoshi et al. [58] performed a retrospective data analysis of 142 patients with both ARDS and DIC and concluded administering a neutrophil elastase inhibitor and an inhibitor of thrombin coagulation activity combined was more beneficial than either drug alone. That Miyoshi et al. found a clotting inhibitor (thrombomodulin) beneficial in treating ARDS is telling of an underlying mechanism related to coagulation and not fully explained in the current body of knowledge for ARDS.

This study proposes a theoretical ARDS disease progression interacting with the DIC pathway to create a cycle indicative of a positive feedback loop. Figure 4.9 outlines how such

a cycle may exist. ARDS may be wholly described by the left side of the figure where cellular lung injury causes the release of cytokines and cytotoxic molecules (such as ROS, TNFα, IL-1β, etc.). The release causes a disruption of the aveolar-capillary membrane and an increased vascular permeability. The permeable membrane allows protein-rich fluid to accumulate in alveoli and eventually surfactant degradation. The excess fluid causes pulmonary edema and eventually alveolar collapse, in turn causing an impaired gas exchange and decreased lung compliance. Proliferation of Type II alveolar cells gradually produces irreversible fibrosis. DIC is initiated when procoagulants cause excess thrombin formation. The excess thrombin can cause: plasmin activation and fobrinolysis (breaking down existing clots) which leads to excess bleeding; consumption of coagulation factors, depleting the ability to form clots, and leads to excess bleeding; microvascular clots in organs causing injury. The two diseases may promote each other where the organ hypoperfusion and injury of DIC causes cellular lung injury needed for the ARDS pathway, and consequently the disruption of the aveolar-capillary membrane causes the release of procoagulants needed for the DIC pathway.

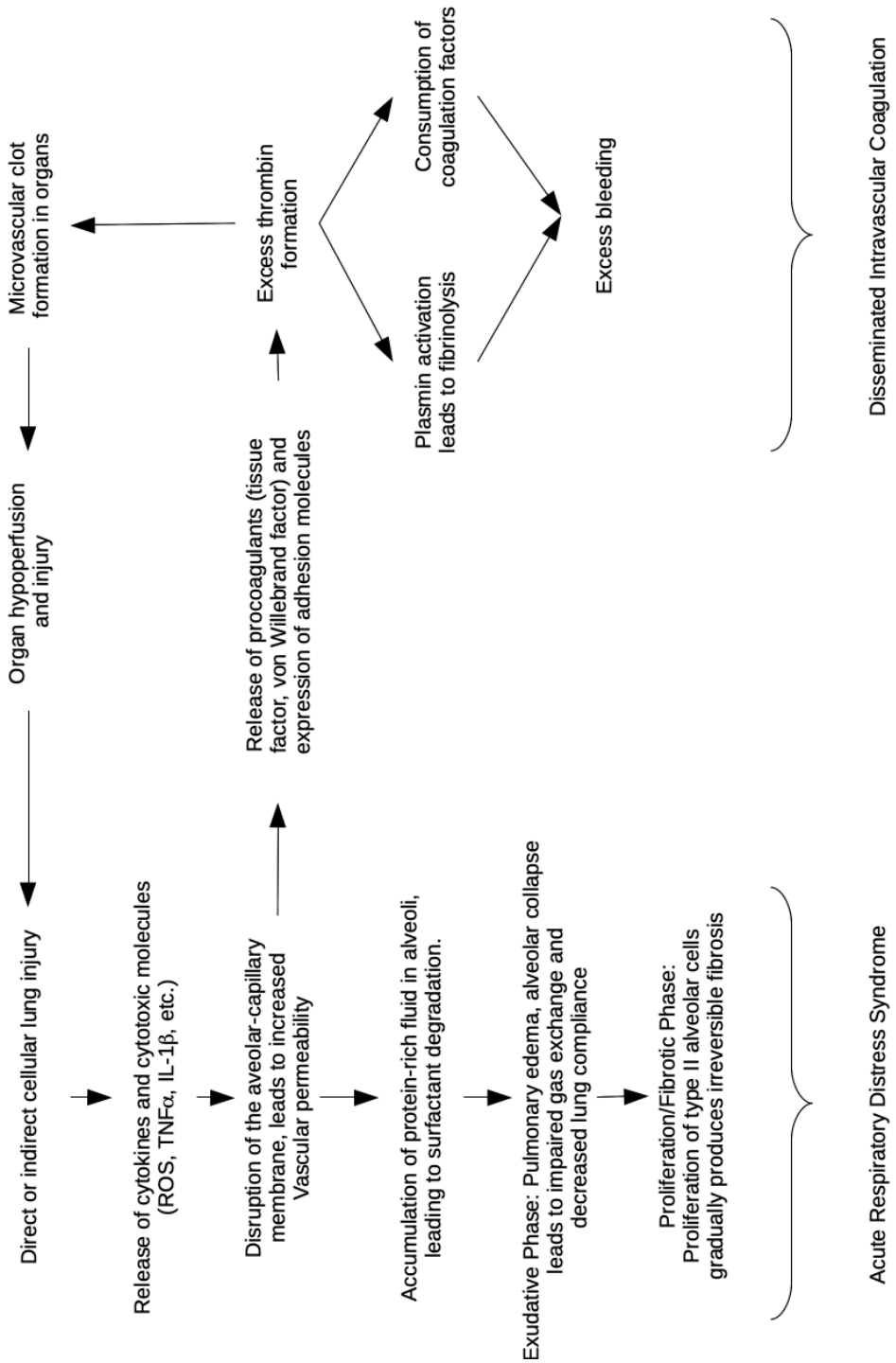


Figure 4.9: Theoretical ARDS pathway merged with DIC pathway to create a feedback loop.

Intelligent agents devised a model to predict the onset of DIC from ARDS patients. The model performs with an AUC of 0.722 (0.689, 0.755) in Figure 4.10. The model degrades over time losing 10.7% of the AUC over six hours as seen in Figure 4.11. The overall performance of the model may be seen in Table 4.9 with a true positive rate (or sensitivity) (Figure D.27) of 0.527, a false positive rate of 0.212, a specificity (Figure D.28) of 0.788, a positive predictive value (Figure D.29) of 0.670, a negative predictive value of 0.671, an accuracy (Figure D.30) of 0.670, and a kappa (Figure D.31) of 0.568.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	217	397	107	195	0.527	0.212	0.527	0.788	0.670	0.671	0.670	0.568
1.50	386	133	371	26	0.937	0.736	0.937	0.264	0.510	0.836	0.567	0.185
1.75	408	51	453	4	0.990	0.899	0.990	0.101	0.474	0.927	0.501	0.057

Table 4.9: DIC FROM ARDS model performance on entire data set.

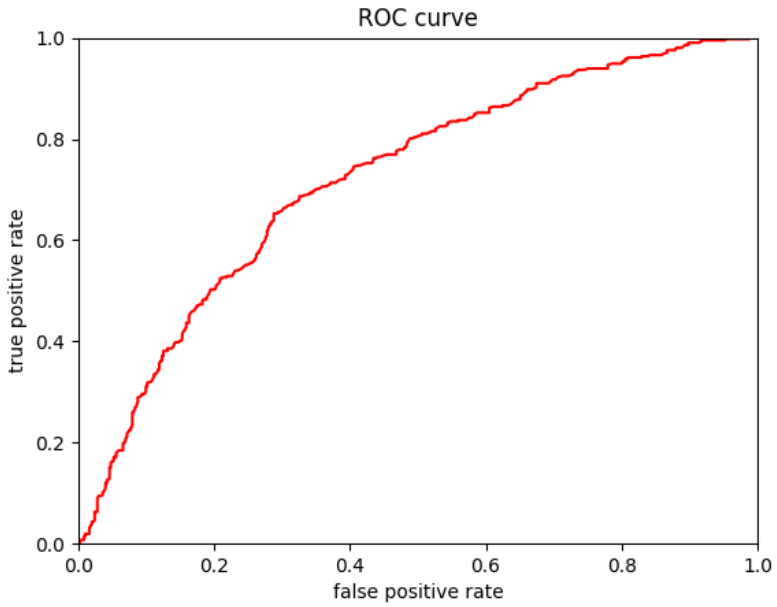


Figure 4.10: DIC from ARDS patients, entire data set AUC=0.722 (0.689, 0.755).

The variables of the model (Table 4.10) have significant difference between ARDS patients that will develop DIC from those who will not (Table D.16), such as: an elevated whole blood chloride (P-value <0.0001); an elongated INR(PT) (P-value <0.0001); a depressed bicarbonate (P-value <0.0001); and a depressed platelet count (P-value <0.0001) suggestive of pre-DIC consumption conditions.

Additionally, ARDS patients who will develop DIC have a higher: in-hospital mortality (36% as opposed to 19%, P-value <0.0001); 30-day mortality (38% as opposed to 21%, P-value <0.0001); and length of stay in the ICU (11.7 days as opposed to 7.2 days, P-value <0.0001) as seen in Table D.14.

Inspecting the comorbidities associated with ARDS patients developing DIC, patients who will develop DIC have a higher incidence of: congestive heart failure; liver disease; and coagulopathy. Pre-DIC patients have a lower incidence of: hypertension, uncomplicated; chronic pulmonary disease; fluid and electrolyte disorders; and depression.

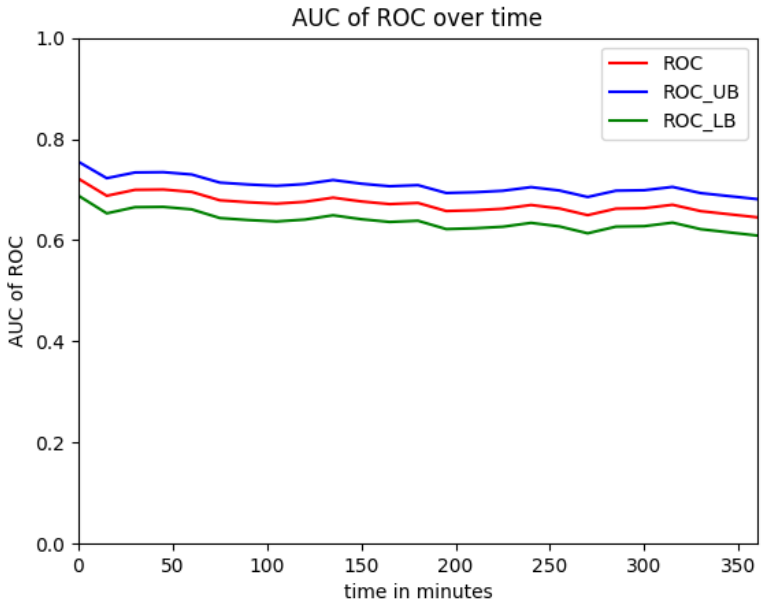


Figure 4.11: DIC from ARDS AUC of the ROC over time. AUC at t-minus=0 is 0.722 (0.689, 0.755). At t-minus six hours, the model's area under the ROC curve is 0.645 (0.609, 0.681).

Intelligent agents devised a model to predict the onset of DIC from ARDS patients. The model performs with an AUC of 0.675 (0.639, 0.711), a sensitivity of 0.478, a specificity of 0.756, and an accuracy of 0.746. The model is based on notable differences between DIC patients who do and do not develop ARDS, namely: higher bands; lower SaO_2; and a higher specific gravity as seen in Tables 4.11 and D.16.

DIC patients who later develop ARDS are notable as significantly trauma patients. In Table D.17 DIC patients who later develop ARDS are over-represented in the Trauma ICU, and only significant comorbidity in Table D.18 is trauma.

Given the proposed pathogenesis of an ARDS-DIC positive feedback cycle and the significant risk increase for ARDS patients to develop DIC, it may be prudent to consider administering low molecular weight heparin (LMWH) [93] to pre-DIC ARDS patients. LMWH

Label	Variable Importance	AUC excluding variable
Chloride	0.36	0.704 (0.670, 0.738)
Oxygen	0.33	0.708 (0.674, 0.742)
Tidal Volume	0.34	0.707 (0.673, 0.741)
Bicarbonate	0.34	0.707 (0.673, 0.741)
Sodium	0.34	0.708 (0.674, 0.742)
INR(PT)	0.34	0.688 (0.653, 0.722)
Gastric Meds	0.27	0.700 (0.665, 0.734)
Chloride, Whole Blood	0.31	0.715 (0.681, 0.748)
Creatinine, Urine	0.24	0.703 (0.669, 0.737)
Alkaline Phosphatase	0.31	0.701 (0.667, 0.735)
Platelet Count	0.31	0.662 (0.627, 0.698)
Cholesterol, LDL	0.22	0.710 (0.676, 0.744)

Table 4.10: Variables and variable importance of best performing DIC from ARDS model generated by intelligent agents.

may inhibit coagulation factors and a below therapeutic dose is a recommended prophylactic treatment for non-symptomatic DIC patients to prevent thrombosis. Acute promyelocytic leukaemia (APML) patients and patients with low platelets should not be given heparins as it would cause bleeding complications. It is certainly prudent to begin monitoring the levels of INR, fibrinogen, and platelets consistent with DIC monitoring. A drop in those levels indicate the consumption effect of DIC. Treatment and management include repleting the consumed products while simultaneously treating the underlying condition.

Label	Variable Importance	AUC excluding variable
Bands	0.28	0.646 (0.61, 0.682)
Unable to assess cognitive / perceptual	0.13	0.643 (0.607, 0.679)
O2 Saturation Alarm - Low	0.2	0.643 (0.607, 0.679)
Specific Gravity	0.33	0.619 (0.583, 0.656)
Alkaline Phosphatase	0.33	0.659 (0.623, 0.695)
Respiratory Rate	0.27	0.648 (0.612, 0.684)
Tidal Volume (Set)	0.24	0.642 (0.606, 0.678)
Differential-Atyps	0.05	0.667 (0.631, 0.703)
SaO2	0.29	0.652 (0.616, 0.688)
O2 Flow (additional cannula)	0.18	0.653 (0.617, 0.689)
Sodium (whole blood)	0.27	0.657 (0.621, 0.693)
Differential-Lymphs	0.25	0.662 (0.626, 0.697)
Iron Binding Capacity, Total	0.17	0.645 (0.609, 0.681)
Heart rate Alarm - High	0.14	0.642 (0.606, 0.678)
High Resp. Rate	0.19	0.644 (0.608, 0.68)
Baseline pain level	0.15	0.648 (0.612, 0.684)
Mixed Venous O2% Sat	0.1	0.642 (0.606, 0.678)
Nitroglycerin	0.15	0.652 (0.616, 0.688)
Metoprolol	0.34	0.636 (0.6, 0.672)
NBP [Systolic]	0.21	0.656 (0.621, 0.692)
Cholesterol Ratio (Total/HDL)	0.16	0.661 (0.625, 0.697)

Table 4.11: Variables and variable importance of best performing ARDS from DIC model generated by intelligent agents.

4.10 Discussion

The intelligent agents produce superior models to LIPS in terms of AUC, accuracy, specificity, ICU based positive predictive value, and false positive rate. The intelligent agent model

shares the same shortfall as the LIPS model, which is the models are better at differentiating patients who will not become afflicted with ARDS than those patients who will (as seen by the positive predictive value and negative predictive value balance). Such a predictive tool remains clinically applicable, though not as useful as a positive predictor for the onset of ARDS.

The intelligent agents produced the only known predictor of Severe Hypoxemic Respiratory Failure. The model has an impressive performance, and the clinical utility for such a predictive model remains in its assistance to diagnose underlying causes.

The characterization and proposed disease pathway of the interaction between ARDS and DIC is potentially the greatest contribution made in this chapter. At 915 patients positive for both ARDS and DIC, this study is six times larger than all previous published studies combined. The mortality rate of the ICU is 13.3%. The mortality rate of ARDS and DIC (JAAM) is 28.5% and 26.4% respectively. The mortality rate for patients afflicted with both is elevated to 38.3%. This dissertation shows DIC is a comorbidity in 78% of all ARDS patients who succumb to in-hospital mortality. Where 64.5% of ARDS patients also have DIC, it is prudent to administer LMWH to ARDS patients, absent of signs or suspicion of bleeding, to curb the consumptive coagulation. Future work is required to verify the proposed disease pathway cycle. If the pathway is verified, it will present a target for therapeutics to break the positive feedback cycle and change the patient trajectory. The predictive model this study presents gives the basis for a powerful clinical tool to predict ARDS patients who will develop DIC (with a sensitivity of 0.788) and potentially target treatments before onset.

It is notable that a key marker for ARDS patients who will eventually develop DIC is acidosis, specifically metabolic acidosis represented by a depressed bicarbonate level. A link between metabolic acidosis and the progression to DIC is not represented in the peer-reviewed literature.

DIC patients who later develop ARDS are significantly trauma patients. The stragtegy for treating DIC is to treat the underlying condition. It is reasonable that tissue factor in

the blood (leading to the DIC cascade) is more more pronounced in trauma patients than other underlying conditions that cause DIC. The difficulty in regulating pro-coagulants, when coagulants are needed in the presence of trauma, is likely the complicating factor that leads to DIC patients developing ARDS. The intelligent agents predict ARDS from DIC positive patients with an AUC of 0.675 (0.639, 0.711), a sensitivity of 0.478, a specificity of 0.756, and an accuracy of 0.746.

4.11 Summary

The intelligent agents were instructed to create predictive models of ARDS based on the Berlin definition [74]. As hypoxemia is a necessary condition this research effort wanted to explore, the intelligent agents were instructed to create a predictive model of Severe Acute Hypoxemic Respiratory Failure (SAHRF). The agents learned from 10% of the MIMIC database to create features. Those features were trained and tested (using 10-fold cross-validation) on 70% of the remaining MIMIC data. The resulting model was validated on the remaining 30% of data. The performance of the direct model on the validating data set is an AUC of 0.861 (0.838, 0.884). The performance of the SAHRF model on the validating data set is an AUC of 0.952 (0.947, 0.957).

A derivative work continued, that examines the relationship between ARDS and DIC. The agents were instructed to create a model to predict the onset of DIC from ARDS patients. Using a 10-fold cross-validation, the model has a performance of an AUC of 0.722 (0.689, 0.755), and notably a sensitivity of 0.788. DIC patients who develop ARDS were additionally examined. The sub-population of trauma patients who develop DIC are significantly at risk for then developing ARDS. The Intelligent agent's performance was an AUC of 0.675 (0.639, 0.711), a sensitivity of 0.478, a specificity of 0.756, and an accuracy of 0.746. DIC is shown to be a comorbidity in 78% of all ARDS patients who die in-hospital.

This chapter demonstrates the predictive power of an intelligent agent approach to predicting the onset of SAHRF, ARDS, ARDS from DIC patients, and DIC from ARDS patients. The ARDS model is superior and competitive with the existing LIPS model. There are no

SAHRF, DIC from ARDS patients, or ARDS from DIC patients predictive models in the peer-reviewed literature to compare.

Six hours prior to the patient meeting the Berlin ARDS criteria, this prediction model offer the basis of a clinical tool. ARDS protocols and DIC therapeutics may have significant beneficial value six hours prior to onset of the diseases, offering the clinician an ability to reduce mortality risk, reduce other adverse outcomes, or redirect the patient's trajectory entirely.

Chapter 5

ACUTE KIDNEY INJURY

5.1 Introduction

Acute Kidney Injury is the sudden decrease of kidney function, which causes waste products, normally filtered and expelled, to remain in the blood. The injury afflicts 7-18% of hospitalize patients, and 50% of ICU patients. Worldwide it is estimated to be the cause of 2 million deaths per year, as described by Chawla et al. [6]. The diagnostic criteria (namely increase in serum creatinine and decrease in urine output) occur after the injury has already affected kidney function. It is difficult to predict AKI events as the progression of the disease is complex. Clinically, it is practical to rely on a myriad of known risk factors and comorbidities to differentiate patients at risk.

This chapter is the culmination of five predictive models which characterize each stage of AKI and the progression of stages. The performances of the models are illustrated in Figure 5.1. The model to predict the onset of AKI Stage 1 is competitive with existing modes, with a superior sensitivity and AUC. The model to predict the onset of AKI Stage 2 is competitive with the only other model in the peer-reviewed literature, with a higher AUC. There is no predictive model in the peer-reviewed literature that predicts the onset of AKI Stage 3; this chapter presents one. This chapter also presents the only known work on machine learning algorithms predicting the progression from AKI Stage 1 to Stage 2 and from AKI Stage 2 to Stage 3.

5.2 Pathogenesis

The causal sources of an AKI can be broken into three categories: pre-renal (an external causal pathway preceding the kidneys); intrinsic (originating within the kidneys themselves);

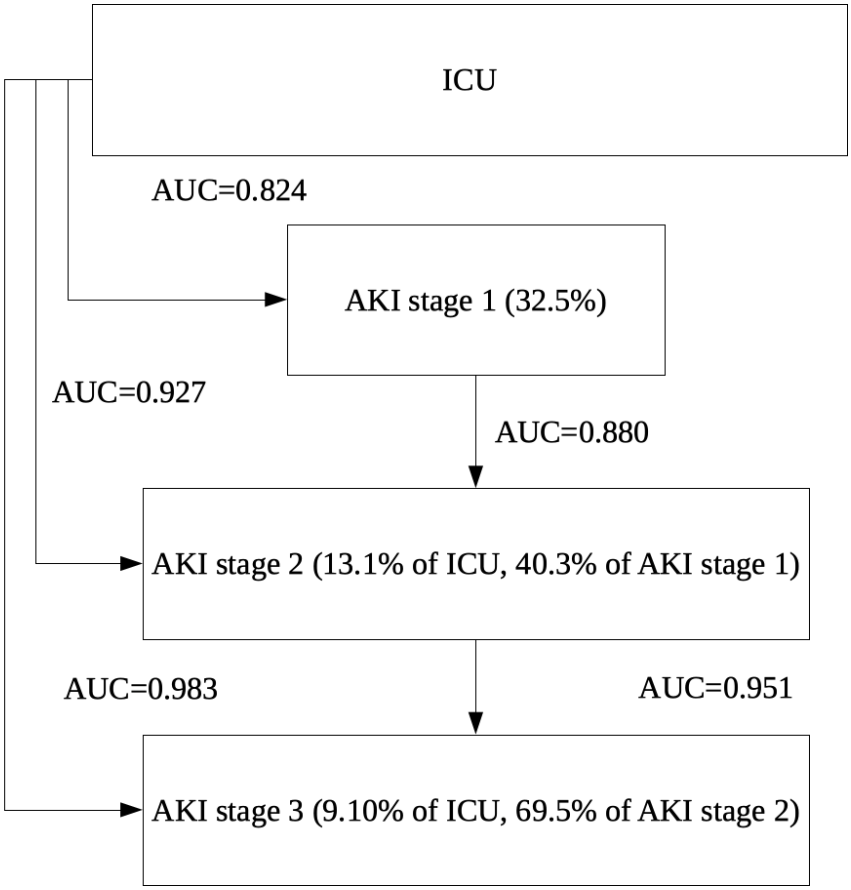


Figure 5.1: Predicting the stages and progression of AKI.

and post-renal (an external causal pathway after renal processing that affects the kidneys). Almost all causes of AKI are due to energy demands (from cellular stress) and oxygen-nutrition imbalance, as described by Makris et al. [51].

Pre-renal causes of AKI are hypoperfusion, or the lack of blood (and therefor nutrients and oxygen) reaching the kidneys, Pre-renal may be caused by: low blood volume (due to bleeding, trauma, or renal fluid loss); impaired heart function (due to congestive heart failure, heart attack, or pulmonary embolism); vasodialation; and increased vascular resistance.

Intrinsic causes of AKI are damage to the kidney structures and may be caused by: bleeding; medication; surgery complications; trauma; infections; and various diseases.

Post-renal causes of AKI are caused by a blockage that backs up fluid and increases pressure in the kidneys. This may be caused by blood clots; tumors; or catheters.

5.3 Value in Early Detection

When an AKI is diagnosed, unless contraindicated, intravenous fluid therapy should be started immediately. Renal replacement therapy (RRT) may be applied to combat hyperkalemia and metabolic acidosis. Oral phosphate binders or intravenous calcium may be administered to combat hypocalcemia and hyperphosphatemia. Three studies have been conducted to examine the effects of early versus delayed RRT. Gaudry et al. [29] found a decrease in the number of medical interventions related to hyperkalemia (elevated potassium in blood serum), metabolic acidosis, and diuretics. Zarbock et al. [96] and Barbar et al. [2] either lacked statistical power or were stopped by a data and safety board.

5.4 Diagnostic Criteria and Implementation

This study uses the Acute Kidney Injury Network (AKIN) definition, as described by Mehta et al. [57], to establish the onset of AKI of any stage. This definition has also been adopted by Kidney Disease Improving Global Guidelines (KDIGO) Clinical Practice Guidelines, as described by Khwaja [40]. This definition differs from another popular guideline, RIFLE. RIFLE considers injury at a higher baseline increase of serum creatinine, excludes an absolute increase in serum creatinine, and includes a glomerular filtration rate (GFR) as a biomarker.

The AKIN Stage 1 definition for injury includes: an absolute increase in serum creatinine by ≥ 0.3 mg/dl within 48 hours; an increase in serum creatinine to between 1.5 and 2.0 times baseline; urine volume ≤ 0.5 ml/kg/h for 6 hours. Creatinine baseline is calculated by the following equation from De Rosa et al. [12]:

$$SerumCreatinine = \frac{75}{186 * age^{-0.203}(0.742\_if\_female)(1.21\_if\_black))}^{-0.887} \tag{5.1}$$

The AKIN Stage 2 definition for injury includes: an increase in serum creatinine to

between 2.0 and 3.0 and times baseline; urine volume $\leq$ 0.5 ml/kg/h for 12 hours. The AKIN Stage 3 definition for injury includes: an absolute increase in serum creatinine by 4.0 mg/dl with an acute increase of 0.5 mg/dl; an increase in serum creatinine to $\geq$ 3.0 times baseline; urine volume $\leq$ 0.3 ml/kg/h for 24 hours.

The diagnostic criteria were computed for all patients in MIMIC. The laboratory test results database (including serum creatinine) encompasses the patients' entire hospital stay. The database table that includes urine volume is only applicable to the patient's ICU stay. It is possible to diagnose an AKI event before the patient arrives in the ICU. Patients whose first discernible AKI event occurs prior to their ICU stay are excluded from this study. Additionally, the modeling requires some data before the event may be predicted. Patients whose first discernible AKI event occurs within the first six hours of their ICU stay were excluded from this study. Patients with a direct kidney trauma (discernible by the ICD9 codes 866.00-866.13), numbering 49 or 0.08% were also excluded. Patients whose AKI event occurs after they leave the ICU were considered AKI negative, as the purpose of this study is to predict the onset, and those patients neither had an existing AKI nor had an AKI onset in the ICU. For AKI Stage progression, only the first progression from either Stage 1 to Stage 2 or from Stage 2 to Stage 3 were considered to eliminate over-representation of borderline patients who thrash between stages.

5.5 Competing Models

There exist three models from the literature with similar aims: to predict the onset of AKI. One additional study is the only AKI Stage 2 predictive model. Kate et al. [39] used a logistic regression on 25,521 hospital stays of 60 years or older patients to produce a predictive model with an AUC of 0.743. The group used the AKIN definition, but chose to exclude the urine condition of AKI due to the fact that the "value is frequently not monitored and can be missed till AKI is established." The study found that 2,258 (8.84%) of encounters included an AKI event. Cheng et al. [7] studied 109,319 hospital encounters and, using the AKIN definition excluding urine, found 4,405 (8.99%) AKI events. The group's best performing

model was a Random Forest which produced an AUC of 0.783 (PPV=0.721,TPR=0.655). He et al. [33] used a cohort of 76,957 to produce a predictive algorithm with a maximum AUC of 0.764 (0.762, 0.766). Mohamadlou et al. [59] use the MIMIC data set (in addition to a Stanford data set). Though the group used the KDIGO definition, they only sought to predict stage 2 and stage 3 AKI. Stages 2 and 3 are much more severe and pronounced than stage 1. This group used the XGBoost python package to produce a model with an AUC 0.841 (sensitivity=0.81, specificity=0.75, and accuracy=0.81). This model is compatible with MIMIC and can be used for the purposes of a valid comparison.

For this dissertation, both Kate et al. and Mohamadlou et al. models were re-implemented on MIMIC. This chapter's AKI definition was used: Stage 1 for Kate et al. and Stage 2 for Mohamadlou et al. Kate et al. model re-implementation had a performance with an AUC of 0.680 (0.661, 0.699) (sensitivity=0.594, specificity=0.687, accuracy=0.669). Mohamadlou et al. model re-implementation had a performance with an AUC of 0.918 (0.901, 0.935).

In contrast to the four models described above (Cheng et al., He et al., Kate et al., and Mohamadlou et al.), this study finds that ignoring the urine output condition of AKIN in the ICU setting would falsely label: 11.8% AKI positive patients as AKI negative patients; 12.4% AKI Stage 2 positive patients as negative; and 11.6% AKI Stage 3 positive patients as negative.

5.6 Intelligent Agent Model - AKI stage 1

5.6.1 Variable Selection

Using the 10% data set, the intelligent agents performed 2240 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model is described by Table 5.1. The variables selected for the model by the intelligent agents fall under categories: oxygen and carbon dioxide exchange; blood chemistry; urine makeup, functional and procedural, hemodynamics, and interventions.

Respiration variables include a low exhaled minimum volume, O_2 flow, and the partial pressure of venous O_2, which is elevated for AKI positive patients (P-value <0.0064 in Table E.1). The plateau pressure (pressure during a breath hold during mechanical ventilation) is slightly, but significantly, higher in condition positive patients (P-value <0.0001).

Several blood chemistry variables are significant in predicting the onset of AKI. Phosphate and potassium, both of which are excreted by the kidneys and are commonly elevated with AKI. Asparate aminotransferase (AST) is a blood assay that is primarily a marker for liver damage, but may also be indicative hypoperfusion, ischemic hepatitis, or damage to other organs that produce the enzyme-like muscle, heart or lung (elevated in condition positive patients P-value <0.0041). Chloride levels (elevated in condition positive patients, P-value <0.0001) are carefully regulated by the kidneys and play an important part in respiration via the Chloride Shift. Troponin is cleared by the kidneys and elevated in condition positive patients (P-value <0.0001) is a marker for heart attack, heart stress injury caused by hypoperfusion or sepsis, or chronic renal disease. Lipase is an enzyme that processes dietary fat. Lactic acid is a marker of hypoperfusion and elevated in condition positive patients (P-value <0.0069). Creatinine is a marker of renal health and elevated in condition positive patients (P-value <0.0001). Hemoglobin, despite males having a normal range 12.3%-16.1% higher than females and males comprising 64% of condition positive patients (P-value <0.0001 from Table E.3), is still significantly depressed in condition positive patients (P-value <0.0001). Lactate dehydrogenase is elevated (P-value <0.0001) and may indicate anemia as observed with a depressed hemoglobin. Basophils and lymphocytes are types of white blood cells, often elevated during inflammation or infection. Elevated urea nitrogen (P-value <0.0001) and glucose are also significant in predicting the onset of AKI. Hemodynamically, the central venous pressure (the pressure of the blood returning to the heart and ready to be pumped through the arteries) is slightly but significantly higher (P-value <0.0001). Diastolic pressure and arterial CO_2 pressure were also chosen variables.

Measurements taken at the bedside were included in the variables, including: impaired skin width; previous weight (elevated for condition positive patients P-value <0.0001); pain

level; visual/hearing deficit; Glasgow Coma Scale; and Left Ventricular Stroke Work Index. Significant interventions applied prior to the AKI onset include: Lasix (a diuretic); normal saline; dextrose.

Label	Variable Importance	AUC excluding variable
Low Exhaled Min Vol	0.39	0.820 (0.810,0.829)
Furosemide (Lasix)	0.4	0.821 (0.812,0.830)
Urea Nitrogen	0.44	0.813 (0.803,0.822)
Phosphate	0.41	0.817 (0.808,0.827)
pO_2	0.39	0.820 (0.811,0.830)
Central Venous Pressure	0.36	0.818 (0.809,0.828)
Glucose	0.38	0.820 (0.811,0.829)
NBP diastolic	0.35	0.821 (0.811,0.830)
AST	0.34	0.821 (0.812,0.830)
Potassium	0.35	0.817 (0.808,0.827)
Chloride (serum)	0.33	0.818 (0.809,0.827)
Impaired Skin Width	0.3	0.821 (0.812,0.831)
Basophils	0.34	0.820 (0.811,0.830)
Lipase	0.27	0.820 (0.810,0.829)
Previous Weight	0.3	0.820 (0.811,0.830)
Lymphocytes	0.31	0.819 (0.810,0.828)
Lactic Acid(0.5-2.0)	0.29	0.820 (0.810,0.829)
.9% Normal Saline	0.29	0.820 (0.810,0.829)
Arterial CO_2 Pressure	0.26	0.821 (0.812,0.830)
Troponin-T	0.26	0.819 (0.810,0.828)
Creatinine	0.29	0.817 (0.808,0.827)
5% Dextrose in water	0.28	0.818 (0.809,0.828)
Hemoglobin	0.27	0.821 (0.812,0.830)
Pain Level	0.26	0.819 (0.810,0.829)
Visual / hearing deficit	0.27	0.821 (0.812,0.831)
Plateau Pressure	0.23	0.821 (0.812,0.830)
Urine Out Foley	0.27	0.815 (0.806,0.825)
GCS Total	0.28	0.820 (0.811,0.829)
Left Ventricular Stroke Work Index	0.26	0.816 (0.807,0.826)
Lactate Dehydrogenase (LD)	0.24	0.822 (0.812,0.831)
O_2 Flow (lpm)	0.18	0.821 (0.811,0.830)

Table 5.1: Variables and variable importance of best performing AKI model generated by intelligent agents.

5.6.2 Model Training and Testing

Using the variables from Table 5.1, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 24,181 patients. At t-minus zero, the time immediately before the onset of AKI, the model has an area under the ROC curve of 0.831 (0.825, 0.837) (see Figure 5.2. The model performance on the training/testing data set is outlined in Table 5.2. With the cutoff set at the maximum value for the F-measure (beta=1.75) the model produces: 7,088 true positives; 8,887 true negatives; 7,329 false positives; 877 false negatives; a true positive rate (sensitivity) of 0.890 (see Figure E.1); a false positive rate of 0.454; a specificity of 0.548 (see Figure E.2); a positive predictive value of 0.492 (see Figure E.3; a negative predictive value of 0.910; an accuracy of 0.661 (see Figure E.4); with a kappa of 0.440 (see Figure E.5).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	3600	15154	1062	4365	0.452	0.065	0.452	0.935	0.772	0.776	0.776	0.736
1.50	6875	9767	6449	1090	0.863	0.398	0.863	0.602	0.516	0.900	0.688	0.450
1.75	7088	8887	7329	877	0.890	0.452	0.890	0.548	0.492	0.910	0.661	0.440

Table 5.2: AKI model performance on the training/testing data set.

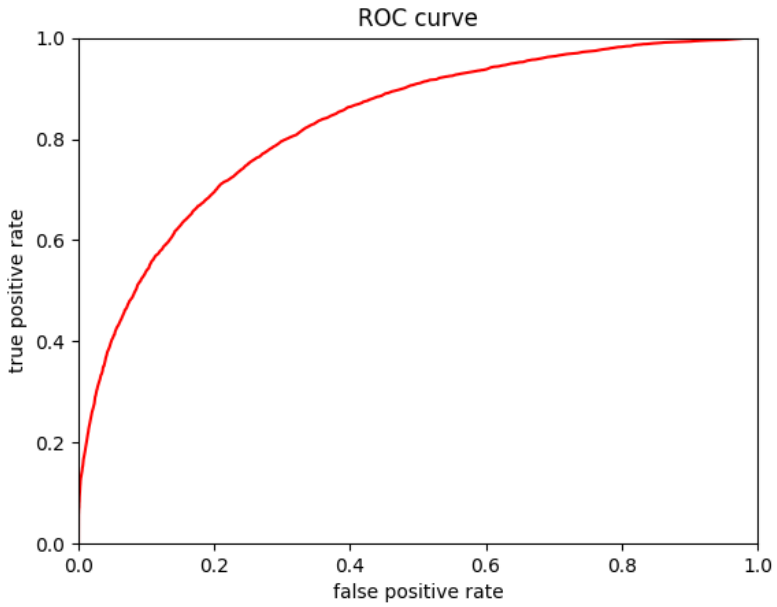


Figure 5.2: AKI patients, training/testing set AUC=0.831 (0.825, 0.837).

5.6.3 Model Validation

The threshold resulting for the maximum F-measure (beta=1.75) was used to set the threshold for the validation set containing 10,408 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of overfitting. The ROC curve in Figure 5.3 has an area under the curve of 0.824 (0.815, 0.833) compared to the training/testing are of 0.831 (0.825, 0.837). The performance of the model on the validation set may be found in Table 5.3 with 2,961 true positives, 3,834 true negatives, 3156 false positives, and 457 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.866 (see Figure E.7) as opposed to 0.890; a false positive rate of 0.452 identical to the training/testing corollary; a specificity of 0.548 (see Figure E.8) identical to the training/testing corollary; a positive predictive value of 0.484 (see Figure E.9) as opposed

to 0.492; a negative predictive value of 0.893 as opposed to 0.910; an accuracy of 0.653 (see Figure E.10) as opposed to 0.661; with a kappa of 0.392 (see Figure E.11) as opposed to 0.440.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	1666	6417	573	1752	0.487	0.082	0.487	0.918	0.744	0.786	0.777	0.732
1.50	2892	4087	2903	526	0.846	0.415	0.846	0.585	0.499	0.886	0.671	0.540
1.75	2961	3834	3156	457	0.866	0.452	0.866	0.548	0.484	0.893	0.653	0.392

Table 5.3: AKI model performance on the validation data set.

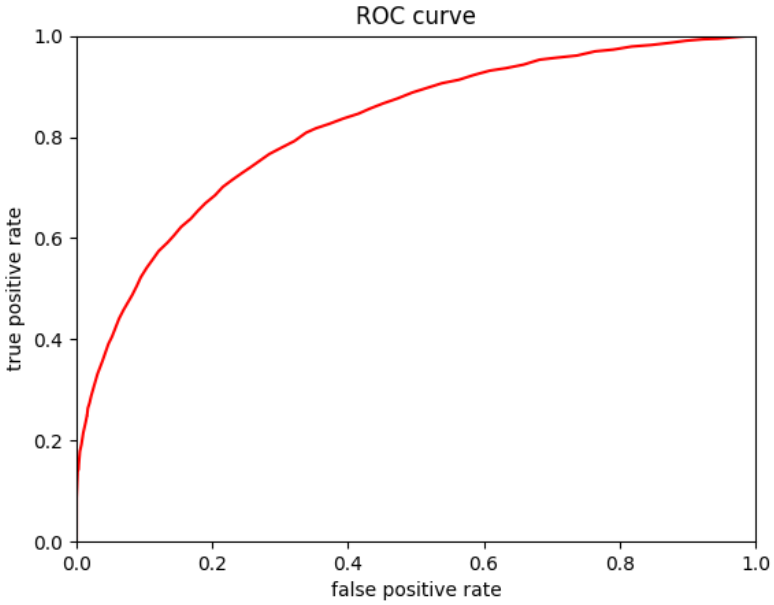


Figure 5.3: AKI patients, validation set AUC=0.824 (0.815, 0.833).

5.6.4 As a Function of Time

The AKI patients model is stable, losing 0.85% AUC with six hours prior variable values. In Figure 5.4 t-minus zero has an AUC of 0.824 (0.815, 0.833), and t-minus six hours, the models area under the ROC curve is 0.817 (0.809, 0.827).

Additional lower performing machine learning algorithms over time may be found in Figure E.13.

5.6.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table E.1. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI positive patients from AKI negative patients. AKI patients have a greater administration of Lasix (diuretic), measured urea nitrogen, phosphate, potassium, chloride, troponin, creatinine, previous weight, plateau pressure,

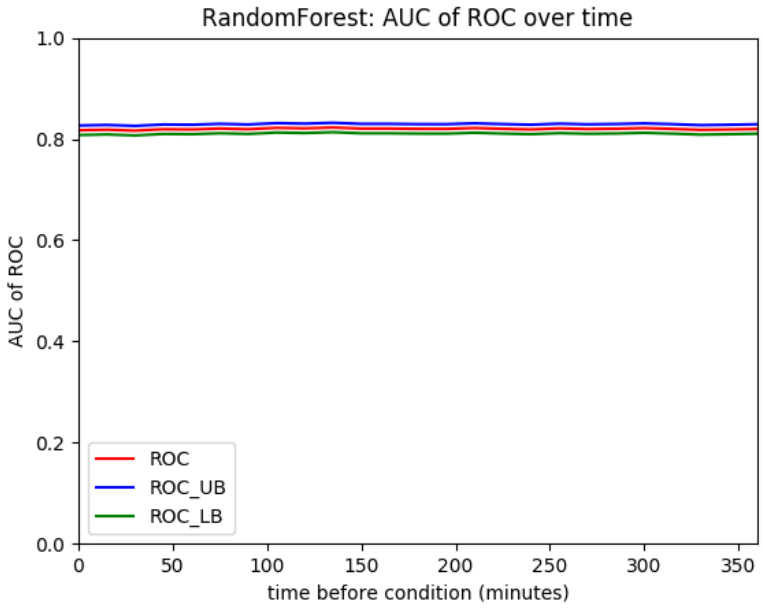


Figure 5.4: AKI validation AUC of the ROC over time. AUC at t-minus=0 is 0.824 (0.815, 0.833). At t-minus six hours, the model's area under the ROC curve is 0.817 (0.809, 0.827).

and lactate dehydrogenase. AKI positive patients had a different distribution of the left ventricular stroke work index, and a depressed hemoglobin.

Variable composition by classification was assessed in Table E.2 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they present with: depressed pO_2; normal central venous pressure; depressed AST; less application of 0.9% normal saline; depressed arterial CO_2 pressure; higher pain level; depressed urine output; and an elevated GCS.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table E.3. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI positive patients from AKI negative patients. Condition positive patients have a higher weight; correspondingly

higher BMI; are mostly male; have twice the in-hospital mortality; increased 30-day mortality; increased ICU length of stay; over represented in the Cardiac Surgery Recovery Unit and under represented in the Medical ICU, Surgical ICU, and Trauma Surgery ICU; more likely on Medicare and less likely on government or private insurance.

Demographic and key variables was assessed by classification in Table E.4. Of the statistically significant variables not included above, there are no significant misclassifications based on demographics.

Comorbidities of condition positive and condition negative patients are assessed in Table E.5. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate AKI positive patients from AKI negative patients. AKI positive patients are overrepresented in: congestive heart faiure; cardiac arrhythmias; valvular disease; pulmonary circulation disorders; peripheral vascular disorders; hypertension; uncomplicated diabetes; renal failure; liver disease; coagulopathy; weight loss; fluid and electrolyte disorders. AKI patients are underrepresented in trauma comorbidities.

Comorbidities by classification is assessed in Table E.6. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their comorbidities include: diabetes, complicated; and under-represented depression.

5.7 Intelligent Agent Model - AKI stage 2

AKI Stage 2 is a more severe form of AKI than Stage 1. AKI Stage 2 occurs when the serum creatinine increases to between 2.0 and 3.0 times baseline or the urine output volume drops to $\leq$ 0.5 ml/kg/h for 12 hours.

5.7.1 Variable Selection

Using the 10% data set, the intelligent agents performed 1269 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model is described by Table 5.4. The variables selected for the model by the intelligent agents include: IV access prior to admission; potas-

sium in the urine; fibrinogen; creatinine; gastro tube flush; administration of crystalloid fluid replacement; amylase; alanine aminotransferase (ALT) where high values are indicative of liver disease; arterial blood $PaCO_2$; alkaline phosphate; lactic acid; creatine kinase; epithelial cells in the urine; total iron binding capacity; percent hemoglobin A1C; urea nitrogen.

Label	Variable Importance	AUC excluding variable
IV access prior to admission	0.29	0.926 (0.915, 0.937)
Potassium, Urine	0.29	0.924 (0.913, 0.935)
Fibrinogen	0.28	0.925 (0.914, 0.937)
Creatinine	0.32	0.906 (0.893, 0.918)
GT Flush	0.23	0.924 (0.913, 0.935)
OR Crystalloid	0.2	0.925 (0.914, 0.936)
Amylase	0.28	0.924 (0.912, 0.935)
ALT	0.29	0.923 (0.912, 0.935)
Arterial $PaCO_2$	0.28	0.926 (0.915, 0.937)
Alk. Phosphate	0.3	0.926 (0.915, 0.937)
Lactic Acid	0.27	0.924 (0.912, 0.935)
Creatine Kinase	0.32	0.924 (0.913, 0.935)
Epithelial Cells, urine	0.18	0.924 (0.913, 0.935)
Iron Binding Capacity, Total	0.23	0.924 (0.913, 0.935)
percent Hemoglobin A1c	0.24	0.925 (0.914, 0.937)
Urea Nitrogen	0.32	0.898 (0.885, 0.911)

Table 5.4: Variables and variable importance of best performing AKI stage 2 model generated by intelligent agents.

5.7.2 Model Training and Testing

Using the variables from Table 5.4, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 18,523 patients. At t-minus zero, the time immediately before the onset of AKI, the model has an area under the ROC curve of 0.928 (0.921, 0.935) (see Figure 5.5. The model performance on the training/testing data set is outlined in Table 5.5. With the cutoff set at the maximum value for the F-measure (beta=1.75) the model produces: 1,903 true positives; 15,043 true negatives; 993 false positives; 584 false negatives; a true positive rate (sensitivity) of 0.765 (see Figure E.14); a false positive rate of 0.062; a specificity of 0.938 (see Figure E.15); a positive predictive value of 0.657 (see Figure E.16; a negative predictive value of 0.963; an accuracy of 0.915 (see Figure E.17); with a kappa of 0.975 (see Figure E.18).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	1407	15838	198	1080	0.566	0.012	0.566	0.988	0.877	0.936	0.931	0.974
1.50	1898	15061	975	589	0.763	0.061	0.763	0.939	0.661	0.962	0.916	0.975
1.75	1903	15043	993	584	0.765	0.062	0.765	0.938	0.657	0.963	0.915	0.975

Table 5.5: AKI stage 2 model performance on the training/testing data set.

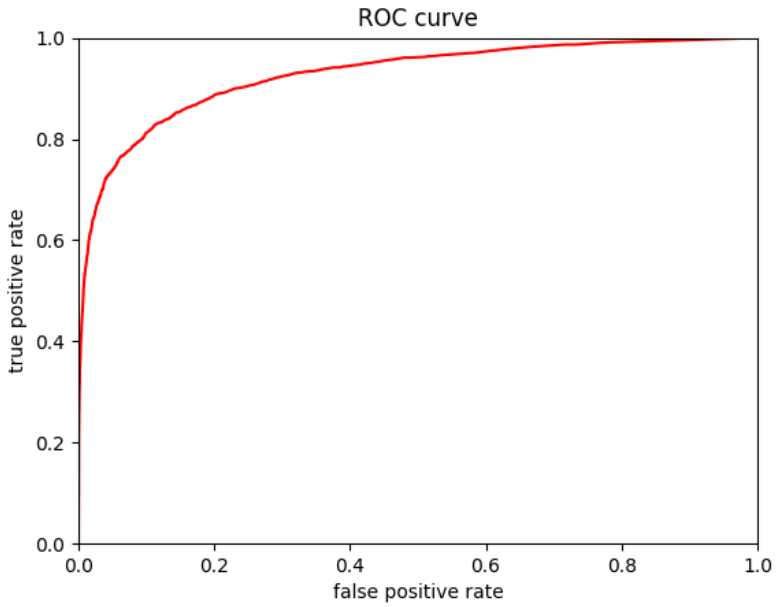


Figure 5.5: AKI stage 2 patients, training/testing set AUC=0.928 (0.921, 0.935).

5.7.3 Model Validation

The threshold resulting for the maximum F-measure (beta=1.75) was used to set the threshold for the validation set containing 7,957 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of overfitting. The ROC curve in Figure 5.6 has an area under the curve of 0.927 (0.916,0.938) compared to the training/testing are of 0.928 (0.921, 0.935). The performance of the model on the validation set may be found in Table 5.6 with 829 true positives, 6,448 true negatives, 456 false positives, and 233 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.779 (see Figure E.20) as opposed to 0.765; a false positive rate of 0.066 as opposed to 0.062; a specificity of 0.934 (see Figure E.21) as opposed to 0.938; a positive predictive value of 0.643 (see Figure E.22) as opposed to 0.657; a negative predictive value

of 0.965 as opposed to 0.963; an accuracy of 0.913 (see Figure E.23) as opposed to 0.915; with a kappa of 0.974 (see Figure E.11) as opposed to 0.975.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	612	6819	85	441	0.581	0.012	0.581	0.988	0.878	0.939	0.934	0.976
1.50	817	6457	447	236	0.776	0.065	0.776	0.935	0.646	0.965	0.914	0.974
1.75	820	6448	456	233	0.779	0.066	0.779	0.934	0.643	0.965	0.913	0.974

Table 5.6: AKI stage 2 model performance on the validation data set.

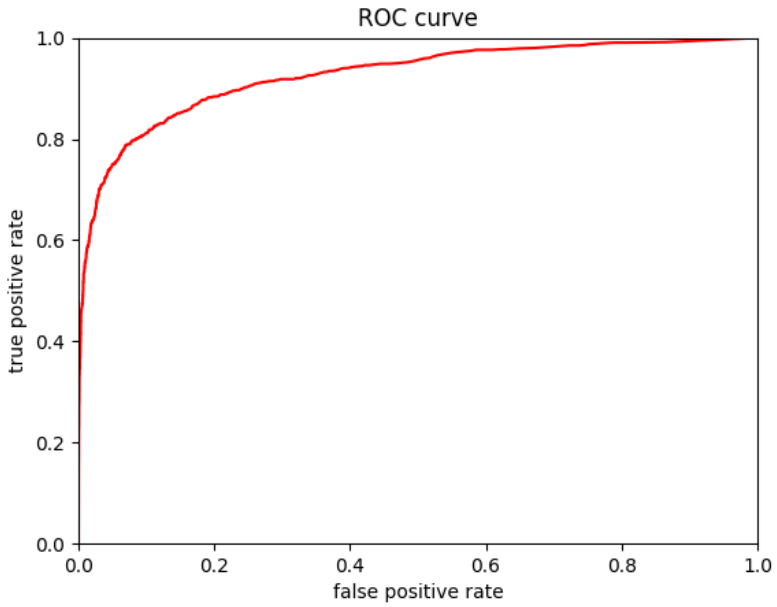


Figure 5.6: AKI stage 2 patients, validation set AUC=0.927 (0.916,0.938).

5.7.4 As a Function of Time

The AKI2 patients model is stable, losing merely 0.76% AUC with six hours prior variable values. In Figure 5.7 t-minus zero has an AUC of 0.927 (0.916, 0.938), and t-minus six hours, the models area under the ROC curve is 0.920 (0.908, 0.931).

Additional lower performing machine learning algorithms over time may be found in Figure E.26.

5.7.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table E.7. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI positive patients from AKI negative patients. AKI Stage 2 patients have a greater: urine potassium concentration, fibrinogen; creatinine; ALT; arterial $PaCO_2$; alkaline phosphate; lactic acid; creatine kinase; urea nitro-

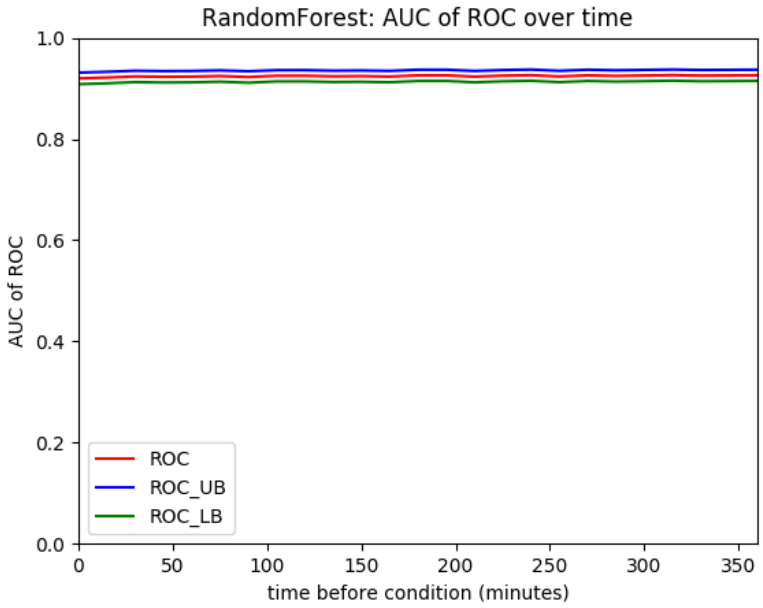


Figure 5.7: AKI stage 2 validation AUC of the ROC over time. AUC at t-minus=0 is 0.927 (0.916,0.938). At t-minus six hours, the model's area under the ROC curve is 0.920 (0.908, 0.931).

gen. AKI Stage 2 positive patients had a different distribution of: the IV access prior to admission; gastro tube flush; epithelial cells in the urine.

Variable composition by classification was assessed in Table E.8 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, there were no significant misclassifications based on variables.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table E.9. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI positive patients from AKI negative patients. Condition positive patients have a higher in-hospital and 30 day mortality; a longer ICU length of stay; are relatively over-represented in the CCU and CSRU;

under-represented in the TSICU; over-represented with Medicare and under-represented in private or government insurance (probably as a result of the age disparity between the two groups, median 71 years old for AKI Stage 2 positive as opposed to median 60 years old for AKI Stage 2 negative).

Demographic and key variables was assessed by classification in Table E.10. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their demographics include under-representation in the MICU.

Comorbidities of condition positive and condition negative patients are assessed in Table E.11. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate AKI Stage 2 positive patients from AKI Stage 2 negative patients. The pattern for statistically significant comorbidities for AKI Stage 2 is almost identical to that of AKI Stage 1. AKI Stage 2 positive patients are over-represented in: congestive heart failure; cardiac arrhythmias; valvular disease; pulmonary circulation disorders; peripheral vascular disorders; complicated hypertension; diabetes (complicated and uncomplicated); renal failure; liver disease; coagulopathy; weight loss; fluid and electrolyte disorders. AKI Stage 2 patients are underrepresented in trauma comorbidities.

Comorbidities by classification is assessed in Table E.12. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their comorbidities include: hypertension, uncomplicated; and under-representing of chronic pulmonary disease.

5.8 Intelligent Agent Model - AKI stage 3

AKI Stage 3 is a more severe form of AKI than Stages 1 and 2. AKI Stage 3 occurs when: serum creatinine is measured $\geq$ 4.0 mg/dL; serum creatinine acutely rises $\geq$ 0.5 mg/dL; serum creatinine is measured $\geq$ 3.0 times baseline; or urine output volume drops $\leq$ 3.0 ml/kg/h for 24 hours.

5.8.1 Variable Selection

Using the 10% data set, the intelligent agents performed 1058 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model is described by Table 5.7. The variables selected for the model by the intelligent agents include: body temperature; PO intake (per os, by mouth administration of medication)l Hayline casts (a cylindrical structure produced and secreted by the kidney concurrent with disease or low urine flow); Gastrotube flush (likely a surrogate variable for intubated patients); gastric medication (likely a surrogate variable for sicker patients, as it is common to administer proton-pump inhibitors, a type of gastric medication, to reduce the risk of stomach ulcers or bleeding); blood gas oxygen; albumin (a blood serum protein); NTproBNP (N-terminal pro b-type Natriuretic Peptide, a marker for heart failure); both blood and urine urea nitrogen (a marker for kidney failure, heart failure, or liver failure).

Label	Variable Importance	AUC excluding variable
Temperature	0.33	0.962 (0.972, 0.952)
PO Intake	0.23	0.966 (0.975, 0.956)
Hyaline Casts	0.31	0.965 (0.975, 0.956)
Sodium, Urine	0.25	0.964 (0.974, 0.954)
Creatinine	0.26	0.933 (0.946, 0.920)
GT Flush	0.21	0.968 (0.977, 0.959)
Oxygen, blood gas	0.22	0.962 (0.972, 0.952)
Urea Nitrogen, blood	0.24	0.963 (0.973, 0.953)
Albumin	0.23	0.968 (0.977, 0.959)
NTproBNP	0.12	0.966 (0.975, 0.956)
Gastric Meds	0.13	0.966 (0.975, 0.956)
Urea Nitrogen, Urine	0.14	0.966 (0.976, 0.957)

Table 5.7: Variables and variable importance of best performing AKI stage 3 model generated by intelligent agents.

5.8.2 Model Training and Testing

Using the variables from Table 5.7, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 17,554 patients. At t-minus zero, the time immediately before the onset of AKI stage 3, the model has an area under the ROC curve of 0.980 (0.975, 0.985) (see Figure 5.8. The model performance on the training/testing data set is outlined in Table 5.8. With the cutoff set at the maximum value for the F-measure (*beta*=1.75) the model produces: 1,459 true positives; 15,877 true negatives; 21 false positives; 197 false negatives; a true positive rate (sensitivity) of 0.881 (see Figure E.27); a false positive rate of 0.001; a specificity of 0.999 (see Figure E.28); a positive predictive value of 0.986 (see Figure E.29; a negative predictive value of 0.988; an

accuracy of 0.988 (see Figure E.30); with a kappa of 0.986 (see Figure E.31).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	1457	15877	21	199	0.880	0.001	0.880	0.999	0.986	0.988	0.987	0.986
1.50	1459	15877	21	197	0.881	0.001	0.881	0.999	0.986	0.988	0.988	0.986
1.75	1459	15877	21	197	0.881	0.001	0.881	0.999	0.986	0.988	0.988	0.986

Table 5.8: AKI stage 3 model performance on the training/testing data set.

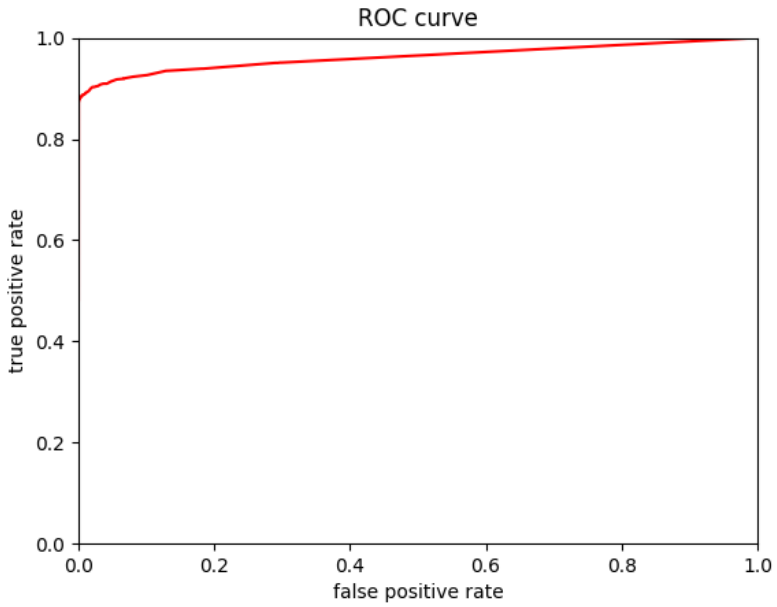


Figure 5.8: AKI stage 3 patients, training/testing set AUC=0.980 (0.975, 0.985).

5.8.3 Model Validation

The threshold resulting for the maximum F-measure (*beta*=1.75) was used to set the threshold for the validation set containing 7,531 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of overfitting. The ROC curve in Figure 5.9 has an area under the curve of 0.983 (0.976, 0.990) compared to the training/testing are of 0.980 (0.975, 0.985). The performance of the model on the validation set may be found in Table 5.9 with 604 true positives, 6,837 true negatives, 11 false positives, and 79 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.884 (see Figure E.33) as opposed to 0.881; a false positive rate of 0.002 as opposed to 0.001; a specificity of 0.998 (see Figure E.34) as opposed to 0.999; a positive predictive value of 0.988 (see Figure E.35), the same as the training/testing corollary; a negative predictive

value of 0.989 as opposed to 0.988; an accuracy of 0.893 (see Figure E.36) as opposed to 0.881; with a kappa of 0.986 (see Figure E.37) which is the same as the training/testing corollary.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	604	6839	9	79	0.884	0.001	0.884	0.999	0.985	0.989	0.988	0.987
1.50	604	6837	11	79	0.884	0.002	0.884	0.998	0.982	0.989	0.988	0.986
1.75	604	6837	11	79	0.884	0.002	0.884	0.998	0.982	0.989	0.988	0.986

Table 5.9: AKI stage 3 model performance on the validation data set.

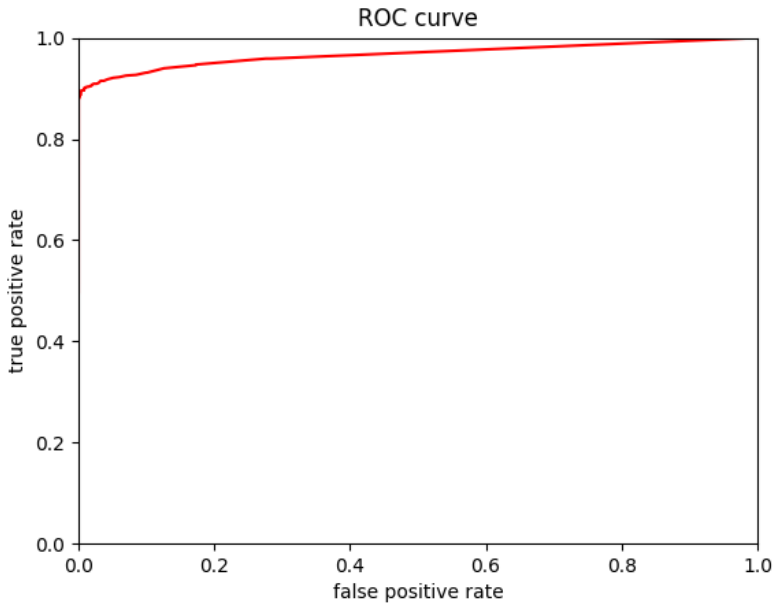


Figure 5.9: AKI stage 3 patients, validation set AUC=0.983 (0.976, 0.990).

5.8.4 As a Function of Time

The AKI stage 3 patients model is stable over time, merely losing 1.83% AUC with six hours prior variable values. In Figure 5.10 t-minus zero has an AUC of 0.983 (0.976, 0.990), and t-minus six hours, the models area under the ROC curve is 0.965 (0.955, 0.975). Additional lower performing machine learning algorithms over time may be found in Figure E.39.

5.8.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table E.13. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI stage 3 progressive positive patients from AKI stage 3 negative patients. AKI stage 3 positive patients have a greater: per os intake of medication; number of Hayline casts; concentration of sodium in urine; creatinine; blood oxygen concentration; urea nitrogen concentration (both in blood and

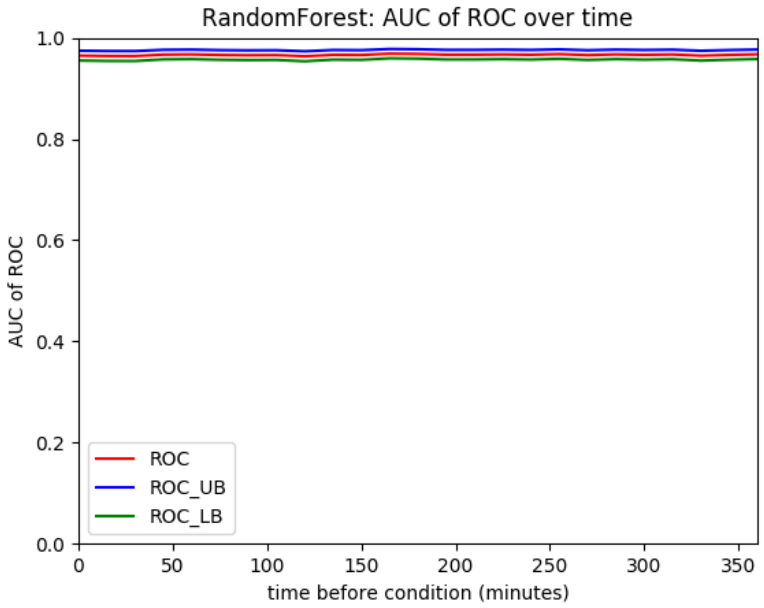


Figure 5.10: AKI stage 3 validation AUC of the ROC over time. AUC at t-minus=0 is 0.983 (0.976, 0.990). At t-minus six hours, the model's area under the ROC curve is 0.965 (0.955, 0.975).

in urine); NTproBNP; number of gastric medications administered. AKI stage 3 positive patients have a depressed temperature and albumin concentration in blood.

Variable composition by classification was assessed in Table E.14 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, there are no significant misclassifications based on variables.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table E.15. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI stage 3 positive patients from AKI stage 3 negative patients. Condition positive patients have: five times as high in-hospital and 30 day mortality; a longer ICU length of stay; over-representation in the

Critical Care Unit; under-represented in the Trauma Surgery ICU; are over-represented in Medicare and under-represented in private insurance (mostly likely due to the age disparity of the two groups, median of 70 years old for condition positive and median of 60 years old for condition negative).

Demographic and key variables was assessed by classification in Table E.16. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their demographics include Black racial designation.

Comorbidities of condition positive and condition negative patients are assessed in Table E.17. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate AKI stage 3 progressive positive patients from stage 3 negative patients. AKI stage 3 positive patients are significantly over-represented in: congestive heart failure; cardiac arrhythmias; valvular disease; pulmonary circulation disorders; complicated hypertension; diabetes; renal failure; liver failure; coagulopathy; weight loss; fluid and electrolyte disorders. The AKI stage 3 positive patients are under-represented in the uncomplicated hypertension comorbidity.

Comorbidities by classification is assessed in Table E.18. Of the statistically significant variables not included above, there are no misclassifications based on comorbidities.

5.9 Intelligent Agent Model - AKI stage 1 progression to AKI stage 2

5.9.1 Variable Selection

Using the 10% data set, the intelligent agents performed 1045 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model is described by Table 5.10. The variables selected for the model by the intelligent agents are: blood-gas oxygen; alanine aminotranserase (ALT); high density lipid cholesterol; urea nitrogen; blood phosphate; Fentanyl (likely a surrogate variable for intubated patients, as it is commonly used as a sedative for those patients to avoid a struggle with the intrusive tubes); gastric medication (likely a

surrogate variable for the sicker and intubated patients commonly given to avoid stomach ulcers and bleeding).

Label	Variable Importance	AUC excluding variable
Fentanyl	0.29	0.846 (0.831, 0.862)
Oxygen, blood gas	0.29	0.853 (0.838, 0.869)
Alanine Aminotransferase (ALT)	0.32	0.849 (0.833, 0.864)
Cholesterol, HDL	0.26	0.859 (0.844, 0.874)
Urea Nitrogen	0.32	0.828 (0.812, 0.844)
Gastric Meds	0.24	0.856 (0.841, 0.872)
Phosphate, blood chemistry	0.3	0.847 (0.832, 0.863)

Table 5.10: Variables and variable importance of best performing AKI progression from stage 1 to stage 2 model generated by intelligent agents.

5.9.2 Model Training and Testing

Using the variables from Table 5.10, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 9,084 patients. At t-minus zero, the time immediately before the progression from AKI stage 1 to AKI stage 2, the model has an area under the ROC curve of 0.875 (0.866, 0.884) (see Figure 5.11. The model performance on the training/testing data set is outlined in Table 5.11. With the cutoff set at the maximum value for the F-measure ($beta$=1.00) the model produces: 1472 true positives; 6233 true negatives; 386 false positives; 993 false negatives; a true positive rate (sensitivity) of 0.597 (see Figure E.40); a false positive rate of 0.058; a specificity of 0.942 (see Figure E.41); a positive predictive value of 0.792 (see Figure E.42; a negative predictive value of 0.863; an accuracy of 0.848 (see Figure E.43); with a kappa of 0.973 (see Figure E.44).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	1472	6233	386	993	0.597	0.058	0.597	0.942	0.792	0.863	0.848	0.973
1.50	1997	5271	1348	468	0.810	0.204	0.810	0.796	0.597	0.918	0.800	1.000
1.75	2062	5000	1619	403	0.837	0.245	0.837	0.755	0.560	0.925	0.777	0.998

Table 5.11: AKI progression from stage 1 to stage 2 model performance on the training/testing data set.

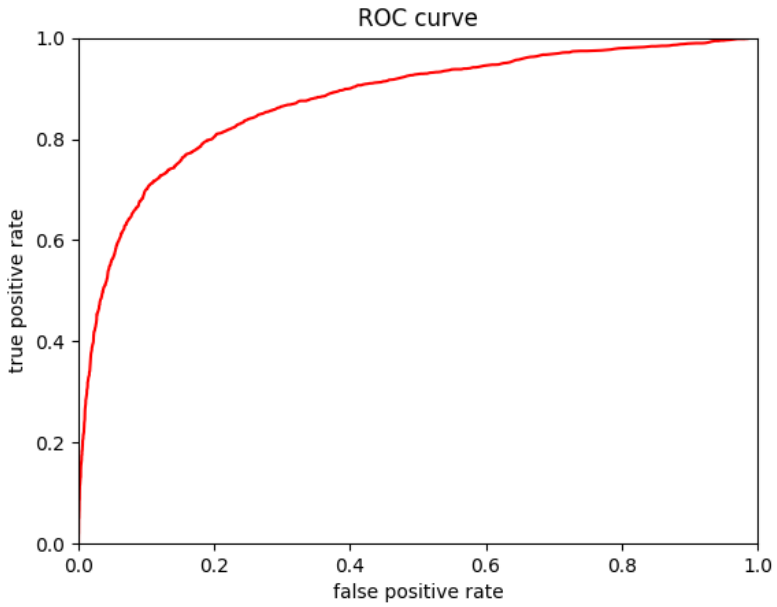


Figure 5.11: AKI progression from stage 1 to stage 2 patients, training/testing set AUC=0.875 (0.866, 0.884).

5.9.3 Model Validation

The threshold resulting for the maximum F-measure (*beta*=1.00) was used to set the threshold for the validation set containing 3,924 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of overfitting. The ROC curve in Figure 5.12 has an area under the curve of 0.880 (0.871, 0.889) compared to the training/testing are of 0.875 (0.866, 0.884). The performance of the model on the validation set may be found in Table 5.12 with 256 true positives, 3,427 true negatives, 157 false positives, and 84 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.753 (see Figure E.46) as opposed to 0.597; a false positive rate of 0.044

as opposed to 0.058; a specificity of 0.956 (see Figure E.47) as opposed to 0.942; a positive predictive value of 0.620 (see Figure E.48) as opposed to 0.792; a negative predictive value of 0.976 as opposed to 0.863; an accuracy of 0.939 (see Figure E.49) as opposed to 0.848; with a kappa of 0.976 (see Figure E.50) as opposed to 0.973.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	256	3427	157	84	0.753	0.044	0.753	0.956	0.620	0.976	0.939	0.976
1.50	306	2863	721	34	0.900	0.201	0.900	0.799	0.298	0.988	0.808	0.911
1.75	316	2548	1036	24	0.929	0.289	0.929	0.711	0.234	0.991	0.730	0.866

Table 5.12: AKI progression from stage 1 to stage 2 model performance on the validation data set.

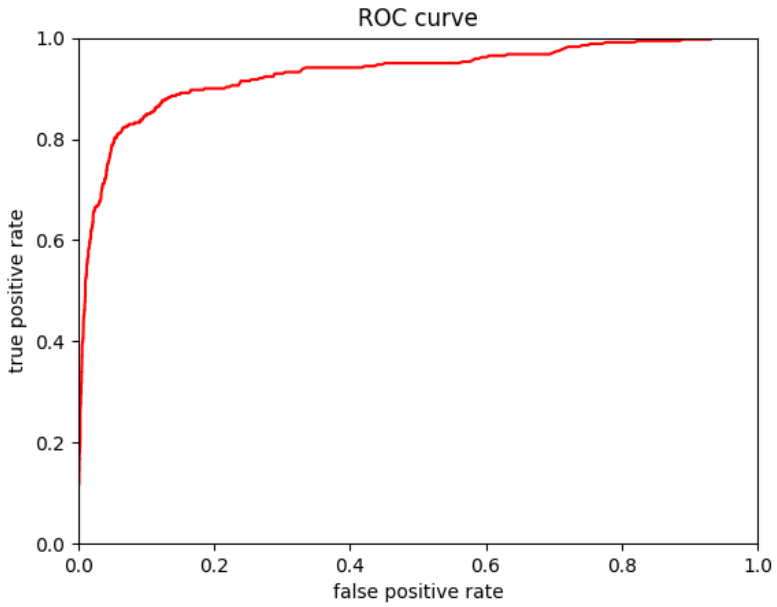


Figure 5.12: AKI progression from stage 1 to stage 2 patients, validation set AUC=0.880 (0.871, 0.889).

5.9.4 As a Function of Time

The AKI Stage 1 to Stage 2 progression model is time sensitive, losing 3.30% AUC with six hours prior variable values. In Figure 5.13 t-minus zero has an AUC of 0.880 (0.871, 0.889)), and t-minus six hours, the models area under the ROC curve is 0.851 (0.836, 0.866). Additional lower performing machine learning algorithms over time may be found in Figure E.52.

5.9.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table E.19. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI stage 1 to stage 2 progressive positive

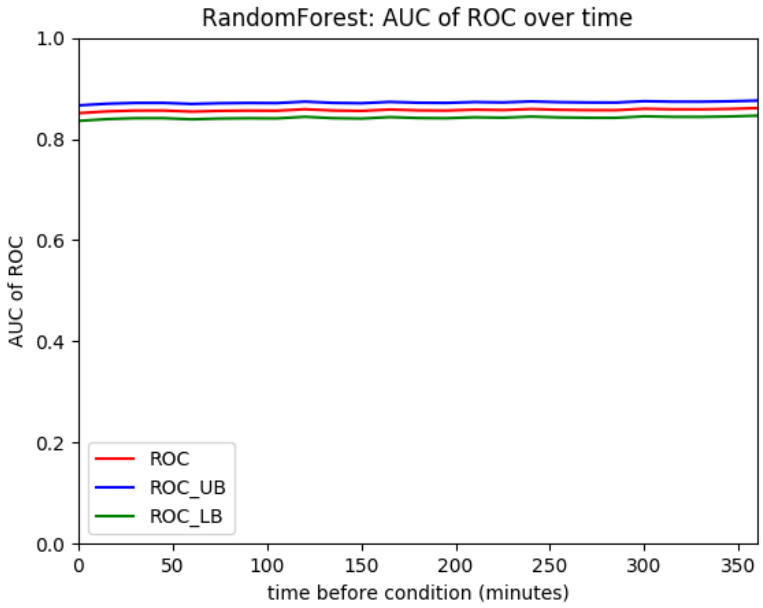


Figure 5.13: AKI progression from stage 1 to stage 2 validation AUC of the ROC over time. AUC at t-minus=0 is 0.880 (0.871, 0.889). At t-minus six hours, the model's area under the ROC curve is 0.851 (0.836, 0.866).

patients from AKI stage 1 to stage 2 negative patients. AKI stage 2 positive patients have a greater: fentanyl administration; blood oxygen; ALT; urea nitrogen; gastric medication administration; blood phosphate.

Variable composition by classification was assessed in Table E.20 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, there are no significant misclassifications based on variables.

In both demographic analysis and comorbidity analysis there were no significant findings (observable in Tables E.21, E.22, E.23, E.24).

5.10 Intelligent Agent Model - AKI stage 2 progression to AKI stage 3

5.10.1 Variable Selection

Using the 10% data set, the intelligent agents performed 1045 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model is described by Table 5.13. The variables selected for the model by the intelligent agents fall under categories: oxygen and carbon dioxide exchange; blood chemistry; administrations of pharmaceuticals.

The variable pCO_2 is the partial pressure of CO_2 in the venous blood. Serum osmolality is a measure of how much substance is dissolved in blood serum, and is a known indicator of kidney damage. Lactate dehydrogenase is an enzyme involved in energy production. Creatinine is a marker for renal health, and is a diagnostic criteria for AKI. Whole blood sodium, urea nitrogen, hemoglobin, calcium and ALT all measure the concentrations present in blood. D5/.45NS is 5% dextrose in 0.45% normal saline solution, indicative of the patient requiring intravenous fluid. Fentanyl is a narcotic to treat severe pain. Misazolam is a sedative used prior to surgery or invasive medical procedures. Cefazolin is an antibiotic used to treat serious bacterial infections.

Label	Variable Importance	AUC excluding variable
pCO_2	0.35	0.936 (0.922, 0.949)
Fentanyl	0.31	0.940 (0.927, 0.953)
Osmolality	0.25	0.933 (0.919, 0.947)
Lactate Dehydrogenase (LD)	0.33	0.932 (0.918, 0.946)
Creatinine	0.43	0.752 (0.726, 0.777)
D5/.45NS	0.2	0.934 (0.920, 0.947)
Sodium (whole blood)	0.31	0.937 (0.924, 0.951)
Urea Nitrogen	0.32	0.932 (0.918, 0.946)
Midazolam (Versed)	0.25	0.933 (0.919, 0.947)
Alanine Aminotransferase (ALT)	0.31	0.935 (0.921, 0.948)
Calcium, Total	0.3	0.933 (0.919, 0.947)
Hemoglobin	0.28	0.936 (0.923, 0.950)
Cefazolin	0.15	0.934 (0.920, 0.948)

Table 5.13: Variables and variable importance of best performing AKI progression from stage 2 to stage 3 model generated by intelligent agents.

5.10.2 Model Training and Testing

Using the variables from Table 5.13, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 3,240 patients. At t-minus zero, the time immediately before the progression from AKI stage 2 to AKI stage 3, the model has an area under the ROC curve of 0.939 (0.930, 0.948) (see Figure 5.14. The model performance on the training/testing data set is outlined in Table 5.14. With the cutoff set at the maximum value for the F- measure ($beta$=1.00) the model produces: 1,331 true positives; 1,522 true negatives; 62 false positives; 325 false negatives; a true positive rate (sensitivity) of 0.804 (see Figure E.53); a false positive rate of 0.039; a specificity of 0.961 (see Figure E.54); a positive predictive value of 0.955 (see Figure E.55; a negative predictive value of

0.824; an accuracy of 0.881 (see Figure E.56); with a kappa of 0.808 (see Figure E.57).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	1331	1522	62	325	0.804	0.039	0.804	0.961	0.955	0.824	0.881	0.808
1.50	1451	1381	203	205	0.876	0.128	0.876	0.872	0.877	0.871	0.874	0.773
1.75	1490	1266	318	166	0.900	0.201	0.900	0.799	0.824	0.884	0.851	0.716

Table 5.14: AKI progression from stage 2 to stage 3 model performance on the training/testing data set.

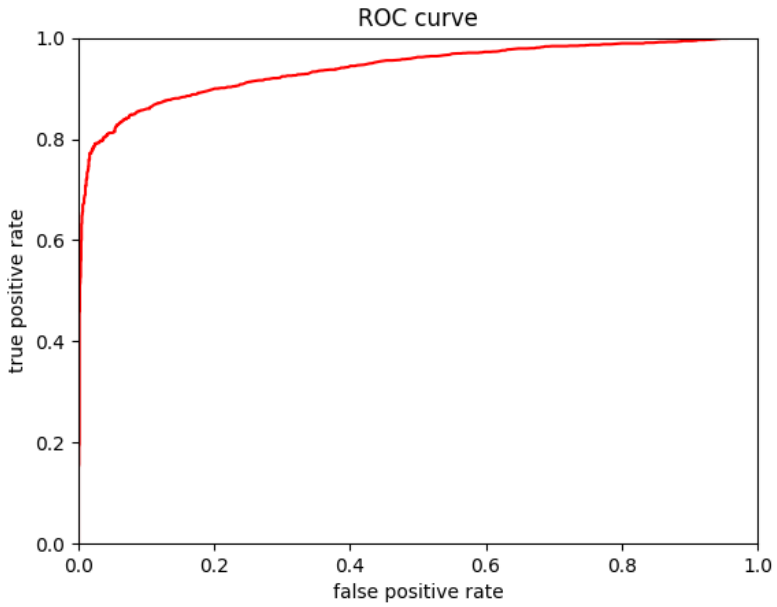


Figure 5.14: AKI progression from stage 2 to stage 3 patients, training/testing set AUC=0.939 (0.930, 0.948).

5.10.3 Model Validation

The threshold resulting for the maximum F-measure (*beta*=1.00) was used to set the threshold for the validation set containing 1,372 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of overfitting. The ROC curve in Figure 5.15 has an area under the curve of 0.951 (0.939, 0.963) compared to the training/testing are of 0.939 (0.930, 0.948). The performance of the model on the validation set may be found in Table 5.15 with 572 true positives, 653 true negatives, 33 false positives, and 114 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.834 (see Figure E.59) as opposed to 0.804; a false positive rate of 0.048

as opposed to 0.039; a specificity of 0.952 (see Figure E.60) as opposed to 0.961; a positive predictive value of 0.945 (see Figure E.61) as opposed to 0.955; a negative predictive value of 0.851 as opposed to 0.824; an accuracy of 0.893 (see Figure E.62) as opposed to 0.881; with a kappa of 0.823 (see Figure E.63) as opposed to 0.808.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	572	653	33	114	0.834	0.048	0.834	0.952	0.945	0.851	0.893	0.823
1.50	614	606	80	72	0.895	0.117	0.895	0.883	0.885	0.894	0.889	0.800
1.75	625	558	128	61	0.911	0.187	0.911	0.813	0.830	0.901	0.862	0.800

Table 5.15: AKI progression from stage 2 to stage 3 model performance on the validation data set.

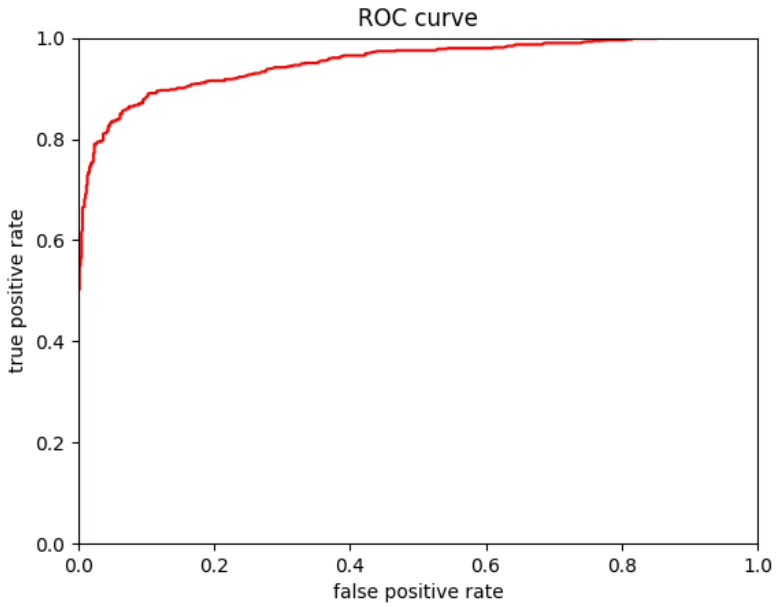


Figure 5.15: AKI progression from stage 2 to stage 3 patients, validation set AUC=0.951 (0.939, 0.963).

5.10.4 As a Function of Time

The AKI stage 2 to stage 3 patients model is stable over time, merely losing 0.736% AUC with six hours prior variable values. In Figure 5.16 t-minus zero has an AUC of 0.951 (0.939, 0.963), and t-minus six hours, the models area under the ROC curve is 0.944 (0.931, 0.956). Additional lower performing machine learning algorithms over time may be found in Figure E.65.

5.10.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table E.25. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI stage 2 to stage 3 progressive positive

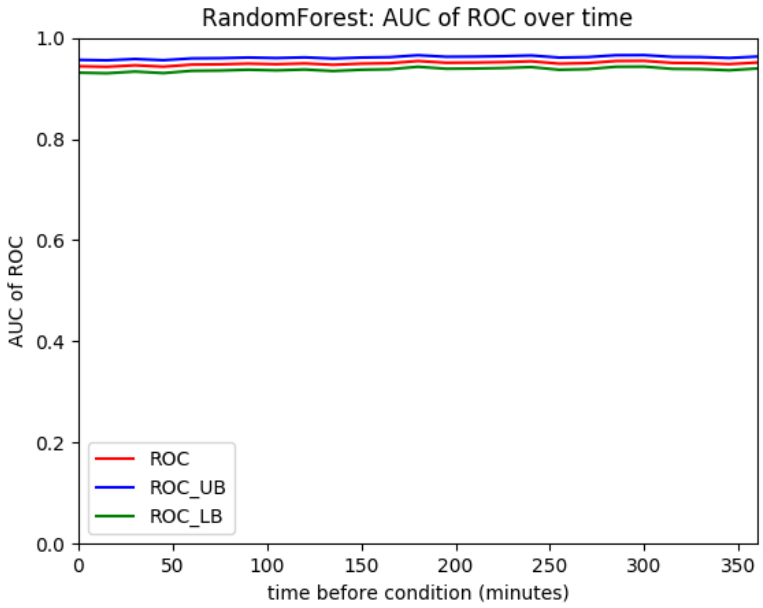


Figure 5.16: AKI progression from stage 2 to stage 3 validation AUC of the ROC over time. AUC at t-minus=0 is 0.951 (0.939, 0.963). At t-minus six hours, the model's area under the ROC curve is 0.944 (0.931, 0.956).

patients from AKI stage 2 to stage 3 negative patients. AKI stage 3 positive patients have a greater lactate dehydrogenase, creatinine, urea nitrogen, and were administered Misazolam.

Variable composition by classification was assessed in Table E.26 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they present with depressed ALT.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table E.27. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate AKI stage 2 to stage 3 progressive positive patients from AKI stage 3 negative patients. Progressive positive patients had a greater ICU length of stay and were under-represented in the Cardiac Surgery

Recovery Unit.

Demographic and key variables was assessed by classification in Table E.28. Of the statistically significant variables not included above, patients are less likely to be misclassified as a false negative if their demographics include a Black racial designation.

Comorbidities of condition positive and condition negative patients are assessed in Table E.29. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate AKI stage 2 to stage 3 progressive positive patients from AKI progressive negative patients. AKI progressive positive patients are overrepresented in: Hypertension, complicated; fluid and electrolyte disorders. AKI progressive patients are underrepresented in uncomplicated hypertension comorbidities.

Comorbidities by classification is assessed in Table E.30. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their comorbidities include: an under-represetnation of renal failure; and an under-representation of liver disease.

5.11 Model Comparisons

This study compares intelligent agent models of AKI to three published predictive model, none of which are ICU specific. Two of the three are re-implement using the MIMIC database. The third, (Cheng et al. and He et al. are effectively the same model with overlapping authors), is not published in enough detail to re-implement. A comparison of the included models with this study's models may be found in Table 5.16.

5.11.1 Stage 1

Kate et al. published a model (for 60 year old patients and older) with an AUC of 0.743. Kate et al. published the best reproducible model when re-implemented using MIMIC produced an AUC of 0.680 (0.661, 0.699). Cheng et al. produced an AUC of 0.783, and He et al. (effectively the same model) produced an AUC of 0.764 (0.762, 0.766). The intelligent agents produce an AUC of 0.824 (0.815, 0.833).

5.11.2 Stage 2

Mohamadlou et al. published a predictive model for specifically Stage 2 and Stage 3 AKI, which are the more pronounced. Mohamadlou et al. produce a model with an AUC of 0.841, while a re-implementation of the model on this studies MIMIC produces an AUC of 0.918 (0.901, 0.935). The intelligent agents produced an AUC of 0.926 (0.915, 0.937).

5.11.3 Stage 3

There is no prediction model of AKI Stage 3 in the peer-reviewed literature. This study presents one with an AUC of 0.983 (0.975, 0.985), a sensitivity of 0.884, a specificity of 0.998, and an accuracy of 0.988.

5.11.4 Progression from Stage 1 to Stage 2

There is no prediction model of AKI Stage 1 to Stage 2 progression in the peer-reviewed literature. This study presents one with an AUC of 0.930 (0.911, 0.949), a sensitivity of 0.75, a specificity of 0.873, and an accuracy of 0.845.

5.11.5 Progression from Stage 2 to Stage 3

There is no prediction model of AKI Stage 2 to Stage 3 progression in the peer-reviewed literature. This study presents one with an AUC of 0.951 (0.939, 0.963), a sensitivity of 0.834, a specificity of 0.952, and an accuracy of 0.893.

	AUC	Kappa	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN	ACC	TPR	FPR
AKI Stage 1													
Kate et al. [39]	0.743	0.489	0.75	0.611	0.158	0.962	1	8.393	5.344	0.333	0.623	0.75	0.389
Kate et al. On MIMIC	0.680 (0.661, 0.699)	0.562	0.594	0.687	0.309	0.878	385	1890	862	263	0.669	0.594	0.313
Cheng et al. [7]	0.783	0.942	0.665	0.975	0.721	0.967	1	14.821	0.387	0.504	0.947	0.665	0.025
He et al. [33]	0.764 (0.762, 0.766)	0.966	0.181	0.983	0.187	0.983	1	254.049	4.348	4.525	0.966	0.181	0.017
Intelligent Agent, this study	0.824 (0.815, 0.833)	0.437	0.866	0.548	0.484	0.893	2961	3834	3156	457	0.653	0.866	0.452
AKI Stage 2													
Mohamadlou et al. [59]	0.841	0.688	0.81	0.75	0.129	0.989	1	20.25	6.75	0.235	0.81	0.81	0.25
Mohamadlou et al. On MIMIC	0.918 (0.901, 0.935)	0.902	0.706	0.969	0.846	0.931	357	2001	65	149	0.917	0.706	0.031
Intelligent Agent, this study	0.926 (0.915, 0.937)	0.899	0.779	0.934	0.643	0.965	820	6448	456	233	0.913	0.779	0.066
AKI Stage 3													
Intelligent Agent, this study	0.983 (0.975, 0.985)	0.987	0.884	0.998	0.982	0.989	604	6837	11	79	0.988	0.884	0.002
AKI progression Stage 1 to 2													
Intelligent Agent, this study	0.930 (0.911, 0.949)	0.8	0.75	0.873	0.631	0.923	662	2654	387	221	0.845	0.750	0.127
AKI progression Stage 2 to 3													
Intelligent Agent, this study	0.951 (0.939, 0.963)	0.823	0.834	0.952	0.945	0.851	572	653	33	114	0.893	0.834	0.048

Table 5.16: AKI model comparison. Values denoted with an asterisk (*) are derived from imputing the confusion matrix ratio. Values denoted with a dagger (†) are ratios relative to true positive value of one, and derived from the equations in the Model Comparison section.

5.12 Discussion

This study covers AKIs in 11,383 patients with an aim to build prediction models, which is almost twice as large as all similarly published predictive studies combined. This dissertation presents a Stage by Stage characterization of AKI with superior performance to existing peer-reviewed models at every Stage. No such Stage by Stage characterization exists in the peer-reviewed literature. The diligence of considering the AKIN urine output criteria in this study is not present in any other models, and accounts for between 11.6% and 12.4% of AKI positive patients at every stage.

AKI Stage 1, 2, and 3 have an in-hospital mortality rate of 18%, 40%, and 47% respectively. A clinical tool that can predict the onset and severity of an AKI or the progression to a higher Stage may allow a clinician to start renal replacement or other therapies before symptoms are present and change the trajectory of the patient.

This study uses two methods of discerning pre-renal (causal circumstance of AKI exists before the renal system), intrinsic (causation of AKI exists within the renal system), or post-renal (causal circumstance of AKI exists after the renal system) AKI. Fractional Excretion of Sodium (FENa) proposed by Espinel [17] calculated by:

$$FENa = \frac{S_{Cr} * U_{Na}}{S_{Na} * U_{Cr}} \tag{5.2}$$

$$S_{Cr} = SerumCreatinine \tag{5.3}$$

$$S_{Na} = SerumSodium \tag{5.4}$$

$$U_{Cr} = UrineCreatinine \tag{5.5}$$

$$U_{Na} = UrineSodium \tag{5.6}$$

Pre-renal is defined by <0.01. Intrinsic is defined ≥ 0.01 and <0.04. Post-renal is defined by ≥ 0.04.

The ratio of blood urea nitrogen and serum creatinine (BUN-to-SCre) was proposed by Morgan et al. [60]. Pre-renal is defined as ≥ 20.0. Intrinsic is ≥ 10.0 and <20.0. Post-renal is <10.0.

The FENa approach categorizes the AKI Stages and Progressions of this chapter's validation set by pre-renal, intrinsic, and post-renal in Table 5.17. The table shows a larger portion of the AKI population are pre-renal AKIs (56.6% pre-renal as opposed to 34.1% intrinsic and 9.3% post-renal), but the intrinsic and post-renal AKIs are more likely to progress to more severe Stages of AKI (39.2% pre-renal as opposed to 39.2% intrinsic and 21.5% post-renal). For AKI Stage 1, the intelligent agent's predictive algorithm performs markedly worse (with a greater portion of false negative) on post-renal. This is is likely due to the disposition of the training/testing set (as well as the greater ICU population) having only roughly 10% post-renal AKIs. In contrast, AKI Stage 3 has the reversed phenomenon where pre-renal false negatives are markedly greater proportionally. The variable predictability between the three patient populations is telling of the physiological differences and implies predictive models designed specifically for each of the three causal conditions (pre-renal, intrinsic, and post-renal) may perform better than a single model. Thee results of the BUN-to-SCre in Table 5.18 echo the results above.

State	Total	Pre-Renal	Intrinsic	Post-Renal	P-value
AKI Stage 1	910	515 (0.566)	310 (0.341)	85 (0.093)	<0.0001
AKI Stage 1 True Positive	760	438 (0.576)	258 (0.339)	64 (0.084)	<0.0001
AKI Stage 1 False Negative	150	77 (0.513)	52 (0.347)	21 (0.140)	<0.0001
AKI Stage 2	274	128 (0.467)	101 (0.369)	45 (0.164)	<0.0001
AKI Stage 2 True Positive	209	105 (0.502)	69 (0.330)	35 (0.167)	<0.0001
AKI Stage 2 False Negative	65	23 (0.354)	32 (0.492)	10 (0.154)	0.0035
AKI Stage 3	130	51 (0.392)	51 (0.392)	28 (0.215)	0.0171
AKI Stage 3 True Positive	116	43 (0.371)	48 (0.414)	25 (0.216)	0.0227
AKI Stage 3 False Negative	14	8 (0.571)	3 (0.214)	3 (0.214)	0.1677
AKI Progression 1 to 2	82	46 (0.561)	28 (0.341)	8 (0.098)	<0.0001
AKI Progression 1 to 2 True Positive	67	37 (0.552)	24 (0.358)	6 (0.090)	<0.0001
AKI Progression 1 to 2 False Negative	15	9 (0.600)	4 (0.267)	2 (0.133)	0.0743
AKI Progression 2 to 3	130	51 (0.392)	51 (0.392)	28 (0.215)	0.0171
AKI Progression 2 to 3 True Positive	109	41 (0.376)	45 (0.413)	23 (0.211)	0.0228
AKI Progression 2 to 3 False Negative	21	10 (0.476)	6 (0.286)	5 (0.238)	0.3679

Table 5.17: AKIs discerned by FENa on validation set.

State	Total	Pre-Renal	Intrinsic	Post-Renal	P-value
AKI Stage 1	3416	2134 (0.625)	1208 (0.354)	57 (0.017)	<0.0001
AKI Stage 1 True Positive	2924	1884 (0.644)	981 (0.335)	46 (0.016)	<0.0001
AKI Stage 1 False Negative	492	250 (0.508)	227 (0.461)	11 (0.022)	<0.0001
AKI Stage 2	528	312 (0.591)	186 (0.352)	27 (0.051)	<0.0001
AKI Stage 2 True Positive	364	247 (0.679)	105 (0.288)	10 (0.027)	<0.0001
AKI Stage 2 False Negative	164	65 (0.396)	81 (0.494)	17 (0.104)	<0.0001
AKI Stage 3	193	100 (0.518)	73 (0.378)	20 (0.104)	<0.0001
AKI Stage 3 True Positive	143	72 (0.503)	59 (0.413)	12 (0.084)	<0.0001
AKI Stage 3 False Negative	50	28 (0.560)	14 (0.280)	8 (0.160)	0.0018
AKI Progression 1 to 2	307	200 (0.651)	101 (0.329)	6 (0.020)	<0.0001
AKI Progression 1 to 2 True Positive	229	144 (0.629)	81 (0.354)	4 (0.017)	<0.0001
AKI Progression 1 to 2 False Negative	78	56 (0.718)	20 (0.256)	2 (0.026)	<0.0001
AKI Progression 2 to 3	196	101 (0.515)	73 (0.372)	22 (0.112)	<0.0001
AKI Progression 2 to 3 True Positive	134	69 (0.515)	54 (0.403)	11 (0.082)	<0.0001
AKI Progression 2 to 3 False Negative	62	32 (0.516)	19 (0.306)	11 (0.177)	0.0044

Table 5.18: AKIs discerned by BUN-to-SCre on validation set.

5.13 Summary

The intelligent agents were instructed to create predictive models of AKI by the AKIN definition. The models consist of Stage 1, Stage 2, Stage 3, the progression from Stage 1 to Stage 2, and the progression from Stage 2 to Stage 3. The agents learned from 10% of the MIMIC database to create features for each model. Those features were trained and tested (using 10-fold cross-validation) on 70% of the remaining MIMIC data. The resulting models were validated on the remaining 30% of data. The performance of the AKI Stage 1 model on the validating data set is an AUC of 0.824 (0.815, 0.833), and is superior to three published models and one re-implementation of a published model using MIMIC. The performance of the AKI Stage 2 model on the validating data set is an AUC of 0.927 (0.916,0.938), and is superior to the only published model and its re-implementation on MIMIC. The performance

of the AKI Stage 3 model on the validating data set is an AUC of 0.983 (0.976, 0.990); there is no basis for comparison. The performance of the progression from AKI Stage 1 to Stage 2 on the validating data set is an AUC of 0.880 (0.871, 0.889); there is no basis for comparison. The performance of the progression from AKI Stage 2 to Stage 3 on the validating data set is an AUC of 0.951 (0.939, 0.963); there is no basis for comparison.

This chapter demonstrates the predictive power of an intelligent agent approach to predicting the onset of AKI by Stage and progression. All models are stable over time having strong predictive power six hours prior to onset or progression. Future work is required to assess the threshold of how much prior to onset or progression models may retain predictive power. A clinical tool that may predict the onset or progression of AKI would allow a clinician to intervene and potentially dramatically change the patients trajectory.

Chapter 6

SEPSIS

6.1 Introduction

Sepsis has a prevalence of 651 cases per 100,000 population, as described by Sutton and Friedman [85]. Severe sepsis has a prevalence of 300 cases over 100,000 population, with a mortality of approximately 25% (or between 200,000 and 250,000 U.S. deaths per year), as described in Mayr et al. [55]. In total, is estimated to cost $24 billion per year in health care costs, and is considered "The most expensive condition treated in U.S. hospitals," as described by Torio and Andrews [88].

This chapter presents a competitive predictive model for sepsis. It is second only to studies who use the InSight algorithm (Mao [52], Rothman [76], McCoy [56], Desautels[13]) or a highly specific and uncommon blood test (Mardi [53]). This study's model presents the highest specificity save McCoy's SOFA (which produces a clinically inapplicable sensitivity) and McCoy's version of the InSight algorithm. When the competitive models are pared down to similar design and inclusion criteria, of those mentioned only the Desautels study is similar and of comparable performance.

6.2 Pathogenesis

Sepsis is, in essence, a dysregulated response to an infection. Stage 1: a pro-inflammatory response assists in destroying damaged tissue and pathogenic organism. Stage 1 is concluded by an anti-inflammatory response to contain the pro-inflammatory response at local site of insult. Stage 2: If Stage 1 is not successful (or biochemically not perceived to be successful) in containing the damage and pathogen, a pro-inflammatory response becomes systemic and recruits system wide immune defenses. An anti-inflammatory response becomes systemic to

down-regulate the systemic pro-inflammatory response. Stage 2 is considered the cytokine storm. Stage 3: the immune system loses the ability to regulate the pro-inflammatory response, potentially resulting in organ dysfunction, organ failure, and shock. Stage 4: an extreme anti-inflammatory response suppresses the immune system. Stage 5: the immunomodulatory system is out-of-balance and results in multiple organ dysfunction and possibly death.

6.3 Value in Early Detection

In 2001 Rivers, et al. [75] discovered that early goal directed therapy reduces mortality as compared to standard-of-care, from 30% to 15%. In 2006 Kumar, et al. [44] discovered that the delay in administering antimicrobials after the onset of hypotension results in a 7.6% increase in patient mortality every hour for the first six hours. In 2005 the Surviving Sepsis Campaign published the first sepsis bundle, mainly comprised of the direction to measure blood culture and lactate, and administer antimicrobials, fluids, and vasopressors. The early application of the sepsis bundle has been validated to decrease patient mortality by half (Gao, et al. from 49% to 23% [28]; Kortgen, et al. from 53% to 27% [43]; Nguyen, et al. from 40% to 21% [64]; Ferrer, et al. from 44% to 40% [22]; Teles, et al. from 56% to 30% [87]; Zambon, et al. from 39% to 10% [95]; Shiramizo, et al. from 54% to 16% [82]).

6.4 Diagnostic Criteria and Implementation

This study utilizes the ICD9 diagnosis coding of sepsis and severe sepsis to identify sepsis positive patients. The time point for the onset of sepsis is defined using the Sepsis-3 definition criteria of suspicion of infection and a Sequential [Sepsis-Related] Organ Failure Assessment (SOFA) Score of two or greater, as described by Singer, et al. [83]. A SOFA score of two or greater is achieved when any of the following criteria are met: PaO_2/FiO_2 <300 mmHg; platelets $<100*10^3/\mu$L; bilirubin $\geq$ 2.0 mg/dL; any administration of dobutamine, Dopamine, epinephrine, or norepinephrine; Glasgow Coma Score $\leq$ 12; Creatinine $\geq$ 2.0 mg/dL; urine output <500 mL/day.

6.5 Competing Models for Sepsis

There exist a large number of competing models in the peer-reviewed literature. A rich sample of 19 studies were chosen as a baseline comparison and are described in Table 6.1. A comparison of the study designs and performance with this chapter's intelligent agent model may be seen in Table 6.1. No peer-reviewed model described here includes patients with missing data for a variable the model considers. This introduces a significant selection bias. Compounding the selection bias models: artificially reduce false positive rates by including patients with positive results for blood culture where negative results may not be represented; require continuous physiological data; require uncommon tests; intervene in the patients they are studying; only include patients already at risk for sepsis; or only examine patients with polytrauma or other specific conditions.

Not included in the comparison samples, the study by Henry et al. [34] predated the current definition of sepsis, though had an impressive performance at the time with an AUC = 0.83 (0.81, 0.85). The authors have not yet re-implemented their work using the current Sepsis-III definition.

Model	Setting	N	sepsis	AUC	Sensitivity	Specificity	Accuracy	Notes
Desautels et al [13]	MIMIC	22,853	2,577	0.88	0.8	0.8	0.8	artificial reduction in false positives by inclusion criteria
Desautels_SIRS	MIMIC	22,853	2,577	0.61	0.72	0.44	0.47	
Desautels_MEWS	MIMIC	22,853	2,577	0.8	0.7	0.77	0.76	
Desautels_qSOFA	MIMIC	22,853	2,577	0.77	0.56	0.84	0.8	
Desautels_SAPSII	MIMIC	22,853	2,577	0.7	0.75	0.52	0.55	
Desautles_SOFA	MIMIC	22,853	2,577	0.73	0.8	0.48	0.52	
van Wyk et al [91]	ICU	1,161	377		0.377			requires continuous physiological data
Vassiliou_Lac [92]	ICU	89	45	0.677 (0.560, 0.788)	0.49	0.82		
Vassiliou_Lac+selectin	ICU	89	45	0.854 (0.775, 0.932)	0.76	0.84		
Gouel_mHLA-DR [31]	T-ICU	100	37	0.79 (0.69, 0.88)	0.826	0.647		
Gouel_IL-6	T-ICU	100	37	0.75 (0.64, 0.84)	0.846	0.725		
Gouel_combined	T-ICU	100	37		0.696	0.902		
Saqib et al [78]	MIMIC	38,270	10,071	0.669	0.476			Uses Angus criteria and excludes patients without measured data
Shanhar-Hari et al [80]	ED/ICU	272	139	0.67 (0.60, 0.74)				Based on Neutrophil CD279
McCoy et al [56]	Hospital	1,665		0.91 (0.90, 0.92)	0.83	0.96		Interventional SIRS screen twice a day
McCoy_SIRS	Hospital	1,665		0.76	0.64	0.88		
McCoy_MEWS	Hospital	1,665		0.55	0.42	0.64		
McCoy_qSOFA	Hospital	1,665		0.55	0.13	0.97		
McCoy_SOFA	Hospital	1,665		0.77	0.67	0.83		
Mao et al [52]	ICU	111,957	1,592	0.92 (0.90, 0.93)	0.98	0.8		excludes patients without measured data
Mao_MEWS	ICU	111,957	1,592	0.76	0.98	0.8		
Mao_SOFA	ICU	111,957	1,592	0.63	0.82	0.8		
Mao_SIRS	ICU	111,957	1,592	0.75	0.82	0.8		
Faisal et al Dev [19]	Hospital	26,247	4,861	0.779 (0.772, 0.786)	0.678	0.732		
Faisal et al Valid	Hospital	30,996	7,773	0.788 (0.782, 0.793)	0.832	0.599		
Nemati et al Dev [63]	ICU	27,527	2,375	0.86	0.85	0.7	0.7	
Nemati et al Valid	ICU	31,179	7,459	0.85	0.85	0.67	0.67	
Shashikumar et al [81]	ICU			0.8	0.85	0.57		
Layios et al [47]	ICU	99	19	0.75				looking at platelet activation markers
Danner et al [11]	Hospital	53,313	884		0.74			
Rothman et al Dev [76]	ICU		511	0.911 (0.906, 0.916)				The difference between development and valid AUC implies overfitting
Rothman et al valid1	ICU		81	0.802 (0.791, 0.812)				
Rothman et al valid2	ICU		826	0.819 (0.814, 0.825)				
Lindner et al [49]	ICU	256	85	0.81	0.82	0.71	0.77	Only applies to polytrauma
Olenick et al [68]				0.842	0.698			
Wang et al [94]				0.703				
Mardi et al [53]		114	37	0.922	0.93	0.76		IL-6 blood test
Lukaszewski et al [50]	ICU	92			0.914	0.802	0.946	Only includes patients already at risk of sepsis

Table 6.1: Sepsis peer-reviewed models.

6.6 Intelligent Agent Model

6.6.1 Variable Selection

Using the 10% data set, the intelligent agents performed 831 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model is described by Table 6.2. The variables selected for the model by the intelligent agents fall under categories: respiration; blood chemistry; functional and procedural; hemodynamics; interventions.

Respiration variables include a low exhaled minimum volume, positive end expiratory pressure, ventilator mode, and pCO_2 (venous). Functional procedural are relegated to Glasgow Coma Score. Hemodynamic variables include a high blood pressure alarm, and stroke volume (the amount of blood pumped by the left ventricle of the heart). Blood chemistry and blood products include: arterial base excess (low pH); bands (immature white blood cells); calculated bicarbonate, whole blood; ALT (alanine aminotransferase, an enzyme required to process energy, and elevation most likley implied liver damage); lactate; creatine kinase; potassium; AST (aspartate aminotanserase, also used to diagnose liver damage); cholesterol. Interventional variables include: magnesium sulfate (bolus); arterial line zero/calibration; nitroglycerine; digoxin; and GT (gastrostomy feeding tube) flush.

6.6.2 Model Training and Testing

Using the variables from Table 6.2, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 16,007 patients. At t-minus zero, the time immediately before the onset of sepsis, the model has an area under the ROC curve of 0.884 (0.867, 0.903) (see Figure 6.1). The model performance on the training/testing data set is outlined in Table 6.3. With the cutoff set at the maximum value for the F-measure (beta=1.75), the model produces: 330 true positives; 15,449 true negatives; 1172 false positives; 219 false negatives; a true positive rate (sensitivity) of 0.601 (see Figure F.1); a false positive rate of 0.076; a specificity of 0.924 (see Figure F.2); a positive predictive

Label	Variable Importance	AUC excluding variable
Magnesium Sulfate (Bolus)	0.23	0.852 (0.821, 0.884)
Bands	0.3	0.869 (0.839, 0.899)
Arterial Line Zero/Calibrate	0.31	0.860 (0.829, 0.890)
Calculated Bicarbonate, Whole Blood	0.32	0.872 (0.842, 0.902)
Nitroglycerin	0.33	0.861 (0.830, 0.892)
Low Exhaled Min Vol	0.31	0.854 (0.823, 0.885)
Non-Invasive Blood Pressure Alarm - High	0.25	0.862 (0.832, 0.893)
Total PEEP Level	0.28	0.868 (0.837, 0.898)
pCO_2	0.35	0.864 (0.834, 0.895)
GCS Total	0.25	0.859 (0.828, 0.890)
Stroke Volume	0.28	0.860 (0.829, 0.890)
ALT	0.29	0.870 (0.840, 0.900)
LVSW	0.26	0.868 (0.838, 0.898)
Arterial Base Excess	0.27	0.859 (0.828, 0.890)
D5/.45NS	0.25	0.865 (0.835, 0.895)
Digoxin	0.26	0.865 (0.834, 0.895)
Lactate	0.31	0.863 (0.832, 0.893)
Ventilator Mode	0.24	0.856 (0.825, 0.887)
Creatine Kinase (CK)	0.31	0.870 (0.840, 0.900)
Potassium	0.3	0.859 (0.828, 0.890)
Asparate Aminotransferase (AST)	0.29	0.865 (0.834, 0.895)
Cholesterol, Total	0.24	0.861 (0.830, 0.892)
GT Flush	0.21	0.861 (0.831, 0.892)

Table 6.2: Variables and variable importance of best performing sepsis model generated by intelligent agents.

value of 0.220 (see Figure F.3; a negative predictive value of 0.985; an accuracy of 0.913 (see Figure F.4); with a kappa of 0.957 (see Figure F.5).

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	21	15449	9	528	0.038	0.001	0.038	0.999	0.700	0.967	0.966	0.983
1.50	253	14912	546	296	0.461	0.035	0.461	0.965	0.317	0.981	0.947	0.974
1.75	330	14286	1172	219	0.601	0.076	0.601	0.924	0.220	0.985	0.913	0.957

Table 6.3: Sepsis model performance on the training/testing data set.

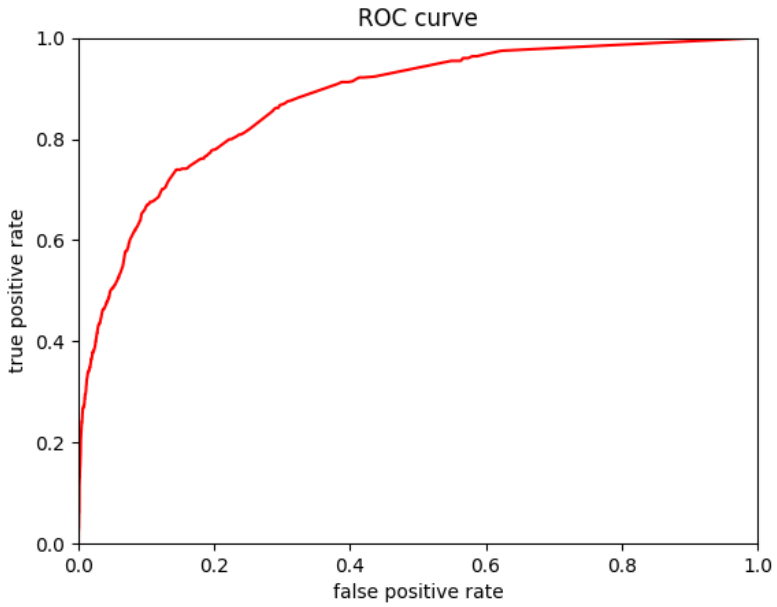


Figure 6.1: Sepsis patients, training/testing set AUC=0.884 (0.867, 0.903).

6.6.3 Model Validation

The threshold resulting for the maximum F-measure (beta=1.75) was used to set the threshold for the validation set containing 6,914 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of overfitting. The ROC curve in Figure 6.2 has an area under the curve of 0.874 (0.844, 0.904) compared to the training/testing are of 0.884 (0.867, 0.903). The performance of the model on the validation set may be found in Table 6.4 with 140 true positives, 6137 true negatives, 550 false positives, and 87 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.617 (see Figure F.7) as opposed to 0.601; a false positive rate of 0.082 as opposed to 0.076; a specificity of 0.918 (see Figure F.8) as opposed to 0.924; a positive predictive value of 0.203 (see Figure F.9) as opposed to 0.220; a negative predictive value of

0.986 as opposed to 0.985; an accuracy of 0.908 (see Figure F.10) as opposed to 0.913; with a kappa of 0.954 (see Figure F.11) as opposed to 0.957.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	21	6685	2	206	0.093	0.000	0.093	1.000	0.913	0.970	0.970	0.985
1.50	116	6420	267	111	0.511	0.040	0.511	0.960	0.303	0.983	0.945	0.973
1.75	140	6137	550	87	0.617	0.082	0.617	0.918	0.203	0.986	0.908	0.954

Table 6.4: Sepsis model performance on the validation data set.

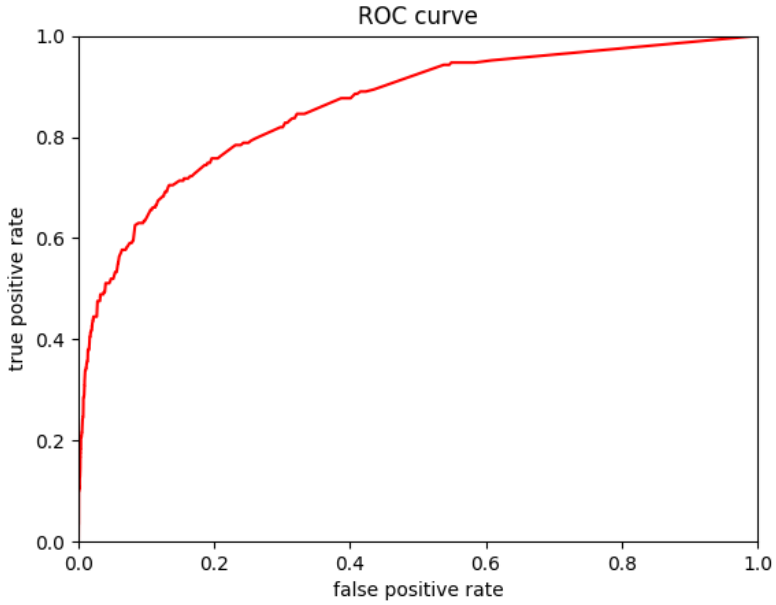


Figure 6.2: Sepsis patients, validation set AUC=0.874 (0.844, 0.904).

6.6.4 As a Function of Time

The sepsis patients model is stable, losing 2.06% AUC with six hours prior variable values. In Figure 6.3 t-minus zero has an AUC of 0.874 (0.844, 0.904), and t-minus six hours, the models area under the ROC curve is 0.856 (0.825, 0.887). Additional lower performing machine learning algorithms over time may be found in Figure F.13.

6.6.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table F.1. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate sepsis positive patients from sepsis negative patients. Sepsis patients have a greater bands, whole blood bicarbonate, PEEP level, ALT, lactate, and AST. Sepsis positive patients had a different distribution of GCS and ventilator mode, and a depressed arterial base excess.

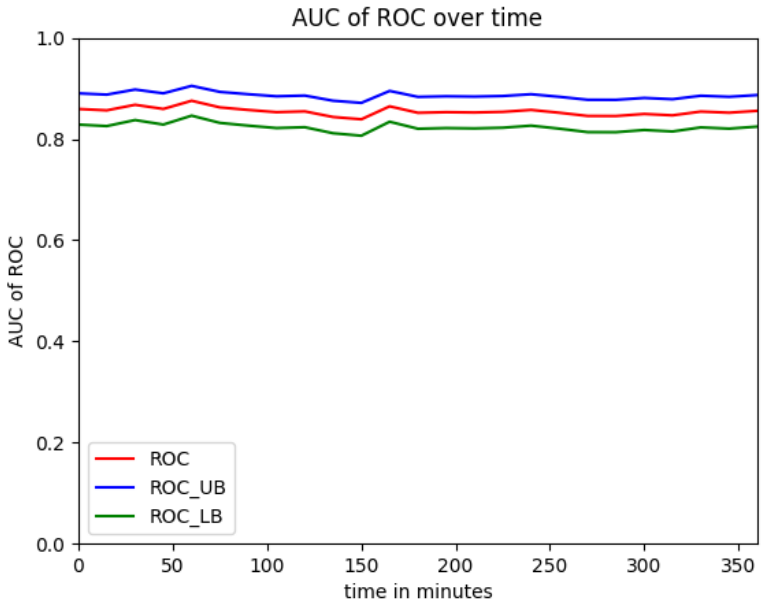


Figure 6.3: Sepsis validation AUC of the ROC over time. AUC at t-minus=0 is 0.874 (0.844, 0.904). At t-minus six hours, the model's area under the ROC curve is 0.856 (0.825, 0.887).

Variable composition by classification was assessed in Table F.2 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they present with: depressed nitroglycerin administration; a lower exhaled minimum volume; a narrow pCO_2; less administration of D5/.45 normal saline; depressed creatinine kinase; and fewer gastro tube flushes.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table F.3. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate sepsis positive patients from sepsis negative patients. Condition positive patients have: a higher in-hospital mortality (44% as opposed to 7%); a higher 30-day mortality (47% as opposed to 9%); a longer ICU LOS (7.05 days as opposed to 1.57 days); were more represented in the Medical ICU (61%

as opposed to 35%); were more likely covered by Medicare (68% as opposed to 48%) likely because of the age difference of the two cohorts (median 70 years old for sepsis patients as opposed to a median of 62 years old for sepsis negative patients.

Demographic and key variables was assessed by classification in Table F.4. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their demographics include: under-representation in the TSICU; and under-represented by private insurance.

Comorbidities of condition positive and condition negative patients was assessed in Table F.5. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate sepsis positive patients from sepsis negative patients. Sepsis positive patients are over represented in: Congestive heart failure (43% as opposed to 20%); liver disease (13% as opposed to 5%); coagulopathy (25% as opposed to 4%); weight loss (12% as opposed to 3%); fluid and electrolyte disorders (52% as opposed to 22%).

Comorbidities by classification is assessed in Table F.6. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they under-represent comorbidities: cardiac arrhythmias; valvular disease; hypertension, complicated; chronic pulmonary disease renal failure; and deficiency anemia.

6.7 Model Comparisons

The intelligent agents produce a model competitive with models from the peer-reviewed literature. The agents achieve an AUC in among the top level performers, as seen in Figure 6.4. This dissertation has focused efforts on minimizing false negatives at the expense of false positives, which weighs specificity more significant that sensitivity. The intelligent agent model is among the best performers for sensitivity with other metrics well balanced, as seen in Figure 6.5. Only McCoy's model and McCoy's qSOFA have a higher specificity, and the qSOFA has a clinically impractical sensitivity.

The similarities between predictive models based on design is included in Table 6.5 and include: if the study was performed in the ICU; if the study is generally applicable to those

patients, without extra inclusion criteria such as polytrauma, already suspected of sepsis, or specific surgeries; if the study includes patients with missing values for included variables; if the study was performed without an intervention; if the study is examining the onset of sepsis as opposed to a hospital-wide sepsis diagnosis; and if the study used Sepsis-3 as the sepsis definition.

Of the models considered for comparison, Rothman et al. [76] is one of the more similar in overall goal. The Rothman study investigate patients who present with sepsis in the ICU who developed sepsis in the ICU. The Rothman model is considered to be subject to over-fitting, where the development model learned and improved performance on the development data set at the expense of general predictive performance. Rothman's predictive model on the development data set, derived from Sarasota Memorial Hospital in Sarasota Florida, performed with an AUC = 0.911 (0.906, 0.916). On the first validation data set, derived from Riverside Regional Medical Center in Newport News Virginia, it performed with an AUC = 0.802 (0.791, 0.812). On the second validation set, derived from Yale New Haven Hospital in New Haven Connecticut, it performed with an AUC = 0.819 (0.814, 0.825). The drop in AUC by 12.0% and 10.1% respectively indicate a model that gained performance on the training set at the expense of losing the ability to be applied generally.

Other similar studies include those form Desautels et al. [13], Nemati, and Shashikumar. All three: are ICU based; are generally applicable; do not employ an intervention; examine the onset of sepsis; and use Sepsis-3 as the definition of sepsis. No peer-reviewed published sepsis prediction model includes patients with missing values for included variables. The inclusion of patients with missing values is unique to this dissertation's approach to predicting sepsis. The Desautels study requires blood culture results to indicate a suspicion of infection, where the lack of results do not necessarily negate the lack of a blood culture test. It is therefor artificially reducing the false positive results by the blood culture inclusion criteria. The intelligent agents perform competitively with all sepsis models including those most similar. The intelligent agents best performing model achieved an AUC of 0.874 (0.844, 0.904) on the validation set, closely matching Desautels' AUC = 0.88 while surpassing Nemati's

AUC = 0.85 and Shashikumar's AUC = 0.80. The intelligent agent's model produces an accuracy of 0.908 and surpasses Desautels' accuracy of 0.80 and Nemati's accuracy of 0.67 (Shashikumar's accuracy not reported). The intelligent agent's model weighs specificity more significantly than sensitivity to minimize false negatives (the patient's whom without intervention would succumb to harm) resulting in a specificity of 0.918 (sensitivity of 0.617) surpassing Desautels' specificity of 0.80 (sensitivity of 0.80), Nemati's specificity of 0.67 (sensitivity of 0.85), and Shashikumar's specificity of 0.57 (sensitivity of 0.85).

	AUC	ICU	Generally applicable	Include patients with missing data	Without intervention	Onset of sepsis	Sepsis-3
Intelligent Agent	0.874 (0.844, 0.904)	✓	✓	✓	✓	✓	✓
Desautels et al [13]	0.88	✓	✓	-	✓	✓	✓
Desautels_SIRS	0.61	✓	✓	-	✓	✓	✓
Desautels_MEWS	0.8	✓	✓	-	✓	✓	✓
Desautels_qSOFA	0.77	✓	✓	-	✓	✓	✓
Desautels_SAPSII	0.7	✓	✓	-	✓	✓	✓
Desautles_SOFA	0.73	✓	✓	-	✓	✓	✓
Vassiliou_Lac [92]	0.677 (0.560, 0.788)	✓	✓	-	✓	✓	-
Vassiliou_Lac+selectin	0.854 (0.775, 0.932)	✓	✓	-	✓	✓	-
Gouel_mHLA-DR [31]	0.79 (0.69, 0.88)	✓	-	-	-	-	-
Gouel_IL-6	0.75 (0.64, 0.84)	✓	-	-	-	-	-
Saqib et al [78]	0.669	✓	✓	-	✓	-	✓
Shanhar-Hari et al [80]	0.67 (0.60, 0.74)	✓	-	-	-	-	-
McCoy et al [56]	0.91 (0.90, 0.92)	-	✓	-	-	-	✓
McCoy_SIRS	0.76	-	✓	-	-	-	✓
McCoy_MEWS	0.55	-	✓	-	-	-	✓
McCoy_qSOFA	0.55	-	✓	-	-	-	✓
McCoy_SOFA	0.77	-	✓	-	-	-	✓
Mao et al [52]	0.92 (0.90, 0.93)	✓	✓	-	✓	✓	-
Mao_MEWS	0.76	✓	✓	-	✓	✓	-
Mao_SOFA	0.63	✓	✓	-	✓	✓	-
Mao_SIRS	0.75	✓	✓	-	✓	✓	-
Faisal et al Dev [19]	0.779 (0.772, 0.786)	-	-	-	✓	-	-
Faisal et al Valid	0.788 (0.782, 0.793)	-	-	-	✓	-	-
Nemati et al Dev [63]	0.86	✓	✓	-	✓	✓	✓
Nemati et al Valid	0.85	✓	✓	-	✓	✓	✓
Shashikumar et al [81]	0.8	✓	✓	-	✓	✓	✓
Layios et al [47]	0.75	✓	-	-	✓	✓	✓
Rothman et al Dev [76]	0.911 (0.906, 0.916)	✓	✓	-	✓	✓	-
Rothman et al valid1	0.802 (0.791, 0.812)	✓	✓	-	✓	✓	-
Rothman et al valid2	0.819 (0.814, 0.825)	✓	✓	-	✓	✓	-
Lindner et al [49]	0.81	✓	-	-	✓	✓	-
Olenick et al [68]	0.842	-	✓	-	✓	✓	-
Wang et al [94]	0.703	-	✓	-	✓	-	-
Mardi et al [53]	0.922	-	✓	-	✓	-	-

Table 6.5: Sepsis model comparison based on AUC.

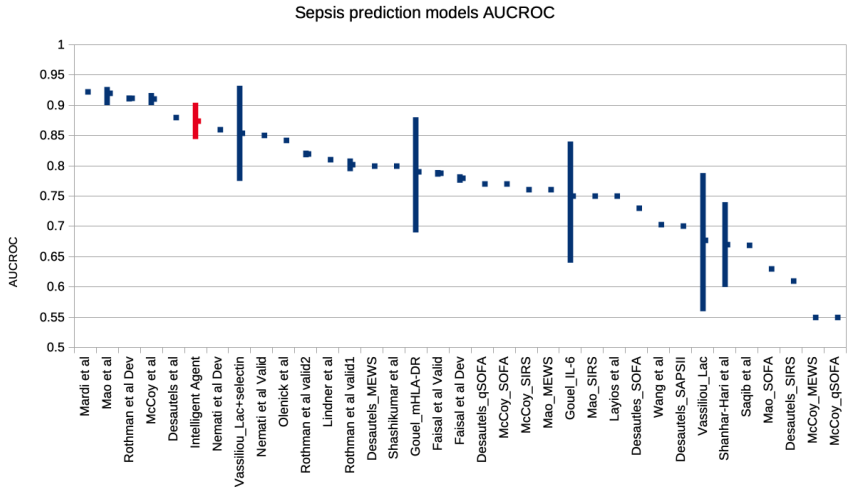


Figure 6.4: Sepsis prediction models ordered by AUC. This studies intelligent agent generated model in red.

6.8 Discussion

The intelligent agent best performing model is comparable to the inSight algorithm used by Desautles. The inSight algorithm makes use of: systolic blood pressure; pulse pressure; heart rate; temperature; respiration rate; white blood cell count; pH; blood oxygen saturation; and age. The intelligent agents have access to those very same variables, and yet found prediction by another route. By employing an intelligent agent approach to sepsis this dissertation produces a perspective on a sub-population of septic patients and can characterize significant variables related to sepsis and hemodynamics. Though septic patients are most significantly from the MICU and under-represented in the CSRU (see Table F.3), they do have a significant comorbidity of congestive heart failure and cardiac arrhythmias (see table F.5). Several of the variables the intelligent agents included in the best performing model include cardiac function related variables: arterial line is commonly used for monitoring high resolution hemodynamic properties; nitroglycerin, a vasodilator; high blood pressure alarm; stroke volume, the volume of blood pumped by a single stroke of the heart; LVSW (left ventricle

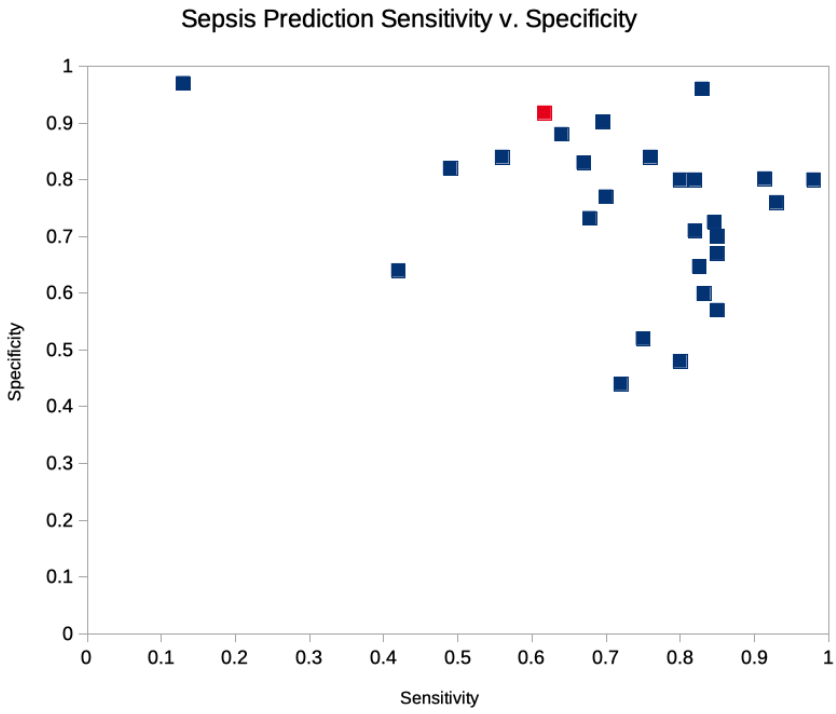


Figure 6.5: Sepsis prediction models scatter chart of sensitivity v. specificity. This study's intelligent-agent-generated model is in red.

stroke work), and Digoxin, used to treat various heart conditions.

6.9 Summary

The intelligent agents were instructed to create a predictive model of sepsis based on the Sepsis-3 definition. The agents learned from 10% of the MIMIC database to create features. Those features were trained and tested (using 10-fold cross-validation) on 70% of the remaining MIMIC data. The resulting model was validated on the remaining 30% of data. The performance of the model on the validating data set is an AUC of 0.874 (0.844, 0.904).

This chapter demonstrates the predictive power of an intelligent agent approach to predicting the onset of sepsis. In addition to out competing 75% of evaluated models, it stands

out from those algorithms with the most similar study design. While matching Desautles' AUC, it surpasses other similar model's in AUC and all similar models in accuracy and specificity. Significantly, the agents highlight significant markers that link sepsis with hemodynamics.

Six hours prior to the patient meeting the Sepsis-3 criteria, the intelligent agent's prediction model has a performance of an AUC of 0.856 (0.825, 0.887). There exists evidence that the early application of measures contained in the Surviving Sepsis Campaign - sepsis bundle can reduce mortality rates by as much as 50%. Kumar et al. [44] identified a 7.6% increase in mortality for every hour antimicrobials are delayed for the first six hours. With a clinical tool that may predict onset of sepsis six hours before with an AUC of 0.856 a clinician may be able to further reduce mortality risk, reduce other adverse outcomes, or redirect the patient's trajectory entirely.

Chapter 7

DISSEMINATED INTRAVASCULAR COAGULATION

7.1 Introduction

Matsuda [54] identifies DIC present in 1% of tertiary care patients. This manifestation does not account for mild and sub-clinical conditions. Mortality rates of DIC afflicted patients vary substantially by severity of DIC and co-occurring conditions.

7.2 Pathogenesis

DIC is initiated when procoagulants cause excess thrombin formation. The excess thrombin can cause: plasmin activation and fibrinolysis (breaking down existing clots) which leads to excess bleeding; consumption of coagulation factors, depleting the ability to form clots, and leads to excess bleeding; microvascular clots in organs causing injury.

7.3 Value in Early Detection

The recommendation by Wada et al. [93] is to begin treatment as soon as diagnosed. The recommended treatment strategy is to treat the underlying condition that provides the procoagulants that perpetuates the DIC cascade. There exists no literature on delayed treatment, and there exists no predictive model to preempt the manifestation of the disease.

7.4 Diagnostic Criteria and Implementation

For the definition of diagnostic criteria this chapter uses JAAM's DIC definition, provided by Gando et al. [27]. JAAM puts forth a scoring system, where a score of greater than or equal to 5 points is considered DIC where: 3+ SIRS criteria is worth one point; platelet count between 80 and 120 $*10^9$/L or a 30%+ decrease within 24 hours is worth one point;

platelet count less than 80 *10^9/L or a 50%+ decrease within 24 hours is worth 3 points; prothrombin time greater or equal to 1.2 is worth one point; fibrinogen less than 3.5 g/L is worth one point; fibrin or fibrinogen degradation product between 10 and 25 mg/L is worth one point; and fibrin or fibrinogen degradation product greater or equal to 25 mg/L is worth 3 points.

7.5 Intelligent Agent Model

7.5.1 Variable Selection

Using the 10% data set, the intelligent agents tested 1,040 viable models. The best performing model variables, corresponding variable importance metric and AUC of a Random Forest ROC if the variable is missing from the model, are described by Table 7.1. The variables selected for the model by the intelligent agents fall under categories: pharmaceuticals; vital signs; laboratory results; procedural or observational; and ventilation.

Pharmaceutical interventions used as variables for predicting the onset of DIC are the administrations of Digoxin and calcium gluconate. Digoxin is a cardio-pulmonary drug that supports blood pressure and is used in the treatment of heart failure and heart arrhythmia. Calcium gloconate is an intravenous medications to treat calcium deficiencies. Vital sign measurements used in the prediction model include: temperature; respiratory rate; weight (both admit and previous); and blood pressure alarms.

Laboratory results used in the prediction model are varied. Bilirubin is used to clear waste products that arise from the breaking down of red blood cells. INR(PT) is telling of how long it takes for the body to form clots. Creatinine is a product of breaking down of muscle tissue. Creatine phosphokinase (CPK and CK-MB) is an enzyme found in muscle (notably heart tissue), brain, and skeletal tissue. Elevated CPK are indicative of the destruction of those tissues. Eosinophils are a type of white blood cell activated for types of cancer, allergic reactions and parasitic infections. Basophils (measured by Differential-Basos) contain heparin, a blood thinner, and may prevent blood from forming clots.

Procedure and observational events include: pre-admission intake; noted discharge needs; difficulty swallowing; Braden nutrition, a note of diminished or imbalanced nutrition; and Braden friction/shear, motion issues that may cause pressure ulcers or cellular and other micro-trauma when moving. The tube feeding residual variable may additionally indicate nutritional issues related to enteral nutrition tolerance and stomach digestive function. Minute ventilation volume is emphasized in three different variables while forming of this model.

7.5.2 Model Training and Testing

Using the variables from Table 7.1, a 10-fold cross validation model (with 90% for training and 10% for testing) was created on a training/testing set of 38,241 patients. At t-minus zero, the time immediately before the onset of DIC, the model has an area under the ROC curve of 0.830 (0.823, 0.837) (see Figure 7.1). The model performance on the training/testing data set is outlined in Table 7.2. With the cutoff set at the maximum value for the F-measure (beta=1.75), the model produces: 4,410 true positives; 23,216 true negatives; 9,409 false positives; 1,206 false negatives; a true positive rate (sensitivity) of 0.785 (see Figure G.1); a false positive rate of 0.288; a specificity of 0.712 (see Figure G.2); a positive predictive value of 0.319 (see Figure G.3; a negative predictive value of 0.951; an accuracy of 0.722 (see Figure G.4); with a kappa of 0.885 (see Figure G.5).

Label	Variable Importance	AUC excluding variable
Admit Wt	0.34	0.835 (0.824, 0.845)
Previous Weight	0.33	0.834 (0.823, 0.844)
Calcium Gluconate	0.3	0.836 (0.826, 0.846)
CPK	0.33	0.835 (0.825, 0.846)
Digoxin	0.29	0.836 (0.826, 0.847)
Eosinophils	0.35	0.834 (0.824, 0.844)
Temperature C	0.31	0.825 (0.814, 0.835)
Total Bilirubin	0.3	0.834 (0.824, 0.845)
Creatinine	0.34	0.834 (0.823, 0.844)
CK-MB	0.28	0.836 (0.825, 0.846)
Pre-Admission Intake	0.26	0.836 (0.826, 0.847)
Differential-Basos	0.24	0.835 (0.824, 0.845)
Braden Friction/Shear	0.27	0.835 (0.824, 0.845)
Braden Nutrition	0.28	0.836 (0.825, 0.846)
Difficulty swallowing	0.18	0.833 (0.822, 0.843)
Arterial Blood Pressure Alarm - High	0.24	0.835 (0.824, 0.845)
Minute Volume Alarm - Low	0.25	0.834 (0.824, 0.844)
Minute Volume(Obser)	0.27	0.834 (0.824, 0.845)
INR(PT)	0.32	0.824 (0.813, 0.834)
Discharge needs	0.18	0.833 (0.822, 0.843)
Central Venous Pressure Alarm - High	0.18	0.836 (0.826, 0.846)
Non-Invasive Blood Pressure Alarm - High	0.22	0.836 (0.826, 0.846)
Minute Volume	0.24	0.834 (0.823, 0.844)
Resp Rate (Total)	0.25	0.838 (0.828, 0.848)
Tube Feeding Residual	0.29	0.836 (0.825, 0.846)

Table 7.1: Variables and variable importance of best performing DIC model generated by intelligent agents.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	901	32310	315	4715	0.160434	0.010	0.160	0.990	0.741	0.873	0.868	0.933
1.50	4047	25265	7360	1569	0.721	0.226	0.721	0.774	0.355	0.942	0.767	0.906
1.75	4410	23216	9409	1206	0.785	0.288	0.785	0.712	0.319	0.951	0.722	0.885

Table 7.2: DIC model performance on the training/testing data set.

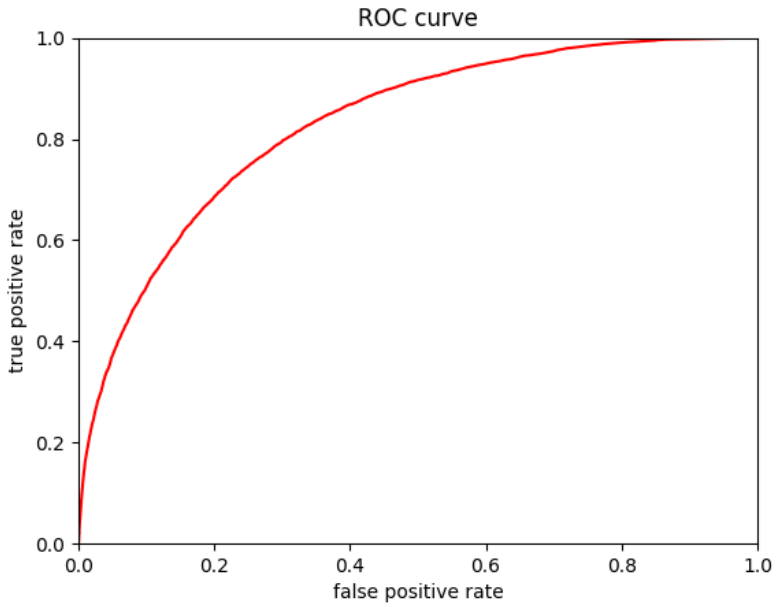


Figure 7.1: DIC patients, training/testing set AUC=0.830 (0.823, 0.837).

7.5.3 Model Validation

The threshold resulting for the maximum F-measure (beta=1.75) was used to set the threshold for the validation set containing 16,399 patients. The performance of the model against the validation set was largely consistent with the performance on the training/testing data set, which indicated a well-balanced model with no evidence of over-fitting. The ROC curve in Figure 7.2 has an area under the curve of 0.838 (0.828, 0.848) compared to the training/testing are of 0.830 (0.823, 0.837). The performance of the model on the validation set may be found in Table 7.3 with 1,881 true positives, 9,975 true negatives, 4,035 false positives, and 508 false negatives. In comparing the validation performance (using the predefined threshold) with the training/testing performance, the validation set has: a true positive rate (sensitivity) of 0.787 (see Figure G.7) as opposed to 0.785; a false positive rate of 0.288, the same as the training/testing counterpart; a specificity of 0.712 (see Figure G.8), the same as the training/testing counterpart; a positive predictive value of 0.318 (see Figure G.9) as

opposed to 0.319; a negative predictive value of 0.952 as opposed to 0.951; an accuracy of 0.723 (see Figure G.10) as opposed to 0.722; with a kappa of 0.885 (see Figure G.11), the same as the training/testing counterpart.

F-measure β	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa
1.00	400	13848	162	1989	0.167	0.012	0.167	0.988	0.712	0.874	0.869	0.933
1.50	1730	10860	3150	659	0.724	0.225	0.724	0.775	0.355	0.943	0.768	0.906
1.75	1881	9975	4035	508	0.787	0.288	0.787	0.712	0.318	0.952	0.723	0.885

Table 7.3: DIC model performance on the validation data set.

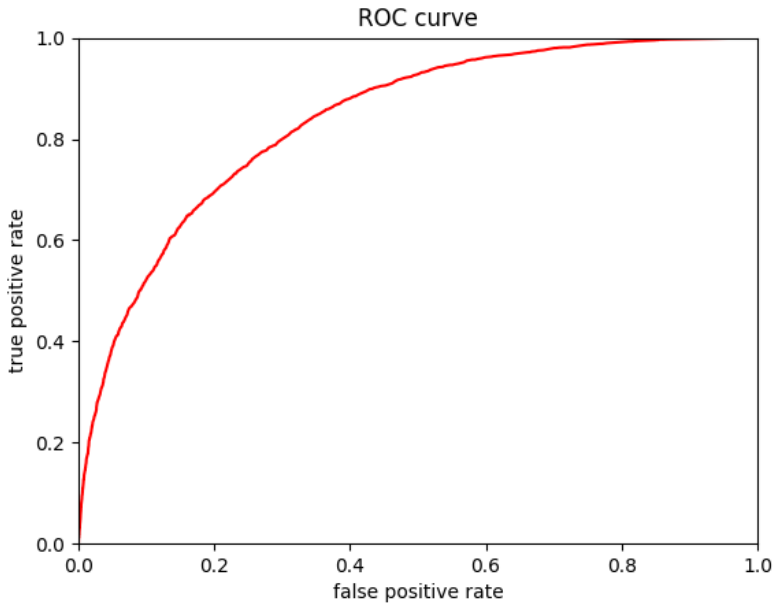


Figure 7.2: DIC patients, validation set AUC=0.838 (0.828, 0.848).

7.5.4 As a Function of Time

The DIC patients model is stable, losing 3.34% AUC with six hours prior variable values. In Figure 7.3 t-minus zero has an AUC of 0.838 (0.828, 0.848), and t-minus six hours, the models area under the ROC curve is 0.810 (0.799, 0.821). Additional lower performing machine learning algorithms over time may be found in Figure G.13.

7.5.5 Variable Composition, Demographics, and Comorbidities

Variable composition for condition positive and condition negative patients in the validation set was was assessed in Table G.1. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate DIC positive patients from DIC negative patients. DIC patients have: a greater weight; greater administration of calcium gluconate; depressed Eosinophils; elevated creatinine; elevated INR(PT) (implying trouble forming blood clots); a higher ventilation minute volume; and an elevated respiration rate.

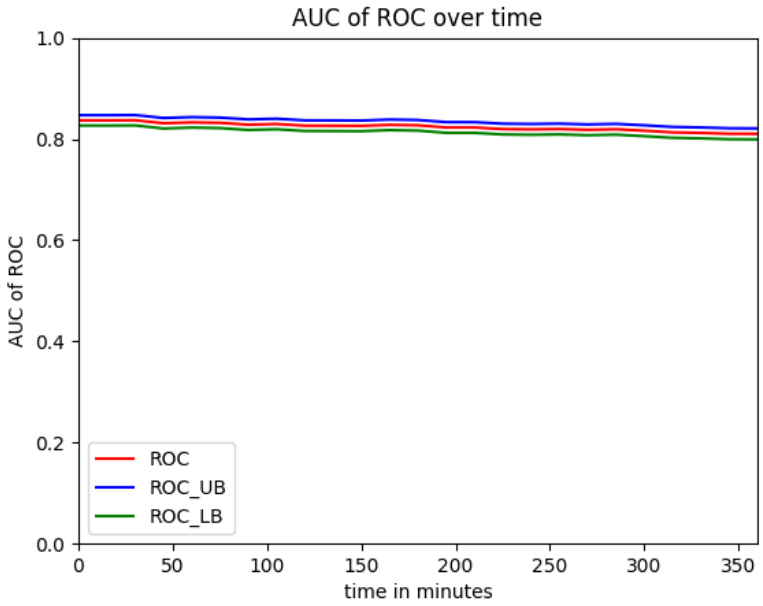


Figure 7.3: DIC validation AUC of the ROC over time. AUC at t-minus=0 is 0.838 (0.828, 0.848). At t-minus six hours, the model's area under the ROC curve is 0.810 (0.799, 0.821).

Variable composition by classification was assessed in Table G.2 to greater understand the models false positives and false negatives in the context of true positives and true negatives. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if they present with: elevated temperature; depressed creatinine; elevated CK-MB; or an elevated observed ventilated minute volume.

The demographic and key variables were assessed for condition positive and condition negative patients in the validation set in Table G.3. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) variables that differentiate DIC positive patients from DIC negative patients. DIC patients have: a higher in-hospital mortality; higher 30 day mortality rate; longer ICU length of stay; are under-represented in the CSRU; and are over-represented with Medicare while being under-represented by private insurance which is indicative of the age different (median of 67 years old for DIC positive patients and median

of 61 years old for DIC negative patients).

Demographic and key variables was assessed by classification in Table G.4. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their demographics include: under-representation in the CCU and MICU; are racially Asian; government and medicaid insurance.

Comorbidities of condition positive and condition negative patients was assessed in Table G.5. There are several statistically significant (Kruskal-Wallis test P-value <0.0001) comorbidities that differentiate DIC positive patients from DIC negative patients. DIC positive patients are over represented in: congestive heart failure; cardiac arrhythmias; valvular disease; peripheral vascular disorders; liver disease; lymphoma; coagulopathy; weight loss; fluid and electrolyte disorders; and alcohol abuse. DIC patients are under-represented in depression.

Comorbidities by classification is assessed in Table G.6. Of the statistically significant variables not included above, patients are more likely to be misclassified as a false negative if their comorbidities are under-represented in: pulmonary circulation disorders; hypertension, complicated; chronic pulmonary disease; diabetes; renal failure; rheumatoid arthritis/collagen vascular diseases; and depression.

7.6 Discussion

The model constructed by the intelligent agents shows some interesting correlations between the pre-DIC patients and their health status. There is focus on: weight (comorbidity, admit, and previous); white blood cell representation (basophils and eosinophils); respiration volume and rate; and nutrition (Braden and feeding tube residual). These variables are likely non-causal correlations to elevated pro-coagulant exposure in blood, which is the initializing of the DIC pathway.

7.7 Summary

The intelligent agents were instructed to create a predictive model of DIC based on the JAAM definition. the agents learned from 10% of the MIMIC database to create features. Those features were trained and tested (using 10-fold cross-validation) on 70% of the remaining MIMIC data. The resulting model was validated on the remaining 30% of data. The performance of the model on the validating data set is an AUC of 0.838 (0.828, 0.848).

This chapter demonstrates the predictive power of an intelligent agent approach to predicting the onset of DIC. There exist no peer-reviewed published predictive model of DIC, so there is no comparison. The agents highlight significant medical markers for the onset of DIC, producing a sensitivity of 0.787, a specificity of 0.712 and an accuracy of 0.723.

Six hours prior to the patient meeting the JAAM DIC criteria, the prediction model has a performance of an AUC of 0.810 (0.799, 0.821). There is no literature demonstrating the effects of early interventions with DIC, but treatment is urged to begin immediately by Wada et al. [93]. With a clinical tool that may predict onset of DIC six hours before with an AUC of 0.810 (0.799, 0.821) a clinician may be able to reduce mortality risk, reduce other adverse outcomes, or redirect the patient's trajectory entirely.

Chapter 8

CONCLUSIONS

The research effort described in this dissertation sought and succeeded to create an intelligent agent framework capable of data mining a database to predict an event. The framework was applied to an ICU EMR to investigate three diseases: ARDS, AKI, and sepsis. Several derivative investigations were devised from the initial results, and the research was expanded to include: SAHRF, DIC, the relationship between DIC and ARDS, and all Stages of AKI including progression to more severe Stages.

8.1 Contributions

By utilizing the intelligent agent framework created for this study and applying it to an ICU EMR several contributions are made to medicine and medical informatics. All four models created by the intelligent agents that have comparable peer-reviewed prediction models (ARDS, AKI Stage 1, AKI Stage 2, and sepsis) have a comparable or superior performance to the published counterpart. This dissertation presents seven novel prediction models not found in the peer-reviewed published literature, including: SAHRF; DIC from ARDS positive patients; ARDS from DIC positive patients; AKI Stage 3; the progression from AKI Stage 1 to Stage 2; the progression from AKI Stage 2 to Stage 3; and DIC.

The ARDS models created by this study have a superior AUC of 0.861 (0.838, 0.884). A published account of LIPS in an ICU setting [4] produces an AUC of 0.79, and even this dissertation's own re-implementation of LIPS on MIMIC produces an AUC of 0.847 (0.823, 0.871). The intelligent agents have a superior sensitivity (0.877 compared to the re-implementation of LIPS of 0.773) and accuracy (0.869 compared to 0.773).

The intelligent agent AKI models considers the urine output criteria of AKIN that no

other AKI prediction model considers, and represents between 11.6% and 12.4% of AKI patients at every Stage. The AKI Stage 1 model created by the intelligent agents of this study produce a superior AUC of 0.824 (0.815, 0.833) compared to the next highest AUC of 0.783 produced by Cheng et al. [7].

The AKI Stage 2 model created by the intelligent agents of this study produce a superior AUC of 0.926 (0.915, 0.937) compared to the only other published model by Mohamadlou et al. [59] with an AUC of 0.841, and this dissertation's re-implementation of Mohamadlou's model on MIMIC with an AUC of 0.918 (0.901, 0.935).

The sepsis model created by the intelligent agents of this study produce a comparable AUC of 0.874 (0.844, 0.904) to that of the highest performing most similar model by Desautels et al. [13] with an AUC of 0.88. In critical comparison, the intelligent agent's model has a superior specificity (0.918 compared to 0.80) and accuracy (0.908 compared to 0.80).

The performance of the novel prediction models demonstrate the impressive predictive capability of the intelligent agents. SAHRF prediction model produced an AUC of 0.952 (0.947, 0.957), a sensitivity of 0.914, a specificity of 0.862, and an accuracy of 0.872. Predicting DIC from ARDS patients produced a performance with an AUC of 0.722 (0.689, 0.755), a sensitivity of 0.527, a specificity of 0.788, and an accuracy of 0.670. Predicting ARDS from DIC positive patients produced a model with an AUC of 0.675 (0.639, 0.711), a sensitivity of 0.478, a specificity of 0.756, and an accuracy of 0.746. Predicting AKI Stage 3 produces a performance with an AUC of 0.983 (0.975, 0.985), a sensitivity of 0.884, a specificity of 0.998, and an accuracy of 0.988. Predicting the progression from AKI Stage 1 to Stage 2 produces an AUC of 0.930 (0.911, 0.949), a sensitivity of 0.750, a specificity of 0.873, and an accuracy of 0.845. Predicting the progression from AKI Stage 2 to Stage 3 produces an AUC of 0.951 (0.939, 0.963), a sensitivity of 0.834, a specificity of 0.952, and an accuracy of 0.893. Predicting DIC produced a model with an AUC of 0.838 (0.828, 0.848), a sensitivity of 0.787, a specificity of 0.712, and an accuracy of 0.723.

The significance of study populations contribute the the successes in this dissertation. The characterization of 915 ARDS and DIC positive patients represents a cohort nearly six

times larger than that of all published work combined. The prediction of AKI involves 11,383 patients, almost twice as large as all similarly published studies combined.

In the course of research this dissertation's efforts have discovered a correlation between DIC and ARDS. DIC and ARDS correlation is a known phenomenon (a sample size of 18 in [67] and [15]), and is suspected a causal phenomenon (a sample size of 142 [58]). By studying the 915 ARDS and DIC positive patients this dissertation has characterized an elevated mortality rate (nearly double the ARDS positive DIC negative prognosis) and incidence (64.5% of ARDS patients are or will be afflicted with DIC). This dissertation proposes a theoretical disease pathway that merges the ARDS pathway with the DIC pathway, and show the phenomenon to cause a positive feedback loop in Figure 4.9. This dissertations shows DIC is a comorbidity in 78% of all ARDS patients who succumb to in-hospital mortality.

This work also inspected the DIC positive patients who later develop ARDS and found them to significantly consist of trauma patients. When trauma is the underlying condition for DIC, the patient is twice as likely to develop ARDS.

The overarching contributions of this dissertation are the models and framework to create models that offer the basis of clinical tools. The clinical tools may serve to predict the onset of disease and allow for actionable early therapeutics to change the patients trajectory

8.2 Meta Analysis

The culmination of this research presents the rare opportunity to perform a meta analysis of unpredictable patients. The framework presented in this dissertation tracks the provenance of every patient through the validation models. For individual models, this research, describes the significant features of misclassified patients, which constitutes part of the limitations for each model. That analysis highlights the patient cohorts which are likely to be misclassified for a particular model.

This dissertation utilizes several models, all with the same patient population, giving the rare opportunity to examine misclassifications. Amassing the false positive patients and false negative patients from four primary diseases in this dissertation (ARDS, AKI, sepsis, and

DIC) it is possible to highlight patients who contribute to those misclassifications. This, meta analysis of unpredictable patients may set the limitations of predictive models in general.

Table H.1 denotes patients who were misclassified as false positive for the four primary diseases, which include the comorbidities of: congestive heart failure; cardiac arrhythmias; valvular disease; pulmonary circulation disorders; hypertension; chronic pulmonary disease; diabetes, uncomplicated; liver disease; solid tumor without metastasis; coagulopathy; blood loss anemia; and trauma. Table H.2 denotes patients who were misclassified as false negative for the four primary diseases, which include the comorbidities of: hypertension, uncomplicated; paralysis; renal failure; and trauma. Identifying patients who are difficult to predict helps to set the clinically contextual limitations of predictive models.

8.3 Future Work

8.3.1 Temporal Boundary Value

Each model devised in this dissertation has been assessed up to six hours prior, with differing stability in the models. Most are stable over time, but some lose a significant amount of predictive power over time such as DIC from ARDS in Table 4.11 that loses 10.7% of it's AUC over six hours prior to onset. Determining how many hours prior to the onset of disease a model is still valid to meet clinician's minimum performance for utility would add to the application of such a tool. The severity, progression, and predictability are different for different disease, the temporal boundary should be assessed per disease.

8.3.2 Probing β

The design of this research emphasized minimizing false negatives at the expense of false positives. The cutoff of each model derived from the general ICU population was set to be at the maximum F-measure with $\beta = 1.75$. This threshold setting was chosen because the risk to a false negative patient (delayed therapeutics) outweighs the risk to a false positive patient (unnecessary intervention). The severity of the diseases chosen for this study warrants early

intervention to avoid mortality and adverse outcomes. Different diseases and disease stages may have a different clinically significant β setting. Probing the balance of false positives and false negatives in the setting where predictive models will be employed is prudent.

8.3.3 Validating

Translating a model to a clinical tool is necessarily complicated and thorough. The next progression of the research conducted in this dissertation is to validate the methods and models in other ICU EMRs.

8.3.4 ARDS-DIC Relationship

Much work is necessary to probe the relationship between ARDS and DIC. This research has definitively shown correlation. To verify causality, as described in Table 4.9, is outside the capability of this retrospective research. The initial results by Miyoshi et al. [58] and the predictive capabilities of the intelligent agents in this dissertation are telling of causality, but neither are definitive.

8.3.5 High Throughput

This dissertation has taken great steps to generalize the approach of data mining databases. The limitations of generating predictive models under this dissertation's framework are mostly computational. The human centered front-end work is limited to carefully inputting the definitions of disease onset. A reasonable next step would be to focus on scaling this research to encompass many disease. To accomplish up-scaling future efforts will create a repository of computational disease onsets. The repository will act as a computational assay. When ICU databases become available for research, this assay may be applied to the database to generate or validate predictive models. As no aspect of the predictive models need identifiable information, any such initiative may qualify for minimal risk for an Institutional Review Board (IRB) under an existing secondary data use exemption.

8.3.6 Learning Events Generalized

With some adaptation to the original intelligent agent framework, agents may be given the tools necessary to define the clinical events they then predict. An agent framework that can define a clinical event (with clustering algorithms) and then predict the onset of the event (with the tools currently in place) can transcend the original focus of disease onset. Such an enhanced framework may find novel correlative relationships that translate into disease: classifications; stages; or risk stratification.

BIBLIOGRAPHY

[1] D. G. Ashbaugh, D. B. Bigelow, T. L. Petty, and B. E. Levine. Acute respiratory distress in adults. *Lancet*, 2(7511):319–323, Aug 1967.

[2] S. D. Barbar, R. Clere-Jehl, A. Bourredjem, R. Hernu, F. Montini, R. Bruyere, C. Lebert, J. Bohe, J. Badie, J. P. Eraldi, J. P. Rigaud, B. Levy, S. Siami, G. Louis, L. Bouadma, J. M. Constantin, E. Mercier, K. Klouche, D. du Cheyron, G. Piton, D. Annane, S. Jaber, T. van der Linden, G. Blasco, J. P. Mira, C. Schwebel, L. Chimot, P. Guiot, M. A. Nay, F. Meziani, J. Helms, C. Roger, B. Louart, R. Trusson, A. Dargent, C. Binquet, J. P. Quenot, L. Arene, L. Delapierre, A. L. Merel, S. Moschietto, G. Pradel, A. Privat, M. Feissel, L. Barrot, N. Belin, E. Belle, F. Belon, G. Capellier, C. Chaignat, M. Claveau, M. Grandperrin, G. Labro, P. Luporsi, C. Manzon, J. C. Navellou, C. Patry, G. Besch, Y. Brunin, L. Carteron, M. Ginet, J. Paillot, S. Pili-Floury, E. Samain, S. Gaillard, A. Robine, E. Sauvadet, N. Sedillot, X. Tchenio, J. Brunet, C. Daubin, B. Sauneuf, A. Seguin, X. Valette, S. Cayot-Constantin, R. Guerin, M. Jabaudon, J. Pascal, S. Perbet, F. Bougerol, P. L. Declercq, S. Gelinotte, P. Andreu, P. E. Charles, M. Hamet, N. Jacquiot, A. Large, S. Mortreux, A. Pavon, S. Prin, A. Roudaut, A. Toitot, J. Aboab, D. Friedman, V. Maxime, X. Ambrosi, N. Heming, C. Ara Somohano, A. Bonadona, J. C. Cartier, R. Hamidfar-Roy, M. Lugosi, C. Minet-Binauld, L. Potton, N. Terzi, B. Vivet, Y. Aihem, G. Colin, M. Fiancette, M. Henry-Lagarrigue, J. C. Lacherade, L. Martin-Lefevre, I. Vinatier, C. Fourdin, M. Ledein, C. Thomas, L. Argaud, T. Baudry, T. Chroboczek, M. Cour, S. De la Salle, E. Faucher, A. Grateau, T. Madeleine, M. Simon, C. Bernet, O. Fontaine-Kesteloot, A. Friggeri, S. Ledochowski, A. Lepape, N. Mottard, M. Rinaudo, E. Rulliat-Giraud, N. Sens, F. Thiolliere, F. Wallet, G. Barberet, M. Egard, F. Ganster, K. Kuteifan, Y. Mootien, A. Poidevin, T. Auchet, N. Ducrocq, A. Kimmoun, E. Novy, P. Perez, C. Thivilier, J. B. Lascarrou, A. Ayral, C. Bengler, C. Boutin, G. Dingemans, L. Elotmani, F. Garnier, J. M. Le Goff, S. Lloret, P. Massanet, B. Jung, L. Platon, M. Cisse, M. Conseil, F. Belafia, G. Chanques, N. Clavieras, Y. Coisel, J. Carr, F. Barbier, D. Benzekri Lefevre, T. Boulain, A. Bretagnol, A. Mathonnet, G. Muller, I. Runge, B. Mourvillier, R. Sonneville, J. F. Timsit, M. Arnaout, W. Bougouin, A. Cariou, B. Champigneulle, J. Charpentier, J. D. Chiche, G. Geri, F. C. Pene, J. Rouche, S. Spagnolo, S. Bedon-Carte, P. H. Dessalles, Y. Monseau, M. Saint-Leger, A. Berger, A. Boivin, X. Delabranche, S. Heenen, T. Khouri, C. Kummerien, O. Martinet, A. Monnier, A. R. Neagu, Y. Rabouel, H. Rahmani, D. Schnell, L. Bodet-Contentin, P. F. Dequin, S. Ehrmann, D. Garot, A. Guillon, A. Joret, Y. Jouan, A. Legras, C. Lhommet, J. Mankikian, E. Rouve, and C. Salmon-Gandonniere. Timing

of Renal-Replacement Therapy in Patients with Acute Kidney Injury and Sepsis. *N. Engl. J. Med.*, 379(15):1431–1442, Oct 2018.

[3] R. Bellomo, C. Ronco, J. A. Kellum, R. L. Mehta, and P. Palevsky. Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care*, 8(4):R204–212, Aug 2004.

[4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] J. S. Calvert, D. A. Price, U. K. Chettipally, C. W. Barton, M. D. Feldman, J. L. Hoffman, M. Jay, and R. Das. A computational approach to early sepsis detection. *Comput. Biol. Med.*, 74:69–73, 07 2016.

[6] L. S. Chawla, R. Bellomo, A. Bihorac, S. L. Goldstein, E. D. Siew, S. M. Bagshaw, D. Bittleman, D. Cruz, Z. Endre, R. L. Fitzgerald, L. Forni, S. L. Kane-Gill, E. Hoste, J. Koyner, K. D. Liu, E. Macedo, R. Mehta, P. Murray, M. Nadim, M. Ostermann, P. M. Palevsky, N. Pannu, M. Rosner, R. Wald, A. Zarbock, C. Ronco, and J. A. Kellum. Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup. *Nat Rev Nephrol*, 13(4):241–257, 04 2017.

[7] P. Cheng, L. R. Waitman, Y. Hu, and M. Liu. Predicting Inpatient Acute Kidney Injury over Different Time Horizons: How Early and Accurate? *AMIA Annu Symp Proc*, 2017:565–574, 2017.

[8] John G. Cleary and Leonard E. Trigg. K*: An instance-based learner using an entropic distance measure. In *12th International Conference on Machine Learning*, pages 108–114, 1995.

[9] William W. Cohen. Fast effective rule induction. In *Twelfth International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann, 1995.

[10] Lucy Vanderwende Eric Horvitz Dae Hyun Lee1, Meliha Yetisgen. Predicting severe clinical events by learning about life-saving actions and outcomes using distant supervision (in press). *JBI*, 2019.

[11] O. K. Danner, S. Hendren, E. Santiago, B. Nye, and P. Abraham. Physiologically-based, predictive analytics using the heart-rate-to-Systolic-Ratio significantly improves the timeliness and accuracy of sepsis prediction compared to SIRS. *Am. J. Surg.*, 213(4):617–621, Apr 2017.

[12] S. De Rosa, S. Samoni, and C. Ronco. Creatinine-based definitions: from baseline creatinine to serum creatinine adjustment in intensive care. *Crit Care*, 20:69, Mar 2016.

[13] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, D. J. Wales, and R. Das. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform*, 4(3):e28, Sep 2016.

[14] Mark A. Hall Eibe Frank and Ian H. Witten. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016.

[15] F. A. el Kassimi, S. Al-Mashhadani, A. K. Abdullah, and J. Akhtar. Adult respiratory distress syndrome and disseminated intravascular coagulation complicating heat stroke. *Chest*, 90(4):571–574, Oct 1986.

[16] A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey. Comorbidity measures for use with administrative data. *Med Care*, 36(1):8–27, Jan 1998.

[17] C. H. Espinel. The FENa test. Use in the differential diagnosis of acute renal failure. *JAMA*, 236(6):579–581, Aug 1976.

[18] Bauman et al. Lung injury prediction score is useful in predicting acute respiratory distress syndrome and mortality in surgical critical care patients. *Critical Care Research and Practice*, 2015(10):8, 2015.

[19] M. Faisal, A. Scally, D. Richardson, K. Beatson, R. Howes, K. Speed, and M. A. Mohammed. Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results. *Crit. Care Med.*, 46(4):612–618, Apr 2018.

[20] E. Fan, D. Brodie, and A. S. Slutsky. Acute Respiratory Distress Syndrome: Advances in Diagnosis and Treatment. *JAMA*, 319(7):698–710, 02 2018.

[21] S. Farooqi and J. G. Dickhout. Major comorbid disease processes associated with increased incidence of acute kidney injury. *World J Nephrol*, 5(2):139–146, Mar 2016.

[22] R. Ferrer, A. Artigas, M. M. Levy, J. Blanco, G. Gonzalez-Diaz, J. Garnacho-Montero, J. Ibanez, E. Palencia, M. Quintana, M. V. de la Torre-Prados, A. Artigas, J. Blanco, G. Gonzalez-Diaz, R. Ferrer, J. Garnacho-Montero, J. Ibanez, M. M. Levy, E. Palencia, M. Quintana, M. V. de la Torre, A. Artigas, R. Ferrer, M. M. Levy, J. Blanco,

G. Gonzalez-Diaz, R. Ferrer, J. Garnacho-Montero, J. Ibanez, E. Palencia, M. Quintana, M. V. de la Torre, G. Goma, D. Suarez, J. Real, I. Martin, A. Navas, R. Ferrer, A. Artigas, M. Alvarez, J. M. Sirvent, S. Herranz Ulldemolins, P. Galdos, G. Balziscueta, P. Marco, I. Azkarate, R. Sierra, J. J. Izua, J. Castano, A. Ambros, J. Ortega, V. Corcoles, L. Tamayo, D. Carriedo, M. Llorente, P. Merino, E. Bustamante, E. Palencia, P. Garcia Olivares, P. Santa Teresa Zamarro, C. Perez, A. Renedo, S. Nicolas-Franco, M. Salome Sanchez, F. J. Gil, M. J. Gomez, E. Piacentini, A. Loza, J. Ibanez, S. Rodriguez, J. A. Berezo, J. Blanco, A. Gaban, M. J. Lopez Cambra, A. Tallet, M. Martinez, J. A. Fernandez, F. Callejo, M. J. Lopez Pueyo, F. Gandia, J. Fernandez, J. C. Ballesteros, M. T. Antuna, S. Herrero, M. Valledor, J. Gutierrez, C. Perez, O. Rodriguez, R. Dominguez, J. Peinado, M. V. de la Torre, C. Salazar, M. de la Cruz Martin, J. Ramon, F. Iglesias Llaca, L. Forcelledo Espina, F. Taboada Costa, J. A. Gonzalo Guerra, F. J. Guerrero, F. Canada, M. Balaguer, I. Mertin, C. Lopez, D. Sanchez, J. Costa, M. Calizaya, A. Arenaza, A. Morillo, D. Del Toro, T. Guzman, A. Blesa, F. Martinez, A. Moneo, J. Broch, J. A. Camacho, F. J. Garcia, X. L. Perez, N. Garcia, J. C. Ruiz, J. Caballero, E. Francisco, T. Requena, A. Ruiz, J. L. Boveda, J. M. Soto, C. Tormo, R. Blancas, M. Quintana, M. A. Taberna, J. M. Anon, J. B. Aranjo, M. Rodriguez, J. M. Garcia, I. Rodriguez, J. Huertos, C. Ortiz, E. Yuste, J. F. Machado, D. Ocana, R. Vegas, and L. Vallejo. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *JAMA*, 299(19):2294–2303, May 2008.

[23] Eibe Frank and Ian H. Witten. Generating accurate rule sets without global optimization. In J. Shavlik, editor, *Fifteenth International Conference on Machine Learning*, pages 144–151. Morgan Kaufmann, 1998.

[24] Y. Freund and R. E. Schapire. Large margin classification using the perceptron algorithm. In *11th Annual Conference on Computational Learning Theory*, pages 209–217, New York, NY, 1998. ACM Press.

[25] O. Gajic, O. Dabbagh, P. K. Park, A. Adesanya, S. Y. Chang, P. Hou, H. Anderson, J. J. Hoth, M. E. Mikkelsen, N. T. Gentile, M. N. Gong, D. Talmor, E. Bajwa, T. R. Watkins, E. Festic, M. Yilmaz, R. Iscimen, D. A. Kaufman, A. M. Esper, R. Sadikot, I. Douglas, J. Sevransky, M. Malinchoc, A. Ahmed, O. Gajic, M. Malinchoc, D. J. Kor, B. Afessa, R. Cartin-Ceba, O. Dabbagh, N. Nagam, S. Patel, A. Karo, B. Hess, P. K. Park, J. Harris, L. Napolitano, K. Raghavendran, R. C. Hyzy, J. Blum, C. Dean, A. Adesanya, S. Hosur, V. Enoh, S. Y. Chang, A. Patrawalla, M. Elie, P. C. Hou, J. M. Barry, I. Shempp, A. Malhotra, G. Frendl, H. Anderson, K. Tchorz, M. C. McCarthy, D. Uddin, J. J. Hoth, B. Yoza, M. Mikkelsen, J. D. Christie, D. F. Gaieski, P. Lanken, N. Meyer, C. Shah, N. T. Gentile, K. Stevenson, B. Freeman, S. Srinivasan, M. N. Gong, D. Talmor, S. P. Bender, M. Garcia, E. Bajwa, A. Malhotra, B. T. Thompson, D. C. Christiani, T. R. Watkins, S. Deem, M. Treggiari, E. Festic, A. Lee, J. Daniels,

M. Cengiz, M. Yilmaz, R. Iscimen, D. Kaufman, A. Esper, G. Martin, R. Sadikot, I. Douglas, and J. Sevransky. Early identification of patients at risk of acute lung injury: evaluation of lung injury prediction score in a multicenter cohort study. *Am. J. Respir. Crit. Care Med.*, 183(4):462–470, Feb 2011.

[26] O. Gajic, S. I. Dara, J. L. Mendez, A. O. Adesanya, E. Festic, S. M. Caples, R. Rana, J. L. St Sauver, J. F. Lymp, B. Afessa, and R. D. Hubmayr. Ventilator-associated lung injury in patients without acute lung injury at the onset of mechanical ventilation. *Crit. Care Med.*, 32(9):1817–1824, Sep 2004.

[27] S. Gando, T. Iba, Y. Eguchi, Y. Ohtomo, K. Okamoto, K. Koseki, T. Mayumi, A. Murata, T. Ikeda, H. Ishikura, M. Ueyama, H. Ogura, S. Kushimoto, D. Saitoh, S. Endo, and S. Shimazaki. A multicenter, prospective validation of disseminated intravascular coagulation diagnostic criteria for critically ill patients: comparing current criteria. *Crit. Care Med.*, 34(3):625–631, Mar 2006.

[28] F. Gao, T. Melody, D. F. Daniels, S. Giles, and S. Fox. The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: a prospective observational study. *Crit Care*, 9(6):R764–770, 2005.

[29] S. Gaudry, D. Hajage, F. Schortgen, L. Martin-Lefevre, B. Pons, E. Boulet, A. Boyer, G. Chevrel, N. Lerolle, D. Carpentier, N. de Prost, A. Lautrette, A. Bretagnol, J. Mayaux, S. Nseir, B. Megarbane, M. Thirion, J. M. Forel, J. Maizel, H. Yonis, P. Markowicz, G. Thiery, F. Tubach, J. D. Ricard, D. Dreyfuss, A. Thiagarajah, M. Fiancette, C. Gisbert-Mora, J. Messika, D. Roux, P. Cronier, F. Tamion, R. Favory, T. Boulain, A. Demoule, M. Adda, A. Goury, G. Thomas, M. Mentec, M. Slama, C. Guerin, E. Messai, J. Richecoeur, D. Combaux, E. Demontmolin, D. Da Silva, N. Bige, B. Guidet, B. Zuber, G. Lacave, N. Chudeau, J. Dellamonica, M. Darmon, P. Fangio, V. Das, C. Blayau, and L. Bouadma. Initiation Strategies for Renal-Replacement Therapy in the Intensive Care Unit. *N. Engl. J. Med.*, 375(2):122–133, Jul 2016.

[30] M. N. Gong, B. T. Thompson, P. Williams, L. Pothier, P. D. Boyce, and D. C. Christiani. Clinical predictors of and mortality in acute respiratory distress syndrome: potential role of red cell transfusion. *Crit. Care Med.*, 33(6):1191–1198, Jun 2005.

[31] A. Gouel-Cheron, B. Allaouchiche, C. Guignant, F. Davin, B. Floccard, and G. Monneret. Early interleukin-6 and slope of monocyte human leukocyte antigen-DR: a powerful association to predict the development of sepsis after major trauma. *PLoS ONE*, 7(3):e33095, 2012.

[32] Robert. Friedman Jerome. Hastie, Trevor. Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York, NY, 2009.

[33] Jianqin He, Lijuan Wu, Xiangzhou Zhang, Yong Hu, Lemuel R Waitman, and Mei Liu. Multi-perspective predictive modeling for acute kidney injury in general hospital populations using electronic medical records. *JAMIA Open*, 2(1):115–122, 11 2018.

[34] K. E. Henry, D. N. Hager, P. J. Pronovost, and S. Saria. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med*, 7(299):299ra122, Aug 2015.

[35] R.C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11:63–91, 1993.

[36] Pat. Iba, Wayne. Langley. Introduction of one-level decision trees. In *Ninth International Conference on Machine Learning*, pages 223–240. Morgan Kaufmann, 1992.

[37] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo, 1995. Morgan Kaufmann.

[38] R. H. Kallet, W. Corral, H. J. Silverman, and J. M. Luce. Implementation of a low tidal volume ventilation protocol for patients with acute lung injury or acute respiratory distress syndrome. *Respir Care*, 46(10):1024–1037, Oct 2001.

[39] Rohit J. Kate, Ruth M. Perez, Debesh Mazumdar, Kalyan S. Pasupathy, and Vani Nilakantan. Prediction and detection models for acute kidney injury in hospitalized older adults. *BMC Medical Informatics and Decision Making*, 16(1):39, Mar 2016.

[40] A. Khwaja. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract*, 120(4):c179–184, 2012.

[41] Y. Koh. How to approach the acute respiratory distress syndrome: Prevention, plan, and prudence. *Respir Investig*, 55(3):190–195, May 2017.

[42] Ron Kohavi. The power of decision tables. In *8th European Conference on Machine Learning*, pages 174–189. Springer, 1995.

[43] A. Kortgen, P. Niederprum, and M. Bauer. Implementation of an evidence-based "standard operating procedure" and outcome in septic shock. *Crit. Care Med.*, 34(4):943–949, Apr 2006.

[44] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar, and M. Cheang. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit. Care Med.*, 34(6):1589–1596, Jun 2006.

[45] Niels Landwehr, Mark Hall, and Eibe Frank. Logistic model trees. *Machine Learning*, 95(1-2):161–205, 2005.

[46] J. LaValle, S. Kuffner. Randomized kinodynamic planning. In *IEEE International Conference on Robotics and Automation*, pages 473–479. IEEE, 1999.

[47] N. Layios, C. Delierneux, A. Hego, J. Huart, C. Gosset, C. Lecut, N. Maes, P. Geurts, A. Joly, P. Lancellotti, A. Albert, P. Damas, A. Gothot, and C. Oury. Sepsis prediction in critically ill patients by platelet activation markers on ICU admission: a prospective pilot study. *Intensive Care Med Exp*, 5(1):32, Dec 2017.

[48] S. le Cessie and J.C. van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201, 1992.

[49] H. A. Lindner, U. Balaban, T. Sturm, C. Weiss, M. Thiel, and V. Schneider-Lindner. An Algorithm for Systemic Inflammatory Response Syndrome Criteria-Based Prediction of Sepsis in a Polytrauma Cohort. *Crit. Care Med.*, 44(12):2199–2207, 12 2016.

[50] R. A. Lukaszewski, A. M. Yates, M. C. Jackson, K. Swingler, J. M. Scherer, A. J. Simpson, P. Sadler, P. McQuillan, R. W. Titball, T. J. Brooks, and M. J. Pearce. Presymptomatic prediction of sepsis in intensive care unit patients. *Clin. Vaccine Immunol.*, 15(7):1089–1094, Jul 2008.

[51] K. Makris and L. Spanou. Acute Kidney Injury: Definition, Pathophysiology and Clinical Phenotypes. *Clin Biochem Rev*, 37(2):85–98, May 2016.

[52] Q. Mao, M. Jay, J. L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Y. Zhou, and R. Das. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*, 8(1):e017833, 01 2018.

[53] D. Mardi, B. Fwity, R. Lobmann, and A. Ambrosch. Mean cell volume of neutrophils and monocytes compared with C-reactive protein, interleukin-6 and white blood cell count for prediction of sepsis and nonsystemic bacterial infections. *Int J Lab Hematol*, 32(4):410–418, Aug 2010.

[54] T. Matsuda. Clinical aspects of DIC–disseminated intravascular coagulation. *Pol J Pharmacol*, 48(1):73–75, 1996.

[55] F. B. Mayr, S. Yende, and D. C. Angus. Epidemiology of severe sepsis. *Virulence*, 5(1):4–11, Jan 2014.

[56] A. McCoy and R. Das. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual*, 6(2):e000158, 2017.

[57] R. L. Mehta, J. A. Kellum, S. V. Shah, B. A. Molitoris, C. Ronco, D. G. Warnock, and A. Levin. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Crit Care*, 11(2):R31, 2007.

[58] S. Miyoshi, R. Ito, H. Katayama, K. Dote, M. Aibiki, H. Hamada, T. Okura, and J. Higaki. Combination therapy with sivelestat and recombinant human soluble thrombomodulin for ARDS and DIC patients. *Drug Des Devel Ther*, 8:1211–1219, 2014.

[59] H. Mohamadlou, A. Lynn-Palevsky, C. Barton, U. Chettipally, L. Shieh, J. Calvert, N. R. Saber, and R. Das. Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data. *Can J Kidney Health Dis*, 5:2054358118776326, 2018.

[60] D. B. Morgan, M. E. Carver, and R. B. Payne. Plasma creatinine and urea: creatinine ratio in patients with raised plasma urea. *Br Med J*, 2(6092):929–932, Oct 1977.

[61] P. Msaouel, T. Kappos, A. Tasoulis, A. P. Apostolopoulos, I. Lekkas, E. S. Tripodaki, and N. C. Keramaris. Assessment of cognitive biases and biostatistics knowledge of medical residents: a multicenter, cross-sectional questionnaire study. *Med Educ Online*, 19:23646, 2014.

[62] S. Negi, D. Koreeda, S. Kobayashi, T. Yano, K. Tatsuta, T. Mima, T. Shigematsu, and M. Ohya. Acute kidney injury: Epidemiology, outcomes, complications, and therapeutic strategies. *Semin Dial*, 31(5):519–527, 09 2018.

[63] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit. Care Med.*, 46(4):547–553, Apr 2018.

[64] H. B. Nguyen, E. L. Lynch, J. A. Mou, K. Lyon, W. A. Wittlake, and S. W. Corbett. The utility of a quality improvement bundle in bridging the gap between research and standard care in the management of severe sepsis and septic shock in the emergency department. *Acad Emerg Med*, 14(11):1079–1086, Nov 2007.

[65] T. J. Nuckton, J. A. Alonso, R. H. Kallet, B. M. Daniel, J. F. Pittet, M. D. Eisner, and M. A. Matthay. Pulmonary dead-space fraction as a risk factor for death in the acute respiratory distress syndrome. *N. Engl. J. Med.*, 346(17):1281–1286, Apr 2002.

[66] Office of Community Health Systems. *Washington State Trauma Registry Users Guide*. Washington State Department of Health, Olympia, WA, 9 2011.

[67] R. Ogawa, Y. Takano, and T. Fujita. Disseminated intravascular coagulation in the pathogenesis of adult respiratory distress syndrome: 2. Experimental study. *Jpn J Surg*, 7(4):223–229, Dec 1977.

[68] E. M. Olenick, K. S. Zimbro, G. M. D?Lima, P. Ver Schneider, and D. Jones. Predicting Sepsis Risk Using the "Sniffer" Algorithm in the Electronic Medical Record. *J Nurs Care Qual*, 32(1):25–31, 2017.

[69] Y. Panahi, M. Mojtahedzadeh, A. Najafi, M. R. Ghaini, M. Abdollahi, M. Sharifzadeh, A. Ahmadi, S. M. Rajaee, and A. Sahebkar. The role of magnesium sulfate in the intensive care unit. *EXCLI J*, 16:464–482, 2017.

[70] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.

[71] Ruth Pordes, Don Petravick, Bill Kramer, Doug Olson, Miron Livny, Alain Roy, Paul Avery, Kent Blackburn, Torre Wenaus, Frank Wrthwein, Ian Foster, Rob Gardner, Mike Wilde, Alan Blatecky, John McGee, and Rob Quick. The open science grid. *Journal of Physics: Conference Series*, 78:012057, jul 2007.

[72] H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J. C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*, 43(11):1130–1139, Nov 2005.

[73] Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[74] V. M. Ranieri, G. D. Rubenfeld, B. T. Thompson, N. D. Ferguson, E. Caldwell, E. Fan, L. Camporota, A. S. Slutsky, V. Ranieri, G. D. Rubenfeld, B. Thompson, N. D. Ferguson, E. Caldwell, E. Fan, L. Camporota, A. S. Slutsky, M. Antonelli, A. Anzueto, R. Beale, L. Brochard, R. Brower, A. Esteban, L. Gattinoni, A. Rhodes, J. L. Vincent, A. Bersten, D. Needham, and A. Pesenti. Acute respiratory distress syndrome: the Berlin Definition. *JAMA*, 307(23):2526–2533, Jun 2012.

[75] E. Rivers, B. Nguyen, S. Havstad, J. Ressler, A. Muzzin, B. Knoblich, E. Peterson, and M. Tomlanovich. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N. Engl. J. Med.*, 345(19):1368–1377, Nov 2001.

[76] M. Rothman, M. Levy, R. P. Dellinger, S. L. Jones, R. L. Fogerty, K. G. Voelker, B. Gross, A. Marchetti, and J. Beals. Sepsis as 2 problems: Identifying sepsis at admission and predicting onset in the hospital using an electronic medical record-based acuity score. *J Crit Care*, 38:237–244, 04 2017.

[77] G. Saposnik, D. Redelmeier, C. C. Ruff, and P. N. Tobler. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*, 16(1):138, 11 2016.

[78] M. Saqib, Y. Sha, and M. D. Wang. Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks. *Conf Proc IEEE Eng Med Biol Soc*, 2018:4038–4041, Jul 2018.

[79] Igor Sfiligoi, Daniel C. Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, and Frank Wurthwein. The pilot way to grid resources using glideinwms. In *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 02*, CSIE '09, pages 428–432, Washington, DC, USA, 2009. IEEE Computer Society.

[80] M. Shankar-Hari, D. Datta, J. Wilson, V. Assi, J. Stephen, C. J. Weir, J. Rennie, J. Antonelli, A. Bateman, J. M. Felton, N. Warner, K. Judge, J. Keenan, A. Wang, T. Burpee, A. K. Brown, S. M. Lewis, T. Mare, A. I. Roy, J. Wright, G. Hulme, I. Dimmick, A. Gray, A. G. Rossi, A. J. Simpson, A. Conway Morris, and T. S. Walsh. Early PREdiction of sepsis using leukocyte surface biomarkers: the ExPRES-sepsis cohort study. *Intensive Care Med*, 44(11):1836–1848, Nov 2018.

[81] S. P. Shashikumar, Q. Li, G. D. Clifford, and S. Nemati. Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas*, 38(12):2235–2248, Nov 2017.

[82] S. C. Shiramizo, A. R. Marra, M. S. Durao, A. T. Paes, M. B. Edmond, and O. F. Pavao dos Santos. Decreasing mortality in severe sepsis and septic shock patients by implementing a sepsis bundle in a hospital setting. *PLoS ONE*, 6(11):e26790, 2011.

[83] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J. D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. L.

Vincent, and D. C. Angus. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8):801–810, Feb 2016.

[84] G. J. Soto, D. J. Kor, P. K. Park, P. C. Hou, D. A. Kaufman, M. Kim, H. Yadav, N. Teman, M. C. Hsu, T. Shvilkina, Y. Grewal, M. De Aguirre, S. Gunda, O. Gajic, and M. N. Gong. Lung Injury Prediction Score in Hospitalized Patients at Risk of Acute Respiratory Distress Syndrome. *Crit. Care Med.*, 44(12):2182–2191, Dec 2016.

[85] Friedman B. Sutton JP. *Trends in Septicemia Hospitalizations and Readmissions in Selected HCUP States, 2005 and 2010: Statistical Brief 161.* Agency for Healthcare Research and Quality, Rockville, MD, 9 2013.

[86] F. B. Taylor, C. H. Toh, W. K. Hoots, H. Wada, and M. Levi. Towards definition, clinical and laboratory criteria, and a scoring system for disseminated intravascular coagulation. *Thromb. Haemost.*, 86(5):1327–1330, Nov 2001.

[87] J. M. Teles, E. Silva, G. Westphal, R. C. Filho, and F. R. Machado. Surviving sepsis campaign in Brazil. *Shock*, 30 Suppl 1:47–52, Oct 2008.

[88] Andrews RM Torio CM. *National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: Statistical Brief 160.* Agency for Healthcare Research and Quality, Rockville, MD, 8 2013.

[89] C. Trillo-Alvarez, R. Cartin-Ceba, D. J. Kor, M. Kojicic, R. Kashyap, S. Thakur, L. Thakur, V. Herasevich, M. Malinchoc, and O. Gajic. Acute lung injury prediction score: derivation and validation in a population-based sample. *Eur. Respir. J.*, 37(3):604–609, Mar 2011.

[90] T. Van der Linden, B. Souweine, L. Dupic, L. Soufir, and P. Meyer. Management of thrombocytopenia in the ICU (pregnancy excluded). *Ann Intensive Care*, 2(1):42, Aug 2012.

[91] F. van Wyk, A. Khojandi, A. Mohammed, E. Begoli, R. L. Davis, and R. Kamaleswaran. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int J Med Inform*, 122:55–62, Feb 2019.

[92] A. G. Vassiliou, Z. Mastora, E. Jahaj, A. Koutsoukou, S. E. Orfanos, and A. Kotanidou. Does serum lactate combined with soluble endothelial selectins at ICU admission predict sepsis development? *In Vivo*, 29(2):305–308, 2015.

[93] H. Wada, T. Matsumoto, and Y. Yamashita. Diagnosis and treatment of disseminated intravascular coagulation (DIC) according to four DIC guidelines. *J Intensive Care*, 2(1):15, 2014.

[94] H. E. Wang, J. P. Donnelly, R. Griffin, E. B. Levitan, N. I. Shapiro, G. Howard, and M. M. Safford. Derivation of Novel Risk Prediction Scores for Community-Acquired Sepsis and Severe Sepsis. *Crit. Care Med.*, 44(7):1285–1294, 07 2016.

[95] M. Zambon, M. Ceola, R. Almeida-de Castro, A. Gullo, and J. L. Vincent. Implementation of the Surviving Sepsis Campaign guidelines for severe sepsis and septic shock: we could go faster. *J Crit Care*, 23(4):455–460, Dec 2008.

[96] A. Zarbock, J. A. Kellum, C. Schmidt, H. Van Aken, C. Wempe, H. Pavenstadt, A. Boanta, J. Gerss, and M. Meersch. Effect of Early vs Delayed Initiation of Renal Replacement Therapy on Mortality in Critically Ill Patients With Acute Kidney Injury: The ELAIN Randomized Clinical Trial. *JAMA*, 315(20):2190–2199, 2016.

Appendix A

MIMIC-III

A.0.1 Admissions

The admissions table (Table A.1) contains information relevant to the patient's stay in the hospital. Every entry of this database table has one unique HADM_ID (Hospital ADMissions IDentifier). This table forges the relationship between a patient (SUBJECT_ID) and a hospital admission (HADM_ID).

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via that Patient table (see subsection A.0.13).
- HADM_ID: A unique identifier to a hospital admissions.
- ADMITTIME: Timestamp for the hospital admit time.
- DISCHTIME: Timestamp for the hospital discharge time.
- DEATHTIME: Timestamp for death if the patient died in-hospital.
- INSURANCE: Insurance classification at the time of hospital admissions.
- LANGUAGE: Primary language spoken by the patient.
- RELIGION: Religious preference of the patient.

- MARITAL_STATUS: Marital status at the time of hospital admissions.
- ETHNICITY: Race and ethnic disposition preference.

Relationships for this table exist as:

- One patient (see subsection A.0.13) can have many admissions keyed by SUBJECT_ID.
- One admission can have multiple ICU stays (see subsection A.0.7).
- One admission can have multiple laboratory events (see subsection A.0.10).
- One admission can have multiple microbiology laboratory events (see subsection A.0.11).

ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	INSURANCE	LANGUAGE	RELIGION	MARITAL_STATUS	ETHNICITY
1	2	163353	2138-07-17 19:04:00	2138-07-21 15:48:00	NULL	Private	NULL	NOT SPECIFIED	NULL	ASIAN
2	3	145834	2101-10-20 19:08:00	2101-10-31 13:58:00	NULL	Medicare	NULL	CATHOLIC	MARRIED	WHITE
3	4	185777	2191-03-16 00:28:00	2191-03-23 18:41:00	NULL	Private	NULL	PROTESTANT QUAKER	SINGLE	WHITE
4	5	178980	2103-02-02 04:31:00	2103-02-04 12:15:00	NULL	Private	NULL	BUDDHIST	NULL	ASIAN
5	6	107064	2175-05-30 07:15:00	2175-06-15 16:00:00	NULL	Medicare	ENGL	NOT SPECIFIED	MARRIED	WHITE
6	7	118037	2121-05-23 15:05:00	2121-05-27 11:57:00	NULL	Private	NULL	CATHOLIC	NULL	WHITE
7	8	159514	2117-11-20 10:22:00	2117-11-24 14:20:00	NULL	Private	NULL	CATHOLIC	NULL	WHITE
8	9	150750	2149-11-09 13:06:00	2149-11-14 10:15:00	2149-11-14 10:15:00	Medicaid	NULL	UNOBTAINABLE	NULL	UNKNOWN/NOT SPECIFIED
9	10	184167	2103-06-28 11:36:00	2103-07-06 12:10:00	NULL	Medicaid	NULL	UNOBTAINABLE	NULL	BLACK/AFRICAN AMERICAN
10	11	194540	2178-04-16 06:18:00	2178-05-11 19:00:00	NULL	Private	NULL	OTHER	MARRIED	WHITE

Table A.1: MIMIC Admissions.

A.0.2 Chartevents

The Chartevents table (Table A.2) houses all events captured in the patient's electronic chart. This includes all data captured at the bedside, like vital signs or fingerstick tests. It also includes settings made at the bedside, like ventilator settings. Additional information is captured in the electronic chart per the discretion of the provider.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- ICUSTAY_ID: An identifier to an ICU stay that links to information specific to the ICU stay via the ICU Stay table (see subsection A.0.7).
- ITEMID: An item identifier that links to item specific information via the Items table (see subsection A.0.4).
- CHARTTIME: Is a record of the timestamp the observation was made.
- STORETIME: Is a record of the timestamp the information was entered into the electronic chart.
- CGID: CGID a CareGiver IDentifier for the individual who validated the measurment or recording.
- VALUE: This field contains the text representation of the value of the measurement or setting.

- VALUENUM: If VALUE is numeric, this field represents that numeric value, otherwise it is left null.

- VALUEUOM: If valid, this field contains the Unit Of Measurement for VALUE.

Relationships for this table exist as:

- One ICU stay (see subsection A.0.7) may have multiple Chartevents keyed by ICUSTAY_ID.

- One Chartevent has exactly one Item (see subsection A.0.4).

ROW_ID	SUBJECT_ID	HADM_ID	ICUSTAY_ID	ITEMID	CHARTTIME	STORETIME	CGID	VALUE	VALUENUM	VALUEUOM	WARNING	ERROR	RESULTSTATUS	STOPPED
321	34	144319	290505	220045	2191-02-23 05:25:00	2191-02-23 10:53:00	17741	72	72	bpm	0	0	NULL	NULL
322	34	144319	290505	220074	2191-02-23 05:25:00	2191-02-23 07:45:00	17741	7	7	mmHg	0	0	NULL	NULL
323	34	144319	290505	220210	2191-02-23 05:25:00	2191-02-23 07:45:00	17741	14	14	insp/min	0	0	NULL	NULL
324	34	144319	290505	220277	2191-02-23 05:25:00	2191-02-23 07:45:00	17741	98	98	%	0	0	NULL	NULL
325	34	144319	290505	220179	2191-02-23 05:30:00	2191-02-23 07:45:00	17741	112	112	mmHg	0	0	NULL	NULL
326	34	144319	290505	220180	2191-02-23 05:30:00	2191-02-23 07:45:00	17741	66	66	mmHg	0	0	NULL	NULL
327	34	144319	290505	220181	2191-02-23 05:30:00	2191-02-23 07:45:00	17741	78	78	mmHg	0	0	NULL	NULL
328	34	144319	290505	223875	2191-02-23 07:00:00	2191-02-23 07:36:00	16924	10	10	insp/min	0	0	NULL	NULL
329	34	144319	290505	224690	2191-02-23 07:00:00	2191-02-23 07:35:00	16924	22	22	insp/min	0	0	NULL	NULL
330	34	144319	290505	224696	2191-02-23 07:00:00	2191-02-23 07:35:00	16924	16	16	cmH2O	0	0	NULL	NULL

Table A.2: MIMIC Chartevents.

A.0.3 Definition for ICD diagnoses

See Table A.3. The definition for ICD diagnoses table houses the ICD9 (International Classification of Disease Version 9) codes, titles, and descriptions of diseases afflicting the patients. There are 14,567 codes, each code relating to a diagnostic concept proposed and maintained by the World Health Organization.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- ICD9_CODE: ICD9 codes are alpha-numeric code relating to a diagnostic concept.
- SHORT_TITLE: The short title is a description of the diagnostic concept limited to 50 characters.
- LONG_TITLE: The long title is a description of the diagnostic concept limited to 300 characters.

Relationships for this table exist as:

- An ICD Diagnosis (see subsection A.0.6) has exactly one definition for the ICD diagnosis.

ROW_ID	ICD9_CODE	SHORT_TITLE	LONG_TITLE
1	1716	Erythem nod tb-oth test	Erythema nodosum with hypersensitivity reaction in tuberculosis, tubercle bacilli not found by bacteriological or histological examination, but tuberculosis confirmed by other methods [inoculation of animals]
2	1720	TB periph lymph-unspec	Tuberculosis of peripheral lymph nodes, unspecified
3	1721	TB periph lymph-no exam	Tuberculosis of peripheral lymph nodes, bacteriological or histological examination not done
4	1722	TB periph lymph-exam unk	Tuberculosis of peripheral lymph nodes, bacteriological or histological examination unknown (at present)
5	1723	TB periph lymph-micro dx	Tuberculosis of peripheral lymph nodes, tubercle bacilli found (in sputum) by microscopy
6	1724	TB periph lymph-cult dx	Tuberculosis of peripheral lymph nodes, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture
7	1725	TB periph lymph-histo dx	Tuberculosis of peripheral lymph nodes, tubercle bacilli not found by bacteriological examination, but tuberculosis confirmed histologically
8	1726	TB periph lymph-oth test	Tuberculosis of peripheral lymph nodes, tubercle bacilli not found by bacteriological or histological examination, but tuberculosis confirmed by other methods [inoculation of animals]
9	1730	TB of eye-unspec	Tuberculosis of eye, unspecified
10	1731	TB of eye-no exam	Tuberculosis of eye, bacteriological or histological examination not done

Table A.3: MIMIC definition for ICD Diagnoses.

A.0.4 Definition for items

See Table A.4. There are 12,487 items from CareVue and Metavision datasets that are defined in this table.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- ITEMID: is a unique identifier for an item, sourced from multiple tables.
- LABEL: describes and defines the item.
- LINKSTO: identifies the database table for which the item is defined.

Relationships for this table exist as:

- One Output event (see subsection A.0.12) has exactly one definition for the item.
- One Chart event (see subsection A.0.2) has exactly one definition for the item.
- One Input event (see subsection A.0.8 and subsection A.0.9) has exactly one definition for the item.
- One procedure event (see subsection A.0.14) has exactly one definition for the item.

ROW_ID	ITEMID	LABEL	LINKSTO
1	1435	Sustained Nystamus	chartevents
2	1436	Tactile Disturbances	chartevents
3	1437	Tremor	chartevents
4	1438	Ulnar Pulse [Right]	chartevents
5	1439	Visual Disturbances	chartevents
6	1447	Transpulmonary Pres	chartevents
7	1448	Vd/Vt:	chartevents
8	1449	Arterial BP(Rad)	chartevents
9	1450	level one	chartevents
10	1451	L girth size	chartevents

Table A.4: MIMIC definition for Items.

A.0.5 Definition for laboratory items

Table A.5 shows part of the definitions for laboratory items. There are 753 different laboratory measurements defined in this table.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- ITEMID: is a unique identifier for the laboratory item.
- LABEL: describes and defines the laboratory item.
- FLUID: identifies the fluid source of the specimen being measured.
- CATEGORY: specifies a type of measurement at a higher level than the label does.

- LOINC_CODE: Logical Observation Identifiers Names and Codes is an ontology for organizing medical laboratory measurements.

Relationships for this table exist as:

- One laboratory event (see subsection A.0.10) has exactly one definition for the laboratory item.

- One microbiology laboratory event (see subsection A.0.11) has exactly one definition for the laboratory item.

ROW_ID	ITEMID	LABEL	FLUID	CATEGORY	LOINC_CODE
1	50800	SPECIMEN TYPE	BLOOD	BLOOD GAS	NULL
2	50801	Alveolar-arterial Gradient	Blood	Blood Gas	19991-9
3	50802	Base Excess	Blood	Blood Gas	11555-0
4	50803	Calculated Bicarbonate, Whole Blood	Blood	Blood Gas	1959-6
5	50804	Calculated Total CO2	Blood	Blood Gas	34728-6
6	50805	Carboxyhemoglobin	Blood	Blood Gas	20563-3
7	50806	Chloride, Whole Blood	Blood	Blood Gas	2069-3
8	50807	Comments	Blood	Blood Gas	NULL
9	50808	Free Calcium	Blood	Blood Gas	1994-3
10	50809	Glucose	Blood	Blood Gas	2339-0

Table A.5: MIMIC definition for Laboratory Items.

A.0.6 ICD diagnoses

ICD diagnoses are assigned to every hospital admission (Table A.6). They are coded, in practice, for billing purposes by coders who review patient charts after the hospital stay.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.

- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- SEQ_NUM: is a number indicating the priority of the diagnoses in relation to other diagnoses for the hospital admission.
- ICD9_CODE: ICD9 codes are alpha-numeric code relating to a diagnostic concept and links to a definition table (see subsection A.0.3).

Relationships for this table exist as:

- One hospital admission (see subsection A.0.1) may have multiple ICD diagnoses.

ROW_ID	SUBJECT_ID	HADM_ID	SEQ_NUM	ICD9_CODE
4	3	145834	1	389
5	3	145834	2	78559
6	3	145834	3	5849
7	3	145834	4	4275
8	3	145834	5	41071
9	3	145834	6	4280
10	3	145834	7	6826
11	3	145834	8	4254
12	3	145834	9	2639
54	12	112213	1	1570

Table A.6: MIMIC ICD Diagnoses.

A.0.7 ICU stays

MIMIC-III contains 61,532 distinct Intensive Care Unit stays (see Table A.7). Each stay is captured within this table. This table constitutes the center of the database, both in the concept of MIMIC and as the table with the most relationships to other tables within the database.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- ICUSTAY_ID: An identifier to an ICU stay .
- DBSOURCE: this value is either CareVue or Metavision, depending on the vendor for the ICU patient database at the time.
- FIRST_CAREUNIT: The first ICU care unit the patient was admitted to during the ICU stay.
- LAST_CAREUNIT: The last ICU care unit the patient was admitted to during the ICU stay.
- FIRST_WARDID: The first ward where the patient was placed during the ICU stay.
- LAST_WARDID: The last ward where the patient was placed during the ICU stay.
- INTIME: is a timestamp for the beginning of the ICU stay.

- OUTTIME: is a timestamp for the end of the ICU stay.
- LOS: is a Length Of Stay calculation in days between INTIME and OUTTIME.

Relationships for this table exist as:

- One ICU stays may have many Output event (see subsection A.0.12).
- One ICU stays may have many Chart event (see subsection A.0.2).
- One ICU stay may have many Input event (see subsection A.0.8 and subsection A.0.9).
- One ICU stay may have many procedure event (see subsection A.0.14).
- One hospital admission (see subsection A.0.1) may have multiple ICU stays.

ROW_ID	SUBJECT_ID	HADM_ID	ICUSTAY_ID	DBSOURCE	FIRST_CAREUNIT	LAST_CAREUNIT	FIRST_WARDID	LAST_WARDID	INTIME	OUTTIME	LOS
2	3	145834	211552	carevue	MICU	MICU	12	12	2101-10-20 19:10:11	2101-10-26 20:43:09	6.0646
12	12	112213	232669	carevue	SICU	SICU	23	23	2104-08-08 02:08:17	2104-08-15 17:22:25	7.6348
18	19	109235	273430	carevue	TSICU	TSICU	23	23	2108-08-05 16:26:09	2108-08-06 23:40:35	1.3017
30	30	104557	225176	carevue	CCU	CCU	57	57	2172-10-14 17:24:00	2172-10-16 13:44:00	1.8472
31	31	128652	254478	carevue	MICU	MICU	15	15	2108-08-22 23:28:42	2108-08-30 21:59:20	7.9379
33	33	176176	296681	carevue	MICU	MICU	12	12	2116-12-23 22:31:53	2116-12-25 11:49:55	1.5542
34	34	115799	263086	carevue	MICU	MICU	23	23	2186-07-18 18:10:49	2186-07-19 11:27:20	0.7198
35	34	144319	290505	metavision	CCU	CCU	7	7	2191-02-23 05:25:32	2191-02-24 19:24:10	1.5824
57	56	181711	275642	carevue	SICU	SICU	57	57	2104-01-02 02:02:39	2104-01-03 22:25:29	1.8492
103	97	127870	288376	carevue	CSRU	CSRU	14	14	2105-04-30 08:37:39	2105-05-04 20:05:55	4.478

Table A.7: MIMIC ICU Stays.

A.0.8 Input events CareVue version

See Table A.8. Input events tables house data on all items ingested, injected, or otherwise put into the patient during their ICU stay.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- ICUSTAY_ID: An identifier to an ICU stay.
- CHARTTIME: is a record of the timestamp the item was administered to the patient.
- ITEMID: is the identifier of the item administered to the patient (see subsection A.0.4).
- AMOUNT: is a numerical amount (if applicable) of the item administered to the patient.
- AMOUNTUOM: is the Unit Of Measurement of the AMOUNT (if applicable).

Relationships for this table exist as:

- One ICU stay (see subsection A.0.7) may have multiple input events.
- One input event has exactly one definition for the item administered (see subsection A.0.4).

ROW_ID	SUBJECT_ID	HADM_ID	ICUSTAY_ID	CHARTTIME	ITEMID	AMOUNT	AMOUNTUOM
221	19872	134153	249265	2106-08-18 21:00:00	30056	60	ml
222	19872	134153	249265	2106-08-19 00:00:00	30056	60	ml
223	19872	134153	249265	2106-08-19 06:00:00	30056	60	ml
224	19872	134153	249265	2106-08-19 10:00:00	30056	120	ml
225	19872	134153	249265	2106-08-19 14:00:00	30056	120	ml
226	19872	134153	249265	2106-08-19 18:00:00	30056	240	ml
227	19872	134153	249265	2106-08-20 00:00:00	30056	60	ml
228	19872	134153	249265	2106-08-20 10:00:00	30056	60	ml
229	19872	134153	249265	2106-08-20 14:00:00	30056	20	ml
230	19872	134153	249265	2106-08-22 09:00:00	30056	100	ml

Table A.8: MIMIC Input Events from CareVue.

A.0.9 Input events Metavision version

Input events tables house data on all items ingested, injected, or otherwise put into the patient during their ICU stay. See Table A.9.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- ICUSTAY_ID: An identifier to an ICU stay.
- STARTTIME: is a record of the timestamp the item was administered to the patient.
- ENDTIME: is a record of the timestamp the item's administration ceased.
- ITEMID: is the identifier of the item administered to the patient (see subsection A.0.4).

- AMOUNT: is a numerical amount (if applicable) of the item administered to the patient.

- AMOUNTUOM: is the Unit Of Measurement of the AMOUNT (if applicable).

Relationships for this table exist as:

- One ICU stay (see subsection A.0.7) may have multiple input events.

- One input event has exactly one definition for the item administered (see subsection A.0.4).

ROW_ID	SUBJECT_ID	HADM_ID	ICUSTAY_ID	STARTTIME	ENDTIME	ITEMID	AMOUNT	AMOUNTUOM
318	23792	150835	227602	2106-12-14 13:00:00	2106-12-14 13:01:00	226452	240	ml
319	23792	150835	227602	2106-12-14 14:16:00	2106-12-14 14:17:00	226452	60	ml
320	23792	150835	227602	2106-12-14 17:37:00	2106-12-14 17:38:00	226452	180	ml
321	23792	150835	227602	2106-12-14 11:00:00	2106-12-14 11:01:00	226452	120	ml
322	23792	150835	227602	2106-12-14 13:00:00	2106-12-14 13:01:00	226452	240	ml
355	18082	181163	267692	2156-02-25 08:23:00	2156-02-25 08:24:00	225799	120	ml
356	18082	181163	267692	2156-02-25 03:02:00	2156-02-25 03:03:00	221744	10	mcg
357	18082	181163	267692	2156-02-29 18:00:00	2156-02-29 18:01:00	225851	1	dose
358	18082	181163	267692	2156-02-29 18:00:00	2156-02-29 18:01:00	220949	50	ml
359	18082	181163	267692	2156-03-04 09:44:00	2156-03-04 15:17:00	226048	250	ml

Table A.9: MIMIC Input Events from Metavision.

A.0.10 Laboratory events

See Table A.10. The laboratory events table contains all laboratory derived measurements excluding microbiology measurements which has a dedicated table (see subsection A.0.11).

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.

- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).

- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- ITEMID: laboratory item that links to the definition table for laboratory items (see subsection A.0.5).
- CHARTTIME: is a record of the timestamp the measurement was observed.
- VALUE: This field contains the text representation of the value of the measurement.
- VALUENUM: If VALUE is numeric, this field represents that numeric value, otherwise it is left null.
- VALUEUOM: If valid, this field contains the Unit Of Measurement for VALUE.
- FLAG: If the measurement is abnormal (outside of established normal boundaries), then the field is given a value of "abnormal", null otherwise.

Relationships for this table exist as:

- One hospital admission (see subsection A.0.1) may have multiple laboratory events.
- One laboratory event has exactly one definition for the laboratory item (see subsection A.0.5).

ROW_ID	SUBJECT_ID	HADM_ID	ITEMID	CHARTTIME	VALUE	VALUENUM	VALUEUOM	FLAG
441	3	145834	50868	2101-10-20 16:40:00	17	17	mEq/L	NULL
442	3	145834	50882	2101-10-20 16:40:00	25	25	mEq/L	NULL
443	3	145834	50893	2101-10-20 16:40:00	8.2	8.2	mg/dL	abnormal
444	3	145834	50902	2101-10-20 16:40:00	99	99	mEq/L	abnormal
445	3	145834	50910	2101-10-20 16:40:00	48	48	IU/L	NULL
446	3	145834	50911	2101-10-20 16:40:00	NotDone	NULL	ng/mL	NULL
447	3	145834	50912	2101-10-20 16:40:00	3.2	3.2	mg/dL	abnormal
448	3	145834	50931	2101-10-20 16:40:00	91	91	mg/dL	NULL
449	3	145834	50960	2101-10-20 16:40:00	2.4	2.4	mg/dL	NULL
450	3	145834	50970	2101-10-20 16:40:00	4.8	4.8	mg/dL	abnormal

Table A.10: MIMIC Laboratory Events.

A.0.11 Microbiology events

See Table A.11. The microbiology laboratory events table contains all laboratory derived measurements relating to microbiological organisms.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- CHARTDATE: is the record of the date the measurement was observed.
- CHARTTIME: is a record of the timestamp the measurement was observed.
- SPEC_ITEMID: specimen item identifier that links to laboratory item definition table A.0.5).

- SPEC_TYPE_DESC: A description of the specimen collected and inspected.
- ORG_ITEMID: organism item identifier that links to laboratory item definition table A.0.5).
- ORG_NAME: The name of the organism identified.

Relationships for this table exist as:

- One hospital admission (see subsection A.0.1) may have multiple microbiological laboratory events.
- One microbiological laboratory event has exactly one definition for the microbiological laboratory organism item (see subsection A.0.5).

ROW_ID	SUBJECT_ID	HADM_ID	CHARTDATE	CHARTTIME	SPEC_ITEMID	SPEC_TYPE_DESC	ORG_ITEMID	ORG_NAME
2	3	145834	2101-10-20 00:00:00	2101-10-20 13:00:00	70070	SWAB	80223	PROBABLE ENTEROCOCCUS
3	3	145834	2101-10-20 00:00:00	2101-10-20 13:00:00	70070	SWAB	80075	YEAST
4	3	145834	2101-10-20 00:00:00	2101-10-20 13:00:00	70070	SWAB	80155	STAPHYLOCOCCUS, COAGULASE NEGATIVE
5	3	145834	2101-10-20 00:00:00	2101-10-20 13:00:00	70070	SWAB	80058	GRAM NEGATIVE ROD(S)
6	3	145834	2101-10-20 00:00:00	2101-10-20 18:00:00	70012	BLOOD CULTURE	NULL	NULL
7	3	145834	2101-10-20 00:00:00	2101-10-20 18:30:00	70012	BLOOD CULTURE	NULL	NULL
8	3	145834	2101-10-20 00:00:00	NULL	70079	URINE	80075	YEAST
9	3	145834	2101-10-21 00:00:00	2101-10-21 04:00:00	70012	BLOOD CULTURE	NULL	NULL
10	3	145834	2101-10-21 00:00:00	2101-10-21 04:00:00	70012	BLOOD CULTURE	NULL	NULL
11	3	145834	2101-10-21 00:00:00	2101-10-21 04:15:00	70012	BLOOD CULTURE	NULL	NULL

Table A.11: MIMIC Microbiology Events.

A.0.12 Output events

Output events table captures data relating to items that exit the patient: eg. urine flush, hemodialysis, thoracentesis. See Table A.12.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).
- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).
- ICUSTAY_ID: An identifier to a unique ICU stay (see subsection A.0.7).
- CHARTTIME: is a record of the timestamp the output was observed.
- ITEMID: output item that links to the definition table for items (see subsection A.0.4).
- VALUE: This field contains the numerical value of the measurement.
- VALUEUOM: this field contains the Unit Of Measurement for VALUE.

Relationships for this table exist as:

- One ICU stay (see subsection A.0.7) may have multiple output events.
- One output event has exactly one definition for the item outputted (see subsection A.0.4).

ROW_ID	SUBJECT_ID	HADM_ID	ICUSTAY_ID	CHARTTIME	ITEMID	VALUE	VALUEUOM
62	19872	134153	249265	2106-08-18 02:00:00	40055	80	ml
63	19872	134153	249265	2106-08-18 04:00:00	40055	100	ml
64	19872	134153	249265	2106-08-18 06:00:00	40055	100	ml
65	19872	134153	249265	2106-08-18 08:00:00	40055	40	ml
66	19872	134153	249265	2106-08-18 10:00:00	40055	200	ml
67	19872	134153	249265	2106-08-18 11:00:00	40055	200	ml
68	19872	134153	249265	2106-08-18 12:00:00	40055	200	ml
69	19872	134153	249265	2106-08-18 13:00:00	40055	180	ml
70	19872	134153	249265	2106-08-18 14:00:00	40055	140	ml
71	19872	134153	249265	2106-08-18 15:00:00	40055	140	ml

Table A.12: MIMIC Output Events.

A.0.13 Patients

The patients table (Table A.13) contains the records of 46,520 patients associated with the ICU. This table is main table for SUBJECT_ID throughout the rest of the database. There is exactly one unique SUBJECT_ID per entry of this table.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.
- SUBJECT_ID: A unique person identifier.
- GENDER: a binary M/F gender determination.
- DOB: Date of Birth, and randomly shifted DOB to comply with HIPAA and maintain age at hospital admission. If the patient is 90 years or older, the DOB was shifted 300 years prior to hospital admissions.
- DOD: Is the known date of death, shifted to maintain consistency with the DOB.
- DOD_HOSP: Date of Death recorded in-hospital.

- DOD_SSN: Date of Death recorded by the social security database.

- EXPIRE_FLAG: 1 if expired, 0 otherwise.

Relationships for this table exist as:

- One patient may have multiple hospital admissions (see subsection A.0.1).

ROW_ID	SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP	DOD_SSN	EXPIRE_FLAG
2	3	M	2025-04-11 00:00:00	2102-06-14 00:00:00	NULL	2102-06-14 00:00:00	1
11	12	M	2032-03-24 00:00:00	2104-08-20 00:00:00	2104-08-20 00:00:00	2104-08-20 00:00:00	1
16	19	M	1808-08-05 00:00:00	2109-08-18 00:00:00	NULL	2109-08-18 00:00:00	1
26	30	M	1872-10-14 00:00:00	NULL	NULL	NULL	0
27	31	M	2036-05-17 00:00:00	2108-08-30 00:00:00	2108-08-30 00:00:00	2108-08-30 00:00:00	1
29	33	M	2034-08-02 00:00:00	NULL	NULL	NULL	0
30	34	M	1886-07-18 00:00:00	2192-01-30 00:00:00	NULL	2192-01-30 00:00:00	1
49	56	F	1804-01-02 00:00:00	2104-01-08 00:00:00	2104-01-08 00:00:00	2104-01-08 00:00:00	1
88	97	M	2031-12-08 00:00:00	NULL	NULL	NULL	0
99	108	M	2037-12-02 00:00:00	NULL	NULL	NULL	0

Table A.13: MIMIC Patients.

A.0.14 Procedure events Metavision version

Table A.14 captures data relevant to the performance of procedures in the ICU.

The table contains the following information relevant to the study:

- ROW_ID: a unique identifier per entry in the table.

- SUBJECT_ID: A person identifier that links to patient information via the Patient table (see subsection A.0.13).

- HADM_ID: An identifier to a hospital admissions that links to the Admissions table (see subsection A.0.1).

- ICUSTAY_ID: An identifier to an ICU stay.

- STARTTIME: is a record of the timestamp the procedure began.
- ENDTIME: is a record of the timestamp the procedure ceased.
- ITEMID: is the identifier of the procedure (see subsection A.0.4).
- VALUE: is a numerical value (if applicable) of the procedure.
- VALUEUOM: is the Unit Of Measurement of the VALUE (if applicable).

Relationships for this table exist as:

- One ICU stay (see subsection A.0.7) may have multiple procedure events.
- One procedure event has exactly one definition for the procedure (see subsection A.0.4).

ROW_ID	SUBJECT_ID	HADM_ID	ICUSTAY_ID	STARTTIME	ENDTIME	ITEMID	VALUE	VALUEUOM
44	23792	150835	227602	2106-12-14 12:00:00	2106-12-14 18:33:00	224275	393	min
45	23792	150835	227602	2106-12-14 12:00:00	2106-12-14 18:33:00	224275	393	min
55	18082	181163	267692	2156-02-24 11:30:00	2156-02-25 02:54:00	224277	924	min
56	18082	181163	267692	2156-02-24 11:35:00	2156-02-24 11:36:00	224385	1	None
57	18082	181163	267692	2156-02-24 11:38:00	2156-02-27 14:00:00	225792	4462	min
58	18082	181163	267692	2156-02-24 11:45:00	2156-02-24 11:46:00	224385	1	None
59	18082	181163	267692	2156-02-24 12:00:00	2156-02-25 15:14:00	224263	1634	min
60	18082	181163	267692	2156-02-24 12:10:00	2156-02-24 12:11:00	225459	1	None
61	18082	181163	267692	2156-02-24 16:32:00	2156-02-24 16:33:00	225459	1	None
62	18082	181163	267692	2156-02-24 20:17:00	2156-02-24 20:18:00	225402	1	None

Table A.14: MIMIC Procedure Events from Metavision.

Appendix B

KNOWLEDGE BASE

B.0.1 Aggregate

The aggregate table (see Table B.1) describes the aggregate functions of SQL used by the intelligent agents for this study.

The table contains the following information:

- aggregate_id: a unique iterated identifier to reference each row.
- aggregate_name: the text used to call the SQL function.
- inputs: the number of inputs to be called for this function.

aggregate_id	aggregate_name	inputs
1	AVG	1
2	MIN	1
3	MAX	1

Table B.1: Knowledge Base Aggregate.

B.0.2 Attribute Reduction

The attribute reduction table (Table B.2) was designed to consolidate duplicative attributes in the MIMIC database under a single attribute identifier. The problem of duplicative attributes manifests in two separate methods. Over the 12 years of collection of ICU data, the

hospital changed vendors for the ICU data recording. This in turn indicates a duplication of many items. Additionally, there are several different means of recording patient information electronically. Some are duplicated per medical providers preference, while some are duplicated systemically as procedural data capture. As described above, this leads to a 6 different non-overlapping identifiers for white blood cell count. In total 103 separate items were consolidated under 34 non-duplicative attributes.

The table contains the following information:

- att_id: is a knowledge base specific attribute identifier.
- dup_att_id: is a duplicative attribute identifier reduced under the single attribute identifier att_id.

Relationships for this table exist as:

- One attribute reduced identifier necessarily has multiple main attribute identifiers (subsection B.0.4) keyed by dup_att_id in attribute reduction is equalled to an att_id in main.

att_id	dup_att_id
10000	1639
10000	13
10000	1566
10001	1602
10001	12
10001	1310
10001	1283
10002	1650
10002	1254
10002	1491

Table B.2: Knowledge Base Attribute Reduction.

B.0.3 Logic Operators

See Table B.3. When an intelligent agent is constructing SQL queries or complex concepts, it requires a reference to the logical operators. When constructing a complex concept from simple concepts A and B, the intelligent agents need options on how A and B can be assembled. A AND B produces a different model than A OR B. The intelligent agent then evaluates the outcome of the different assembled models and uses this table to track what logical operator was used to assemble the superior model.

The table contains the following information:

- logic_id: a unique iterated identifier to reference each row.
- logic_name: the text used to call the SQL function or otherwise identify the function to be called.
- inputs: the number of inputs to be called for this function.

logic_id	logic_name	inputs
1	OR	2
2	AND	2

Table B.3: Knowledge Base Logic Operators.

B.0.4 Main

The main table (Table B.4) describes MIMIC items used in this study. It is effectively a map that consolidates the various MIMIC attributes relevant to this study into a single table. There are 575 different attributes populated here that are present in a minimum of 5% of patients. There are many other attributes in MIMIC, but to par down the computational load the decision was made to limit attributes that appear in a minimum of 5% of patients. A lower percentage would dramatically increase the feature space for attributes that would not prove statistically significant.

The table contains the following information:

- label: a short name of the item.
- itemid: the MIMIC ITEMID referenced in output events (subsection A.0.12), chart events (subsection A.0.2), input events (subsections A.0.8 and A.0.9), procedure events (subsection A.0.14), laboratory events (subsection A.0.10), and microbiological laboratory events (subsection A.0.11).
- category: references the MIMIC table where the itemid is directly referenced.
- att_id: a unique iterated identifier to reference each row.

Relationships for this table exist as:

- One main attribute may exist in the attribute reduce table (subsection B.0.2).

- One main attribute has many statistical distribution records in the statistics table (subsection B.0.7).

label	itemid	category	att_id
Arterial BP Mean	52	CHARTEVENTS	1
FiO2 Set	190	CHARTEVENTS	2
Heart Rate	211	CHARTEVENTS	3
Mean Airway Pressure	444	CHARTEVENTS	4
Respiratory Rate	618	CHARTEVENTS	5
SpO2	646	CHARTEVENTS	6
Temperature F	678	CHARTEVENTS	7
Arterial PaCO2	778	CHARTEVENTS	8
Arterial PaO2	779	CHARTEVENTS	9
Arterial pH	780	CHARTEVENTS	10

Table B.4: Knowledge Base Main.

B.0.5 Metric

The metrics table (Table B.5) defines and indexes some basic metrics intelligent agents use to evaluate a model's performance.

The table contains the following information:

- metric_id: a unique iterated identifier to reference each row.

- metric: the name of the metric.

metric_id	metric
1	TP
2	TN
3	FP
4	FN

Table B.5: Knowledge Base Metrics.

B.0.6 Operator

The operator table (Table B.6) defines and indexes basic operators used in setting parameters in SQL queries and value comparisons.

The table contains the following information:

- operator_id: a unique iterated identifier to reference each row.
- operator_name: the name of the operator necessary to call in a function.
- inputs: the number of inputs the operator takes to evaluate against.

operator_id	operator_name	inputs
1	$>$	1
2	$<$	1
3	between	2

Table B.6: Knowledge Base Operators.

B.0.7 Statistics

The statistics table (Table B.7) contains previously computed statistical distributions of the value of items in the MIMIC database. Previously computed statistical distributions are

important for intelligent agents that will be selecting and adjusting parameters of models based on these items. For example, if an intelligent agent were inspecting heart rate for a possible addition to a selection model, it is valuable to know the median heart rate recorded is 82 beats per minute, and most of the heart rates sampled (60% of them) are between 70 beats per minute and 98 beats per minute. The intelligent agent would otherwise have to compute the statistical distribution upon selecting an attribute to include, or select a random number and use a computationally costly gradient descent algorithm to converge on a reasonable selection criteria.

The table contains the following information:

- att_id: attribute identifier found in the main table (subsection B.0.4).
- statistical_id: a statistical identifier found in the statistics identifier table (subsection B.0.8).
- statistical_value: the value of the statistical function for the attribute.

Relationships for this table exist as:

- One main attribute (subsection B.0.4) has many statistical values.
- One statistical value has exactly one statistical function (subsection B.0.8)

att_id	statistical_id	statistical_value
1	1	1
1	2	300
1	3	60
1	4	66
1	5	69
1	6	73
1	7	77
1	8	81
1	9	86
1	10	92

Table B.7: Knowledge Base Statistics.

B.0.8 Statistics Identifiers

The statistical identifiers table (Table B.8) defines and indexes statistical metrics used in describing data, and is most notably used in describing statistical distributions such as in the statistics table (subsection B.0.7) of this knowledge base.

The table contains the following information:

- stat_id: a unique iterated identifier to reference each row.
- stat_name: the name of the statistical metric used in calling the function.
- code: a placeholder for specialized code needed to call the function.

Relationships for this table exist as:

- One statistical value from the statistics table (subsection B.0.7) has exactly one statistic identifier.

stat_id	stat_name	code
1	min	NULL
2	max	NULL
3	q10	NULL
4	q20	NULL
5	q30	NULL
6	q40	NULL
7	q50	NULL
8	q60	NULL
9	q70	NULL
10	q80	NULL

Table B.8: Knowledge Base Statistics Identifiers.

B.0.9 MLA Classname

The MLA Classname table (Table B.9) identifies and facilitates the utilization of 38 machine learning algorithms used by the intelligent agents.

The table contains the following information:

- cn_id: a unique iterated identifier to reference each row.
- classname: the classname which describes and calls the machine learning algorithm function.

Relationships for this table exist as:

- A MLA classname may have multiple options found in the classname options table (subsection B.0.10).

cn_id	classname
1	weka.classifiers.bayes.BayesianLogisticRegression
2	weka.classifiers.bayes.BayesNet
3	weka.classifiers.bayes.ComplementNaiveBayes
4	weka.classifiers.bayes.DMNBtext
5	weka.classifiers.bayes.NaiveBayes
6	weka.classifiers.bayes.NaiveBayesSimple
7	weka.classifiers.functions.Logistic
8	weka.classifiers.functions.MultilayerPerceptron
9	weka.classifiers.functions.RBFNetwork
10	weka.classifiers.functions.SimpleLogistic

Table B.9: Machine Learning Algorithm Classname.

B.0.10 MLA Class Options

Each algorithm (subsection B.0.9) has a set of definable options that direct the performance of the algorithm and are amassed in the MLA Class Option table (Table B.10).

The table contains the following information:

- cn_id: a unique iterated identifier to reference a MLA classname (subsection B.0.9).
- opt_id: a unique iterated identifier to reference each row.

Relationships for this table exist as:

- A MLA classname (subsection B.0.9) may have multiple class options.
- A class option has exactly one type found in the MLA options table (subsection B.0.11).

cn_id	opt_id
1	1
1	2
2	3
2	4
8	5
8	6
8	7
11	8
11	9
12	10

Table B.10: Machine Learning Algorithm Class Options.

B.0.11 MLA Options

MLA options table (Table B.11) distinguishes the type of option. Types are either a range, as defined in the mla_opt_range table (subsection B.0.13) or enumerated text in the mla_opt_enum table (subsection B.0.12).

The table contains the following information:

- opt_id: a unique iterated identifier to reference MLA class options (subsection B.0.10).
- opttype: type which assigns the option to either a range or an enumeration text.
- opttext: the text required to designate the option when executing the algorithm.

Relationships for this table exist as:

- A MLA class option (subsection B.0.10) has exactly one option identifier.

- An option may have exactly one range entry (subsection B.0.13).
- An option may have multiple enumerated text options (subsection B.0.12).

opt_id	opttype	opttext
1	enum	-H
2	enum	-P
3	enum	-E
4	enum	-Q
5	range	-H
6	range	-L
7	range	-M
8	range	-C
9	enum	-K
10	enum	-F

Table B.11: Machine Learning Algorithm Options.

B.0.12 MLA Options Enumeration

The enumeration table (Table B.12) contains the text necessary to call an algorithm with text based options. Commonly these are other algorithms being called for the purpose of filters.

The table contains the following information:

- opt_id: a unique iterated identifier to reference MLA options (subsection B.0.11).
- enum_id: a unique iterated identifier that represents an enumerated text option.

- enum_pos: a unique iterated identifier that represents the position the enum_text will be placed in the text string for this called option.

- enum_text: the text string that constitutes the parameters of the option.

Relationships for this table exist as:

- A MLA option (subsection B.0.11) may have multiple enumerated text options.

opt_id	enum_id	enum_pos	enum_text
1	1	1	1
1	2	1	2
1	3	1	3
2	4	1	1
2	5	1	2
3	6	1	weka.classifiers.bayes.net.estimate.BayesNetEstimator
3	6	2	–
3	6	3	-A
3	6	4	0.5
3	7	1	weka.classifiers.bayes.net.estimate.BMAEstimator

Table B.12: Machine Learning Algorithm Options Enumeration.

B.0.13 MLA Options Range

Many options for machine learning algorithms come in the form of ranges. Those ranges are defined in the MLA Options Range table (Table B.13).

The table contains the following information:

- opt_id: a unique iterated identifier to reference MLA options (subsection B.0.11).

- int1float0: a flagged field, 1 if the value is an integer, 0 if the value is a float data type.
- LB: lower bound of the range.
- UP: upper bound of the range.

Relationships for this table exist as:

- A MLA option (subsection B.0.11) may have at most one ranged parameter.

opt_id	int1float0	LB	UP
5	1	1	10
6	0	0	1
7	0	0	1
8	0	0	5
11	1	1	5

Table B.13: Machine Learning Algorithm Options Range.

Appendix C

SUPPORT DATABASE

C.0.1 Concept Database

The concept database table (Table C.1) stores previously assembled patients, attributes, values of attributes, and timestamps for every concept. This table exists to reduce computation time by intelligent agents which would otherwise be required to do on-the-fly assembly of the data contained herein.

The table contains the following information:

- concept_id: a concept identifier relating to the concept table (subsection C.0.9).
- patient_id: a patient identifier relating to the disease definition table (subsection C.0.15).
- att_id: an attribute identifier relating to the attributes table (subsection B.0.4).
- val: the value of the attribute.
- pt_timestamp: the time that the attribute has the stated value.

Relationships for this table exist as:

- One hypothesis (subsection C.0.11) has many cdb entries.

concept_id	patient_id	att_id	val	pt_timestamp
3	211552	NULL	NULL	2101-10-20 18:45:00
3	211552	NULL	NULL	2101-10-20 19:30:00
3	211552	NULL	NULL	2101-10-20 19:45:00
3	211552	NULL	NULL	2101-10-20 20:00:00
3	211552	NULL	NULL	2101-10-20 20:15:00
3	211552	NULL	NULL	2101-10-20 20:30:00
3	211552	NULL	NULL	2101-10-20 20:45:00
3	211552	NULL	NULL	2101-10-20 21:00:00
3	211552	NULL	NULL	2101-10-20 21:15:00
3	211552	NULL	NULL	2101-10-20 21:30:00

Table C.1: Concept Database.

C.0.2 Feature Database

See Table C.2. The feature space per disease is calculated from MIMIC and populated in the Feature Database table (Table C.2). For every disease inspected it contains an attribute value and the timestamp for that value for all attributes in the main table of the knowledge base (see subsection B.0.4). The table also contains a t-minus calculation which indicates the time before the onset of disease as defined in the sdb_target_class table (subsection C.0.15).

The table contains the following information:

- target_id: a disease identifier for the disease definition table (subsection C.0.15).
- patient_id: a patient identifier for the disease definition table (subsection C.0.15).
- att_id: an attribute identifier relating to the attributes table (subsection B.0.4).
- val: the value of the attribute.

- pt_timestamp: the time that the attribute has the stated value.

- t_minus: the difference between pt_timestamp and the onset of disease as defined in the disease definition table (subsection C.0.15).

target_id	patient_id	att_id	val	pt_timestamp	t_minus
1	211552	1	76	2101-10-20 21:45:00	0
1	211552	1	71	2101-10-20 21:30:00	15
1	211552	1	66	2101-10-20 21:15:00	30
1	211552	1	55	2101-10-20 21:00:00	45
1	211552	1	67	2101-10-20 20:45:00	60
1	211552	1	61	2101-10-20 20:30:00	75
1	211552	1	58	2101-10-20 20:15:00	90
1	211552	1	60	2101-10-20 20:00:00	105
1	211552	1	259	2101-10-20 19:15:00	120
1	211552	1	259	2101-10-20 19:15:00	135

Table C.2: Feature Database.

C.0.3 Feature Selection

See Table C.3. Using this feature space table (subsection C.0.2) in conjunction with the knowledge base main table the intelligent agents create a feature selection instance in the Feature Selection table (Table C.3). This table maintains provenance of feature selection instances via a parent-child framework.

The table contains the following information:

- fs_id: a unique feature selection identifier.

- inst_id: an instance identifier.

- parent_fs_id: the feature selection identifier of the parent (if one exists).
- rfroc: a random forest receiver operator characteristic area under curve metric for the model derived from this feature selection instance.
- target_id: a disease identifier for the disease definition table (subsection C.0.15).

Relationships for this table exist as:

- A feature selection - attribute entry (subsection C.0.4) entry has exactly one relationship to a feature selection entry.
- A feature selection - patient entry (subsection C.0.5) entry has exactly one relationship to a feature selection entry.
- A feature selection statistics entry has exactly (subsection C.0.6) one relationship to a feature selection entry.
- A feature selection statistics threshold entry has exactly (subsection C.0.7) one relationship to a feature selection entry.

fs_id	inst_id	parent_fs_id	rfroc	target_id
1	0	0	NULL	1
2	0	0	0.473783	1
3	0	0	NULL	1
4	0	0	NULL	1
5	0	0	NULL	1
6	0	0	0.509334	1
7	0	0	NULL	1
8	0	0	0.558632	1
9	0	0	0.473289	1
10	0	0	0.558888	1

Table C.3: Feature Selection.

C.0.4 Feature Selection - Attribute

An instance is associated with a set of attributes found in the Feature Selection Attribute Table (Table C.4) and constitutes an iteration the intelligent agents are inspecting (subsection C.0.3).

The table contains the following information:

- fs_id: a feature selection identifier defined in the feature selection table (subsection C.0.3).
- att_id: an attribute identifier relating to the attributes table (subsection B.0.4).
- ord: the position of the attribute in an ordered list of all attributes under the feature selection identifier which describes minimum redundancy and maximum relevance.

- attimp: the numeric attribute importance (relative to other attributes under this feature selection identifier) calculated from running a random forest algorithm on the data assembled for this instance, using the disease described by the target_id in the feature selection table (subsection C.0.3).

Relationships for this table exist as:

- A feature selection identifier (subsection C.0.3) has many feature selection - attributes associated with it.

fs_id	att_id	ord	attimp
1	1518	1	NULL
2	1541	1	NULL
3	1356	1	NULL
4	1426	1	NULL
5	1701	1	NULL
6	1550	1	NULL
7	1475	1	NULL
8	10031	1	NULL
9	1608	1	NULL
10	10016	21	NULL

Table C.4: Feature Selection - Attribute.

C.0.5 Feature Selection - Patient

The Feature Selection Patient Table (Table C.5) is designed to declare the patients for whom the feature selection is valid. The feature space - patient table saves computation time when amassing data files from the fdb table data, but also allows for inclusion/exclusion concepts to be applied.

The table contains the following information:

- fs_id: a feature selection identifier defined in the feature selection table (subsection C.0.3).
- patient_id: a patient identifier for the disease definition table (subsection C.0.15).

Relationships for this table exist as:

- A feature selection identifier (subsection C.0.3) may have many feature selection - patients associated with it.

fs_id	patient_id
1	265227
2	206577
3	203674
4	298003
5	250409
6	278820
7	251696
8	260834
9	278569
10	257806

Table C.5: Feature Selection - Patient.

C.0.6 Feature Selection Statistics

The selected attributes (subsection C.0.4) and patients (subsection C.0.5) (with or without inclusion/exclusion concepts (subsection C.0.11)) constitute a model. The statistics and

description of the machine learning algorithm that ran the model to predict the onset of the disease are stored in the Feature Selection Statistics Table (Table C.6).

The table contains the following information:

- inst_id: an instance identifier that references the feature selection table (subsection C.0.3).
- batch_id: a batch identifier.
- mla_id: a machine learning algorithm identifier that references the machine learning algorithms table where the options are defined (subsection C.0.13).
- TP: true positive count.
- TN: true negative count.
- FP: false positive count.
- FN: false negative count.
- ROC: the area under the receiver operating characteristic curve.
- kappa: Cohen's kappa coefficient statistic that measures the agreement between the model and the disease definition (subsection C.0.15).

Relationships for this table exist as:

- A feature selection identifier (subsection C.0.3) has exactly one feature selection - statistics associated with it.

inst_id	batch_id	mla_id	TP	TN	FP	FN	ROC	kappa
1000	0	35	1016	4431	303	446	0.93429	0.653006
1000	15	35	983	4433	301	479	0.929974	0.635524
1000	30	35	954	4437	297	508	0.923558	0.620754
1000	45	35	920	4448	286	542	0.915179	0.605502
1000	60	35	879	4454	280	583	0.910522	0.583905
1000	75	35	879	4444	290	583	0.908687	0.580153
1000	90	35	865	4441	293	597	0.904985	0.570778
1000	105	35	869	4466	268	593	0.90397	0.582531
1000	120	35	849	4467	267	613	0.904369	0.571012
1000	135	35	864	4462	272	598	0.90164	0.578059

Table C.6: Feature Selection Statistics.

C.0.7 Feature Selection Statistics Threshold

The Feature Selection Statistics Threshold Table (Table C.7) contains the detailed data necessary to analyze thresholding of the model. Thresholding is performed on the training set before being applied to the test set. The model of a machine learning algorithm outputs a list of patients in order of how closely they match the model relative to other patients, from greatest to least model match. We iterate down the list of patients, and at every iteration assign all patients above the iterative mark as predicted positive and all patients below as predicted negative. The set of predictive positive patients who are also disease positive as defined in the disease definition table (subsection C.0.15) become true-positive patients. The set of predictive positive who are disease negative become false-positive. The set of predictive negative patients who are disease positive become false-negatives, and those who are disease negative become true-negatives. This table contains the entries to reconstruct an ROC curve for each model. That is, an entry for every non-repetitive true positive or every

non-repetitive false-negative.

The table contains the following information:

- inst_id: an instance identifier that references the feature selection table (subsection C.0.3).
- batch_id: a batch identifier.
- mla_id: a machine learning algorithm identifier that references the machine learning algorithms table where the options are defined (subsection C.0.13).
- TP: true positive count.
- TN: true negative count.
- FP: false positive count.
- FN: false negative count.
- TPR: true positive rate (TP/(TP+FN)).
- FPR: false positive rate (FN/(FN+TP)).
- sensitivity: TPR
- specificity: equivalent to the true negative rate (TN/(TN+FP)).
- PPV: positive predictive value (TP/(TP+FP)).
- NPV: negative predictive value (TN/(TN+FN)).
- accuracy: (TP+TN)/(TP+TN+FP+FN).

- kappa: Cohen's kappa coefficient statistic that measures the agreement between the model and the disease definition (subsection C.0.15).

- fmeasure: the harmonic mean between PPV and sensitivity (2*TP/(2*TP+FP+FN)).

- ord: the order of iteration.

- f1_5: the fmeasure where $\beta = 1.5, (1 + \beta^2 * TP)/(1 + \beta^2 * TP + \beta^2 * FN + FP)$.

- f1_75: the fmeasure where $\beta = 1.75, (1 + \beta^2 * TP)/(1 + \beta^2 * TP + \beta^2 * FN + FP)$.

inst_id	batch_id	mla_id	TP	TN	FP	FN	TPR	FPR	sensitivity	specificity	PPV	NPV	accuracy	kappa	fmeasure	ord	f1_5	f1_75
1001	0	35	164	0	4572	0	1	1	1	0	0.0346284	NULL	0.0346284	0.071741	0.0669388	0	0.104407	0.127189
1001	0	35	161	2124	2448	3	0.981707	0.535433	0.981707	0.464567	0.0617095	0.99859	0.482475	0.661097	0.11612	1	0.175705	0.210225
1001	0	35	155	2920	1652	9	0.945122	0.36133	0.945122	0.63867	0.0857775	0.996927	0.649282	0.795247	0.157281	2	0.231503	0.272681
1001	0	35	152	3339	1233	12	0.926829	0.269685	0.926829	0.730315	0.109747	0.996419	0.73712	0.854969	0.196256	3	0.281642	0.327196
1001	0	35	148	3600	972	16	0.902439	0.212598	0.902439	0.787402	0.132143	0.995575	0.791385	0.888809	0.23053	4	0.323036	0.370627
1001	0	35	144	3752	820	20	0.878049	0.179353	0.878049	0.820647	0.149378	0.994698	0.822635	0.90733	0.255319	5	0.351088	0.398977
1001	0	35	139	3866	706	25	0.847561	0.154418	0.847561	0.845582	0.164497	0.993575	0.84565	0.920539	0.27552	6	0.372117	0.419141
1001	0	35	132	3956	616	32	0.804878	0.134733	0.804878	0.865267	0.176471	0.991976	0.863176	0.930351	0.289474	7	0.384064	0.428914
1001	0	35	125	4026	546	39	0.762195	0.119423	0.762195	0.880577	0.186289	0.990406	0.876478	0.937649	0.299401	8	0.390625	0.432825
1001	0	35	121	4088	484	43	0.737805	0.105862	0.737805	0.894138	0.2	0.989591	0.888725	0.944261	0.314694	9	0.403747	0.443949

Table C.7: Feature Selection Statistics Threshold.

C.0.8 Batch processing

The Batch Processing Table (Table C.8) is a table to help organize multiple iterations of different model runs.

The table contains the following information:

- batch_id: a unique batch identifier.
- measure: the value of measurement.
- metric: the text description of the metric used.
- metric_id: the metric identifier that relates to a metric from the knowledge base metric table (subsection B.0.5).

batch_id	measure	metric	metric_id
1	410	NULL	1
1	1052	NULL	4
1	901	NULL	3
1	4899	NULL	2

Table C.8: Support Database Batch Processing.

C.0.9 Concept

The Root Concepts Table (Table C.9) contains the necessary parameters to write a SQL query and include a sub-population that meets the criteria of the concept.

A few examples:

- Example 1: A concept that requires selected patients to have an average mean arterial pressure greater than 65 mmHg for their entire stay. Att_id = 12 sets the attribute to

mean arterial pressure from the knowledge base main table (subsection B.0.4). Aggregate_binary = 1 indicates the concept uses an aggregate. Agg_id = 5 sets the aggregate function to 'average' from the knowledge base aggregate table (subsection B.0.1). Operator_id = 1 sets the operator to 'greater than' from the knowledge base operator table (subsection B.0.6). Number_of_values = 1 is the appropriate number of values for the operator. Value1 = 65 is the value of mean arterial pressure in this concept. Internal_or_external_binary = 1 indicates it is an internal constraint, internal to this concept and only applied to the attribute described by att_id. Internal_all_or_window_binary = 0 indicates an examination of the whole of the patient's stay.

- Example 2: A concept that requires selected patients to have an average mean arterial pressure greater than 65 mmHg for any two hour time window. The same as Example 1 except internal_all_or_window_binary = 1 indicating the examination of a time window. Internal_window = 120 to examine a two hour window.

- Example 3: A concept that requires selected patients to have an average mean arterial pressure greater than 65 mmHg within 2 hours after an event where the heart rate was recorded above 90 beats per minute. The same parameters apply from Example 1 except the internal_or_external_binary = 0, meaning external. The external_event_id points to an external concept. External_window is set to 120 for the external window to look back two hours. External_delay is zero, because there is no time delay in the query.

  The external event pointed to by external_event_id has: an att_id = 13 indicating heart rate; aggregate = 0 indicating no aggregate function simply evaluate the value; number_of_values = 1; value1 = 90 for 90 beats per minute; internal_or_external_binary = 1 indicating internal constraint; internal_all_or_window_binary = 1 indicating the whole of the patient's stay.

The table contains the following information:

- concept_id: a unique identifier for each concept.
- att_id: an attribute identifier relating to the attributes table (subsection B.0.4).
- aggregate_binary: A binary to determine if the concept contains an aggregate function. The value is 1 if using an aggregate, 0 otherwise.
- agg_id: the aggregate identifier relating to the knowledge base aggregate table (subsection B.0.1).
- operator_id: an operator identifier relating to the knowledge base operator table (subsection B.0.6).
- number_of_values: the number of values the operator takes.
- value1: the first value the operator takes as a function.
- value2: the second, if any, value the operator takes as a function.
- internal_or_external_binary: a binary to determine if there is an internal or external constraint upon this concept. A value of 1 denotes an internal constraint, 0 indicates an external constraint.
- internal_all_or_window_binary: a binary to determine if the internal constraint applies to the whole of patient stay or a more restricted time window of consideration. A value of 1 denotes the whole of the patient stay, 0 indicates a more restricted time window of consideration. This field is only applicable if the internal_or_external_binary is equal to 1 meaning internal.
- internal_window: is the size of the time window in minutes. This field is only applicable if the internal_all_or_window_binary has the value of 0, meaning internal time window.

- external_event_id: is a different concept that represents the external constraint. This field is only valid if the internal_or_external_binary is set to 0, indicating an external constraint.

- external_window: is the size of the time window in minutes. This field is only applicable if the internal_or_external_binary is set to 0, indicating an external constraint.

- external_delay: is a preceding time shift of the external time window, in minutes.

Relationships for this table exist as:

- A concept logic entry (subsection C.0.10) has two concepts associated with it.

- A hypothesis (subsection C.0.11) may have one concept associated with it.

concept_id	att_id	aggregate_binary	agg_id	operator_id	number_of_values	value1	value2	internal_or_external_binary	internal_all_or_windown_binary	internal_window	external_event_id	external_wondow	external_delay
1	1	0	NULL	1	1	60	NULL	1	1	NULL	NULL	NULL	NULL
2	2	0	NULL	1	1	1	NULL	1	1	NULL	NULL	NULL	NULL
3	3	0	NULL	1	1	92	NULL	1	1	NULL	NULL	NULL	NULL
4	4	0	NULL	1	1	10.2	NULL	1	1	NULL	NULL	NULL	NULL
5	5	0	NULL	1	1	22	NULL	1	1	NULL	NULL	NULL	NULL
6	6	0	NULL	1	1	100.25	NULL	1	1	NULL	NULL	NULL	NULL
7	7	0	NULL	1	1	96.5	NULL	1	1	NULL	NULL	NULL	NULL
8	8	0	NULL	1	1	39	NULL	1	1	NULL	NULL	NULL	NULL
9	9	0	NULL	1	1	71	NULL	1	1	NULL	NULL	NULL	NULL
10	10	0	NULL	1	1	7.36	NULL	1	1	NULL	NULL	NULL	NULL

Table C.9: Support Database Concepts.

C.0.10 Concept Logic

Concepts may be combined using logical operations in the Concept Logic Table (Table C.10). The concept logic table creates new concepts from two concepts and a logical operator. For example: one could create a concept from two of the examples from the concept table (subsection C.0.9): all patients whose average mean arterial pressure was greater 65 mmHg AND all patients whose heart rate was recorded above 90 beats per minute. The new concept would have: its own unique concept identifier; a logic_id = 1 indicating the logical operator AND from the knowledge base operator table (subsection B.0.6); concept1_id would point to the concept where patients had an average mean arterial pressure greater than 65 mmHg; concept2_id would point to the concept where all patients had a heart rate recorded greater than 90 beats per minute.

The table contains the following information:

- concept_id: a unique identifier for each concept.
- logic_id: a logical operator relating to the knowledge base operator table (subsection B.0.6).
- concept1_id: an identifier for a concept in the concept table (subsection C.0.9).
- concept2_id: an identifier for a concept in the concept table (subsection C.0.9).

Relationships for this table exist as:

- A concept logic entry must be made up of two concepts from the concept table (subsection C.0.9).
- A hypothesis (subsection C.0.11) may have one concept associated with it.

concept_id	logic_id	concept1_id	concept2_id
16	2	3	2
17	2	15	2
18	2	3	4
19	1	15	10

Table C.10: Support Database Concept Logic.

C.0.11 Hypothesis

The Hypothesis Table (Table C.11) contains the concept_id associated with the instance of the hypothesis.

The table contains the following information:

- inst_id: a unique instance identifier.
- concept_id: a concept identifier relating to a concept in the concept logic table (subsection C.0.10).

Relationships for this table exist as:

- An entry from the main table (subsection C.0.12) has exactly one hypothesis associated with it.
- A hypothesis has many concept database (subsection C.0.1) entries.
- A hypothesis has one concept logic (subsection C.0.10) entry.
- A hypothesis has many machine learning algorithms (subsection C.0.13) associated with it.
- A hypothesis has many statistics descriptions (subsection C.0.14) associated with it,

inst_id	concept_id
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10

Table C.11: Support Database Hypothesis.

C.0.12 Main

The Main Table (Table C.12) defines an instance of a hypothesis, and tracks to provenance of that instance through a parent-child schema, as well as the target_id that links to the definition of disease (subsection C.0.15).

The table contains the following information:

- target_id: a disease identifier for the disease definition table (subsection C.0.15).
- inst_id: a unique instance identifier.
- parent_inst_id: the instance identifier of the parent (if one exists).

Relationships for this table exist as:

- A main table entry has exactly one hypothesis (subsection C.0.11) associated with it.

target_id	inst_id	parent_inst_id
1	1	0
1	2	0
1	3	0
1	4	0
1	5	0
1	6	0
1	7	0
1	8	0
1	9	0
1	10	0

Table C.12: Support Database Main.

C.0.13 Machine Learning Algorithms

The Machine Learning Algorithms Table (Table C.13) contains the identifiers and options necessary to call the algorithm to evaluate a model. Thee field values contained herein are associated with the five knowledge base machine learning algorithm tables to define and execute the algorithm.

The table contains the following information:

- mla_id: a unique machine learning algorithm identifier.
- cn_id: a classname identifier that links to the knowledge base mla_classname table (subsection B.0.9).
- opt_id: an option identifier that linke to the knowledge base mla_classname_opt table (subsection B.0.10).

- opttype: an options type that links to the knowledge base mla_opt table (subsection B.0.11) and disposes the option to be either a range or an enumeration of text.

- value: the value of the option.

Relationships for this table exist as:

- A hypothesis (subsection C.0.11) has a machine learning algorithm definition for every different algorithm that is run against the hypothesis.

- Every model run by a defined machine learning algorithm has one entry in the statistics table (subsection C.0.14).

mla_id	cn_id	opt_id	opttype	value
1	1	1	enum	3
1	1	2	enum	4
2	2	3	enum	6
2	2	4	enum	15
3	3	NULL	NULL	NULL
4	4	NULL	NULL	NULL
5	5	NULL	NULL	NULL
6	6	NULL	NULL	NULL
7	7	NULL	NULL	NULL
8	8	5	range	5

Table C.13: Support Database Machine Learning Algorithms.

C.0.14 Statistics

The evaluation of a model creates several statistical measures, which are contained in the database Statistics Table (Table C.14).

The table contains the following information:

- inst_id: an instance identifier that references the feature selection table (subsection C.0.3).
- batch_id: a batch identifier.
- mla_id: a machine learning algorithm identifier that references the machine learning algorithms table where the options are defined (subsection C.0.13).
- TP: true positive count.
- TN: true negative count.
- FP: false positive count.
- FN: false negative count.
- ROC: the area under the receiver operating characteristic curve.
- kappa: Cohen's kappa coefficient statistic that measures the agreement between the model and the disease definition (subsection C.0.15).

Relationships for this table exist as:

- Every model run by a defined machine learning algorithm (subsection C.0.13) has one entry in the statistics table.

inst_id	batch_id	mla_id	TP	TN	FP	FN	ROC	kappa
1	0	5	196	550	72	190	0.818059	0.416151
1	0	7	263	549	73	123	0.863211	0.578192
1	0	8	268	519	103	118	0.823497	0.532632
1	0	10	261	544	78	125	0.864298	0.563787
1	0	11	267	551	71	119	0.788781	0.591516
1	0	13	121	558	64	265	0.698037	0.23366
1	0	16	154	536	86	232	0.704855	0.280853
1	0	18	294	519	103	92	0.847808	0.592858
1	0	20	312	505	117	74	0.822481	0.607352
1	0	22	207	572	50	179	0.727942	0.486743

Table C.14: Support Database Statistics.

C.0.15 Target Class

Target Class (Table C.15) is the problem declaration table that holds data surrounding the onset of disease, breaking out positive patients and control patients. For disease positive patients, the timestamp records the earliest conditions necessary for the disease to be diagnosed during their ICU stay, or a randomly generated timestamp within the bounds of their ICU stay for patients who do not contract the disease.

The table contains the following information:

- target_id: an identifier for a disease.
- patient_id: a patient identifier that links to the MIMIC database specific to ICU stays (subsection A.0.7).
- pt_timestamp: a timestamp for the onset of disease (if disease positive) or a randomly chosen time during the patient stay (if disease negative).

- discriminator_binary: a discriminator that separates patients into either disease positive (1) or disease negative (0).

Relationships for this table exist as:

- There is a many-to-many relationship between patients in the target class table and the feature selection database table (subsection C.0.2).

- There is a many-to-many relationship between patients in the target class table and the concept database table (subsection C.0.1).

target_id	patient_id	pt_timestamp	discriminator_binary
1	211552	2101-10-20 21:51:00	1
1	232669	2104-08-08 04:21:00	1
1	254478	2108-08-26 03:01:00	1
1	277421	2109-10-28 16:12:00	1
1	251972	2196-09-28 04:22:00	1
1	260172	2119-12-25 01:11:00	1
1	227964	2107-09-08 17:11:00	1
1	263211	2145-05-10 13:09:00	1
1	264885	2106-06-17 21:49:00	1
1	218740	2116-12-30 16:31:00	1

Table C.15: Support Database Target Class.

Appendix D

ACUTE RESPIRATORY DISTRESS SYNDROME

D.1 ARDS Model

D.1.1 Training and Testing

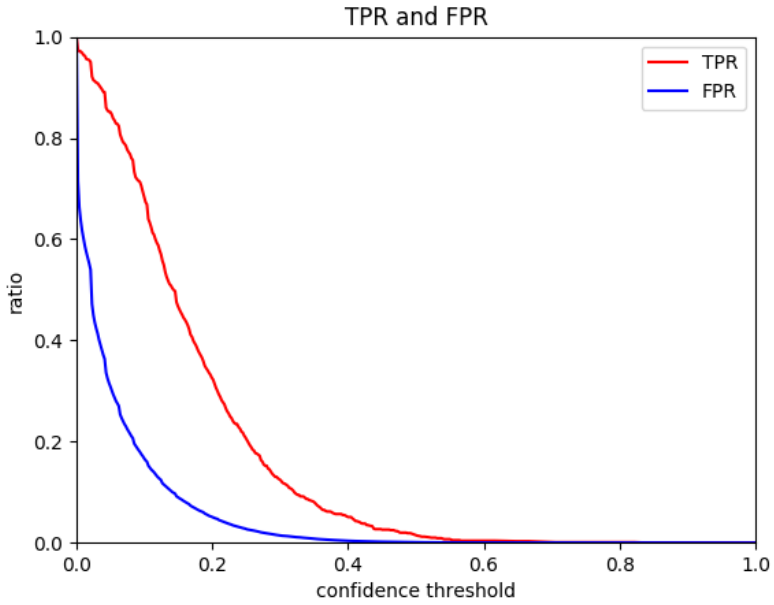


Figure D.1: ARDS training/testing true positive and false positive rates.

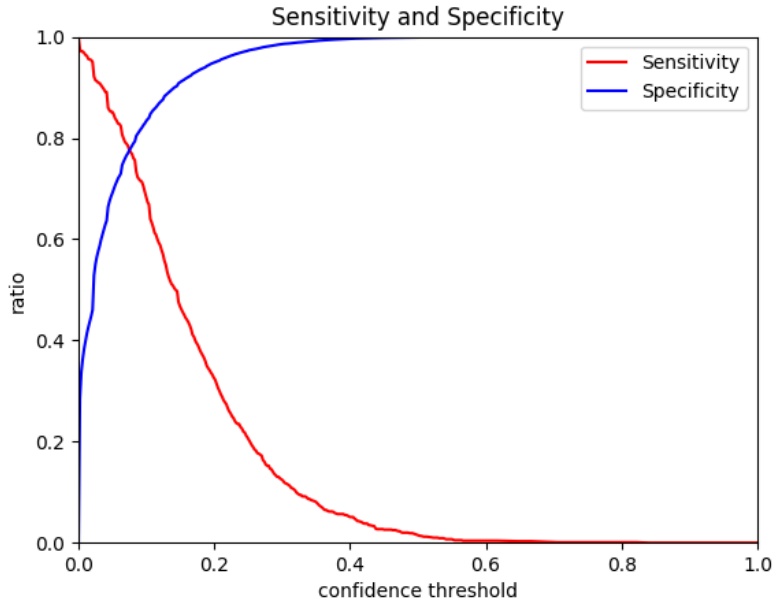


Figure D.2: ARDS training/testing sensitivity and specificity.

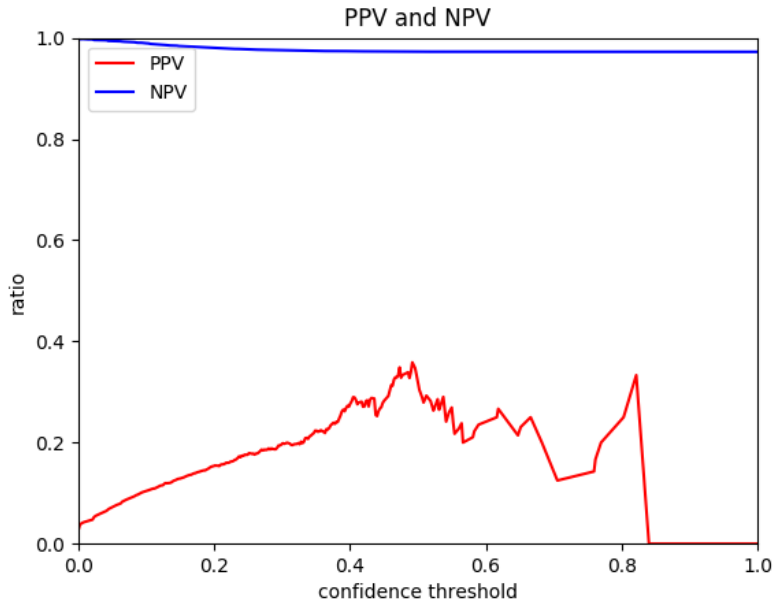


Figure D.3: ARDS training/testing positive and negative predictive value.

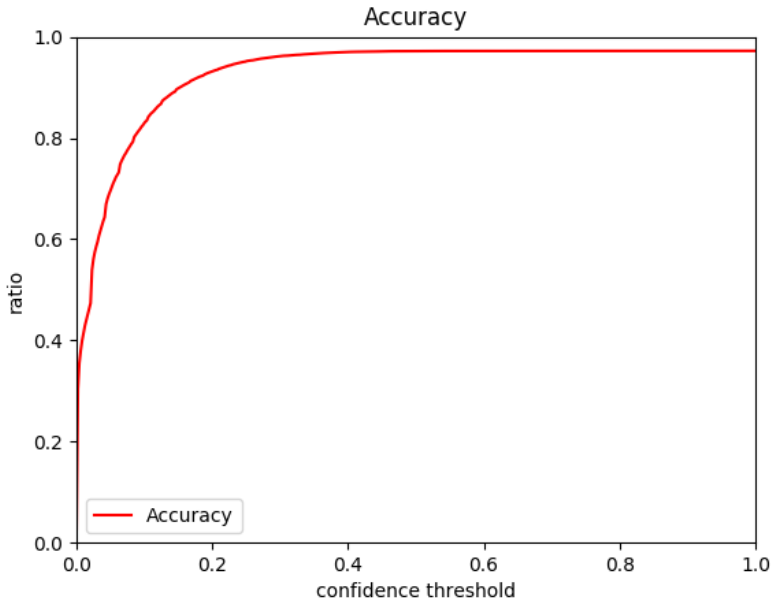


Figure D.4: ARDS training/testing accuracy.

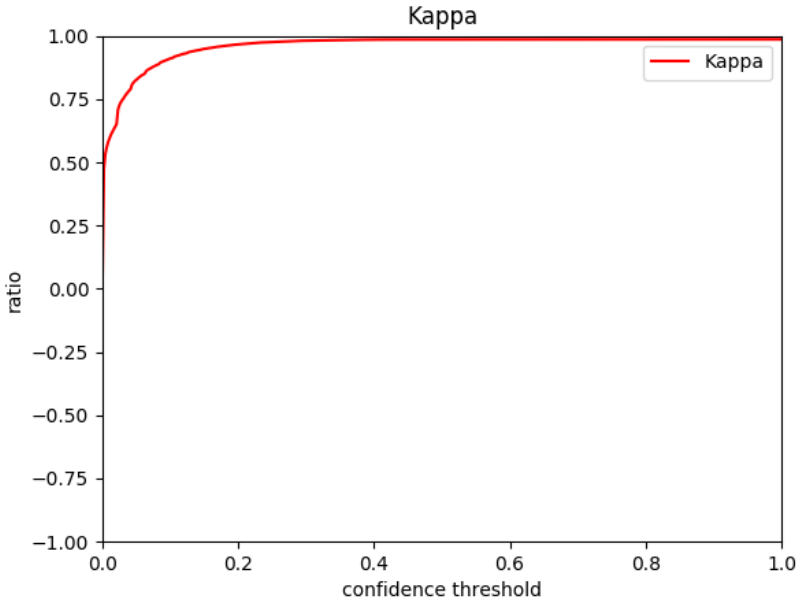


Figure D.5: ARDS training/testing kappa.

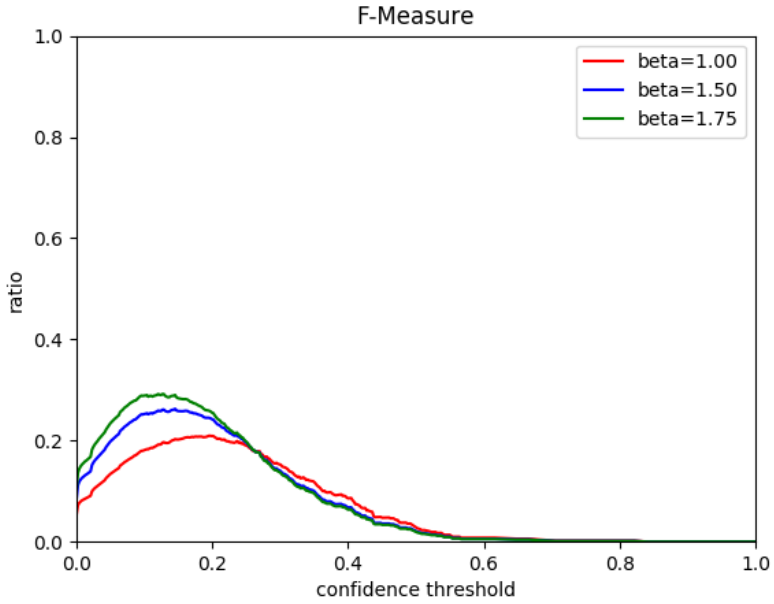


Figure D.6: ARDS training/testing F-measures.

D.1.2 Validation

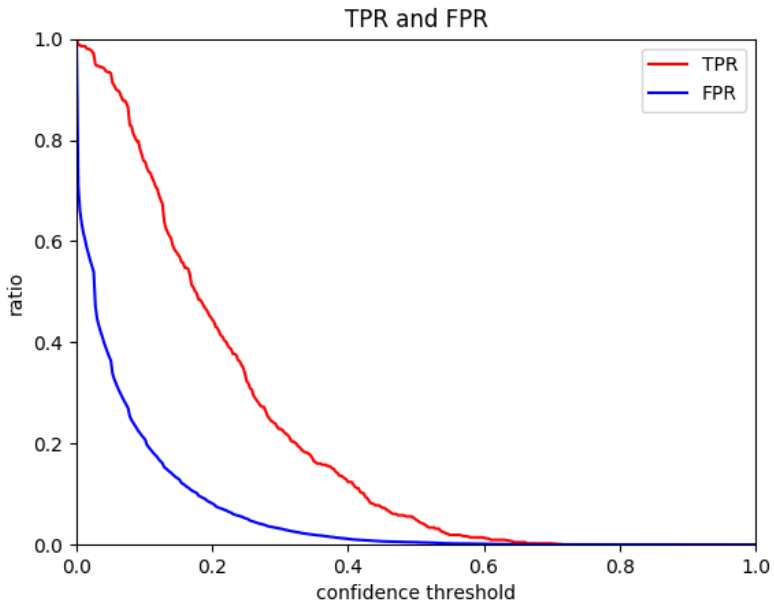


Figure D.7: ARDS validation true and false positive rates.

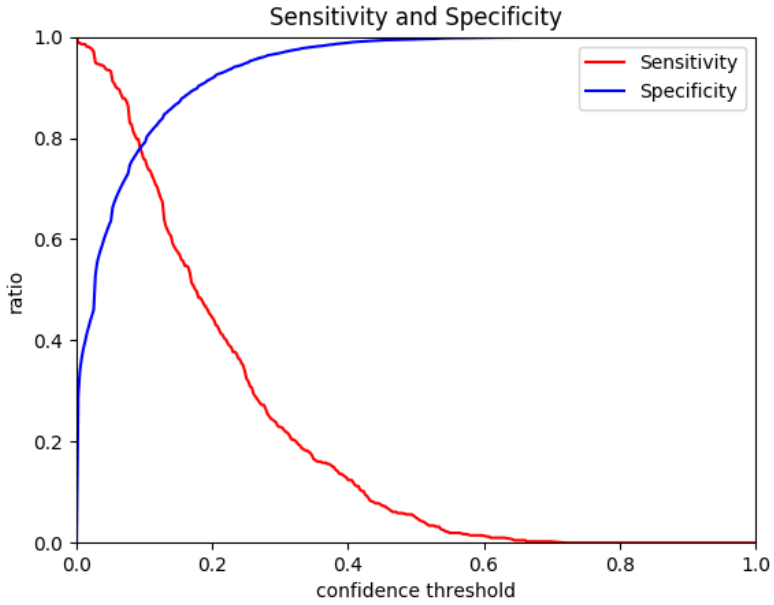


Figure D.8: ARDS validation sensitivity and specificity.

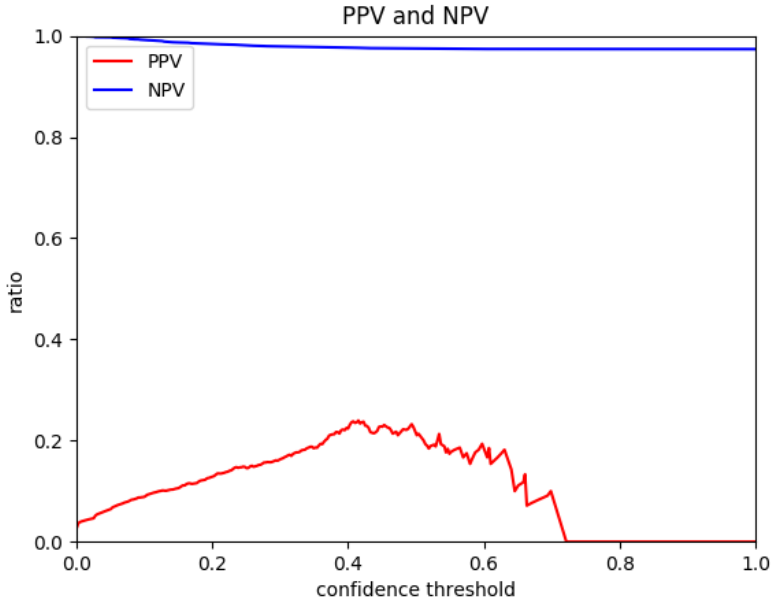


Figure D.9: ARDS validation positive and negative predictive values.

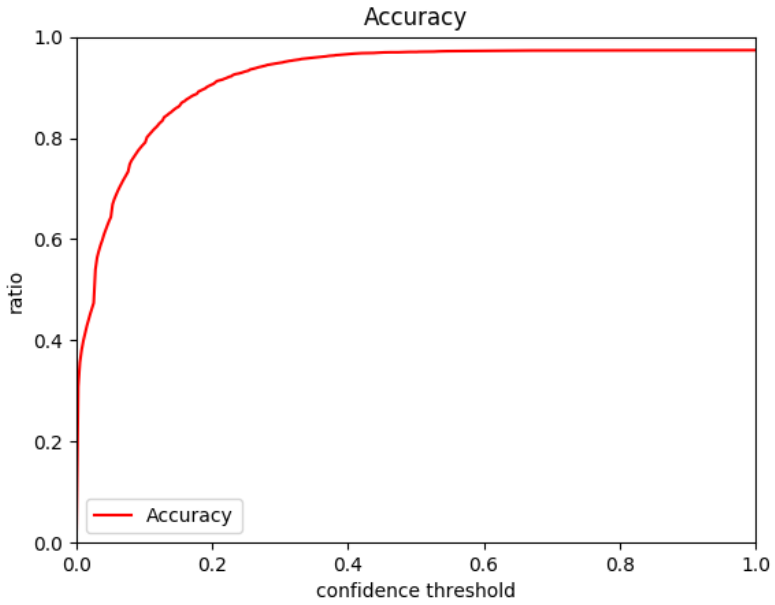


Figure D.10: ARDS validation accuracy.

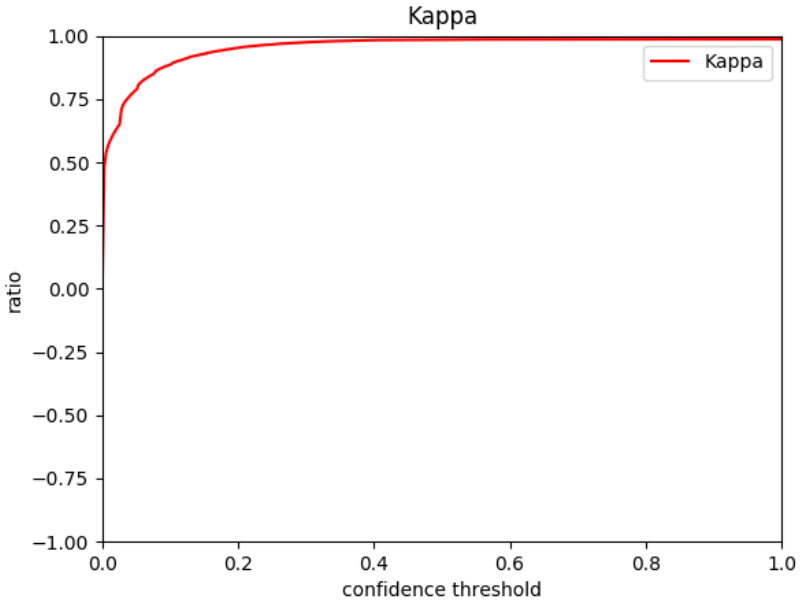


Figure D.11: ARDS validation kappa.

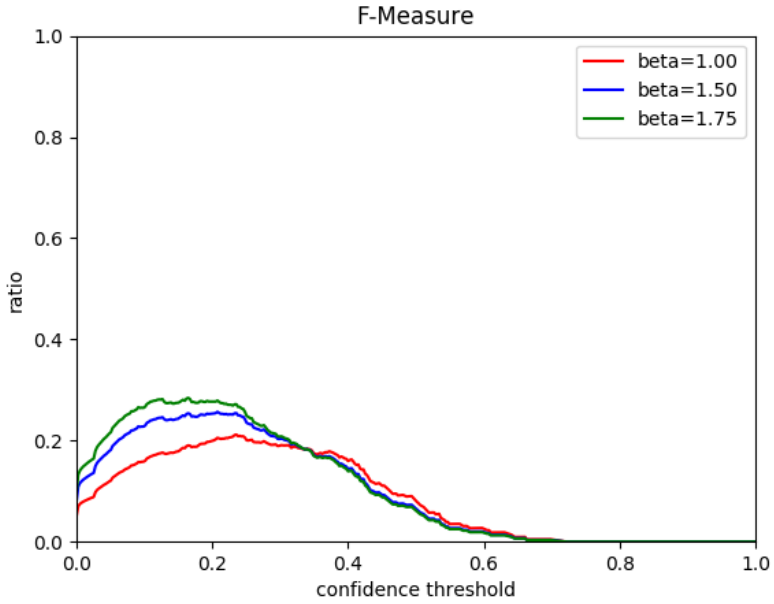
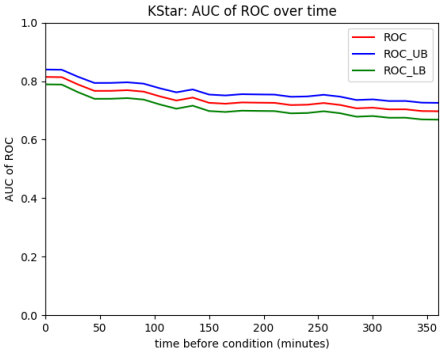
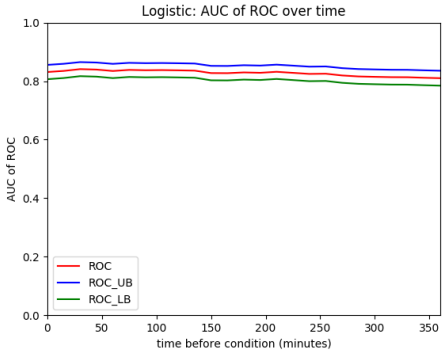


Figure D.12: ARDS validation F-measures.

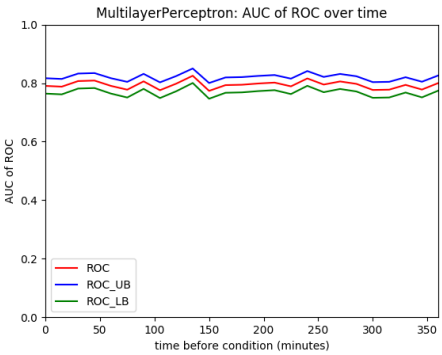
D.1.3 Machine Learning Algorithms over time



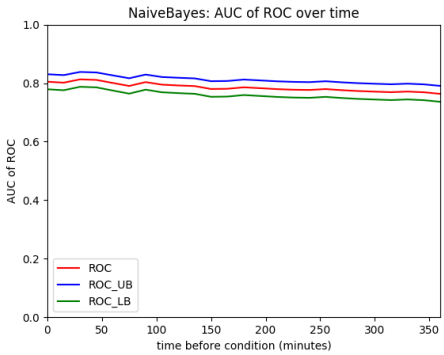
(a) KStar



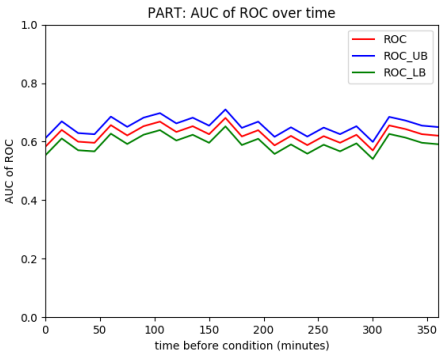
(b) (Logistic Regression



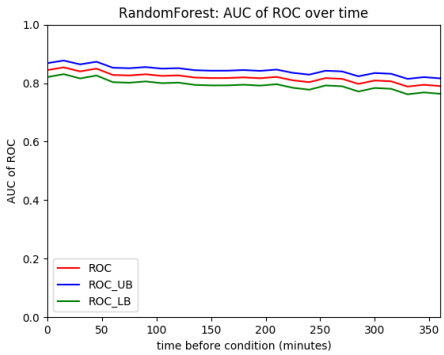
(c) Multilayer Perceptron



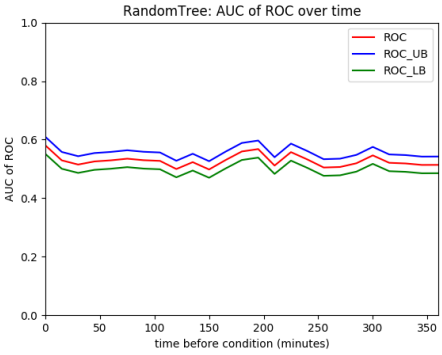
(d) Naive Bayes



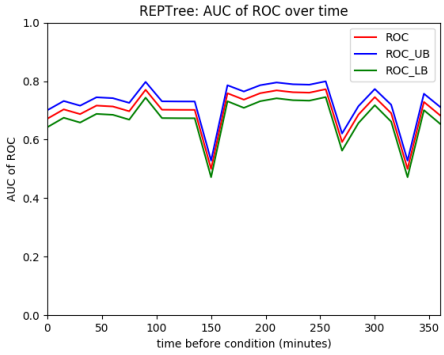
(e) PART



(f) Random Forest



(g) Random Tree



(h) REP Tree

Figure D.13: Algorithm's performance over time on the ARDS selection criteria.

D.1.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Tidal Volume	235	500	600	24	250	500	<0.0001
O_2 Flow (additional cannula)	0	0	0	0	0	0	0.4838
PCO_2 (arterial)	36	42	49	32	39	45	<0.0001
20 Gauge placed in outside facility	0	0	0	0	0	0	0.5293
Compliance	22.222	26.923	33.176	23.077	27.5	32.660	0.3976
Metered Dose Inhaler	2	4	6	2	4	6	0.8523
Mean Airway Pressure	8.5	10.7	13	8	9	11	<0.0001
Creatine Kinase, MB Isoenzyme	3	5	8	2	3	5	<0.0001
Urine Out Foley	35	80	150	25	60	120	<0.0001
Chloride, Whole Blood	102	106	110	99	104	109	<0.0001
IVlock	20	20	20	20	20	20	0.1420
SpO_2	96	98	100	96	98	100	0.7315
Eye Opening	1	3	4	3	4	4	<0.0001
Magnesium Sulfate	0.033	0.967	2	0.033	1	2	0.1020
Lactic Acid	0.9	1.4	1.6	1.2	1.4	1.6	0.1363
Asparate Aminotransferase (AST)	19	30	58	18	26	46	0.0011
O_2 Flow (lpm)	2	3	6	2	2	4	<0.0001
Albumin	2.3	2.8	3.3	2.4	2.9	3.5	0.0010
Hematocrit	28	31.6	34.85	27	30.8	35.6	0.1862

Table D.1: ARDS model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Tidal Volume	400	500	600	24	220	500	420	521	650	70	450	550	<0.0001
O_2 Flow (additional cannula)	0	0	0	0	0	0	0	0	0	0	0	0	0.5140
PCO_2 (arterial)	37	43	52	32	38	44	38	43	51	35	41	48	<0.0001
20 Gauge placed in outside facility	0	0	0	0	0	0	0	0	0	0	0	0	0.4769
Compliance	20.690	25.807	33.333	23.077	27.778	32.609	20	25.596	33.333	23.716	27.500	32.463	<0.0001
Metered Dose Inhaler	2	4	6	2	4	6	2	4	6	2	4	6	<0.0001
Mean Airway Pressure	10	12	16	8	9	11	8	11	14	8	10	11.95	<0.0001
Creatine Kinase, MB Isoenzyme	3	6	9	2	3	5	3	6	10	3	4	7	<0.0001
Urine Out Foley	40	90	190	25	60	110	40	100	200	35	70	136.25	<0.0001
Chloride, Whole Blood	103	107	111	98	104	108	103	107	110	101	105	108	<0.0001
IV/Saline lock	20	20	20	20	20	20	20	20	20	20	20	20	0.0522
SpO_2	96	98	100	96	98	100	96	99	100	96	98	100	0.0993
Eye Opening	1	1	3	3	4	4	1	2	4	1	3	4	<0.0001
Magnesium Sulfate	0.033	0.967	2	0.033	1	2	0.033	0.967	2	0.033	0.967	2	<0.0001
Lactic Acid	0.9	1.3	1.6	1.2	1.4	1.6	0.9	1.4	1.6	1	1.4	1.6	0.5055
Asparate Aminotransferase (AST)	21	34	65	18	25	44	21	32	66	19	28.5	50.5	<0.0001
O_2 Flow (lpm)	2	3	6	2	2	4	2	3	10	2	3	6	<0.0001
Albumin	2.1	2.7	3.2	2.4	2.9	3.5	2.4	2.9	3.4	2.3	2.8	3.4	0.0009
Hematocrit	28	32.1	35.2	26.9	30.7	35.5	28.7	32.2	36.3	28.225	31.35	34.1	<0.0001

Table D.2: ARDS model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	15638	411	15227	
Age, median	64	64	64	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.6167
Weight, median (Q1,Q3)	78.00 (65.20, 92.00)	80.00 (66.00, 95.05)	78.00 (65.20, 91.90)	0.1044
BMI, median (Q1,Q3)	28.08 (24.56, 32.57)	28.10 (24.28, 33.01)	28.07 (24.57, 32.50)	0.8140
Male	8879 (0.57)	238 (0.58)	8641 (0.57)	0.7582
In-Hospital Mortality	2734 (0.17)	117 (0.28)	2617 (0.17)	<0.0001
30 Day Mortality	3169 (0.20)	123 (0.30)	3046 (0.20)	<0.0001
ICU LOS, median(Q1,Q3)	2.27 (1.27, 4.85)	9.22 (4.45, 17.73)	2.21 (1.25, 4.60)	<0.0001
ICU				
CCU	2176 (0.14)	34 (0.08)	2142 (0.14)	0.0019
CSRU	2684 (0.17)	91 (0.22)	2593 (0.17)	0.0136
MICU	5880 (0.38)	108 (0.26)	5772 (0.38)	0.0001
SICU	2433 (0.16)	92 (0.22)	2341 (0.15)	0.0004
TSICU	1837 (0.12)	86 (0.21)	1751 (0.11)	<0.0001
ethnicity				
Asian	402 (0.03)	10 (0.02)	392 (0.03)	0.8601
Black	1435 (0.09)	31 (0.08)	1404 (0.09)	0.2678
Hispanic	540 (0.03)	10 (0.02)	530 (0.03)	0.2594
White	11119 (0.71)	284 (0.69)	10835 (0.71)	0.6256
insurance				
Government	413 (0.03)	9 (0.02)	404 (0.03)	0.5684
Medicaid	1385 (0.09)	34 (0.08)	1351 (0.09)	0.6868
Medicare	8445 (0.54)	219 (0.53)	8226 (0.54)	0.8408
Private	5206 (0.33)	145 (0.35)	5061 (0.33)	0.4788
Self Pay	189 (0.01)	4 (0.01)	185 (0.01)	0.6601

Table D.3: ARDS patients' demographics of validation set by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	15638	230	13359	1868	181	
Age, median	64	61	64	65	66.0	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (63.00, 70.00)	0.9497
Weight, median (Q1,Q3)	78.00 (65.20, 92.00)	80.00 (66.10, 98.75)	77.20 (65.00, 91.00)	79.00 (66.10, 93.05)	79.90 (65.00, 91.50)	0.0077
BMI, median (Q1,Q3)	28.08 (24.56, 32.57)	27.89 (23.97, 32.67)	28.12 (24.64, 32.35)	28.00 (24.49, 32.79)	28.54 (24.77, 33.44)	0.9094
Male	8879 (0.57)	134 (0.58)	7546 (0.56)	1095 (0.59)	104 (0.57)	0.7011
In-Hospital Mortality	2734 (0.17)	70 (0.30)	2066 (0.15)	551 (0.29)	47 (0.26)	<0.0001
30 Day Mortality	3169 (0.20)	72 (0.31)	2455 (0.18)	591 (0.32)	51 (0.28)	<0.0001
ICU LOS, median(Q1,Q3)	2.27 (1.27, 4.85)	10.50 (4.94, 17.83)	2.12 (1.21, 4.11)	3.80 (1.85, 7.82)	7.87 (3.96, 16.05)	<0.0001
ICU						
CCU	2176 (0.14)	14 (0.06)	1885 (0.14)	257 (0.14)	20 (0.11)	0.0089
CSRU	2684 (0.17)	47 (0.20)	2161 (0.16)	432 (0.23)	44 (0.24)	<0.0001
MICU	5880 (0.38)	63 (0.27)	5134 (0.38)	638 (0.34)	45 (0.25)	<0.0001
SICU	2433 (0.16)	47 (0.20)	2080 (0.16)	261 (0.14)	45 (0.25)	0.0009
TSICU	1837 (0.12)	59 (0.26)	1471 (0.11)	280 (0.15)	27 (0.15)	<0.0001
ethnicity						
Asian	402 (0.03)	8 (0.03)	355 (0.03)	37 (0.02)	2 (0.01)	0.1598
Black	1435 (0.09)	18 (0.08)	1277 (0.10)	127 (0.07)	13 (0.07)	0.0019
Hispanic	540 (0.03)	5 (0.02)	478 (0.04)	52 (0.03)	5 (0.03)	0.2243
White	11119 (0.71)	155 (0.67)	9547 (0.71)	1288 (0.69)	129 (0.71)	0.5914
insurance						
Government	413 (0.03)	5 (0.02)	366 (0.03)	38 (0.02)	4 (0.02)	0.3321
Medicaid	1385 (0.09)	22 (0.10)	1185 (0.09)	166 (0.09)	12 (0.07)	0.7654
Medicare	8445 (0.54)	106 (0.46)	7216 (0.54)	1010 (0.54)	113 (0.62)	0.1681
Private	5206 (0.33)	94 (0.41)	4440 (0.33)	621 (0.33)	51 (0.28)	0.1445
Self Pay	189 (0.01)	3 (0.01)	152 (0.01)	33 (0.02)	1 (0.01)	0.1102

Table D.4: ARDS patients' Demographics of validation set in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	Condition Positive	condition Negative	P-value
N	15638	411	15227	
Comorbidity				
Congestive heart failure	4518 (0.29)	150 (0.36)	4368 (0.29)	0.0037
Cardiac arrhythmias	5441 (0.35)	137 (0.33)	5304 (0.35)	0.6111
Valvular disease	2392 (0.15)	55 (0.13)	2337 (0.15)	0.3147
Pulmonary circulation disorders	833 (0.05)	13 (0.03)	820 (0.05)	0.0541
Peripheral vascular disorders	1784 (0.11)	75 (0.18)	1709 (0.11)	<0.0001
Hypertension, uncomplicated	6214 (0.40)	138 (0.34)	6076 (0.40)	0.0447
Hypertension, complicated	2011 (0.13)	38 (0.09)	1973 (0.13)	0.0384
Paralysis	401 (0.03)	10 (0.02)	391 (0.03)	0.8663
Other neurological disorders	1009 (0.06)	22 (0.05)	987 (0.06)	0.3739
Chronic pulmonary disease	3494 (0.22)	83 (0.20)	3411 (0.22)	0.3504
Diabetes, uncomplicated	3198 (0.20)	68 (0.17)	3130 (0.21)	0.0760
Diabetes, complicated	1133 (0.07)	29 (0.07)	1104 (0.07)	0.8852
Hypothyroidism	1626 (0.10)	22 (0.05)	1604 (0.11)	0.0013
Renal failure	2409 (0.15)	45 (0.11)	2364 (0.16)	0.0197
Liver disease	1677 (0.11)	56 (0.14)	1621 (0.11)	0.0687
Peptic ulcer disease excluding bleeding	108 (0.01)	1 (0.00)	107 (0.01)	0.2688
AIDS/HIV	153 (0.01)	6 (0.01)	147 (0.01)	0.3173
Lymphoma	225 (0.01)	5 (0.01)	220 (0.01)	0.7034
Metastatic cancer	149 (0.01)	1 (0.00)	148 (0.01)	0.1354
Solid tumor without metastasis	1202 (0.08)	41 (0.10)	1161 (0.08)	0.0898
Rheumatoid arthritis/collagen vascular diseases	543 (0.03)	8 (0.02)	535 (0.04)	0.0925
Coagulopathy	1646 (0.11)	63 (0.15)	1583 (0.10)	0.0024
Weight loss	726 (0.05)	26 (0.06)	700 (0.05)	0.1084
Fluid and electrolyte disorders	4524 (0.29)	100 (0.24)	4424 (0.29)	0.0790
Blood loss anemia	331 (0.02)	15 (0.04)	316 (0.02)	0.0304
Deficiency anemia	447 (0.03)	7 (0.02)	440 (0.03)	0.1604
Alcohol abuse	727 (0.05)	17 (0.04)	710 (0.05)	0.6252
Drug abuse	635 (0.04)	14 (0.03)	621 (0.04)	0.5047
Psychoses	261 (0.02)	5 (0.01)	256 (0.02)	0.4718
Depression	1379 (0.09)	12 (0.03)	1367 (0.09)	<0.0001
Trauma	1949 (0.12)	95 (0.23)	1854 (0.12)	<0.0001

Table D.5: ARDS patients' Comorbidities of validation set by condition positive or negative. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	15638	230	13359	1868	181	
Comorbidity						
Congestive heart failure	4518 (0.29)	83 (0.36)	3700 (0.28)	668 (0.36)	67 (0.37)	<0.0001
Cardiac arrhythmias	5441 (0.35)	75 (0.33)	4569 (0.34)	735 (0.39)	62 (0.34)	0.0051
Valvular disease	2392 (0.15)	24 (0.10)	2042 (0.15)	295 (0.16)	31 (0.17)	0.2355
Pulmonary circulation disorders	833 (0.05)	6 (0.03)	735 (0.06)	85 (0.05)	7 (0.04)	0.0787
Peripheral vascular disorders	1784 (0.11)	47 (0.20)	1474 (0.11)	235 (0.13)	28 (0.15)	<0.0001
Hypertension, uncomplicated	6214 (0.40)	70 (0.30)	5459 (0.41)	617 (0.33)	68 (0.38)	<0.0001
Hypertension, complicated	2011 (0.13)	16 (0.07)	1839 (0.14)	134 (0.07)	22 (0.12)	<0.0001
Paralysis	401 (0.03)	6 (0.03)	356 (0.03)	35 (0.02)	4 (0.02)	0.2517
Other neurological disorders	1009 (0.06)	16 (0.07)	894 (0.07)	93 (0.05)	6 (0.03)	0.0160
Chronic pulmonary disease	3494 (0.22)	48 (0.21)	2930 (0.22)	481 (0.26)	35 (0.19)	0.0086
Diabetes, uncomplicated	3198 (0.20)	43 (0.19)	2759 (0.21)	371 (0.20)	25 (0.14)	0.1846
Diabetes, complicated	1133 (0.07)	14 (0.06)	1005 (0.08)	99 (0.05)	15 (0.08)	0.0078
Hypothyroidism	1626 (0.10)	11 (0.05)	1472 (0.11)	132 (0.07)	11 (0.06)	<0.0001
Renal failure	2409 (0.15)	18 (0.08)	2187 (0.16)	177 (0.09)	27 (0.15)	<0.0001
Liver disease	1677 (0.11)	34 (0.15)	1372 (0.10)	249 (0.13)	22 (0.12)	0.0004
Peptic ulcer disease excluding bleeding	108 (0.01)	0 (0.00)	100 (0.01)	7 (0.00)	1 (0.01)	0.1728
AIDS/HIV	153 (0.01)	3 (0.01)	124 (0.01)	23 (0.01)	3 (0.02)	0.4458
Lymphoma	225 (0.01)	2 (0.01)	194 (0.01)	26 (0.01)	3 (0.02)	0.8911
Metastatic cancer	149 (0.01)	0 (0.00)	123 (0.01)	25 (0.01)	1 (0.01)	0.1354
Solid tumor without metastasis	1202 (0.08)	19 (0.08)	995 (0.07)	166 (0.09)	22 (0.12)	0.0257
Rheumatoid arthritis/collagen vascular diseases	543 (0.03)	5 (0.02)	471 (0.04)	64 (0.03)	3 (0.02)	0.3987
Coagulopathy	1646 (0.11)	42 (0.18)	1330 (0.10)	253 (0.14)	21 (0.12)	<0.0001
Weight loss	726 (0.05)	16 (0.07)	612 (0.05)	88 (0.05)	10 (0.06)	0.3789
Fluid and electrolyte disorders	4524 (0.29)	59 (0.26)	3951 (0.30)	473 (0.25)	41 (0.23)	0.0034
Blood loss anemia	331 (0.02)	7 (0.03)	275 (0.02)	41 (0.02)	8 (0.04)	0.1251
Deficiency anemia	447 (0.03)	3 (0.01)	407 (0.03)	33 (0.02)	4 (0.02)	0.0087
Alcohol abuse	727 (0.05)	13 (0.06)	608 (0.05)	102 (0.05)	4 (0.02)	0.1253
Drug abuse	635 (0.04)	11 (0.05)	540 (0.04)	81 (0.04)	3 (0.02)	0.3575
Psychoses	261 (0.02)	2 (0.01)	213 (0.02)	43 (0.02)	3 (0.02)	0.1213
Depression	1379 (0.09)	5 (0.02)	1308 (0.10)	59 (0.03)	7 (0.04)	<0.0001
Trauma	1949 (0.12)	67 (0.29)	1516 (0.11)	338 (0.18)	28 (0.15)	<0.0001

Table D.6: ARDS patients' Comorbidities of validation set in context of model performance. P-values for patient counts calculated by Pearson Chi-Squared test.

D.2 Severe Acute Hypoxemic Respiratory Failure

D.2.1 Training and Testing

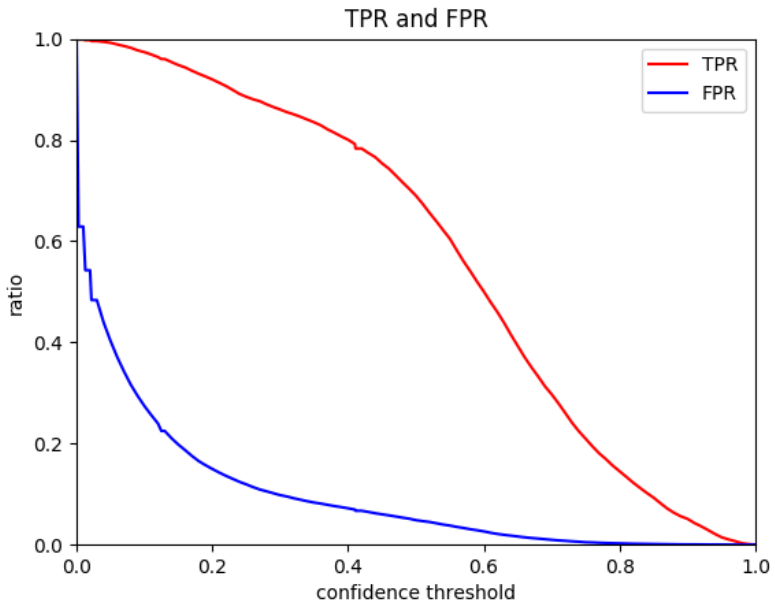


Figure D.14: Hypoxemic training/testing true positive and false positive rates.

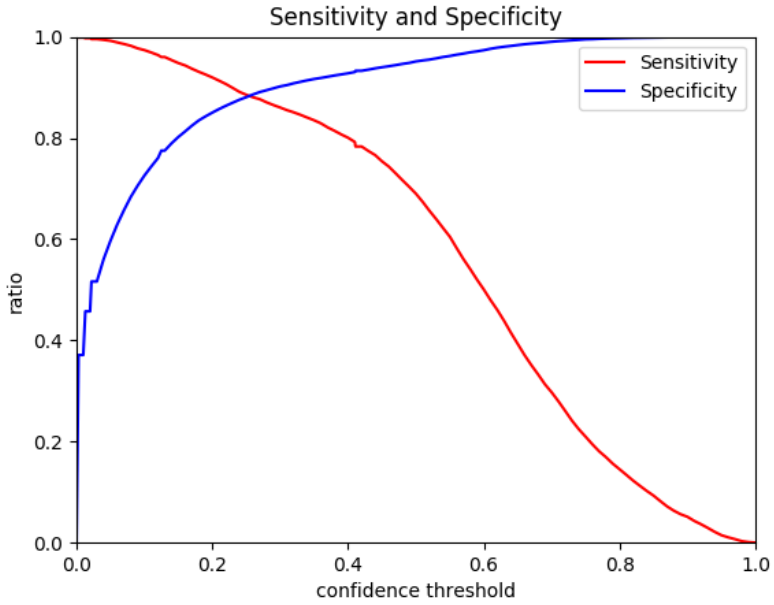


Figure D.15: Hypoxemic training/testing sensitivity and specificity.

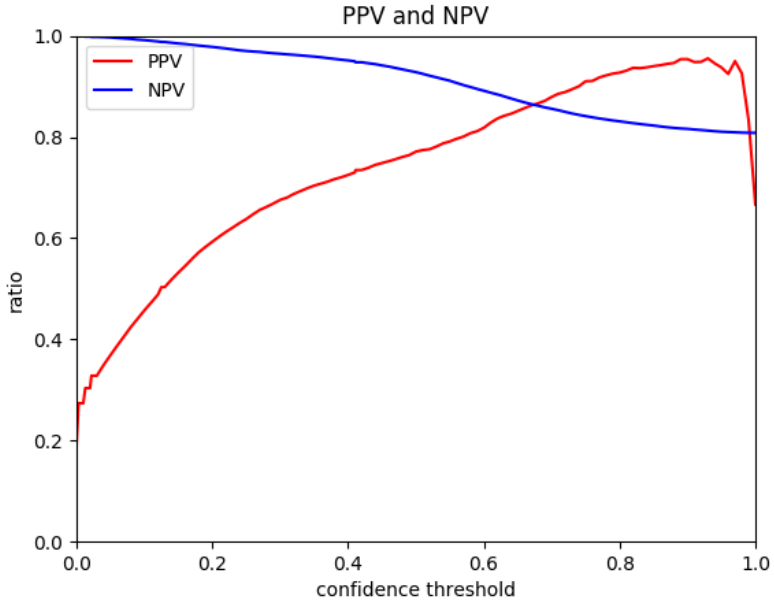


Figure D.16: Hypoxemic training/testing positive and negative predictive value.

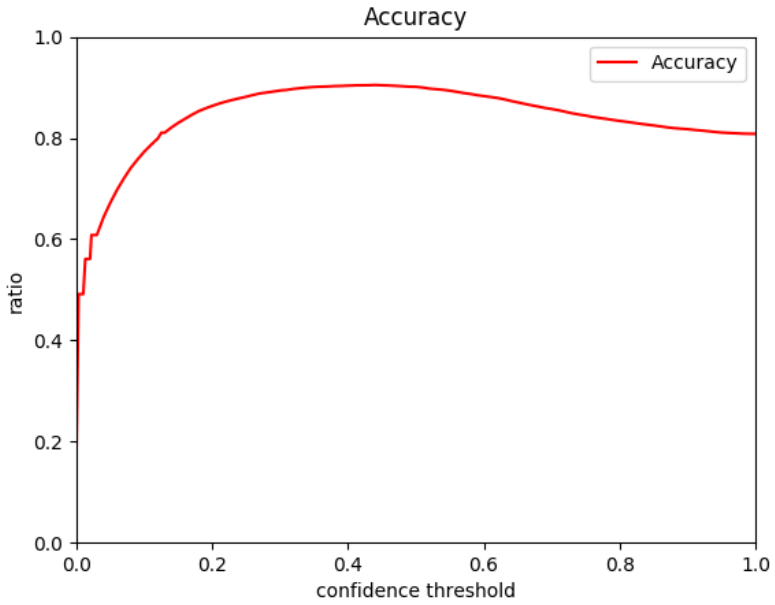


Figure D.17: Hypoxemic training/testing accuracy.

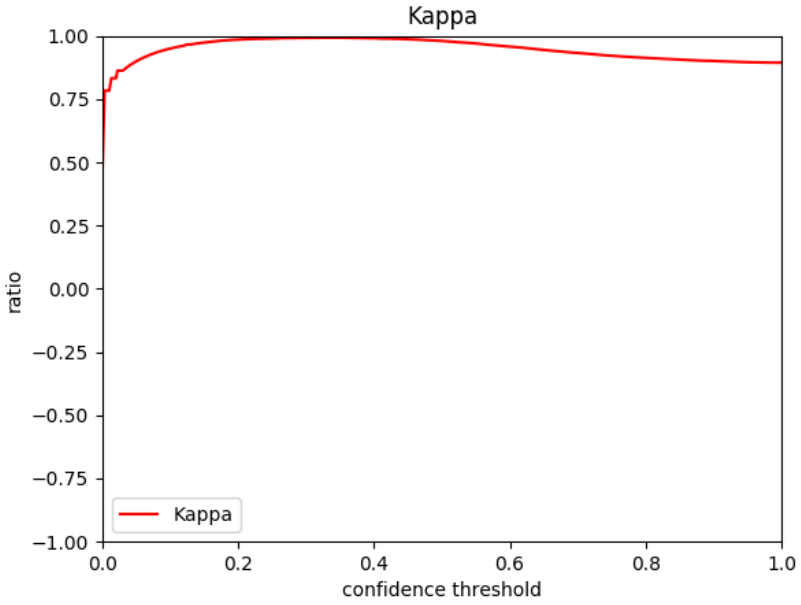


Figure D.18: Hypoxemic training/testing kappa.

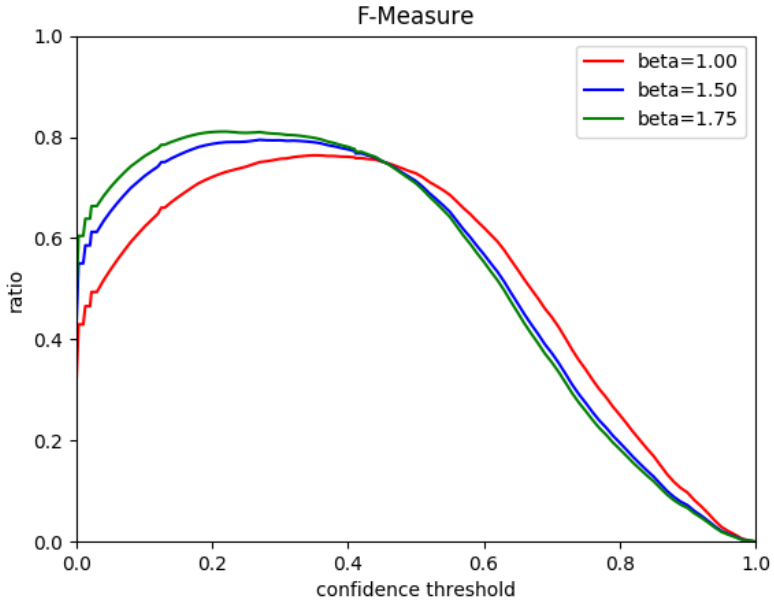


Figure D.19: Hypoxemic training/testing F-measures.

D.2.2 Validation

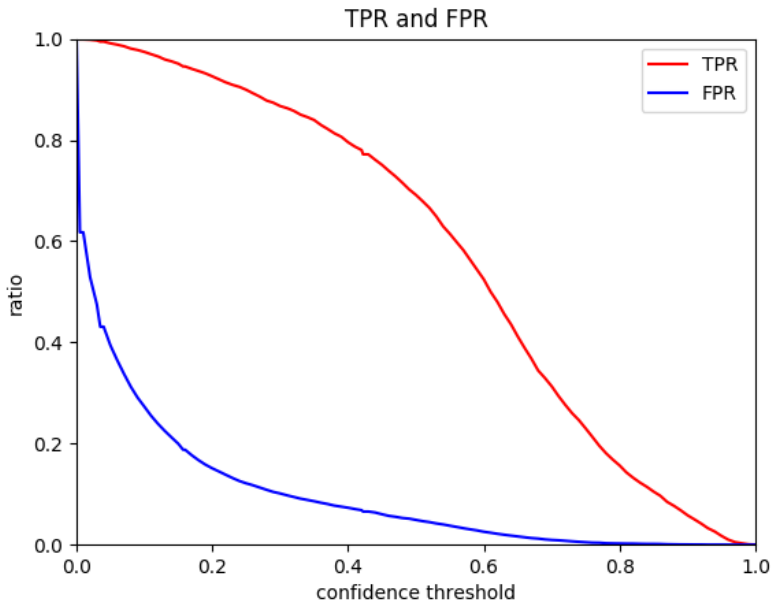


Figure D.20: Hypoxemic validation true and false positive rates.

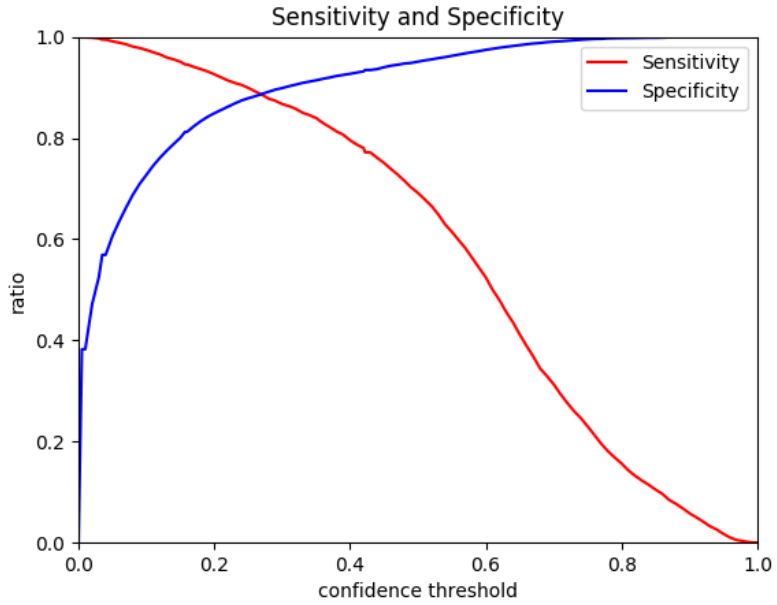


Figure D.21: Hypoxemic validation sensitivity and specificity.

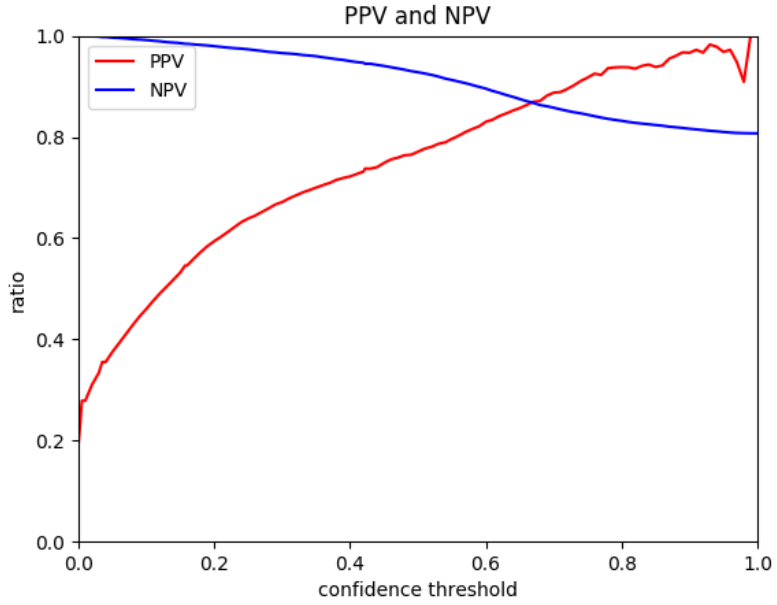


Figure D.22: Hypoxemic validation positive and negative predictive values.

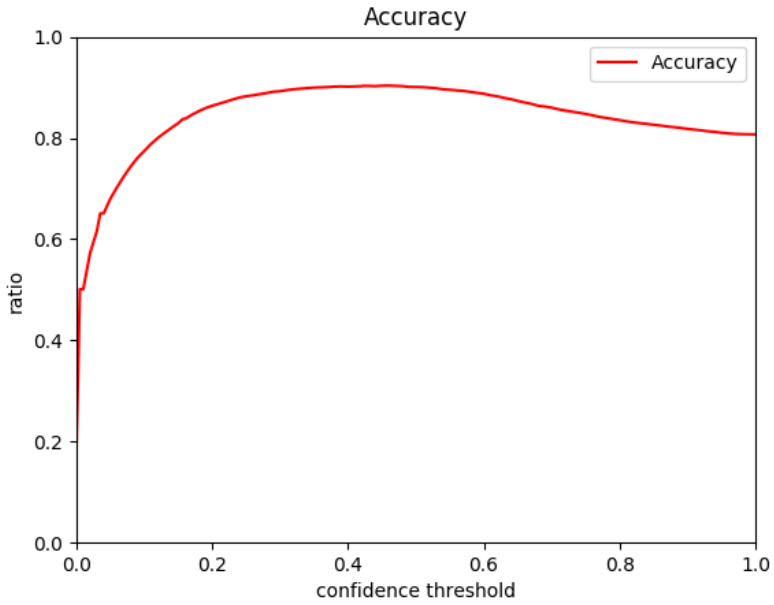


Figure D.23: Hypoxemic validation accuracy.

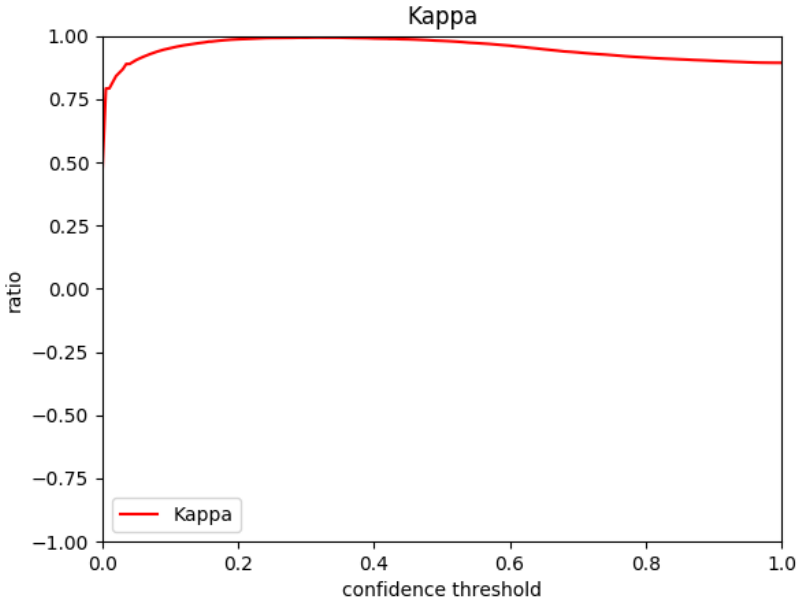


Figure D.24: Hypoxemic validation kappa.

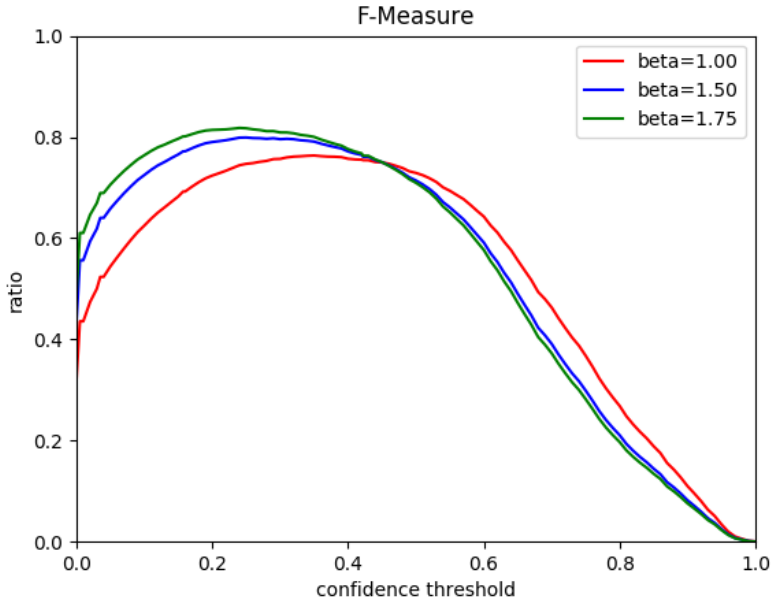
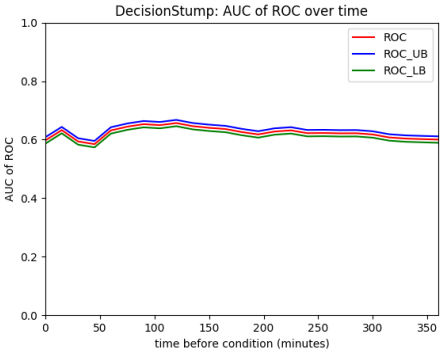
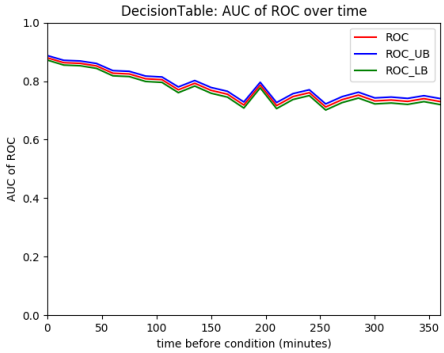


Figure D.25: Hypoxemic validation F-measures.

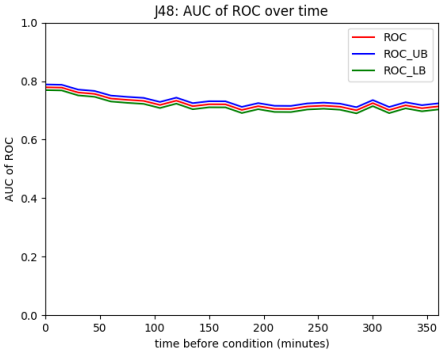
D.2.3 Machine Learning Algorithms over time



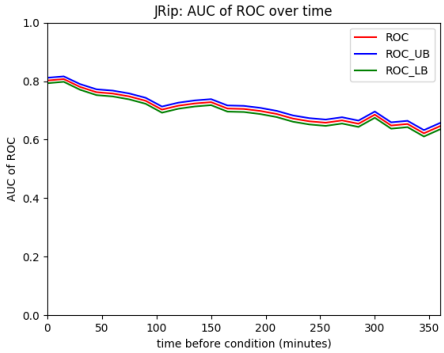
(a) Decision Stump



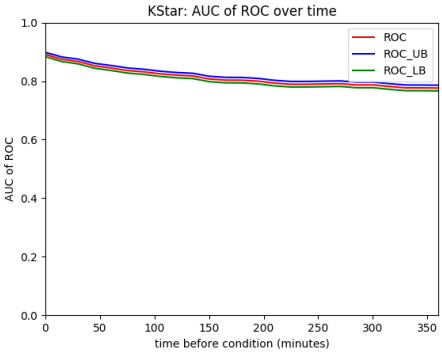
(b) Decision Table



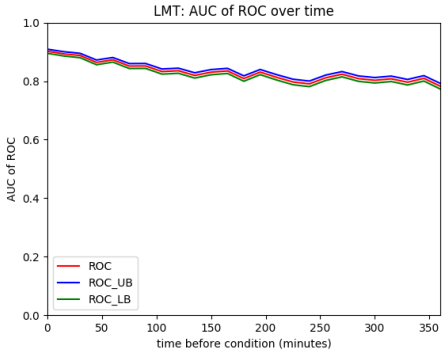
(c) J48



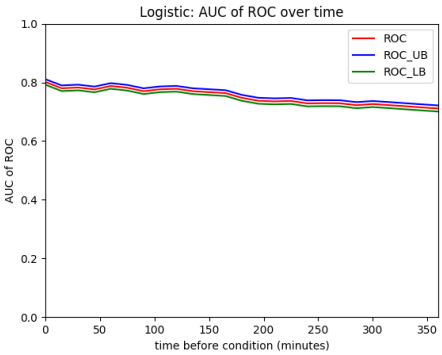
(d) JRip



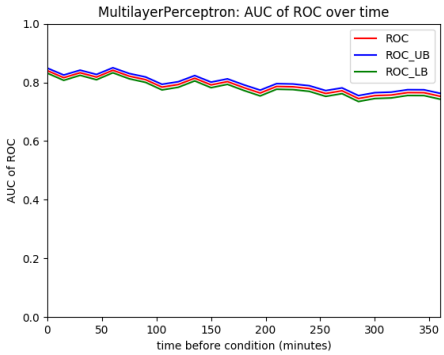
(e) KStar



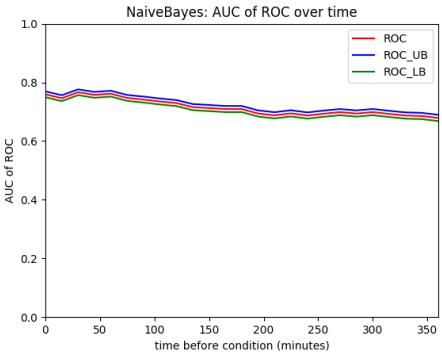
(f) LMT



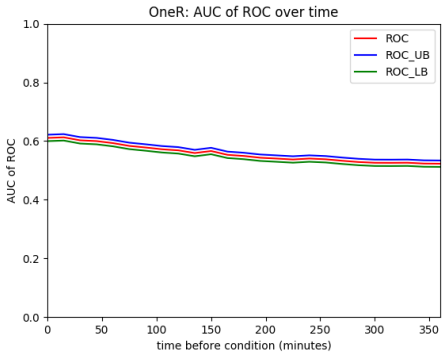
(g) Logistic Regression



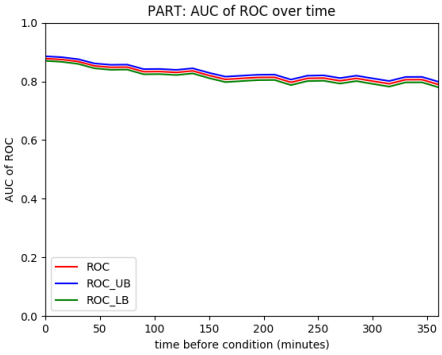
(h) Multilayer Perceptron



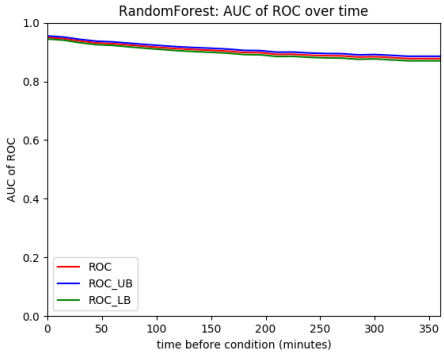
(i) Naive Bayes



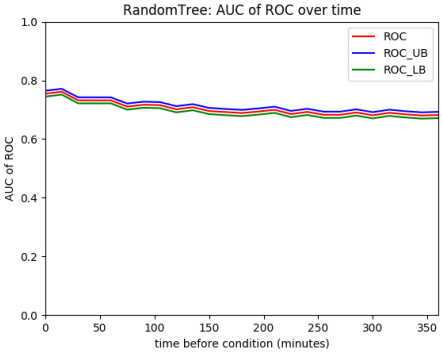
(j) One R



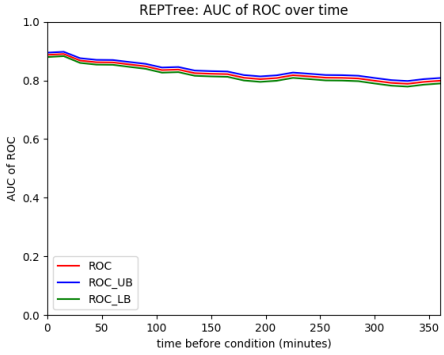
(k) PART



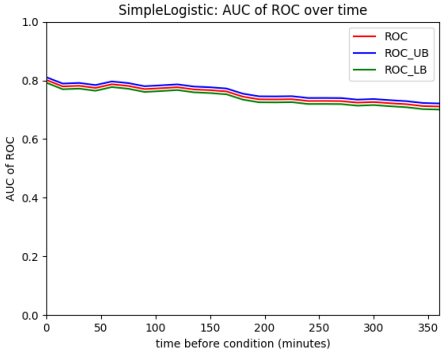
(l) Random Forest



(m) Random Tree



(n) REP Tree



(o) Simple Logistic Regression

Figure D.26: Algorithm's performance over time on the Hypoxemic selection criteria.

D.2.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Motor Response	1	5	6	1	5	6	0.5546
SpO_2	96	99	100	96	98	100	<0.0001
Admit Wt	65.875	78.75	92	63	77	88.9	<0.0001
Admit Ht	63	66	70	62	66	70	0.2461
BSA	1.736	1.95	2.123	1.739	1.972	2.085	0.8319
Carbon Dioxide	21	23	26	19	22	25	<0.0001
WBC (4-11,000)	8.2	11.8	15.8	6.4	9.5	12.7	<0.0001
Braden Score	11	13	14	12	13	16	<0.0001
GCS Total	3	9	13	3	8	15	<0.0001
NBP Mean	62.667	72.333	83.333	61	72.333	83.667	0.828
NBP [Diastolic]	45	54	65	43	51	63	<0.0001
Eye Opening	1	3	4	1	2	4	0.0543
NBP [Systolic]	96	108	126	100	110	130	<0.0001
HR Alarm [Low]	50	60	60	50	55	60	<0.0001
Creatinine	0.7	0.9	1.4	0.6	0.8	1.1	<0.0001
Hemoglobin	9	10.3	11.7	9.3	11.3	14	<0.0001
Verbal Response	1	1	4	1	1	5	<0.0001
Platelet Count	134	187	258	156	210	282	<0.0001
Hematocrit	27.4	31	35	26.7	31.6	38	<0.0001
HR Alarm [High]	110	120	120	120	120	120	<0.0001
Chloride	102	106	110	102	105	108	<0.0001
PTT	27.2	31.6	39.025	24.3	27.6	32.6	<0.0001
Magnesium	1.7	2	2.2	1.7	1.9	2.1	<0.0001
O_2 Flow (lpm)	2	3	6	2	3	4	<0.0001
BUN (6-20)	12	17	29	11	17	27	<0.0001
SpO_2 Alarm [High]	100	100	100	100	100	100	0.0005
INR(PT)	13.2	14.2	15.7	12.3	13.2	14.4	<0.0001
Potassium	3.7	4.1	4.5	3.8	4.1	4.5	<0.0001
Respiratory Rate	14	17	22	15	17	21	0.0131
SpO_2 Alarm [Low]	90	90	92	90	90	92	0.0126
Red Blood Cells	3.1275	3.55	4.01	3.06	3.58	4.18	0.0146
Sodium	136	139	141	136	139	141	0.6039
Heart Rate	112	123	130	112	123	130	0.4346

Table D.7: Hypoxemic model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Motor Response	1	5	6	1	5	6	5	6	6	6	6	6	<0.0001
SpO_2	96	99	100	97	98	100	96	98	100	95.75	98	100	<0.0001
Admit Wt	66	79	92.6	63.2	77	88.9	60.3	72	85.65	63	75.6	185	<0.0001
Admit Ht	63	66	70	62	66	70	62	66	69	63	66	165	0.0069
BSA	1.739	1.952	2.125	1.751	1.995	2.085	1.676	1.87	2.085	1.723	1.941	2.881	<0.0001
Carbon Dioxide	21	23	26	19	22	25	20	24	26	21	24	45	<0.0001
WBC (4-11,000)	8.4	11.9	16.1	6.3	9.4	12.4	7.9	11.1	15.2	7.1	10.2	56.3	<0.0001
Braden Score	11	13	14	12	13	16	12	13	15	13	15	23	<0.0001
GCS Total	3	8	11	3	8	15	8	12	15	14	15	15	<0.0001
NBP Mean	62.333	72.167	83	61	73	83.667	61	71.333	83	64.25	74.167	128.667	0.0077
NBP [Diastolic]	45	54	65	43	51	63	43	53	64	47	55	109	<0.0001
Eye Opening	1	2	4	1	2	4	2	4	4	4	4	4	<0.0001
NBP [Systolic]	95	107	126	101	110	130	95	110	129	98	114	206	<0.0001
HR Alarm [Low]	50	60	60	50	55	60	50	55	60	50	55	150	<0.0001
Creatinine	0.7	0.9	1.4	0.6	0.8	1.1	0.7	1	1.6	0.675	0.9	8.6	<0.0001
Hemoglobin	9	10.3	11.7	9.4	11.4	14.6	9.2	10.3	11.7	9.1	10.5	18.2	<0.0001
Verbal Response	1	1	1	1	1	5	1	3	5	5	5	5	<0.0001
Platelet Count	133.75	186	257	157	213	283	143	201	274	139.75	202.5	736	<0.0001
Hematocrit	27.4	31	35	26.5	31.7	38.6	27.9	30.9	35.1	27.775	30.8	53.8	<0.0001
HR Alarm [High]	110	120	120	120	120	120	120	120	120	120	120	180	<0.0001
Chloride	102	107	110	102	105	108	101	105	110	101	104	119	<0.0001
PTT	27.275	31.7	39.5	24	27.3	31.7	26.7	30.9	38	26.575	30	150	<0.0001
Magnesium	1.7	2	2.3	1.7	1.9	2.1	1.775	2	2.2	1.7	1.9	4.2	<0.0001
O_2 Flow (lpm)	2	4	10	2	3	4	2	4	10	2	3	15	<0.0001
BUN (6-20)	12	18	29	11	16	26	12	19	34	11	16	132	<0.0001
SpO_2 Alarm [High]	100	100	100	100	100	100	100	100	100	100	100	100	<0.0001
INR(PT)	13.2	14.2	15.8	12.2	13.1	14.2	13.075	14	15.6	12.9	13.75	31.6	<0.0001
Potassium	3.7	4.1	4.5	3.8	4.1	4.5	3.7	4.1	4.425	3.7	4	6.5	<0.0001
Respiratory Rate	14	17	22	15	17	20	15	19	23	14	18	35	<0.0001
SpO_2 Alarm [Low]	90	92	92	90	90	92	90	90	92	90	90	94	<0.0001
Red Blood Cells	3.12	3.55	4.01	3.06	3.6	4.22	3.08	3.48	3.9525	3.15	3.49	21	<0.0001
Sodium	136	139	141	136	139	141	137	139	142	136	139	150	<0.0001
Heart Rate	112	123	130	112	123	130	112	123	130	110	123	148	0.2822

Table D.8: Hypoxemic model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition_Positive	Condition_Negative	P-value
N	18074	3484	14590	
Age, median	62	64	61	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (63.00, 70.00)	0.0168
Weight, median (Q1,Q3)	78.00 (65.20, 92.00)	80.65 (68.30, 95.15)	75.00 (63.00, 88.80)	<0.0001
BMI, median (Q1,Q3)	28.06 (24.56, 32.55)	28.93 (25.13, 33.37)	27.11 (23.88, 31.11)	<0.0001
Male	10151 (0.56)	2140 (0.61)	8011 (0.55)	<0.0001
In-Hospital Mortality	2796 (0.15)	804 (0.23)	1992 (0.14)	<0.0001
30 Day Mortality	3250 (0.18)	877 (0.25)	2373 (0.16)	<0.0001
ICU LOS, median(Q1,Q3)	2.11 (1.13, 4.56)	4.13 (2.10, 8.92)	1.90 (1.04, 3.69)	<0.0001
ICU				
CCU	2285 (0.13)	395 (0.11)	1890 (0.13)	0.0159
CSRU	2716 (0.15)	1143 (0.33)	1573 (0.11)	<0.0001
MICU	6250 (0.35)	973 (0.28)	5277 (0.36)	<0.0001
SICU	2654 (0.15)	535 (0.15)	2119 (0.15)	0.2494
TSICU	1923 (0.11)	438 (0.13)	1485 (0.10)	<0.0001
ethnicity				
Asian	589 (0.03)	63 (0.02)	526 (0.04)	<0.0001
Black	1708 (0.09)	221 (0.06)	1487 (0.10)	<0.0001
Hispanic	646 (0.04)	89 (0.03)	557 (0.04)	0.0004
White	12708 (0.70)	2466 (0.71)	10242 (0.70)	0.7128
insurance				
Government	515 (0.03)	75 (0.02)	440 (0.03)	0.0067
Medicaid	1710 (0.09)	277 (0.08)	1433 (0.10)	0.0013
Medicare	8837 (0.49)	1815 (0.52)	7022 (0.48)	0.0026
Private	6805 (0.38)	1277 (0.37)	5528 (0.38)	0.2855
Self Pay	207 (0.01)	40 (0.01)	167 (0.01)	0.9862

Table D.9: Hypoxemic patients' Demographics of validation set by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	True_Positive	True_Negative	False_Positive	False_Negative	P-value
N	18074	2365	13911	679	1119	
Age, median	62.0	64.0	60.0	74.0	64.0	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	66.00 (63.00, 69.00)	67.00 (64.00, 70.00)	0.0009
Weight, median (Q1,Q3)	78.00 (65.20, 92.00)	82.00 (70.00, 97.00)	75.00 (63.40, 89.05)	71.70 (61.40, 85.00)	77.00 (65.00, 90.00)	<0.0001
Male	10151 (0.56)	1493 (0.63)	7681 (0.55)	330 (0.49)	647 (0.58)	<0.0001
In-Hospital Mortality	2796 (0.15)	517 (0.22)	1725 (0.12)	267 (0.39)	287 (0.26)	<0.0001
30 Day Mortality	3250 (0.18)	556 (0.24)	2082 (0.15)	291 (0.43)	321 (0.29)	<0.0001
ICU LOS, median(Q1,Q3)	2.11 (1.13, 4.56)	4.12 (2.08, 8.94)	1.89 (1.04, 3.64)	2.19 (1.15, 4.25)	4.17 (2.15, 8.87)	<0.0001
ICU						
CCU	2285 (0.13)	252 (0.11)	1799 (0.13)	91 (0.13)	143 (0.13)	0.0345
CSRU	2716 (0.15)	901 (0.38)	1473 (0.11)	100 (0.15)	242 (0.22)	<0.0001
MICU	6250 (0.35)	598 (0.25)	5015 (0.36)	262 (0.39)	375 (0.34)	<0.0001
SICU	2654 (0.15)	333 (0.14)	1991 (0.14)	128 (0.19)	202 (0.18)	0.0003
TSICU	1923 (0.11)	281 (0.12)	1387 (0.10)	98 (0.14)	157 (0.14)	<0.0001
ethnicity						
Asian	589 (0.03)	50 (0.02)	507 (0.04)	19 (0.03)	13 (0.01)	<0.0001
Black	1708 (0.09)	150 (0.06)	1431 (0.10)	56 (0.08)	71 (0.06)	<0.0001
Hispanic	646 (0.04)	62 (0.03)	540 (0.04)	17 (0.03)	27 (0.02)	0.0011
White	12708 (0.70)	1663 (0.70)	9772 (0.70)	470 (0.69)	803 (0.72)	0.9281
insurance						
Government	515 (0.03)	54 (0.02)	430 (0.03)	10 (0.01)	21 (0.02)	0.0033
Medicaid	1710 (0.09)	193 (0.08)	1388 (0.10)	45 (0.07)	84 (0.08)	0.0004
Medicare	8837 (0.49)	1226 (0.52)	6579 (0.47)	443 (0.65)	589 (0.53)	<0.0001
Private	6805 (0.38)	871 (0.37)	5362 (0.39)	166 (0.24)	406 (0.36)	<0.0001
Self Pay	207 (0.01)	21 (0.01)	152 (0.01)	15 (0.02)	19 (0.02)	0.0098

Table D.10: Hypoxemic patients' Demographics of validation set in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	Condition_Positive	Condition_Negative	P-value
N	18074	3484	14590	
Comorbidity				
Congestive heart failure	4661 (0.26)	1171 (0.34)	3490 (0.24)	<0.0001
Cardiac arrhythmias	5651 (0.31)	1306 (0.37)	4345 (0.30)	<0.0001
Valvular disease	2460 (0.14)	683 (0.20)	1777 (0.12)	<0.0001
Pulmonary circulation disorders	865 (0.05)	147 (0.04)	718 (0.05)	0.0889
Peripheral vascular disorders	1878 (0.10)	436 (0.13)	1442 (0.10)	<0.0001
Hypertension, uncomplicated	6537 (0.36)	1443 (0.41)	5094 (0.35)	<0.0001
Hypertension, complicated	2123 (0.12)	288 (0.08)	1835 (0.13)	<0.0001
Paralysis	419 (0.02)	62 (0.02)	357 (0.02)	0.0201
Other neurological disorders	1083 (0.06)	172 (0.05)	911 (0.06)	0.0046
Chronic pulmonary disease	3663 (0.20)	778 (0.22)	2885 (0.20)	0.0026
Diabetes, uncomplicated	3330 (0.18)	743 (0.21)	2587 (0.18)	<0.0001
Diabetes, complicated	1175 (0.07)	221 (0.06)	954 (0.07)	0.6844
Hypothyroidism	1723 (0.10)	245 (0.07)	1478 (0.10)	<0.0001
Renal failure	2528 (0.14)	372 (0.11)	2156 (0.15)	<0.0001
Liver disease	1713 (0.09)	367 (0.11)	1346 (0.09)	0.0242
Peptic ulcer disease excluding bleeding	113 (0.01)	13 (0.00)	100 (0.01)	0.0362
AIDS/HIV	163 (0.01)	42 (0.01)	121 (0.01)	0.0357
Lymphoma	240 (0.01)	45 (0.01)	195 (0.01)	0.8362
Metastatic cancer	155 (0.01)	25 (0.01)	130 (0.01)	0.3206
Solid tumor without metastasis	1264 (0.07)	270 (0.08)	994 (0.07)	0.0603
Rheumatoid arthritis/collagen vascular diseases	574 (0.03)	95 (0.03)	479 (0.03)	0.0978
Coagulopathy	1684 (0.09)	355 (0.10)	1329 (0.09)	0.0605
Weight loss	741 (0.04)	138 (0.04)	603 (0.04)	0.6524
Fluid and electrolyte disorders	4673 (0.26)	787 (0.23)	3886 (0.27)	<0.0001
Blood loss anemia	344 (0.02)	75 (0.02)	269 (0.02)	0.2350
Deficiency anemia	476 (0.03)	65 (0.02)	411 (0.03)	0.0019
Alcohol abuse	750 (0.04)	136 (0.04)	614 (0.04)	0.4275
Drug abuse	659 (0.04)	126 (0.04)	533 (0.04)	0.9189
Psychoses	275 (0.02)	58 (0.02)	217 (0.01)	0.4456
Depression	1490 (0.08)	137 (0.04)	1353 (0.09)	<0.0001
Trauma	2081 (0.12)	452 (0.13)	1629 (0.11)	0.0047

Table D.11: Hypoxemic patients' Comorbidities of validation set by condition positive or negative. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	True$_{Positive}$	True$_{Negative}$	False$_{Positive}$	False$_{Negative}$	P-value
N	18074	2365	13911	679	1119	
Comorbidity						
Congestive heart failure	4661 (0.26)	776 (0.33)	3276 (0.24)	214 (0.32)	395 (0.35)	<0.0001
Cardiac arrhythmias	5651 (0.31)	904 (0.38)	4082 (0.29)	263 (0.39)	402 (0.36)	<0.0001
Valvular disease	2460 (0.14)	494 (0.21)	1669 (0.12)	108 (0.16)	189 (0.17)	<0.0001
Pulmonary circulation disorders	865 (0.05)	98 (0.04)	693 (0.05)	25 (0.04)	49 (0.04)	0.1532
Peripheral vascular disorders	1878 (0.10)	313 (0.13)	1368 (0.10)	74 (0.11)	123 (0.11)	<0.0001
Hypertension, uncomplicated	6537 (0.36)	1025 (0.43)	4831 (0.35)	263 (0.39)	418 (0.37)	<0.0001
Hypertension, complicated	2123 (0.12)	188 (0.08)	1763 (0.13)	72 (0.11)	100 (0.09)	<0.0001
Paralysis	419 (0.02)	42 (0.02)	345 (0.02)	12 (0.02)	20 (0.02)	0.0778
Other neurological disorders	1083 (0.06)	112 (0.05)	851 (0.06)	60 (0.09)	60 (0.05)	0.0009
Chronic pulmonary disease	3663 (0.20)	502 (0.21)	2750 (0.20)	135 (0.20)	276 (0.25)	0.0037
Diabetes, uncomplicated	3330 (0.18)	525 (0.22)	2462 (0.18)	125 (0.18)	218 (0.19)	<0.0001
Diabetes, complicated	1175 (0.07)	146 (0.06)	920 (0.07)	34 (0.05)	75 (0.07)	0.3823
Hypothyroidism	1723 (0.10)	153 (0.06)	1415 (0.10)	63 (0.09)	92 (0.08)	<0.0001
Renal failure	2528 (0.14)	239 (0.10)	2071 (0.15)	85 (0.13)	133 (0.12)	<0.0001
Liver disease	1713 (0.09)	250 (0.11)	1296 (0.09)	50 (0.07)	117 (0.10)	0.0528
Peptic ulcer disease excluding bleeding	113 (0.01)	10 (0.00)	98 (0.01)	2 (0.00)	3 (0.00)	0.0930
AIDS/HIV	163 (0.01)	25 (0.01)	115 (0.01)	6 (0.01)	17 (0.02)	0.1007
Lymphoma	240 (0.01)	27 (0.01)	186 (0.01)	9 (0.01)	18 (0.02)	0.7314
Metastatic cancer	155 (0.01)	12 (0.01)	120 (0.01)	10 (0.01)	13 (0.01)	0.0553
Solid tumor without metastasis	1264 (0.07)	174 (0.07)	939 (0.07)	55 (0.08)	96 (0.09)	0.0772
Rheumatoid arthritis/collagen vascular diseases	574 (0.03)	63 (0.03)	460 (0.03)	19 (0.03)	32 (0.03)	0.3394
Coagulopathy	1684 (0.09)	256 (0.11)	1266 (0.09)	63 (0.09)	99 (0.09)	0.0809
Weight loss	741 (0.04)	89 (0.04)	579 (0.04)	24 (0.04)	49 (0.04)	0.6760
Fluid and electrolyte disorders	4673 (0.26)	501 (0.21)	3718 (0.27)	168 (0.25)	286 (0.26)	<0.0001
Blood loss anemia	344 (0.02)	46 (0.02)	259 (0.02)	10 (0.01)	29 (0.03)	0.3088
Deficiency anemia	476 (0.03)	42 (0.02)	396 (0.03)	15 (0.02)	23 (0.02)	0.0123
Alcohol abuse	750 (0.04)	97 (0.04)	597 (0.04)	17 (0.03)	39 (0.03)	0.0974
Drug abuse	659 (0.04)	74 (0.03)	521 (0.04)	12 (0.02)	52 (0.05)	0.0083
Psychoses	275 (0.02)	36 (0.02)	210 (0.02)	7 (0.01)	22 (0.02)	0.4680
Depression	1490 (0.08)	93 (0.04)	1319 (0.09)	34 (0.05)	44 (0.04)	<0.0001
Trauma	2081 (0.12)	294 (0.12)	1534 (0.11)	95 (0.14)	158 (0.14)	0.0020

Table D.12: Hypoxemic patients' Comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by Pearson Chi-Squared test.

D.3 DIC from ARDS patients

D.3.1 Model Performance

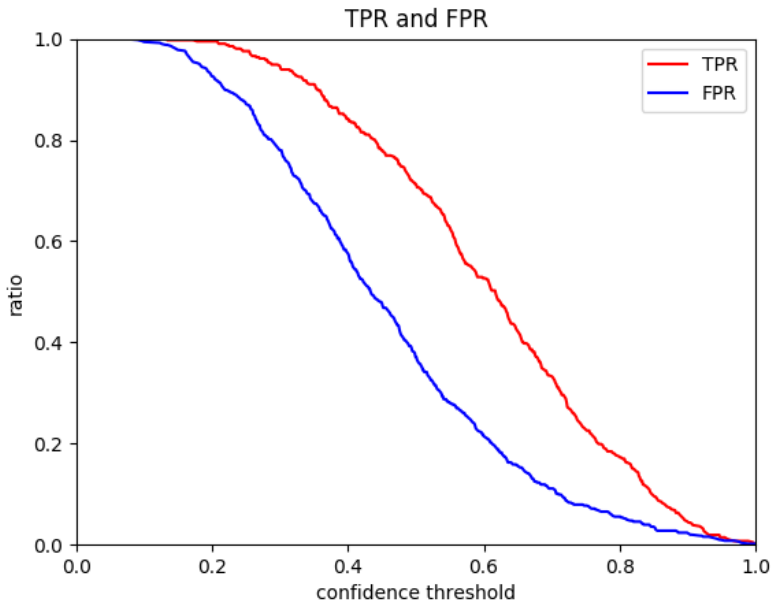


Figure D.27: DIC from ARDS entire set true and false positive rates.

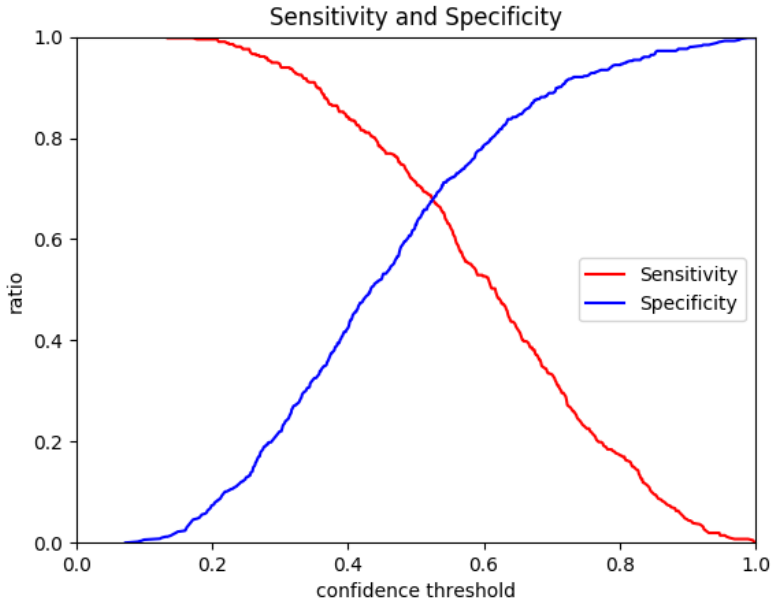


Figure D.28: DIC from ARDS entire set sensitivity and specificity.

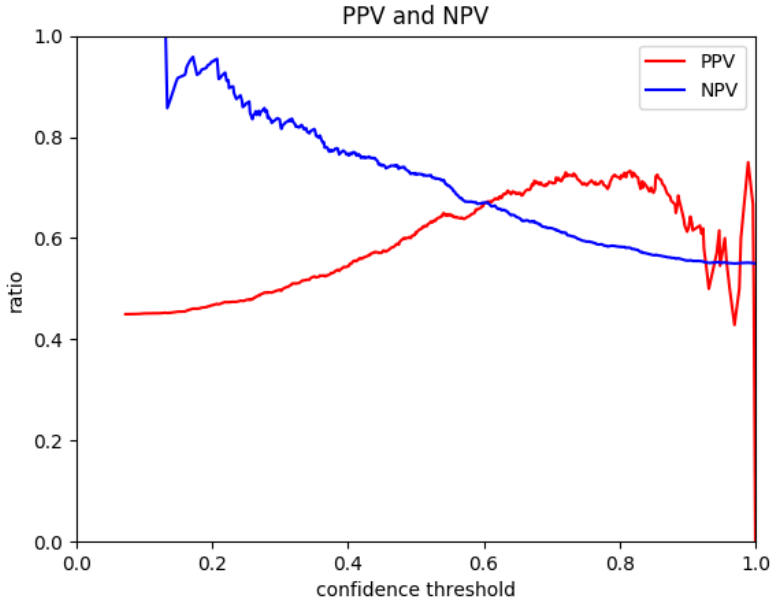


Figure D.29: DIC from ARDS entire set positive and negative predictive values.

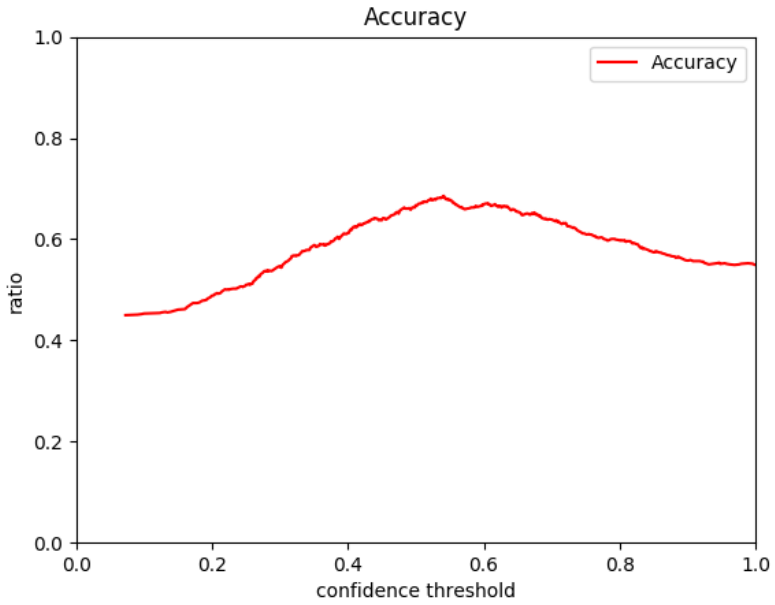


Figure D.30: DIC from ARDS entire set accuracy.

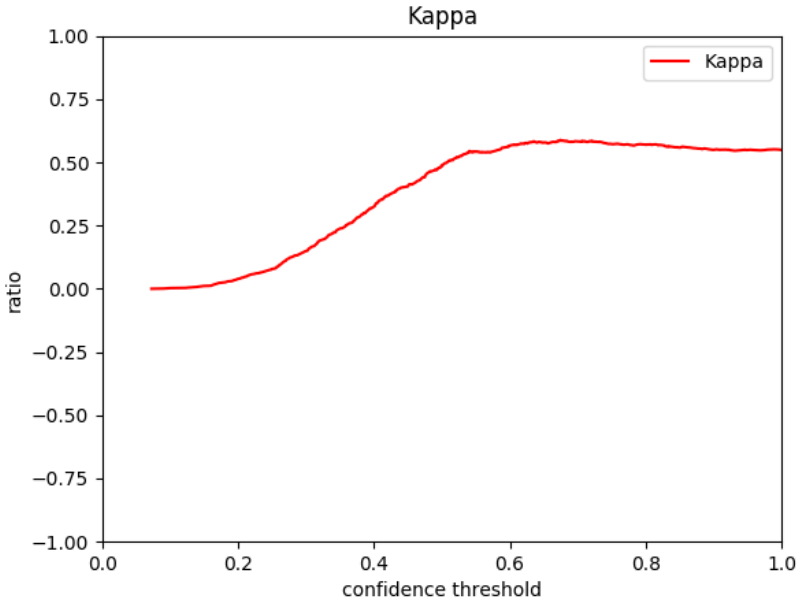


Figure D.31: DIC from ARDS entire set kappa.

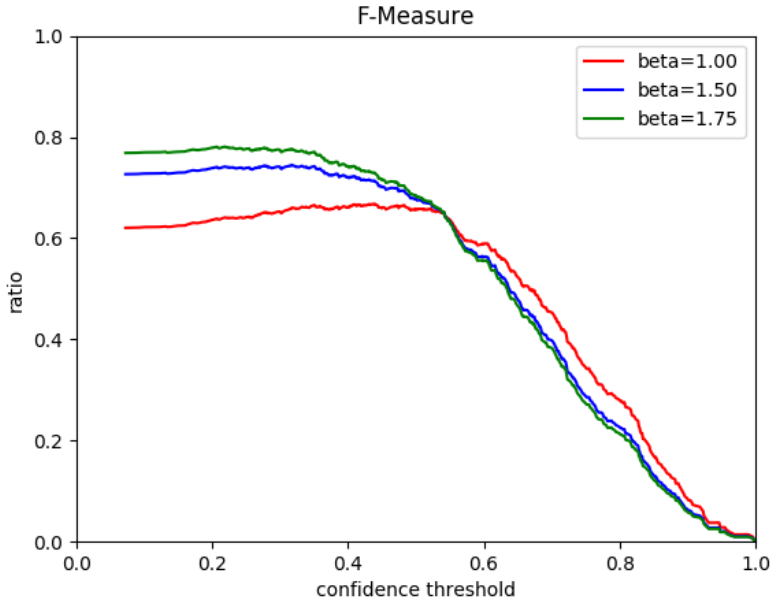


Figure D.32: DIC from ARDS entire set F-measures.

D.3.2 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Chloride	103	106	110	101	104	109	0.0002
Oxygen	50	55	100	47.75	55	100	0.1263
Tidal Volume	550	649	700	550	644	700	0.8181
Bicarbonate	20	23	26	21	24	28	<0.0001
Sodium	136	139	141	136	139	142	0.4576
INR(PT)	13.4	14.4	15.9	12.8	13.5	14.5	<0.0001
Gastric Meds	40	40	40	40	40	40	0.0946
Chloride, Whole Blood	103	106	110	100	105	109	<0.0001
Creatinine, Urine	118	118	118	118	118	118	0.4307
Alkaline Phosphatase	63	77	118	65	81	107	0.8886
Platelet Count	137	190	263.25	183.75	235.5	320	<0.0001
Cholesterol, LDL	117	117	117	117	117	117	0.1526

Table D.13: DIC from ARDS model variable composition of the entire data set. P-values calculated by Kruskal-Wallis test.

Variable N	Total 1419	Condition Positive 915	Condition Negative 504	P-value
Age, median	65	66	63	*
Height, median (Q1,Q3)	67.00 (63.75, 70.00)	67.00 (63.50, 70.00)	67.00 (64.00, 70.00)	0.8720
Weight, median (Q1,Q3)	80.00 (66.62, 94.48)	79.50 (66.00, 92.00)	80.00 (68.20, 97.00)	0.0275
BMI, median (Q1,Q3)	28.47 (24.59, 33.56)	28.20 (24.69, 32.88)	29.15 (24.34, 35.62)	0.1268
Male	799 (0.56)	534 (0.58)	265 (0.53)	0.1649
In-Hospital Mortality	425 (0.30)	329 (0.36)	96 (0.19)	<0.0001
30 Day Mortality	457 (0.32)	351 (0.38)	106 (0.21)	<0.0001
ICU LOS, median(Q1,Q3)	10.10 (4.91, 19.16)	11.66 (6.09, 21.48)	7.16 (3.23, 15.08)	<0.0001
ICU				
CCU	131 (0.09)	87 (0.10)	44 (0.09)	0.6443
CSRU	319 (0.22)	225 (0.25)	94 (0.19)	0.0239
MICU	377 (0.27)	221 (0.24)	156 (0.31)	0.0174
SICU	296 (0.21)	192 (0.21)	104 (0.21)	0.8905
TSICU	296 (0.21)	190 (0.21)	106 (0.21)	0.9162
ethnicity				
Asian	17 (0.01)	14 (0.02)	3 (0.01)	0.1236
Black	93 (0.07)	57 (0.06)	36 (0.07)	0.5201
Hispanic	37 (0.03)	21 (0.02)	16 (0.03)	0.3261
White	1040 (0.73)	667 (0.73)	373 (0.74)	0.8149
insurance				
Government	27 (0.02)	18 (0.02)	9 (0.02)	0.8125
Medicaid	111 (0.08)	66 (0.07)	45 (0.09)	0.2688
Medicare	768 (0.54)	500 (0.55)	268 (0.53)	0.7186
Private	503 (0.35)	325 (0.36)	178 (0.35)	0.9513
Self Pay	10 (0.01)	6 (0.01)	4 (0.01)	0.7671

Table D.14: Demographics of entire set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	Condition Positive	condition Negative	P-value
N	3200	1332	1868	
Comorbidity				
Congestive heart failure	1075 (0.34)	520 (0.39)	555 (0.30)	<0.0001
Cardiac arrhythmias	1275 (0.40)	587 (0.44)	688 (0.37)	0.0014
Valvular disease	494 (0.15)	231 (0.17)	263 (0.14)	0.0206
Pulmonary circulation disorders	173 (0.05)	62 (0.05)	111 (0.06)	0.1226
Peripheral vascular disorders	519 (0.16)	227 (0.17)	292 (0.16)	0.3288
Hypertension, uncomplicated	1192 (0.37)	404 (0.30)	788 (0.42)	<0.0001
Hypertension, complicated	363 (0.11)	127 (0.10)	236 (0.13)	0.0103
Paralysis	106 (0.03)	26 (0.02)	80 (0.04)	0.0004
Other neurological disorders	180 (0.06)	70 (0.05)	110 (0.06)	0.4565
Chronic pulmonary disease	795 (0.25)	261 (0.20)	534 (0.29)	<0.0001
Diabetes, uncomplicated	600 (0.19)	222 (0.17)	378 (0.20)	0.0215
Diabetes, complicated	227 (0.07)	77 (0.06)	150 (0.08)	0.0185
Hypothyroidism	290 (0.09)	94 (0.07)	196 (0.10)	0.0015
Renal failure	437 (0.14)	156 (0.12)	281 (0.15)	0.0120
Liver disease	411 (0.13)	225 (0.17)	186 (0.10)	<0.0001
Peptic ulcer disease excluding bleeding	19 (0.01)	8 (0.01)	11 (0.01)	0.9661
AIDS/HIV	35 (0.01)	17 (0.01)	18 (0.01)	0.4045
Lymphoma	47 (0.01)	26 (0.02)	21 (0.01)	0.0568
Metastatic cancer	27 (0.01)	8 (0.01)	19 (0.01)	0.2061
Solid tumor without metastasis	301 (0.09)	98 (0.07)	203 (0.11)	0.0014
Rheumatoid arthritis/collagen vascular diseases	87 (0.03)	34 (0.03)	53 (0.03)	0.6302
Coagulopathy	539 (0.17)	306 (0.23)	233 (0.12)	<0.0001
Weight loss	248 (0.08)	111 (0.08)	137 (0.07)	0.3169
Fluid and electrolyte disorders	1083 (0.34)	375 (0.28)	708 (0.38)	<0.0001
Blood loss anemia	76 (0.02)	35 (0.03)	41 (0.02)	0.4336
Deficiency anemia	77 (0.02)	28 (0.02)	49 (0.03)	0.3490
Alcohol abuse	154 (0.05)	76 (0.06)	78 (0.04)	0.0518
Drug abuse	127 (0.04)	37 (0.03)	90 (0.05)	0.0043
Psychoses	39 (0.01)	19 (0.01)	20 (0.01)	0.3689
Depression	238 (0.07)	42 (0.03)	196 (0.10)	<0.0001
Trauma	685 (0.21)	286 (0.21)	399 (0.21)	0.9463

Table D.15: Comorbidities of entire set patients by condition positive or negative. P-values for patient counts calculated by Pearson's chi-squared test.

D.4 DIC from ARDS patients

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Bands	0	23	50	0	11	50	0.2184
Unable to assess cognitive / perceptual	1	97.1	97.5	1	95.9	97.5	0.9895
O2 Saturation Alarm - Low	48	55.5	90	48	58	90	0.2743
Specific Gravity	1.018	98	107	1.017	90	106	0.0983
Alkaline Phosphatase	1.02	1.03	90	1.02	1.04	91	0.9070
Respiratory Rate	18	30	54	19	29	54	0.9658
Tidal Volume (Set)	14	21	550	14	22	500	0.7109
Differential-Atyps	0	310	310	0	41	310	0.9845
SaO2	96.875	98	99	97	98	99	0.0644
O2 Flow (additional cannula)	1	1	4	1	1	4	0.9192
Sodium (whole blood)	137	266	267	139	250	267	0.6780
Differential-Lymphs	7.7	52.2	73.125	7.7	45.5	72.5	0.6952
Iron Binding Capacity, Total	0	185	290	0	185	277	0.5922
Heart rate Alarm - High	90	100	135	90	100	135	0.8685
High Resp. Rate	3.54	6.285	20	3.54	8	20	0.8220
Baseline pain level	2	2	2	2	2	2	0.4101
Mixed Venous O2% Sat	5	5	68	5	12	68	0.4325
Nitroglycerin	2.243	5	5	2.243	4	5	0.1761
Metoprolol	5	91.5	104.25	5	88	104	0.4248
NBP [Systolic]	3.611	7.798	113	3.611	9.5676	111	0.6500
Cholesterol Ratio (Total/HDL)	1	1	3.2	1	1	3.2	0.3762

Table D.16: DIC from ARDS model variable composition of the entire data set. P-values calculated by Kruskal-Wallis test.

Variable N	Total 7827	Condition Positive 272	Condition Negative 7555	P-value
Age, median	68	66	68	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.9107
Weight, median (Q1,Q3)	76.60 (65.00, 90.00)	77.30 (66.00, 91.08)	76.60 (65.00, 90.00)	0.3945
BMI, median (Q1,Q3)	27.68 (24.22, 32.16)	28.22 (24.56, 32.71)	27.68 (24.21, 32.13)	0.4760
Male	4589 (0.59)	168 (0.62)	4421 (0.59)	0.4920
In-Hospital Mortality	2063 (0.26)	86 (0.32)	1977 (0.26)	0.0854
30 Day Mortality	2363 (0.30)	93 (0.34)	2270 (0.30)	0.2216
ICU LOS, median(Q1,Q3)	3.25 (1.86, 6.92)	11.28 (6.05, 20.96)	3.16 (1.83, 6.41)	<0.0001
ICU				
CCU	1061 (0.14)	28 (0.10)	1033 (0.14)	0.1370
CSRU	1557 (0.20)	62 (0.23)	1495 (0.20)	0.2748
MICU	3078 (0.39)	66 (0.24)	3012 (0.40)	<0.0001
SICU	1169 (0.15)	50 (0.18)	1119 (0.15)	0.1343
TSICU	921 (0.12)	66 (0.24)	855 (0.11)	<0.0001
ethnicity				
Asian	165 (0.02)	5 (0.02)	160 (0.02)	0.7550
Black	712 (0.09)	14 (0.05)	698 (0.09)	0.0279
Hispanic	240 (0.03)	8 (0.03)	232 (0.03)	0.9045
White	5633 (0.72)	207 (0.76)	5426 (0.72)	0.4133
insurance				
Government	185 (0.02)	9 (0.03)	176 (0.02)	0.3020
Medicaid	643 (0.08)	20 (0.07)	623 (0.08)	0.6136
Medicare	4681 (0.60)	148 (0.54)	4533 (0.60)	0.2417
Private	2240 (0.29)	95 (0.35)	2145 (0.28)	0.0478
Self Pay	78 (0.01)	0 (0.00)	78 (0.01)	0.0938

Table D.17: Demographics of entire set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	7827	272	7555	
Comorbidity				
Congestive heart failure	2949 (0.38)	104 (0.38)	2845 (0.38)	0.8787
Cardiac arrhythmias	3333 (0.43)	116 (0.43)	3217 (0.43)	0.9869
Valvular disease	1506 (0.19)	38 (0.14)	1468 (0.19)	0.0437
Pulmonary circulation disorders	454 (0.06)	10 (0.04)	444 (0.06)	0.1388
Peripheral vascular disorders	1029 (0.13)	43 (0.16)	986 (0.13)	0.2178
Hypertension, uncomplicated	2861 (0.37)	68 (0.25)	2793 (0.37)	0.0013
Hypertension, complicated	1057 (0.14)	20 (0.07)	1037 (0.14)	0.0050
Paralysis	168 (0.02)	9 (0.03)	159 (0.02)	0.1829
Other neurological disorders	430 (0.05)	16 (0.06)	414 (0.05)	0.7808
Chronic pulmonary disease	1706 (0.22)	54 (0.20)	1652 (0.22)	0.4847
Diabetes, uncomplicated	1592 (0.20)	46 (0.17)	1546 (0.20)	0.2020
Diabetes, complicated	585 (0.07)	15 (0.06)	570 (0.08)	0.2289
Hypothyroidism	757 (0.10)	16 (0.06)	741 (0.10)	0.0408
Renal failure	1330 (0.17)	26 (0.10)	1304 (0.17)	0.0025
Liver disease	1168 (0.15)	34 (0.12)	1134 (0.15)	0.2924
Peptic ulcer disease excluding bleeding	53 (0.01)	1 (0.00)	52 (0.01)	0.5278
AIDS/HIV	111 (0.01)	3 (0.01)	108 (0.01)	0.6568
Lymphoma	151 (0.02)	5 (0.02)	146 (0.02)	0.9124
Metastatic cancer	90 (0.01)	2 (0.01)	88 (0.01)	0.5163
Solid tumor without metastasis	622 (0.08)	22 (0.08)	600 (0.08)	0.9329
Rheumatoid arthritis/collagen vascular diseases	294 (0.04)	7 (0.03)	287 (0.04)	0.3057
Coagulopathy	1412 (0.18)	45 (0.17)	1367 (0.18)	0.5543
Weight loss	457 (0.06)	18 (0.07)	439 (0.06)	0.5884
Fluid and electrolyte disorders	2613 (0.33)	64 (0.24)	2549 (0.34)	0.0042
Blood loss anemia	215 (0.03)	7 (0.03)	208 (0.03)	0.8606
Deficiency anemia	217 (0.03)	4 (0.01)	213 (0.03)	0.1894
Alcohol abuse	424 (0.05)	9 (0.03)	415 (0.05)	0.1284
Drug abuse	256 (0.03)	11 (0.04)	245 (0.03)	0.4728
Psychoses	118 (0.02)	4 (0.01)	114 (0.02)	0.9596
Depression	459 (0.06)	3 (0.01)	456 (0.06)	0.0010
Trauma	1041 (0.13)	71 (0.26)	970 (0.13)	<0.0001

Table D.18: Comorbidities of entire set patients by condition positive or negative. P-values for patient counts calculated by Pearson's chi-squared test.

Appendix E

ACUTE KIDNEY INJURY

E.1 Acute Kidney Injury Stage 1

E.1.1 Training and Testing

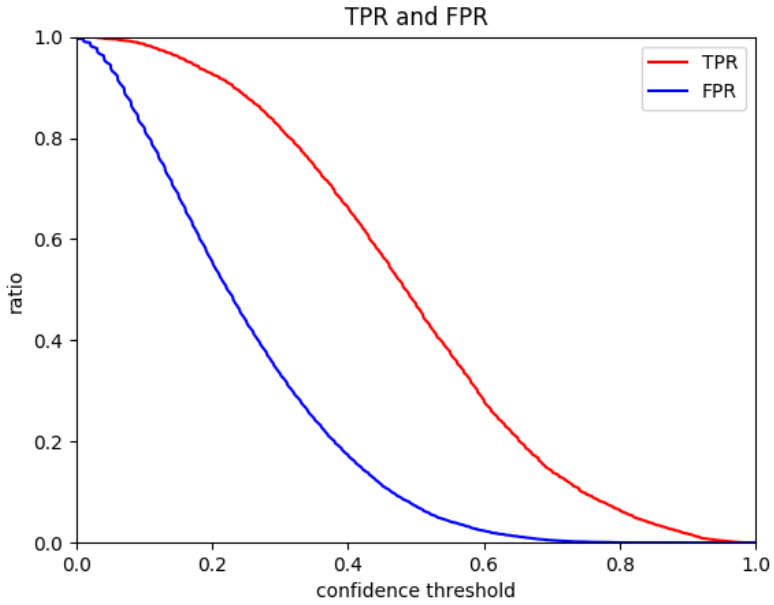


Figure E.1: AKI training/testing true positive and false positive rates.

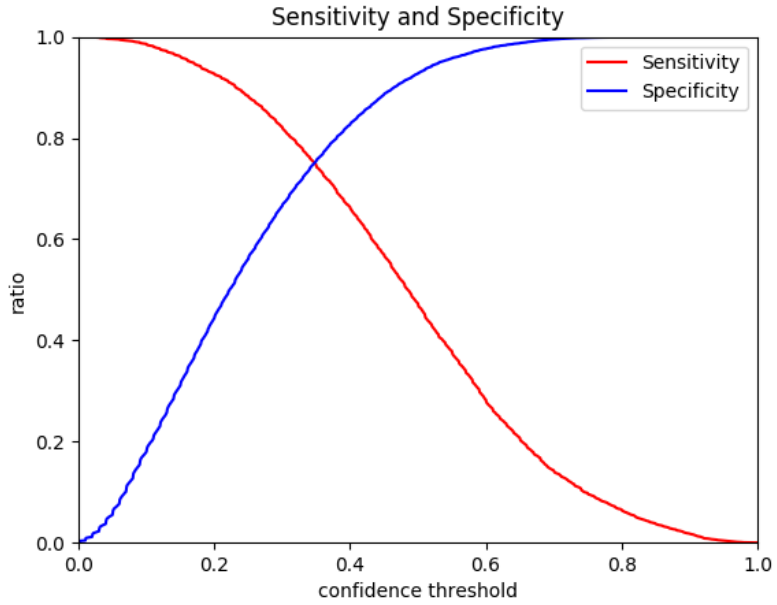


Figure E.2: AKI training/testing sensitivity and specificity.

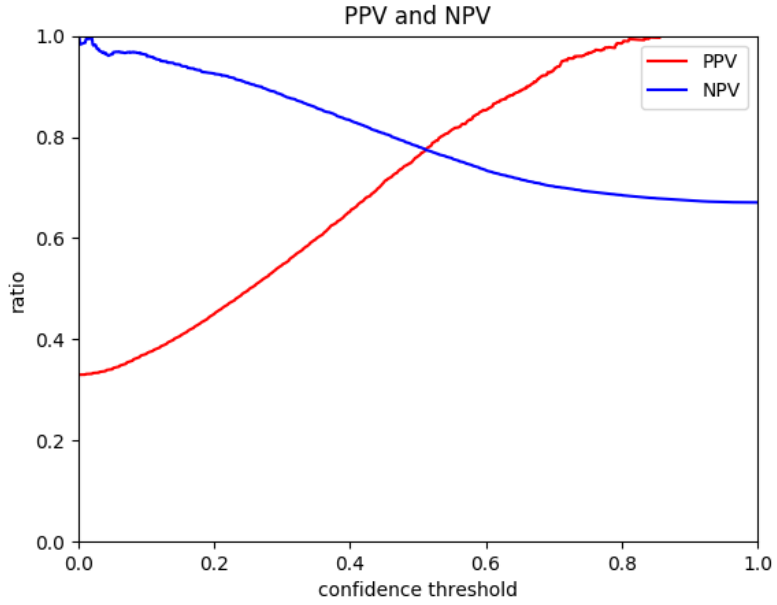


Figure E.3: AKI training/testing positive and negative predictive value.

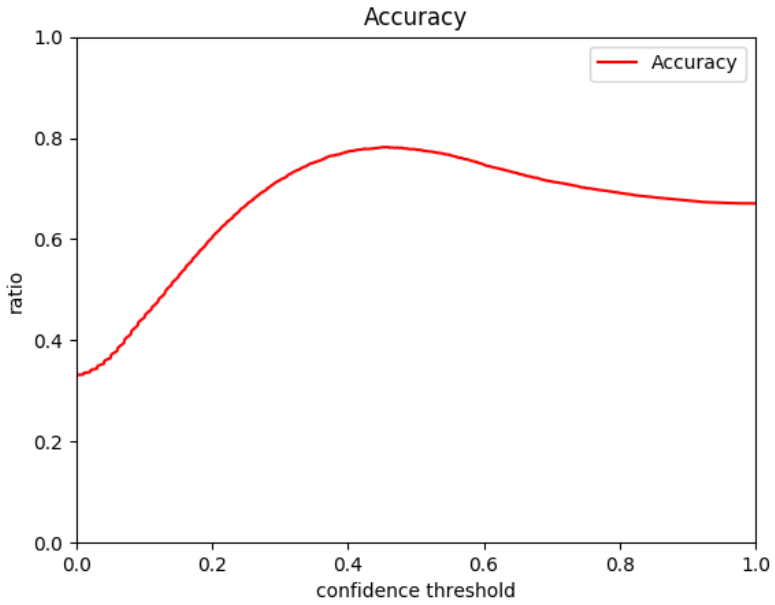


Figure E.4: AKI training/testing accuracy.

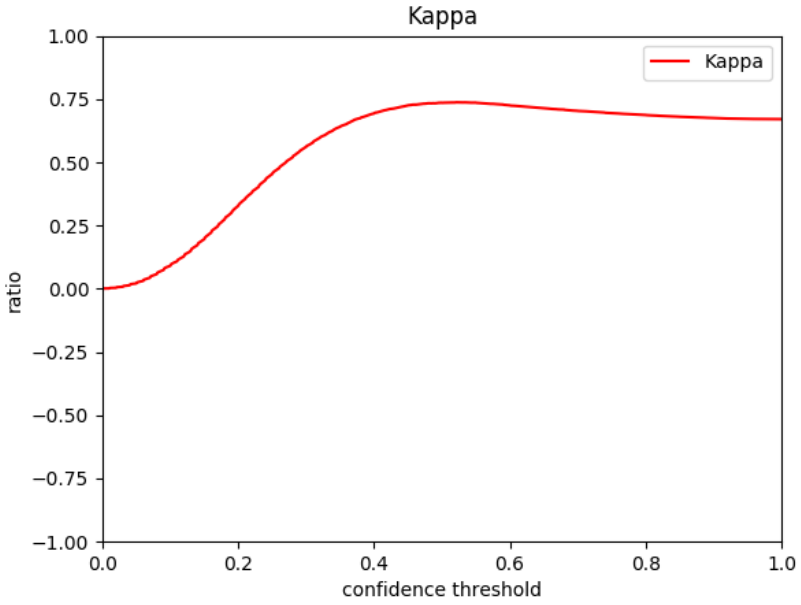


Figure E.5: AKI training/testing kappa.

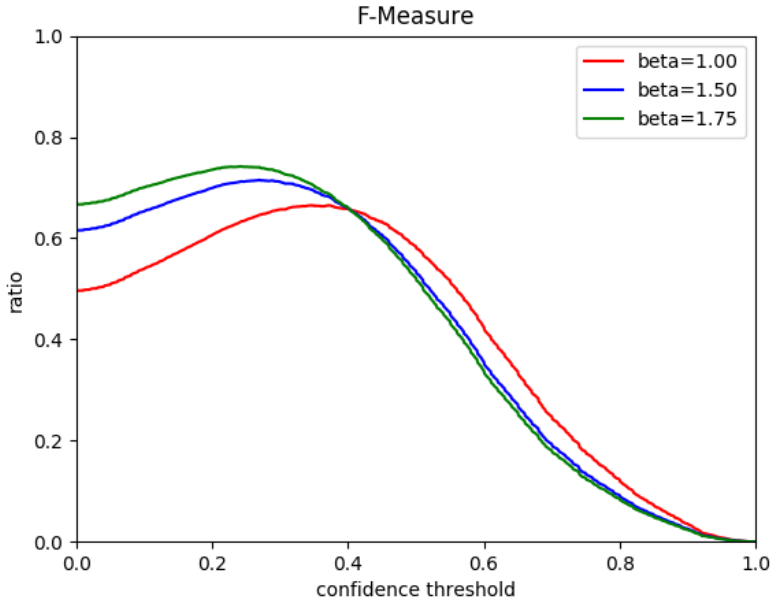


Figure E.6: AKI training/testing F-measures.

E.1.2 Validation

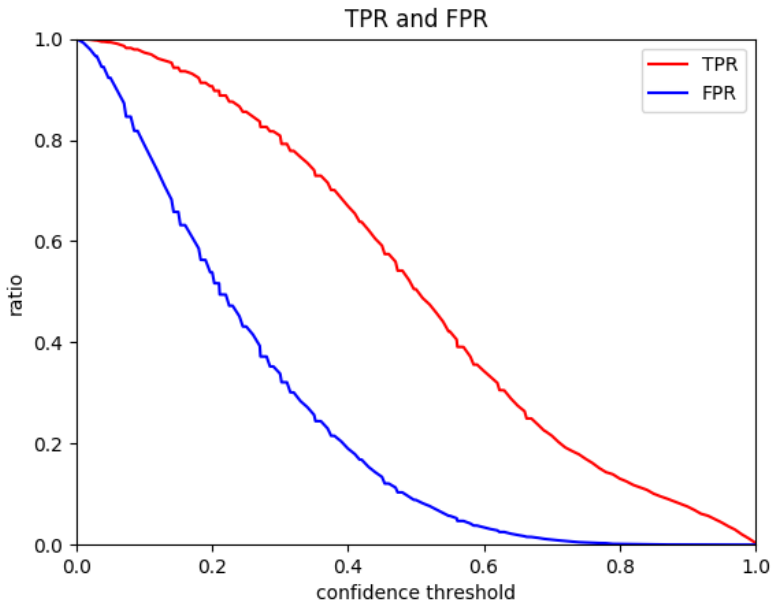


Figure E.7: AKI validation true and false positive rates.

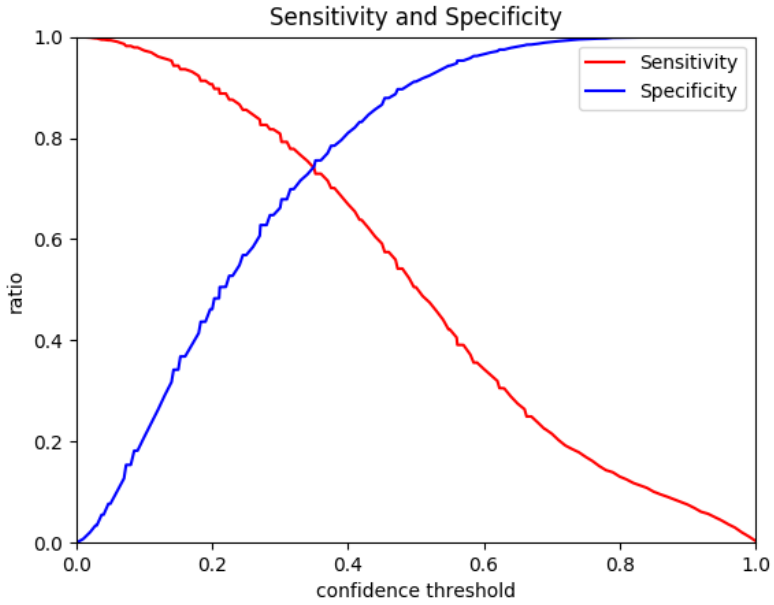


Figure E.8: AKI validation sensitivity and specificity.

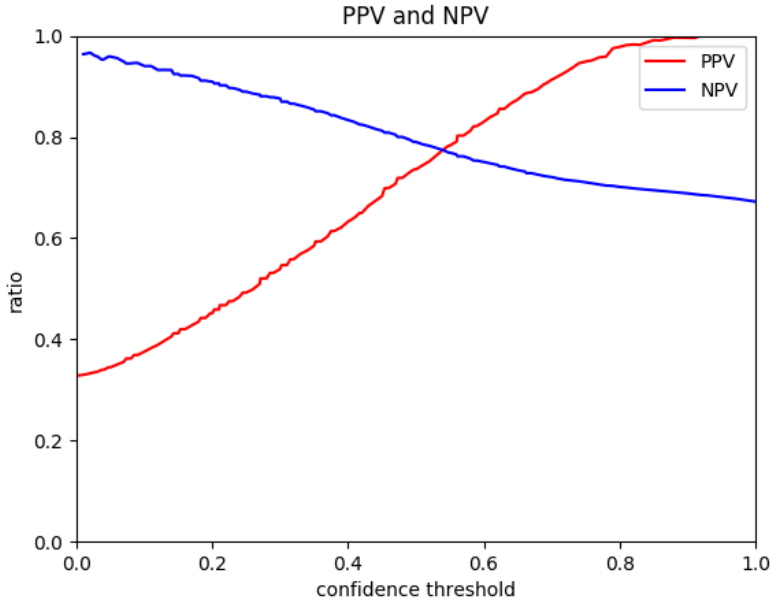


Figure E.9: AKI validation positive and negative predictive values.

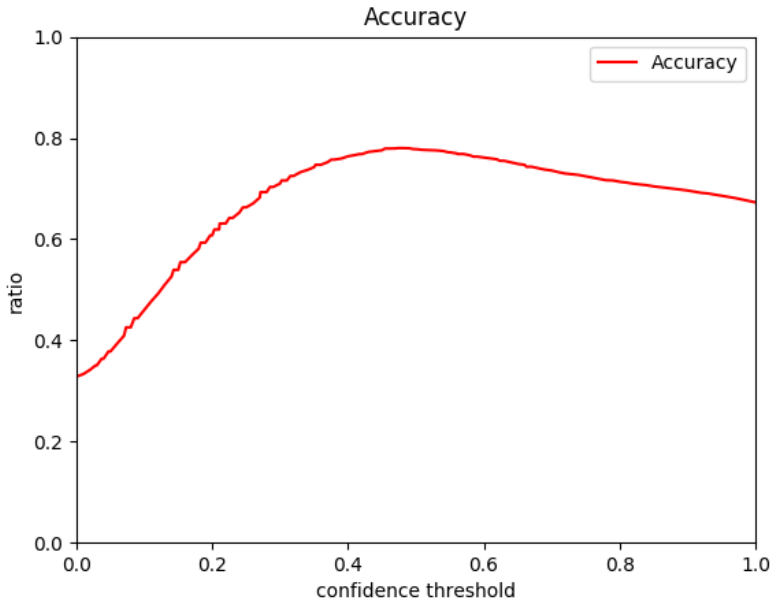


Figure E.10: AKI validation accuracy.

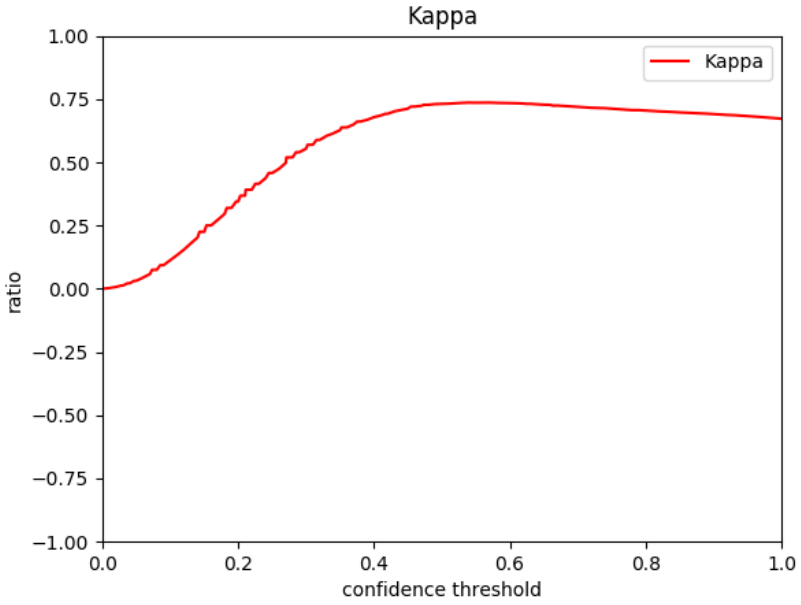


Figure E.11: AKI validation kappa.

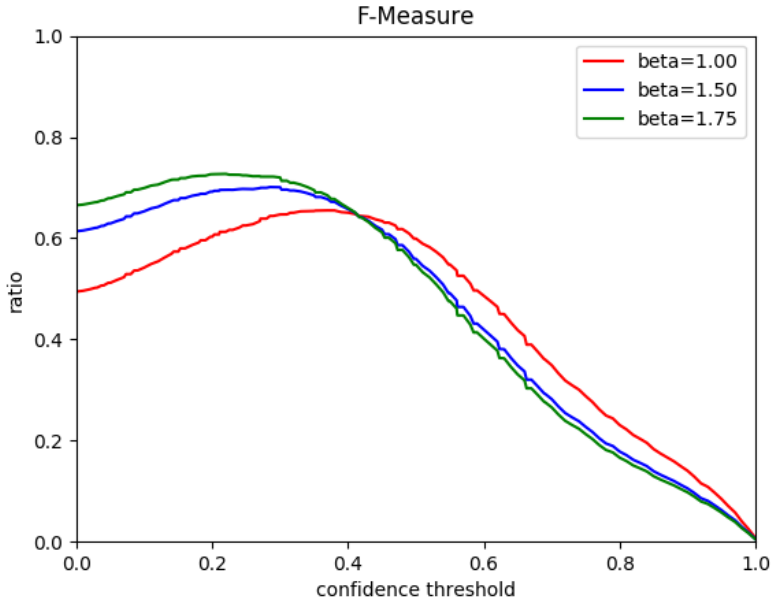
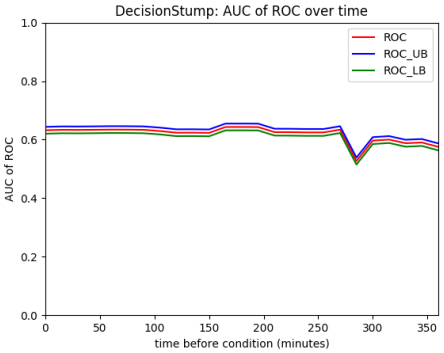
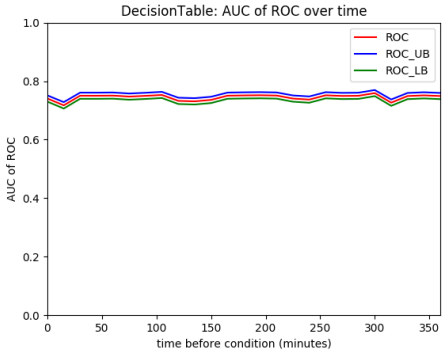


Figure E.12: AKI validation F-measures.

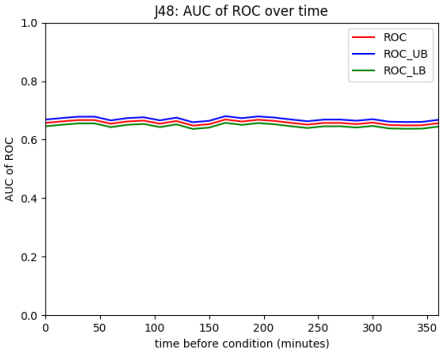
E.1.3 Machine Learning Algorithms over time



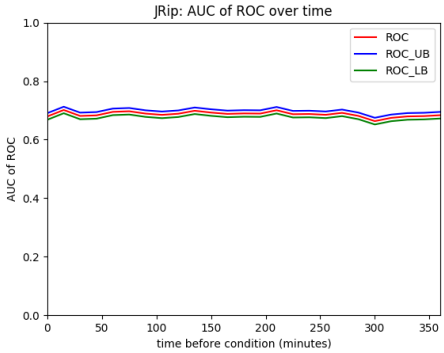
(a) Decision Stump



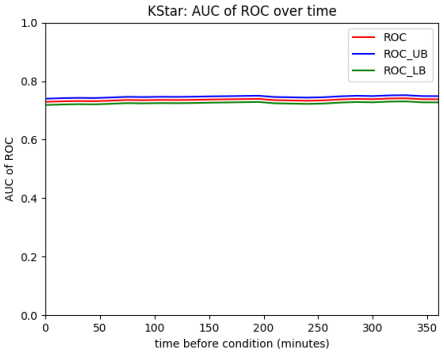
(b) Decision Table



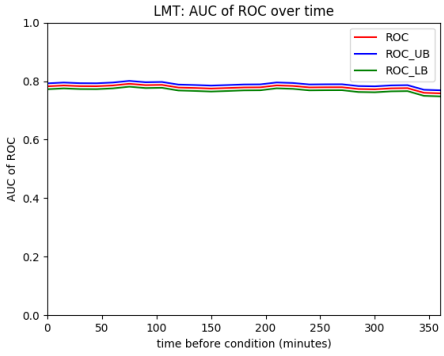
(c) J48



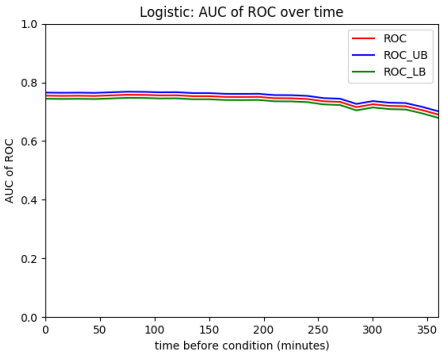
(d) JRip



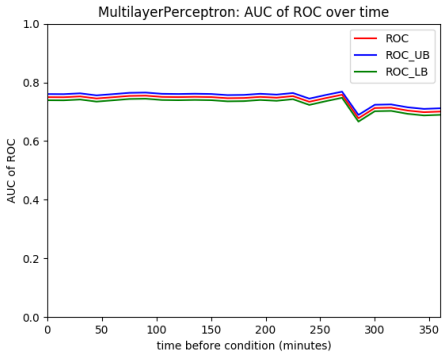
(e) KStar



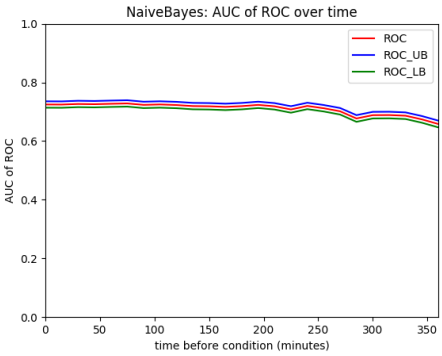
(f) LMT



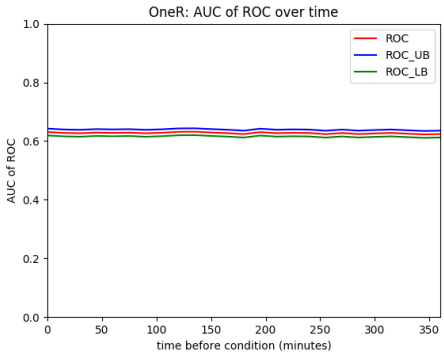
(g) Logistic Regression



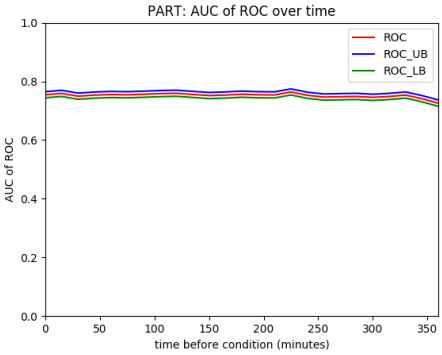
(h) Multilayer Perceptron



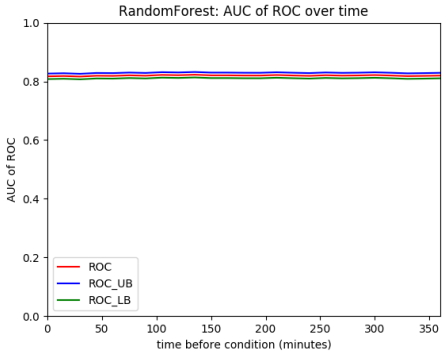
(i) Naive Bayes



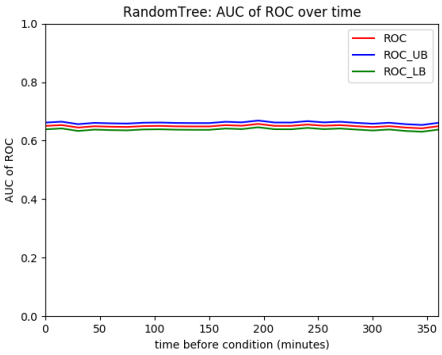
(j) One R



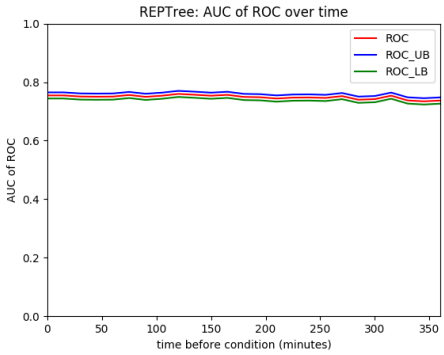
(k) PART



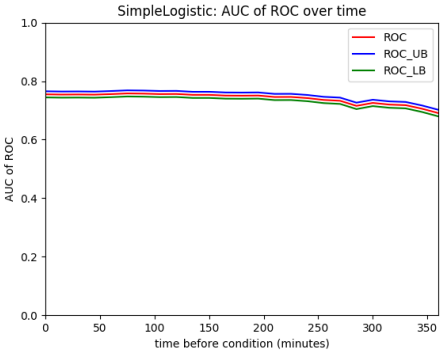
(l) Random Forest



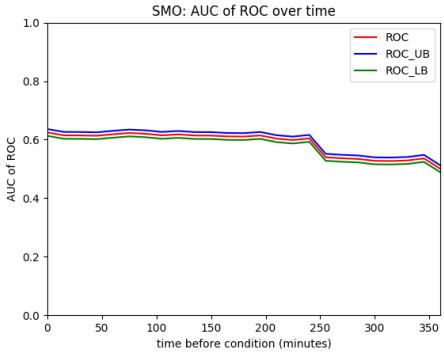
(m) Random Tree



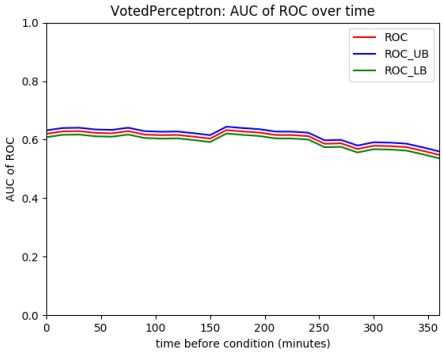
(n) REP Tree



(o) Simple Logistic Regression



(p) SMO



(q) Voted Perceptron

Figure E.13: Algorithm's performance over time on the AKI selection criteria.

E.1.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Low Exhaled Min Vol	3	4	4	3	4	4	0.1212
Furosemide (Lasix)	8.702	10	10	3.155	10	10	<0.0001
Urea Nitrogen	12	17	23	10	14	19	<0.0001
Phosphate	2.8	3.3	3.9	2.6	3.1	3.7	<0.0001
pO_2	76	99	132	70	95	138	0.0064
Central Venous Pressure	0	4	9	0	3	9	0.0021
Glucose	91	103	118	89	101	120	0.0032
NBP diastolic	55	64	73	54	63	73	0.0259
AST	17	23	37	17	23	33	0.0041
Potassium	3.8	4.1	4.4	3.7	4	4.3	<0.0001
Chloride (serum)	103	106	110	102	106	109	<0.0001
Impaired Skin Width	0.25	0.5	1	0.25	0.5	1	0.0383
Basophils	0	0.2	0.4	0.1	0.2	0.4	0.1338
Lipase	10	11	17	10	11	18	0.3575
Previous Weight	62.1	73.2	89	60	70.4	78	<0.0001
Lymphocytes	6.1	10.3	17	6.1	10.4	17	0.6774
Lactic Acid(0.5-2.0)	0.9	1.1	1.6	0.8	1.1	1.6	0.0069
.9% Normal Saline	0	7.5	10	0	9.03125	20	0.6765
Arterial CO_2 Pressure	33	37	42	32	38	45	0.1417
Troponin-T	0.01	0.01	0.03	0.01	0.01	0.02	<0.0001
Creatinine	0.6	0.8	1	0.6	0.7	0.9	<0.0001
5% Dextrose in water	9.573	26.85	50	11.046	22.5	50	0.2507
Hemoglobin	9.1	10.3	11.4	9.2	10.5	12.1	<0.0001
Pain Level	0	0	0	0	0	1	0.0047
Visual / hearing deficit	0	0	0	0	0	0	0.4026
Plateau Pressure	14	17	19	13	17	19	<0.0001
Urine Out Foley	30	50	90	30	60	120	0.0008
GCS Total	14	15	15	14	15	15	0.0014
Left Ventricular Stroke Work Index	20	24.885	31.304	20.574	23.158	31.672	<0.0001
Lactate Dehydrogenase (LD)	165	201	266	151	182	232	<0.0001
O_2 Flow (lpm)	0.4	0.5	0.7	0.4	0.5	1	0.9685

Table E.1: AKI model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Low Exhaled Min Vol	3	4	4	3	4	4	3	4	4	3	3.75	4	0.0583
Furosemide (Lasix)	6.513	10	10	3.782	10	10	3.342	10	10	10	10	10	<0.0001
Urea Nitrogen	13	17	23	9	12	16	10	14	20	8	12	17.25	<0.0001
Phosphate	2.8	3.3	4	2.6	3.1	3.5	2.6	3.1	3.7	2.7	3.1	3.3	<0.0001
pO_2	76	100	132	60	79	121.5	73	99	141	65.5	85.5	121.5	<0.0001
Central Venous Pressure	0	4	9	2	3	6	0	3	9	3	3	3	<0.0001
Glucose	91	103	118	86	101	124	89	101	120	82.25	101.5	116.25	0.0082
NBP diastolic	55	64	73	55	62	73	53	63	73	53.75	62.5	67	0.0418
AST	17	23	38	18	23	26	17	23	34	18	22	24	<0.0001
Potassium	3.8	4.1	4.4	3.7	3.9	4.2	3.7	4	4.3	3.675	3.9	4.125	<0.0001
Chloride (serum)	103	106	110	102	105	107	103	106	109	103	106	108	<0.0001
Impaired Skin Width	0.25	0.5	1	0.2	0.5	0.5	0.25	0.5	1	0.25	0.5	1	<0.0001
Basophils	0	0.2	0.4	0.1	0.2	0.4	0	0.2	0.4	0.1	0.2	0.4	<0.0001
Lipase	10	11	17	10	10	27	10	11	17	9.75	17	28	0.0896
Previous Weight	62.3	73.6	89.2	60.8	70	72.8	59.8	70.4	80	60.8	63.7	72.8	<0.0001
Lymphocytes	6.125	10.4	17	6.5	11.5	19	6	10.1	16.8	6	9.45	17.1	0.0002
Lactic Acid(0.5-2.0)	0.9	1.1	1.6	0.9	1.1	1.7	0.8	1.1	1.5	0.8	1.05	1.5	0.0001
.9% Normal Saline	0	7.5	10	0	0	10	0	10	20	0	0	10	<0.0001
Arterial CO_2 Pressure	33	37	42	29	36	46	33	38	45	30	39	48	<0.0001
Troponin-T	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	<0.0001
Creatinine	0.6	0.8	1	0.6	0.7	0.8	0.6	0.7	0.9	0.6	0.7	0.9	<0.0001
5% Dextrose in water	9.234	28.929	50	10.56	16.667	50	11.5	40	50	13	18	50	<0.0001
Hemoglobin	9.1	10.3	11.4	9.3	10.8	12.9	9.2	10.4	11.9	9.2	10.45	12.625	<0.0001
Pain Level	0	0	0	0	0	2	0	0	1	0	0	2	<0.0001
Visual / hearing deficit	0	0	0	0	0	0	0	0	0	0	0	0	0.7648
Plateau Pressure	14	17	19	13	17	17	13	16	19	13	14	17	<0.0001
Urine Out Foley	30	50	90	25	50	125	30	60	120	25	60	105	<0.0001
GCS Total	14	15	15	14	15	15	14	15	15	15	15	15	<0.0001
Left Ventricular Stroke Work Index	18.932	24.885	31.304	20.574	31.304	33.186	20.574	22.857	31.304	20.574	20.574	31.304	<0.0001
Lactate Dehydrogenase (LD)	166	203	268	137	168	201	154	187	238	137	160	189.25	<0.0001
O_2 Flow (lpm)	0.4	0.5	0.7	0.4	0.5	0.5	0.4	0.5	1	0.5	0.5	1	0.3349

Table E.2: AKI model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition_Positive	Condition_Negative	P-value
N	10408	3418	6990	
Age, median	62	66	60	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	68.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.4295
Weight, median (Q1,Q3)	79.40 (66.57, 93.00)	81.00 (68.30, 95.20)	77.00 (65.00, 90.00)	<0.0001
BMI, median (Q1,Q3)	28.18 (24.69, 32.37)	28.65 (25.00, 33.06)	27.78 (24.26, 31.68)	<0.0001
Male	6019 (0.58)	2182 (0.64)	3837 (0.55)	<0.0001
In-Hospital Mortality	1259 (0.12)	609 (0.18)	650 (0.09)	<0.0001
30 Day Mortality	1500 (0.14)	673 (0.20)	827 (0.12)	<0.0001
ICU LOS, median(Q1,Q3)	2.03 (1.16, 3.90)	3.07 (1.64, 6.70)	1.79 (1.08, 2.97)	<0.0001
ICU				
CCU	1356 (0.13)	482 (0.14)	874 (0.13)	0.0339
CSRU	2090 (0.20)	1032 (0.30)	1058 (0.15)	<0.0001
MICU	3685 (0.35)	1034 (0.30)	2651 (0.38)	<0.0001
SICU	1816 (0.17)	500 (0.15)	1316 (0.19)	<0.0001
TSICU	1461 (0.14)	370 (0.11)	1091 (0.16)	<0.0001
ethnicity				
Asian	250 (0.02)	75 (0.02)	175 (0.03)	0.3390
Black	785 (0.08)	241 (0.07)	544 (0.08)	0.2018
Hispanic	371 (0.04)	99 (0.03)	272 (0.04)	0.0116
White	7506 (0.72)	2432 (0.71)	5074 (0.73)	0.4176
insurance				
Government	327 (0.03)	66 (0.02)	261 (0.04)	<0.0001
Medicaid	995 (0.10)	284 (0.08)	711 (0.10)	0.0039
Medicare	5053 (0.49)	1916 (0.56)	3137 (0.45)	<0.0001
Private	3872 (0.37)	1118 (0.33)	2754 (0.39)	<0.0001
Self Pay	161 (0.02)	34 (0.01)	127 (0.02)	0.0015

Table E.3: AKI patients' Demographics of validation set by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	True_Positive	True_Negative	False_Positive	False_Negative	P-value
N	10408	1752	573	6417	1666	
Age, median	62	66	64	60	66	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.25 (64.00, 70.00)	68.00 (64.00, 70.00)	67.00 (64.00, 70.00)	68.00 (64.00, 70.00)	0.5819
Weight, median (Q1,Q3)	79.40 (66.57, 93.00)	79.40 (65.05, 92.45)	82.40 (70.00, 96.00)	75.80 (64.02, 89.67)	83.29 (71.33, 97.07)	<0.0001
BMI, median (Q1,Q3)	28.18 (24.69, 32.37)	27.81 (24.55, 31.75)	29.12 (25.17, 33.05)	27.41 (24.02, 31.22)	29.36 (25.52, 33.70)	<0.0001
Male	6019 (0.58)	1049 (0.60)	370 (0.65)	3467 (0.54)	1133 (0.68)	<0.0001
In-Hospital Mortality	1259 (0.12)	322 (0.18)	89 (0.16)	561 (0.09)	287 (0.17)	<0.0001
30 Day Mortality	1500 (0.14)	362 (0.21)	106 (0.18)	721 (0.11)	311 (0.19)	<0.0001
ICU LOS, median(Q1,Q3)	2.03 (1.16, 3.90)	3.16 (1.79, 6.64)	2.65 (1.50, 4.75)	1.73 (1.06, 2.89)	3.02 (1.38, 6.76)	<0.0001
ICU						
CCU	1356 (0.13)	286 (0.16)	61 (0.11)	813 (0.13)	196 (0.12)	0.0002
CSRU	2090 (0.20)	350 (0.20)	217 (0.38)	841 (0.13)	682 (0.41)	<0.0001
MICU	3685 (0.35)	600 (0.34)	171 (0.30)	2480 (0.39)	434 (0.26)	<0.0001
SICU	1816 (0.17)	291 (0.17)	60 (0.10)	1256 (0.20)	209 (0.13)	<0.0001
TSICU	1461 (0.14)	225 (0.13)	64 (0.11)	1027 (0.16)	145 (0.09)	<0.0001
ethnicity						
Asian	250 (0.02)	42 (0.02)	7 (0.01)	168 (0.03)	33 (0.02)	0.1217
Black	785 (0.08)	145 (0.08)	41 (0.07)	503 (0.08)	96 (0.06)	0.0279
Hispanic	371 (0.04)	51 (0.03)	11 (0.02)	261 (0.04)	48 (0.03)	0.0043
White	7506 (0.72)	1250 (0.71)	414 (0.72)	4660 (0.73)	1182 (0.71)	0.8766
insurance						
Government	327 (0.03)	39 (0.02)	16 (0.03)	245 (0.04)	27 (0.02)	<0.0001
Medicaid	995 (0.10)	152 (0.09)	42 (0.07)	669 (0.10)	132 (0.08)	0.0028
Medicare	5053 (0.49)	998 (0.57)	283 (0.49)	2854 (0.44)	918 (0.55)	<0.0001
Private	3872 (0.37)	546 (0.31)	228 (0.40)	2526 (0.39)	572 (0.34)	<0.0001
Self Pay	161 (0.02)	17 (0.01)	4 (0.01)	123 (0.02)	17 (0.01)	0.0017

Table E.4: AKI patients' Demographics of validation set in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	Condition_Positive	Condition_Negative	P-value
N	10408	3418	6990	
Comorbidity				
Congestive heart failure	2198 (0.21)	1092 (0.32)	1106 (0.16)	<0.0001
Cardiac arrhythmias	3389 (0.33)	1456 (0.43)	1933 (0.28)	<0.0001
Valvular disease	1534 (0.15)	721 (0.21)	813 (0.12)	<0.0001
Pulmonary circulation disorders	447 (0.04)	212 (0.06)	235 (0.03)	<0.0001
Peripheral vascular disorders	1068 (0.10)	464 (0.14)	604 (0.09)	<0.0001
Hypertension, uncomplicated	4891 (0.47)	1748 (0.51)	3143 (0.45)	<0.0001
Hypertension, complicated	313 (0.03)	157 (0.05)	156 (0.02)	<0.0001
Paralysis	330 (0.03)	98 (0.03)	232 (0.03)	0.2241
Other neurological disorders	768 (0.07)	220 (0.06)	548 (0.08)	0.0133
Chronic pulmonary disease	2320 (0.22)	828 (0.24)	1492 (0.21)	0.0035
Diabetes, uncomplicated	1973 (0.19)	773 (0.23)	1200 (0.17)	<0.0001
Diabetes, complicated	368 (0.04)	146 (0.04)	222 (0.03)	0.0052
Hypothyroidism	1009 (0.10)	326 (0.10)	683 (0.10)	0.7195
Renal failure	389 (0.04)	195 (0.06)	194 (0.03)	<0.0001
Liver disease	904 (0.09)	375 (0.11)	529 (0.08)	<0.0001
Peptic ulcer disease excluding bleeding	68 (0.01)	21 (0.01)	47 (0.01)	0.7310
AIDS/HIV	98 (0.01)	44 (0.01)	54 (0.01)	0.0110
Lymphoma	156 (0.01)	63 (0.02)	93 (0.01)	0.0448
Metastatic cancer	128 (0.01)	35 (0.01)	93 (0.01)	0.1855
Solid tumor without metastasis	881 (0.08)	291 (0.09)	590 (0.08)	0.9042
Rheumatoid arthritis/collagen vascular diseases	308 (0.03)	95 (0.03)	213 (0.03)	0.4557
Coagulopathy	828 (0.08)	409 (0.12)	419 (0.06)	<0.0001
Weight loss	387 (0.04)	172 (0.05)	215 (0.03)	<0.0001
Fluid and electrolyte disorders	2391 (0.23)	951 (0.28)	1440 (0.21)	<0.0001
Blood loss anemia	190 (0.02)	62 (0.02)	128 (0.02)	0.9512
Deficiency anemia	288 (0.03)	75 (0.02)	213 (0.03)	0.0140
Alcohol abuse	485 (0.05)	170 (0.05)	315 (0.05)	0.2997
Drug abuse	468 (0.04)	129 (0.04)	339 (0.05)	0.0151
Psychoses	183 (0.02)	55 (0.02)	128 (0.02)	0.4223
Depression	1000 (0.10)	272 (0.08)	728 (0.10)	0.0001
Trauma	1551 (0.15)	433 (0.13)	1118 (0.16)	<0.0001

Table E.5: AKI patients' Comorbidities of validation set by condition positive or negative. P-values for patient counts calculated by Pearson Chi-Squared test.

Variable	Total	True_Positive	True_Negative	False_Positive	False_Negative	P-value
N	10408	1752	573	6417	1666	
Comorbidity						
Congestive heart failure	2198 (0.21)	504 (0.29)	174 (0.30)	932 (0.15)	588 (0.35)	<0.0001
Cardiac arrhythmias	3389 (0.33)	689 (0.39)	218 (0.38)	1715 (0.27)	767 (0.46)	<0.0001
Valvular disease	1534 (0.15)	276 (0.16)	137 (0.24)	676 (0.11)	445 (0.27)	<0.0001
Pulmonary circulation disorders	447 (0.04)	98 (0.06)	31 (0.05)	204 (0.03)	114 (0.07)	<0.0001
Peripheral vascular disorders	1068 (0.10)	213 (0.12)	73 (0.13)	531 (0.08)	251 (0.15)	<0.0001
Hypertension, uncomplicated	4891 (0.47)	877 (0.50)	305 (0.53)	2838 (0.44)	871 (0.52)	<0.0001
Hypertension, complicated	313 (0.03)	74 (0.04)	19 (0.03)	137 (0.02)	83 (0.05)	<0.0001
Paralysis	330 (0.03)	66 (0.04)	12 (0.02)	220 (0.03)	32 (0.02)	0.0035
Other neurological disorders	768 (0.07)	122 (0.07)	29 (0.05)	519 (0.08)	98 (0.06)	0.0029
Chronic pulmonary disease	2320 (0.22)	422 (0.24)	126 (0.22)	1366 (0.21)	406 (0.24)	0.0337
Diabetes, uncomplicated	1973 (0.19)	366 (0.21)	130 (0.23)	1070 (0.17)	407 (0.24)	<0.0001
Diabetes, complicated	368 (0.04)	56 (0.03)	25 (0.04)	197 (0.03)	90 (0.05)	<0.0001
Hypothyroidism	1009 (0.10)	176 (0.10)	56 (0.10)	627 (0.10)	150 (0.09)	0.7806
Renal failure	389 (0.04)	93 (0.05)	27 (0.05)	167 (0.03)	102 (0.06)	<0.0001
Liver disease	904 (0.09)	183 (0.10)	36 (0.06)	493 (0.08)	192 (0.12)	<0.0001
Peptic ulcer disease excluding bleeding	68 (0.01)	11 (0.01)	1 (0.00)	46 (0.01)	10 (0.01)	0.4760
AIDS/HIV	98 (0.01)	17 (0.01)	5 (0.01)	49 (0.01)	27 (0.02)	0.0157
Lymphoma	156 (0.01)	27 (0.02)	5 (0.01)	88 (0.01)	36 (0.02)	0.0692
Metastatic cancer	128 (0.01)	20 (0.01)	7 (0.01)	86 (0.01)	15 (0.01)	0.5285
Solid tumor without metastasis	881 (0.08)	171 (0.10)	42 (0.07)	548 (0.09)	120 (0.07)	0.0570
Rheumatoid arthritis/collagen vascular diseases	308 (0.03)	52 (0.03)	20 (0.03)	193 (0.03)	43 (0.03)	0.7049
Coagulopathy	828 (0.08)	210 (0.12)	41 (0.07)	378 (0.06)	199 (0.12)	<0.0001
Weight loss	387 (0.04)	96 (0.05)	14 (0.02)	201 (0.03)	76 (0.05)	<0.0001
Fluid and electrolyte disorders	2391 (0.23)	541 (0.31)	107 (0.19)	1333 (0.21)	410 (0.25)	<0.0001
Blood loss anemia	190 (0.02)	34 (0.02)	9 (0.02)	119 (0.02)	28 (0.02)	0.9073
Deficiency anemia	288 (0.03)	37 (0.02)	20 (0.03)	193 (0.03)	38 (0.02)	0.0871
Alcohol abuse	485 (0.05)	98 (0.06)	18 (0.03)	297 (0.05)	72 (0.04)	0.0883
Drug abuse	468 (0.04)	63 (0.04)	16 (0.03)	323 (0.05)	66 (0.04)	0.0073
Psychoses	183 (0.02)	24 (0.01)	6 (0.01)	122 (0.02)	31 (0.02)	0.2618
Depression	1000 (0.10)	142 (0.08)	35 (0.06)	693 (0.11)	130 (0.08)	<0.0001
Trauma	1551 (0.15)	278 (0.16)	62 (0.11)	1056 (0.16)	155 (0.09)	<0.0001

Table E.6: AKI patients' Comorbidities of validation set in context of model performance. P-values for patient counts calculated by Pearson Chi-Squared test.

E.2 AKI Stage 2

E.2.1 Training and Testing

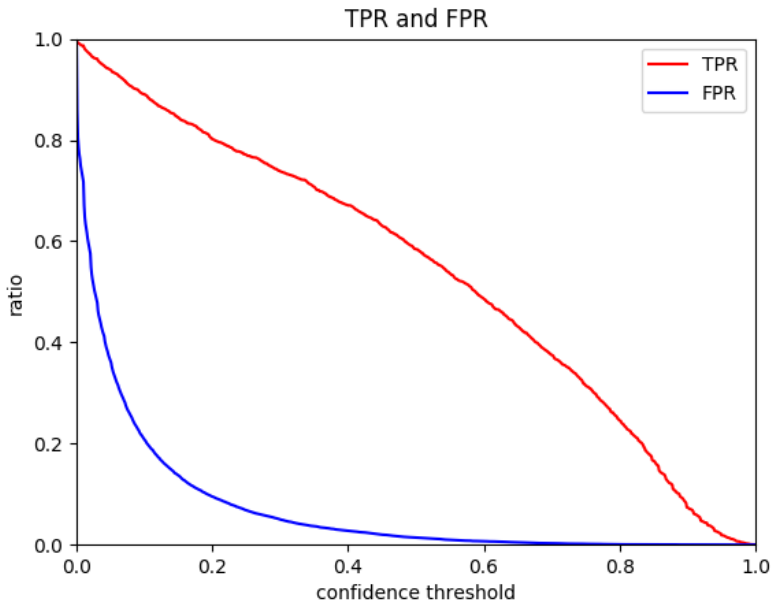


Figure E.14: AKI Stage 2 training/testing true positive and false positive rates.

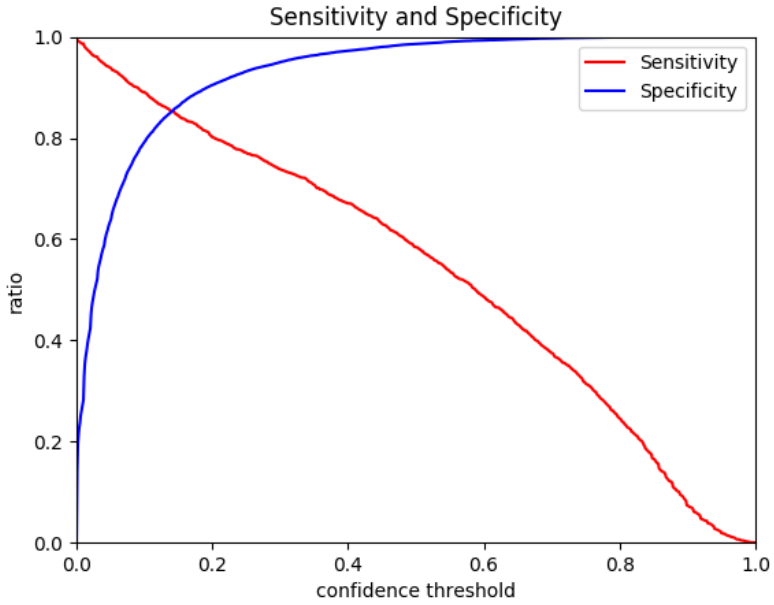


Figure E.15: AKI Stage 2 training/testing sensitivity and specificity.

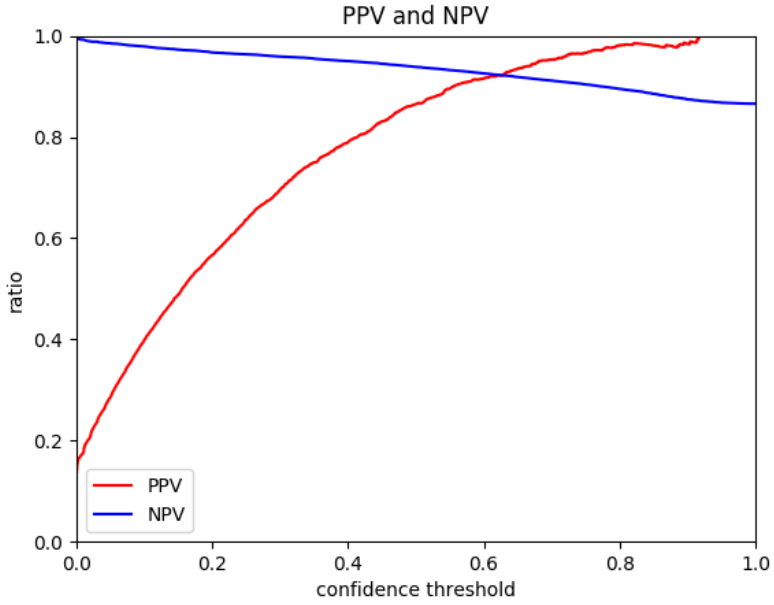


Figure E.16: AKI Stage 2 training/testing positive and negative predictive value.

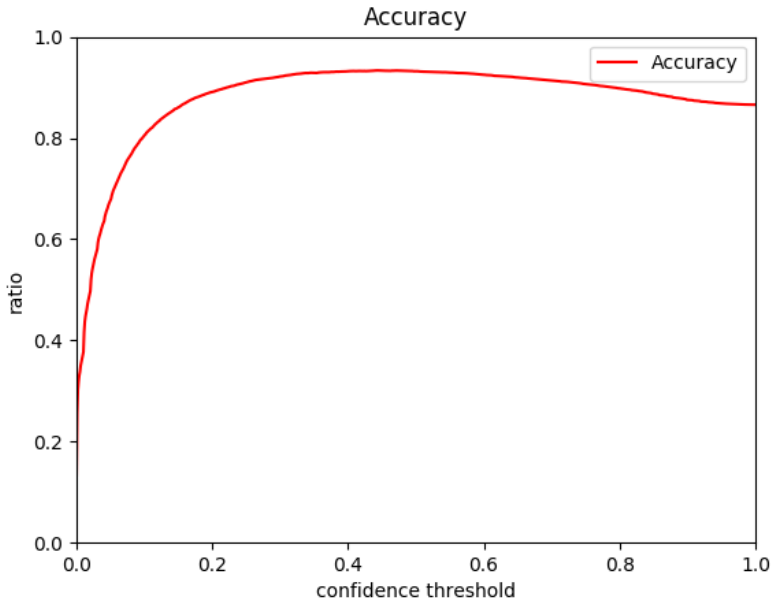


Figure E.17: AKI Stage 2 training/testing accuracy.

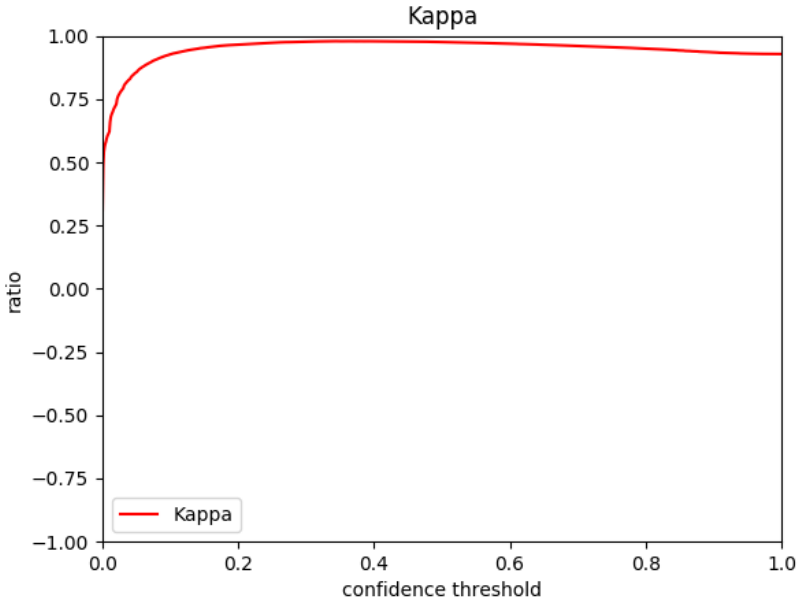


Figure E.18: AKI Stage 2 training/testing kappa.

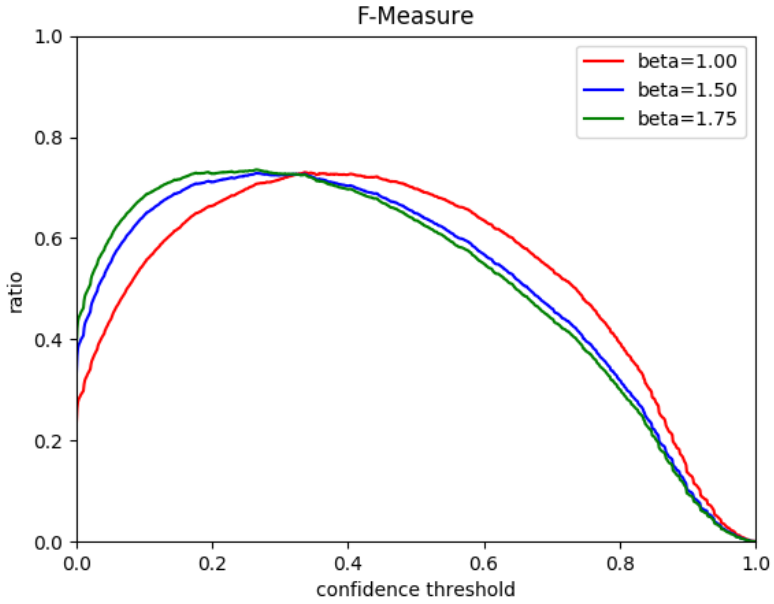


Figure E.19: AKI Stage 2 training/testing F-measures.

E.2.2 Validation

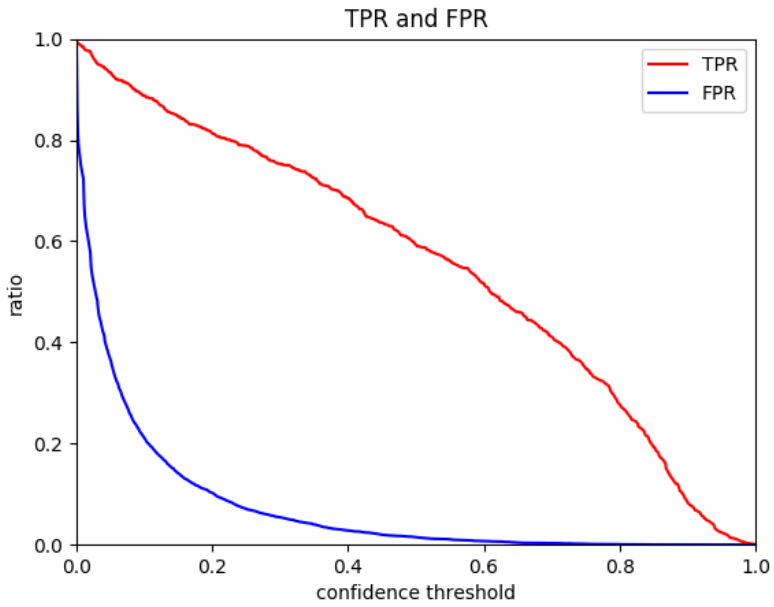


Figure E.20: AKI Stage 2 validation true and false positive rates.

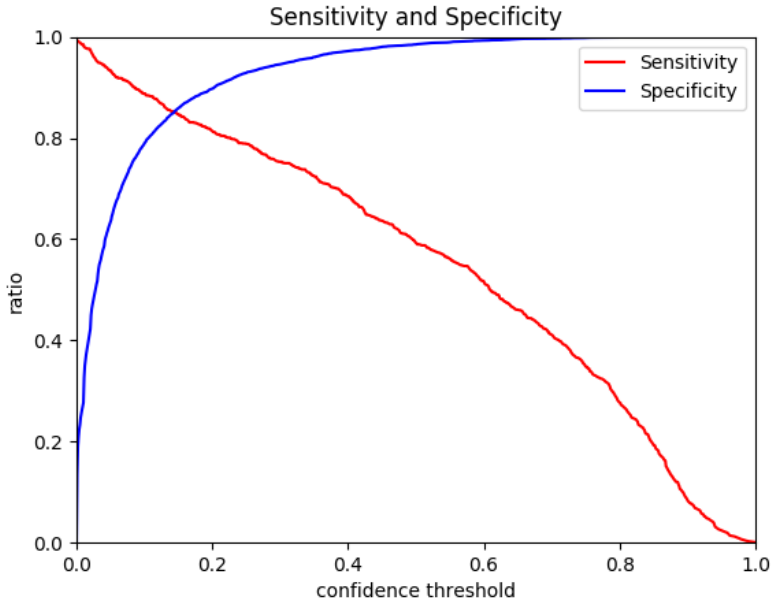


Figure E.21: AKI Stage 2 validation sensitivity and specificity.

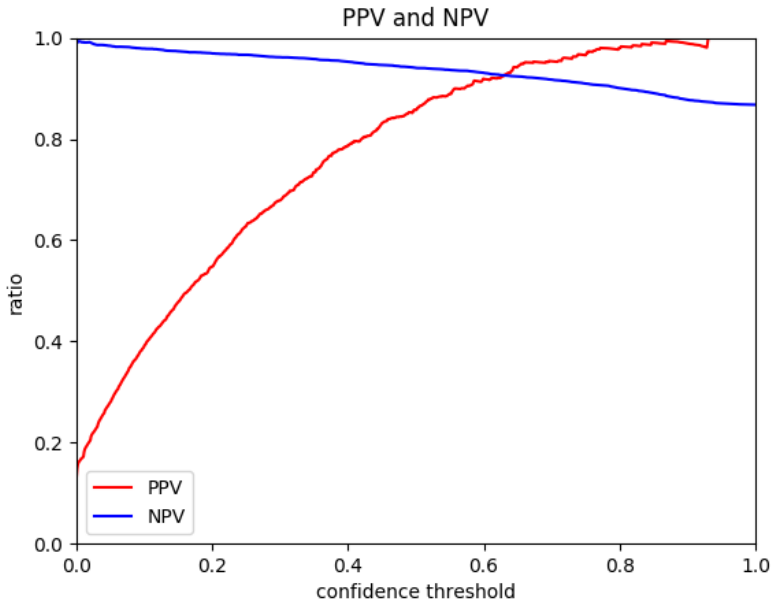


Figure E.22: AKI Stage 2 validation positive and negative predictive values.

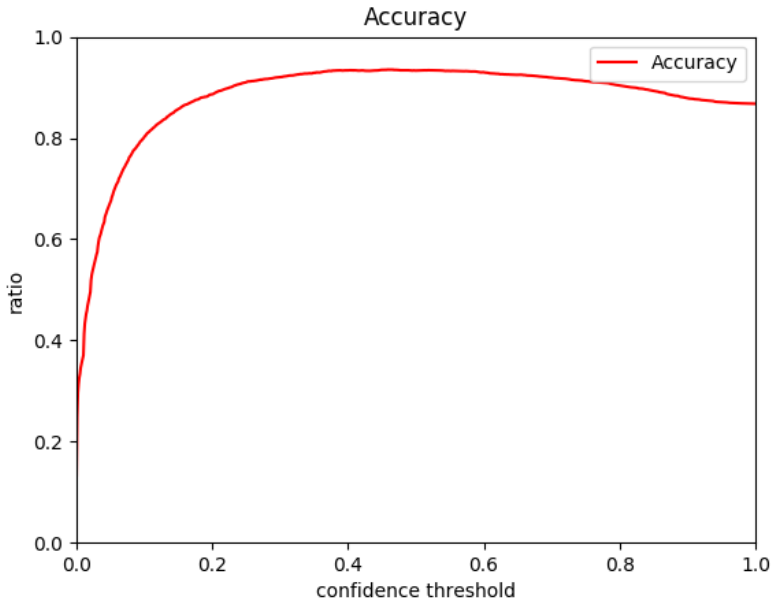


Figure E.23: AKI Stage 2 validation accuracy.

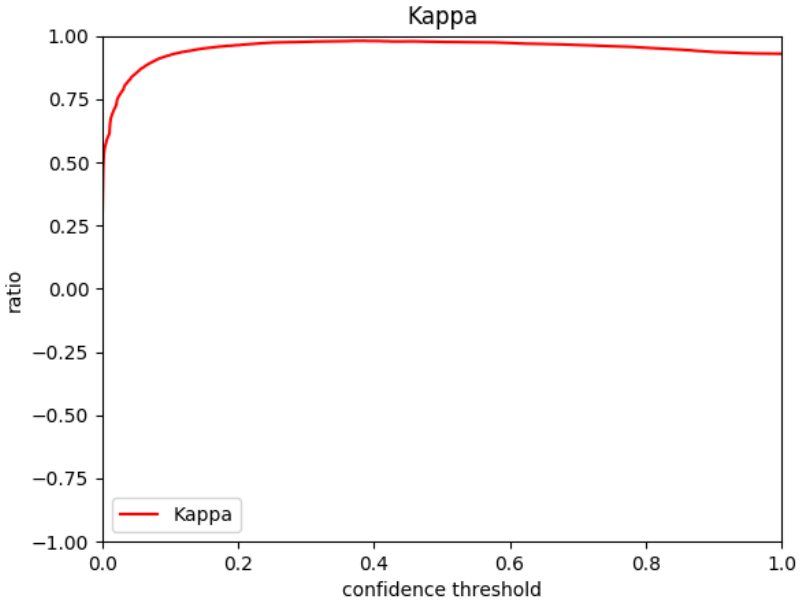


Figure E.24: AKI Stage 2 validation kappa.

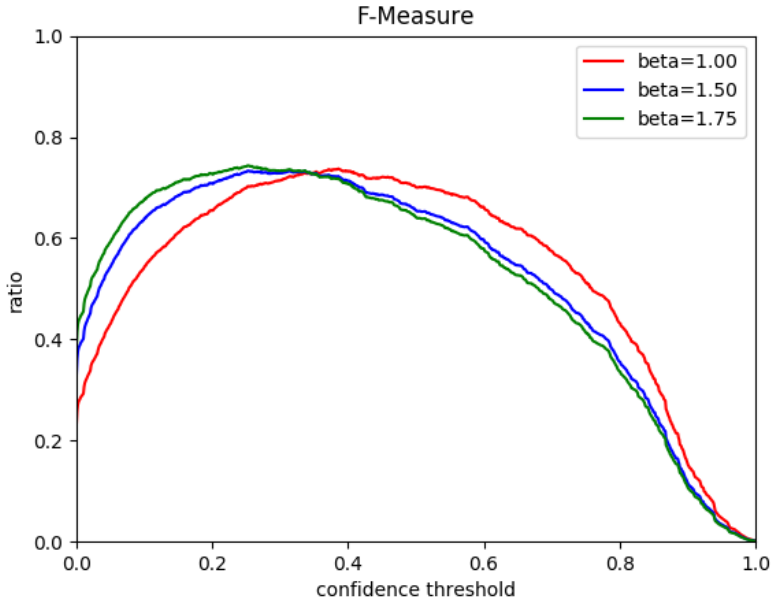
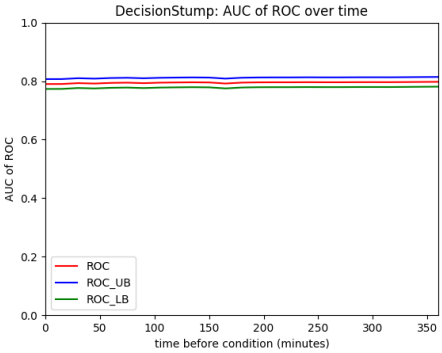
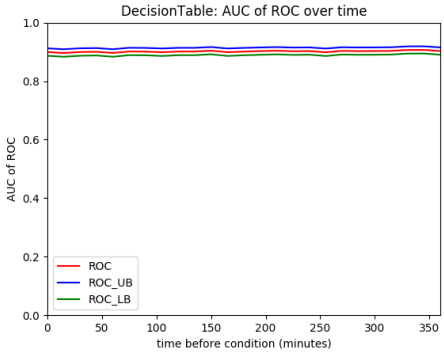


Figure E.25: AKI Stage 2 validation F-measures.

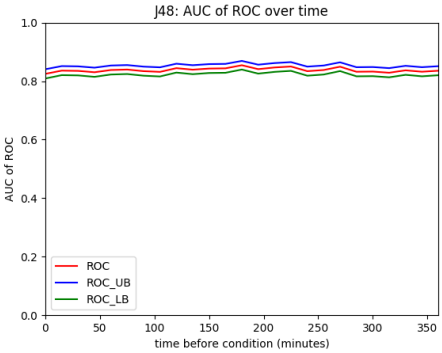
E.2.3 Machine Learning Algorithms over time



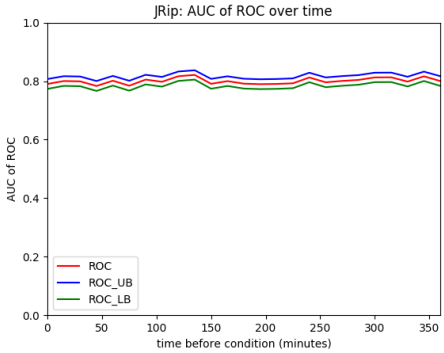
(a) Decision Stump



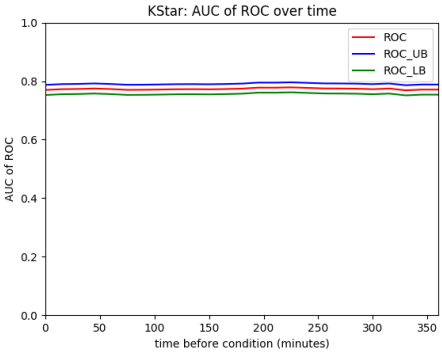
(b) Decision Table



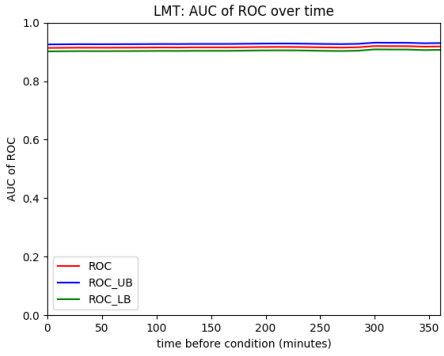
(c) J48



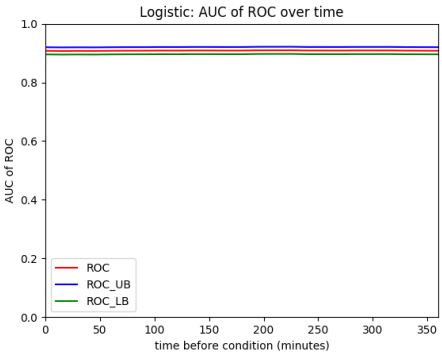
(d) JRip



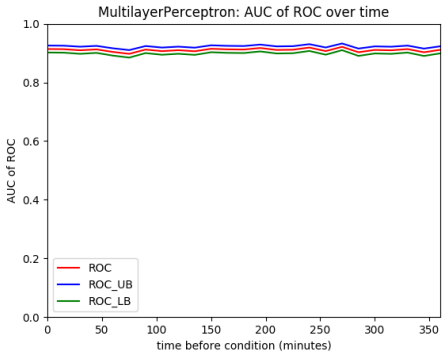
(e) KStar



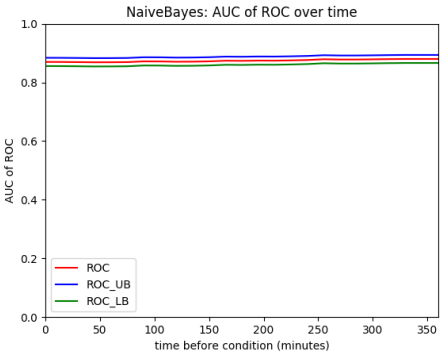
(f) LMT



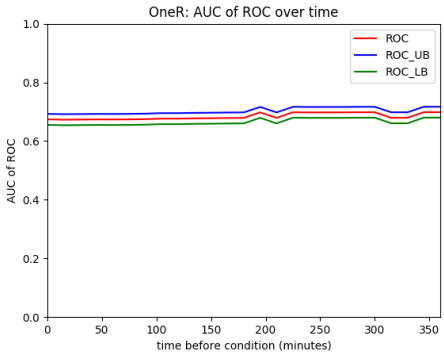
(g) Logistic Regression



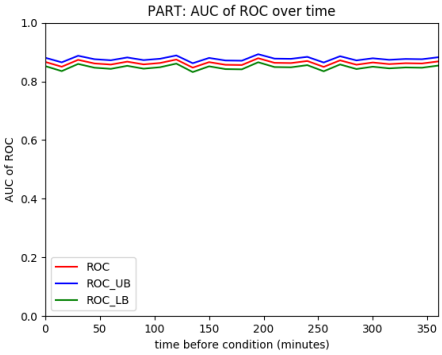
(h) Multilayer Perceptron



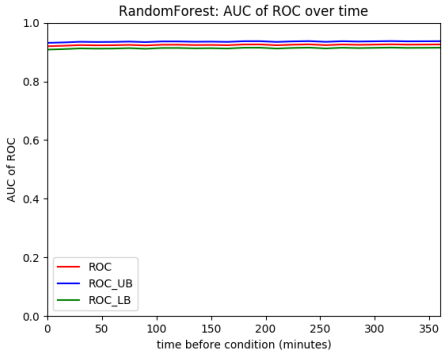
(i) Naive Bayes



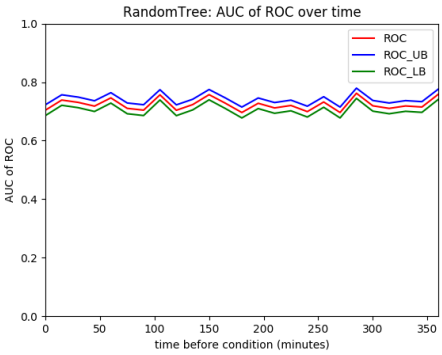
(j) One R



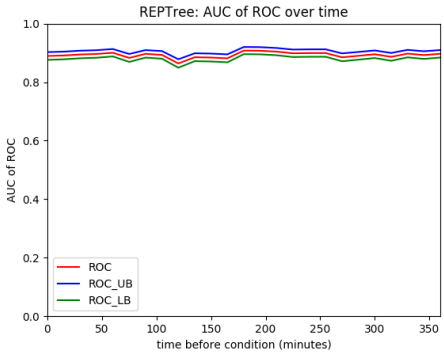
(k) PART



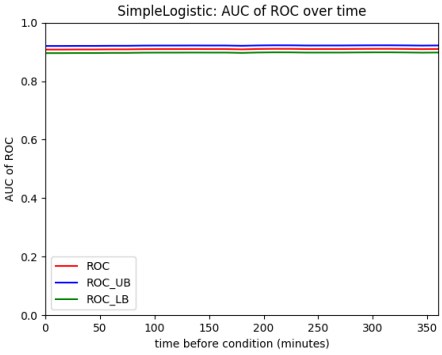
(l) Random Forest



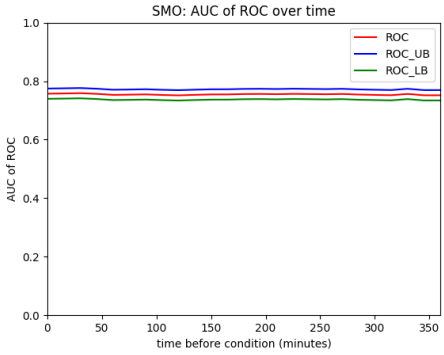
(m) Random Tree



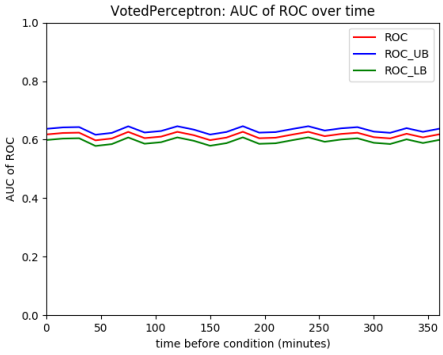
(n) REP Tree



(o) Simple Logistic Regression



(p) SMO



(q) Voted Perceptron

Figure E.26: Algorithm's performance over time on the AKI Stage 2 selection criteria.

E.2.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
IV access prior to admission	0	0	1	0	0	1	<0.0001
Potassium, Urine	4	7	15	4	7	7	<0.0001
Fibrinogen	104	156	184	104	151	184	<0.0001
Creatinine	0.6	1.1	1.7	0.5	0.7	0.9	<0.0001
GT Flush	0	0	0	0	0	0	<0.0001
OR Crystalloid	0	0	0	0	0	0	0.1865
Amylase	15	26	42	17	29	42	0.7211
ALT	13	18	30	11	16	25	<0.0001
Arterial $PaCO_2$	31	37	43	26	36	43	<0.0001
Alk. Phosphate	42	69	97	38	51	74	<0.0001
Lactic Acid	0.9	1.2	1.5	0.9	1.1	1.5	<0.0001
Creatine Kinase	39	84	220	28	55	131	<0.0001
Epithelial Cells, urine	0	0	0	0	0	0	<0.0001
Iron Binding Capacity, Total	126	137	190	137	146	161	0.0006
percent Hemoglobin A1c	5.6	5.8	6	5.6	5.8	6	0.0259
Urea Nitrogen	23	32	45	10	14	19	<0.0001

Table E.7: AKI Stage 2 model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
IV access prior to admission	0	0	1	0	0	1	0	0	1	0	0	1	<0.0001
Potassium, Urine	4	7	22	4	6.5	7	4	7	14	4	7	7	<0.0001
Fibrinogen	104	156	188	104	149	184	104	156	184	104	144	169	<0.0001
Creatinine	0.6	1.4	1.8	0.5	0.7	0.8	0.5	0.9	1.3	0.4	0.6	1	<0.0001
GT Flush	0	0	0	0	0	0	0	0	0	0	0	0	<0.0001
OR Crystalloid	0	0	0	0	0	0	0	0	0	0	0	0	0.5245
Amylase	14	26	42	17	31	42	13.75	26	36	17	26	42	0.0022
ALT	13	18	33	10	16	24	13	17	28.25	11	16	28	<0.0001
Arterial $PaCO_2$	31	37	43	24	36	43	30	37	43	31	37	43	<0.0001
Alk. Phosphate	44	71	101.25	38	50	72	43	68.5	99	39	57	81	<0.0001
Lactic Acid	0.9	1.3	1.7	0.9	1.1	1.5	0.9	1.2	1.5	0.8	1	1.4	<0.0001
Creatine Kinase	41	87.5	240	28	53	127	36	74	206.25	36	68	186	<0.0001
Epithelial Cells, urine	0	0	0	0	0	0	0	0	0	0	0	0	<0.0001
Iron Binding Capacity, Total	116	137	197	137	146	161	137	142.5	195	137	157	165	0.0016
percent Hemoglobin A1c	5.6	5.8	6	5.6	5.8	6	5.6	5.8	6	5.6	5.8	6	0.1544
Urea Nitrogen	28	37	48	10	14	18	22	28	36	14	18	23	<0.0001

Table E.8: AKI Stage 2 model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	7957	1053	6904	
Age, median	62	71	60	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.2718
Weight, median (Q1,Q3)	77.90 (65.00, 91.00)	81.00 (68.00, 96.60)	77.00 (65.00, 90.00)	0.0001
BMI, median (Q1,Q3)	27.99 (24.46, 32.08)	28.86 (24.86, 34.40)	27.82 (24.24, 31.68)	0.0011
Male	4398 (0.55)	602 (0.57)	3796 (0.55)	0.3738
In-Hospital Mortality	1053 (0.13)	416 (0.40)	637 (0.09)	<0.0001
30 Day Mortality	1263 (0.16)	450 (0.43)	813 (0.12)	<0.0001
ICU LOS, median(Q1,Q3)	1.93 (1.13, 3.40)	4.42 (2.11, 10.05)	1.80 (1.09, 2.99)	<0.0001
ICU				
CCU	1062 (0.13)	211 (0.20)	851 (0.12)	<0.0001
CSRU	1286 (0.16)	245 (0.23)	1041 (0.15)	<0.0001
MICU	2986 (0.38)	360 (0.34)	2626 (0.38)	0.0576
SICU	1459 (0.18)	156 (0.15)	1303 (0.19)	0.0042
TSICU	1164 (0.15)	81 (0.08)	1083 (0.16)	<0.0001
ethnicity				
Asian	195 (0.02)	21 (0.02)	174 (0.03)	0.3098
Black	619 (0.08)	82 (0.08)	537 (0.08)	0.9921
Hispanic	301 (0.04)	31 (0.03)	270 (0.04)	0.1330
White	5757 (0.72)	743 (0.71)	5014 (0.73)	0.4632
insurance				
Government	267 (0.03)	8 (0.01)	259 (0.04)	<0.0001
Medicaid	779 (0.10)	79 (0.08)	700 (0.10)	0.0109
Medicare	3800 (0.48)	709 (0.67)	3091 (0.45)	<0.0001
Private	2982 (0.37)	253 (0.24)	2729 (0.40)	<0.0001
Self Pay	129 (0.02)	4 (0.00)	125 (0.02)	0.0007

Table E.9: AKI Stage 2 demographics of validation set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	7957	820	6448	456	233	
Age, median	62	71	60	72	69	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.4306
Weight, median (Q1,Q3)	77.90 (65.00, 91.00)	81.50 (68.00, 96.38)	77.00 (65.00, 90.00)	79.00 (67.90, 90.00)	80.00 (66.70, 97.00)	0.0017
BMI, median (Q1,Q3)	27.99 (24.46, 32.08)	29.14 (25.21, 34.74)	27.78 (24.38, 31.80)	27.92 (23.93, 31.18)	28.48 (24.53, 33.26)	0.0096
Male	4398 (0.55)	490 (0.60)	3506 (0.54)	290 (0.64)	112 (0.48)	0.0080
In-Hospital Mortality	1053 (0.13)	362 (0.44)	542 (0.08)	95 (0.21)	54 (0.23)	<0.0001
30 Day Mortality	1263 (0.16)	387 (0.47)	694 (0.11)	119 (0.26)	63 (0.27)	<0.0001
ICU LOS, median(Q1,Q3)	1.93 (1.13, 3.40)	4.85 (2.18, 11.15)	1.78 (1.09, 2.95)	2.13 (1.19, 4.23)	3.67 (1.97, 6.98)	<0.0001
ICU						
CCU	1062 (0.13)	175 (0.21)	793 (0.12)	58 (0.13)	36 (0.15)	<0.0001
CSRU	1286 (0.16)	176 (0.21)	994 (0.15)	47 (0.10)	69 (0.30)	<0.0001
MICU	2986 (0.38)	289 (0.35)	2376 (0.37)	250 (0.55)	71 (0.30)	<0.0001
SICU	1459 (0.18)	119 (0.15)	1231 (0.19)	72 (0.16)	37 (0.16)	0.0121
TSICU	1164 (0.15)	61 (0.07)	1054 (0.16)	29 (0.06)	20 (0.09)	<0.0001
ethnicity						
Asian	195 (0.02)	18 (0.02)	161 (0.02)	13 (0.03)	3 (0.01)	0.6022
Black	619 (0.08)	68 (0.08)	483 (0.07)	54 (0.12)	14 (0.06)	0.0090
Hispanic	301 (0.04)	26 (0.03)	261 (0.04)	9 (0.02)	5 (0.02)	0.0549
White	5757 (0.72)	581 (0.71)	4676 (0.73)	338 (0.74)	162 (0.70)	0.8653
insurance						
Government	267 (0.03)	4 (0.00)	247 (0.04)	12 (0.03)	4 (0.02)	<0.0001
Medicaid	779 (0.10)	60 (0.07)	659 (0.10)	41 (0.09)	19 (0.08)	0.0636
Medicare	3800 (0.48)	561 (0.68)	2790 (0.43)	301 (0.66)	148 (0.64)	<0.0001
Private	2982 (0.37)	192 (0.23)	2632 (0.41)	97 (0.21)	61 (0.26)	<0.0001
Self Pay	129 (0.02)	3 (0.00)	120 (0.02)	5 (0.01)	1 (0.00)	0.0045

Table E.10: AKI Stage 2 demographics of validation set patients in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	Condition Positive	condition Negative	P-value
N	7957	1053	6904	
Comorbidity				
Congestive heart failure	1570 (0.20)	470 (0.45)	1100 (0.16)	<0.0001
Cardiac arrhythmias	2432 (0.31)	517 (0.49)	1915 (0.28)	<0.0001
Valvular disease	1013 (0.13)	207 (0.20)	806 (0.12)	<0.0001
Pulmonary circulation disorders	329 (0.04)	96 (0.09)	233 (0.03)	<0.0001
Peripheral vascular disorders	772 (0.10)	182 (0.17)	590 (0.09)	<0.0001
Hypertension, uncomplicated	3532 (0.44)	431 (0.41)	3101 (0.45)	0.0705
Hypertension, complicated	331 (0.04)	176 (0.17)	155 (0.02)	<0.0001
Paralysis	254 (0.03)	22 (0.02)	232 (0.03)	0.0315
Other neurological disorders	600 (0.08)	59 (0.06)	541 (0.08)	0.0140
Chronic pulmonary disease	1769 (0.22)	285 (0.27)	1484 (0.21)	0.0004
Diabetes, uncomplicated	1446 (0.18)	260 (0.25)	1186 (0.17)	<0.0001
Diabetes, complicated	311 (0.04)	91 (0.09)	220 (0.03)	<0.0001
Hypothyroidism	800 (0.10)	124 (0.12)	676 (0.10)	0.0585
Renal failure	417 (0.05)	224 (0.21)	193 (0.03)	<0.0001
Liver disease	733 (0.09)	210 (0.20)	523 (0.08)	<0.0001
Peptic ulcer disease excluding bleeding	54 (0.01)	7 (0.01)	47 (0.01)	0.9532
AIDS/HIV	62 (0.01)	9 (0.01)	53 (0.01)	0.7657
Lymphoma	111 (0.01)	19 (0.02)	92 (0.01)	0.2273
Metastatic cancer	102 (0.01)	9 (0.01)	93 (0.01)	0.1887
Solid tumor without metastasis	690 (0.09)	106 (0.10)	584 (0.08)	0.0989
Rheumatoid arthritis/collagen vascular diseases	252 (0.03)	41 (0.04)	211 (0.03)	0.1549
Coagulopathy	640 (0.08)	226 (0.21)	414 (0.06)	<0.0001
Weight loss	282 (0.04)	68 (0.06)	214 (0.03)	<0.0001
Fluid and electrolyte disorders	1853 (0.23)	418 (0.40)	1435 (0.21)	<0.0001
Blood loss anemia	158 (0.02)	31 (0.03)	127 (0.02)	0.0178
Deficiency anemia	249 (0.03)	36 (0.03)	213 (0.03)	0.5686
Alcohol abuse	377 (0.05)	64 (0.06)	313 (0.05)	0.0320
Drug abuse	363 (0.05)	29 (0.03)	334 (0.05)	0.0032
Psychoses	142 (0.02)	15 (0.01)	127 (0.02)	0.3477
Depression	806 (0.10)	83 (0.08)	723 (0.10)	0.0139
Trauma	1223 (0.15)	113 (0.11)	1110 (0.16)	<0.0001

Table E.11: AKI Stage 2 comorbidities of validation set patients by condition positive or negative. P-values for patient counts calculated by chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	7957	820	6448	456	233	
Comorbidity						
Congestive heart failure	1570 (0.20)	408 (0.50)	897 (0.14)	203 (0.45)	62 (0.27)	<0.0001
Cardiac arrhythmias	2432 (0.31)	420 (0.51)	1710 (0.27)	205 (0.45)	97 (0.42)	<0.0001
Valvular disease	1013 (0.13)	175 (0.21)	711 (0.11)	95 (0.21)	32 (0.14)	<0.0001
Pulmonary circulation disorders	329 (0.04)	85 (0.10)	194 (0.03)	39 (0.09)	11 (0.05)	<0.0001
Peripheral vascular disorders	772 (0.10)	148 (0.18)	523 (0.08)	67 (0.15)	34 (0.15)	<0.0001
Hypertension, uncomplicated	3532 (0.44)	298 (0.36)	2912 (0.45)	189 (0.41)	133 (0.57)	<0.0001
Hypertension, complicated	331 (0.04)	172 (0.21)	51 (0.01)	104 (0.23)	4 (0.02)	<0.0001
Paralysis	254 (0.03)	18 (0.02)	217 (0.03)	15 (0.03)	4 (0.02)	0.1901
Other neurological disorders	600 (0.08)	48 (0.06)	512 (0.08)	29 (0.06)	11 (0.05)	0.0512
Chronic pulmonary disease	1769 (0.22)	225 (0.27)	1343 (0.21)	141 (0.31)	60 (0.26)	<0.0001
Diabetes, uncomplicated	1446 (0.18)	213 (0.26)	1075 (0.17)	111 (0.24)	47 (0.20)	<0.0001
Diabetes, complicated	311 (0.04)	83 (0.10)	181 (0.03)	39 (0.09)	8 (0.03)	<0.0001
Hypothyroidism	800 (0.10)	98 (0.12)	610 (0.09)	66 (0.14)	26 (0.11)	0.0025
Renal failure	417 (0.05)	219 (0.27)	71 (0.01)	122 (0.27)	5 (0.02)	<0.0001
Liver disease	733 (0.09)	184 (0.22)	461 (0.07)	62 (0.14)	26 (0.11)	<0.0001
Peptic ulcer disease excluding bleeding	54 (0.01)	4 (0.00)	42 (0.01)	5 (0.01)	3 (0.01)	0.3983
AIDS/HIV	62 (0.01)	8 (0.01)	50 (0.01)	3 (0.01)	1 (0.00)	0.8352
Lymphoma	111 (0.01)	18 (0.02)	87 (0.01)	5 (0.01)	1 (0.00)	0.1266
Metastatic cancer	102 (0.01)	8 (0.01)	84 (0.01)	9 (0.02)	1 (0.00)	0.3023
Solid tumor without metastasis	690 (0.09)	80 (0.10)	538 (0.08)	46 (0.10)	26 (0.11)	0.2011
Rheumatoid arthritis/collagen vascular diseases	252 (0.03)	37 (0.05)	194 (0.03)	17 (0.04)	4 (0.02)	0.0659
Coagulopathy	640 (0.08)	207 (0.25)	343 (0.05)	71 (0.16)	19 (0.08)	<0.0001
Weight loss	282 (0.04)	54 (0.07)	185 (0.03)	29 (0.06)	14 (0.06)	<0.0001
Fluid and electrolyte disorders	1853 (0.23)	355 (0.43)	1254 (0.19)	181 (0.40)	63 (0.27)	<0.0001
Blood loss anemia	158 (0.02)	28 (0.03)	117 (0.02)	10 (0.02)	3 (0.01)	0.0181
Deficiency anemia	249 (0.03)	31 (0.04)	192 (0.03)	21 (0.05)	5 (0.02)	0.1399
Alcohol abuse	377 (0.05)	52 (0.06)	291 (0.05)	22 (0.05)	12 (0.05)	0.1558
Drug abuse	363 (0.05)	24 (0.03)	321 (0.05)	13 (0.03)	5 (0.02)	0.0043
Psychoses	142 (0.02)	12 (0.01)	121 (0.02)	6 (0.01)	3 (0.01)	0.6450
Depression	806 (0.10)	60 (0.07)	667 (0.10)	56 (0.12)	23 (0.10)	0.0321
Trauma	1223 (0.15)	88 (0.11)	1063 (0.16)	47 (0.10)	25 (0.11)	<0.0001

Table E.12: AKI Stage 2 comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by chi-squared test.

E.3 AKI Stage 3

E.3.1 Training and Testing

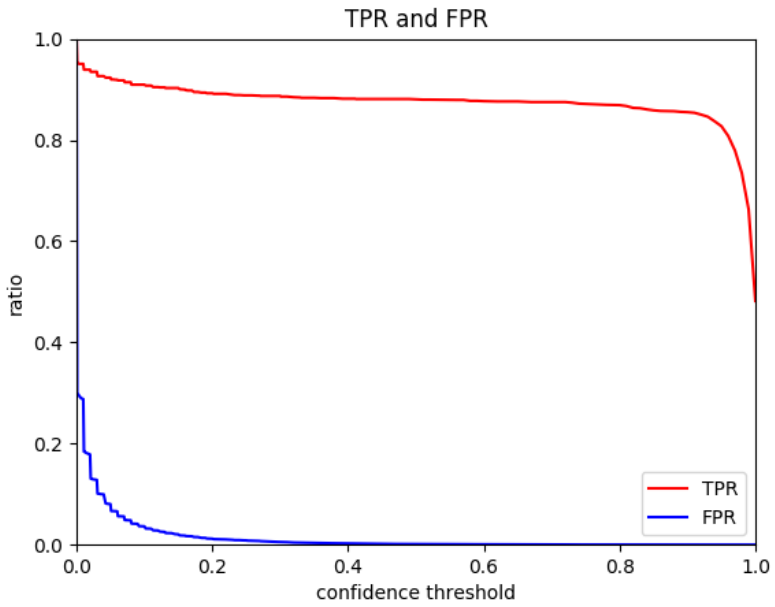


Figure E.27: AKI Stage 3 training/testing true positive and false positive rates.

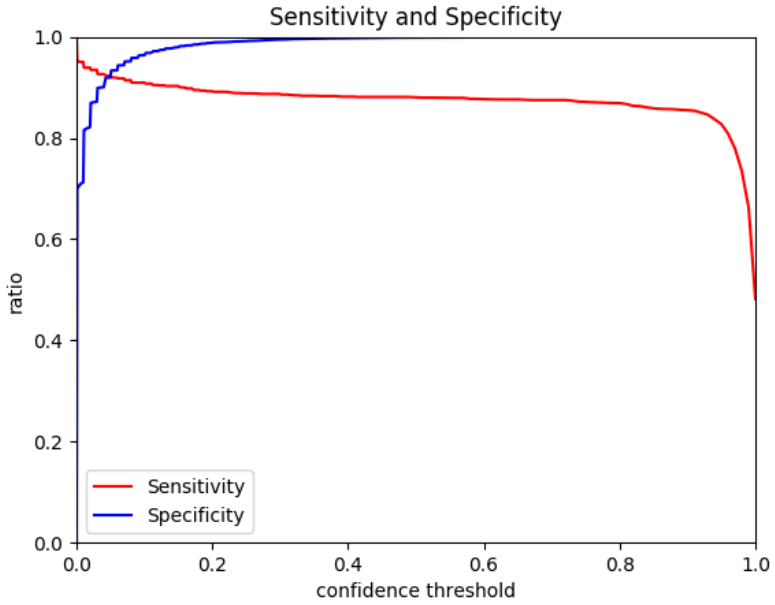


Figure E.28: AKI Stage 3 training/testing sensitivity and specificity.

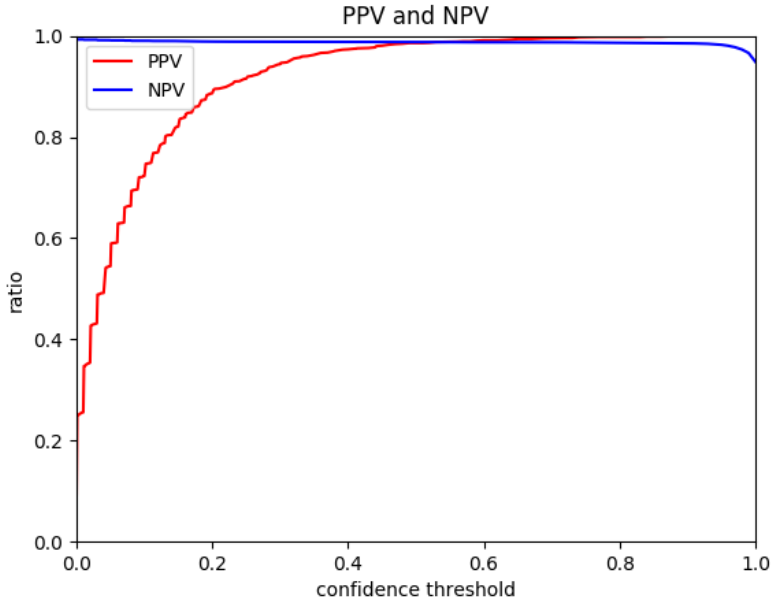


Figure E.29: AKI Stage 3 training/testing positive and negative predictive value.

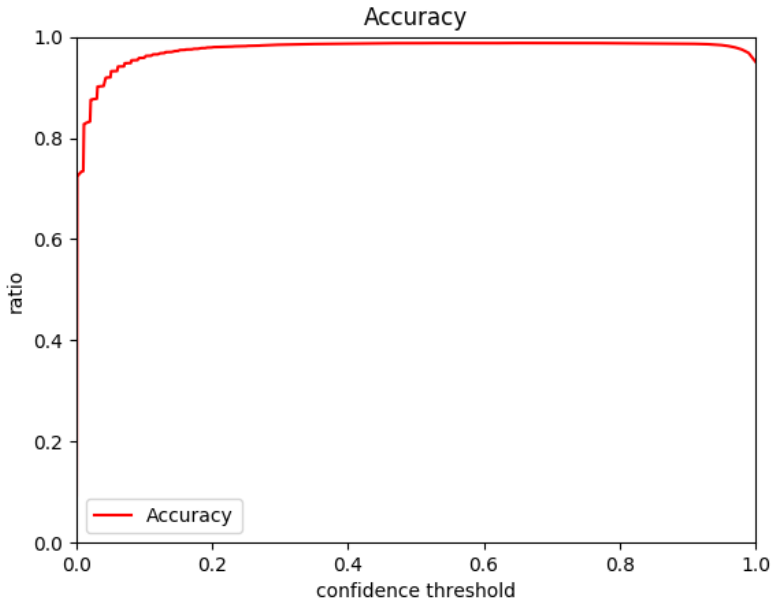


Figure E.30: AKI Stage 3 training/testing accuracy.

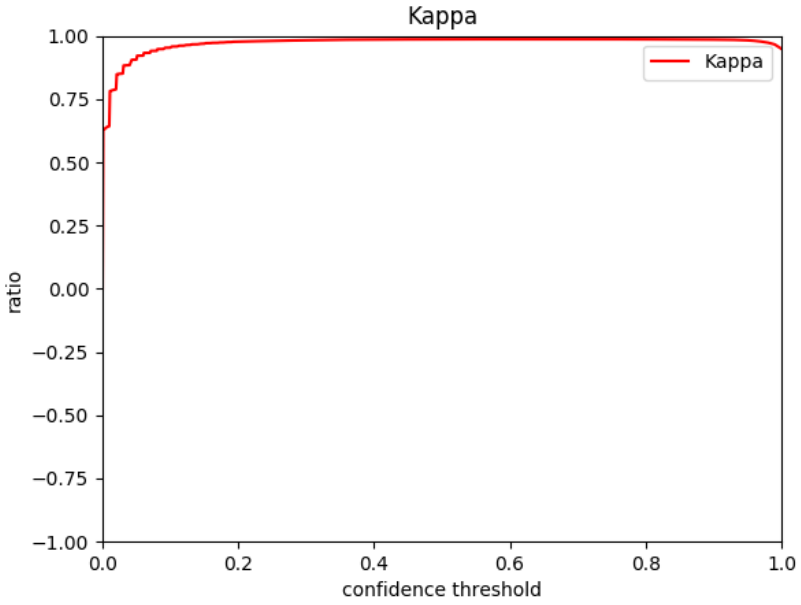


Figure E.31: AKI Stage 3 training/testing kappa.

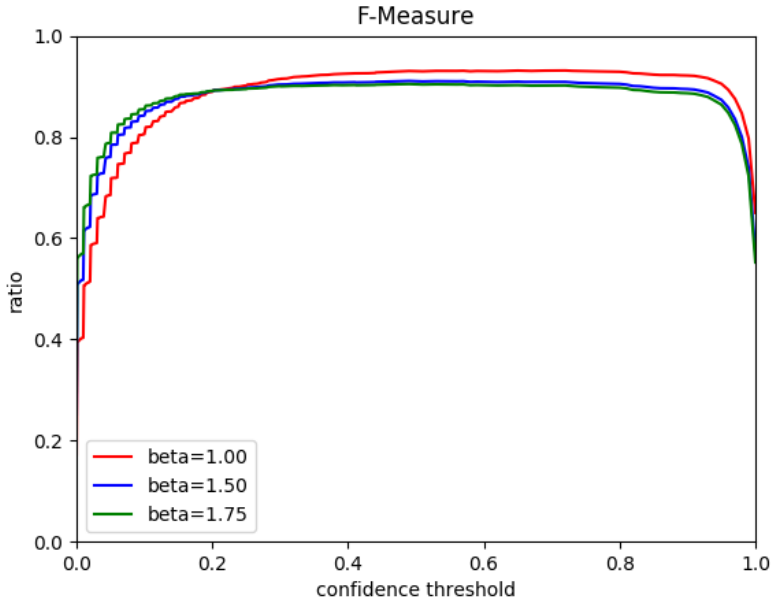


Figure E.32: AKI Stage 3 training/testing F-measures.

E.3.2 Validation

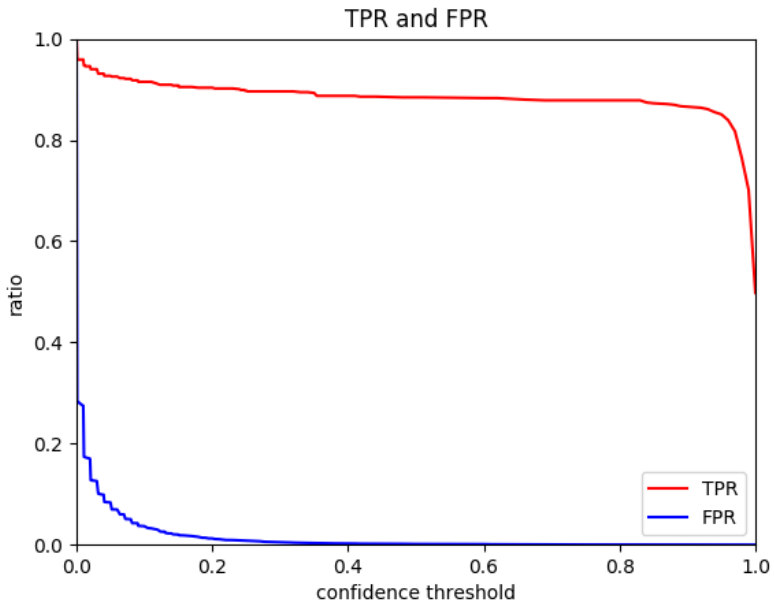


Figure E.33: AKI Stage 3 validation true and false positive rates.

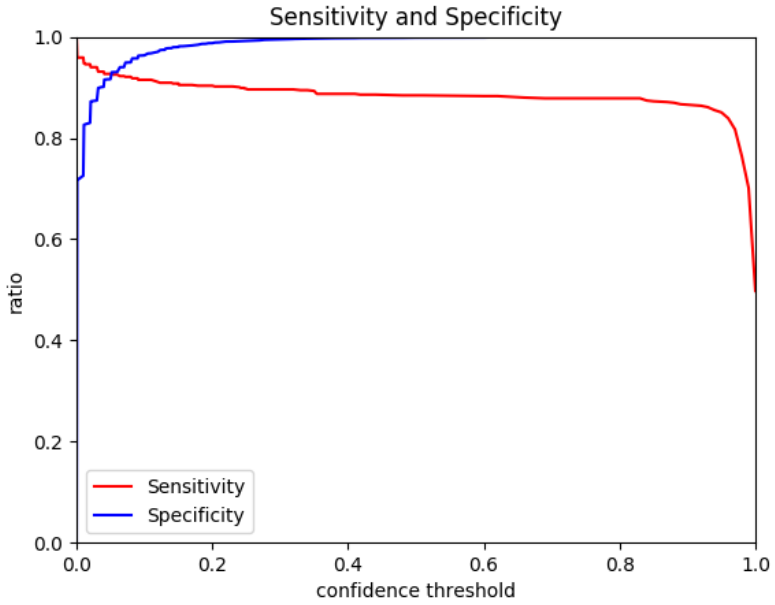


Figure E.34: AKI Stage 3 validation sensitivity and specificity.

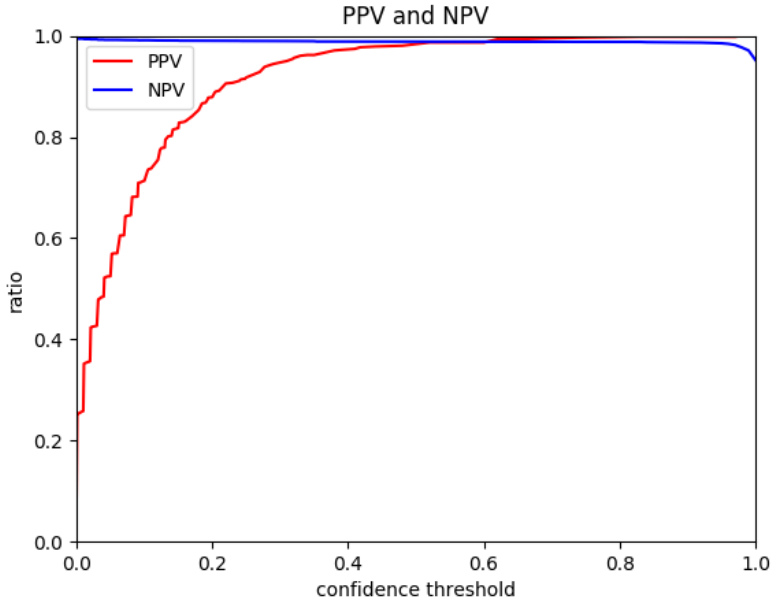


Figure E.35: AKI Stage 3 validation positive and negative predictive values.

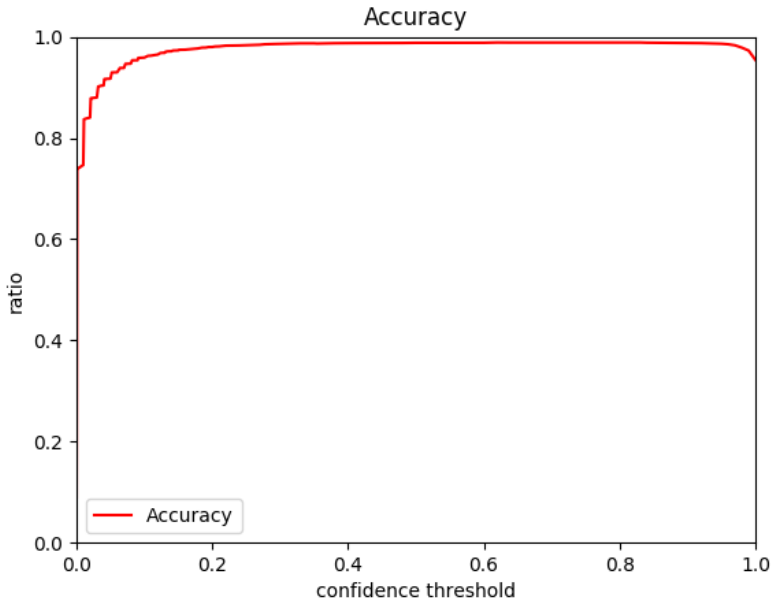


Figure E.36: AKI Stage 3 validation accuracy.

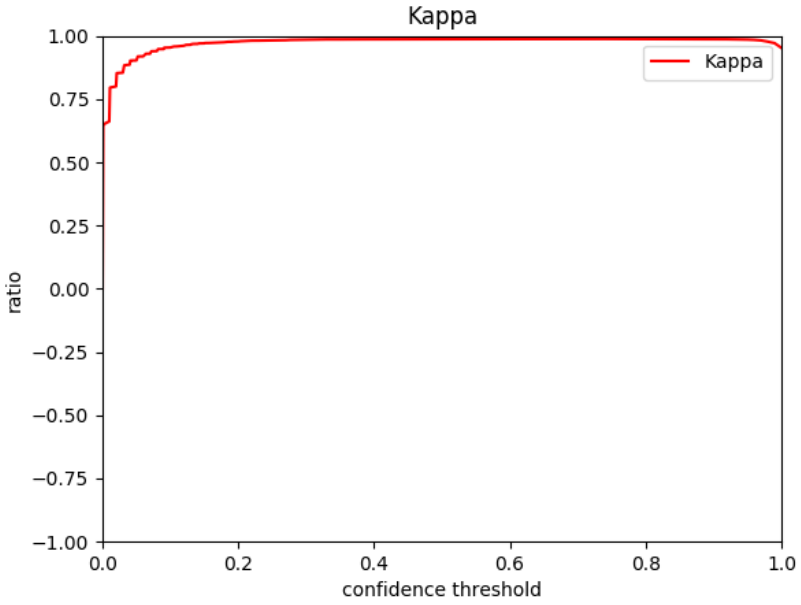


Figure E.37: AKI Stage 3 validation kappa.

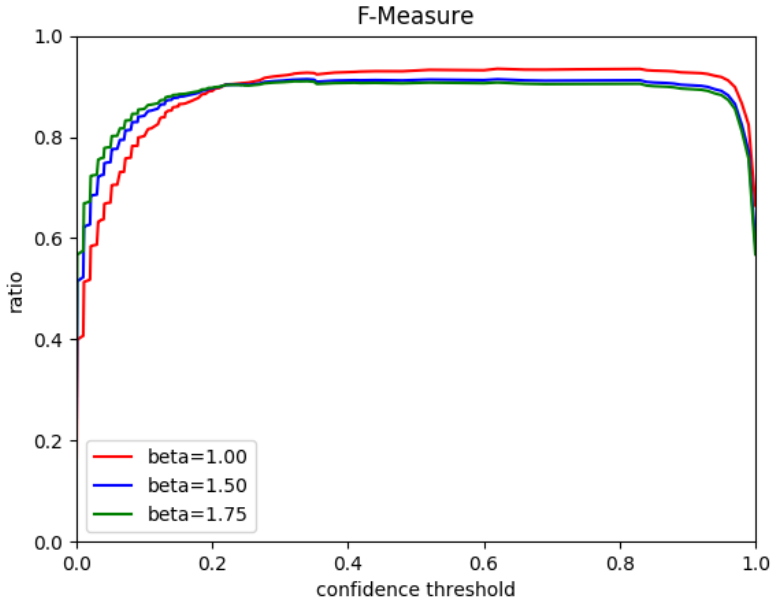
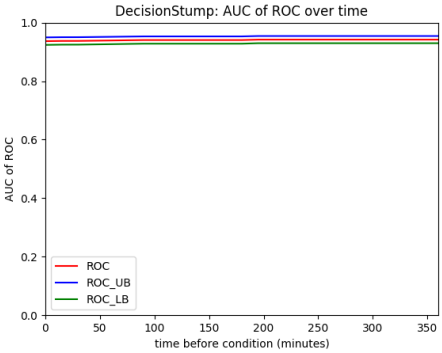
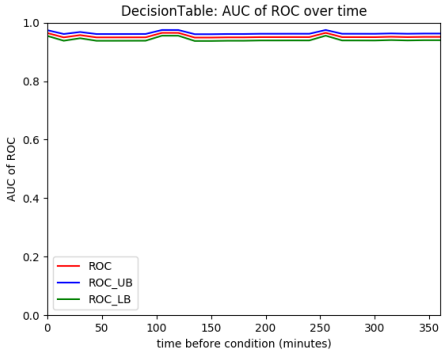


Figure E.38: AKI Stage 3 validation F-measures.

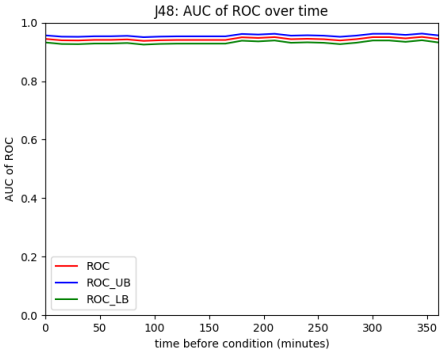
E.3.3 Machine Learning Algorithms over time



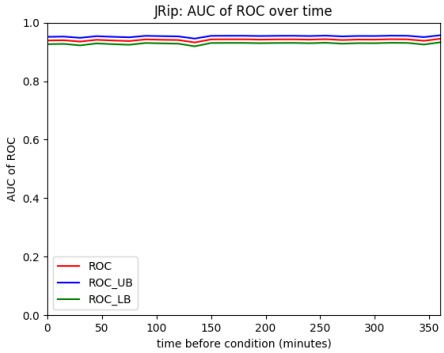
(a) Decision Stump



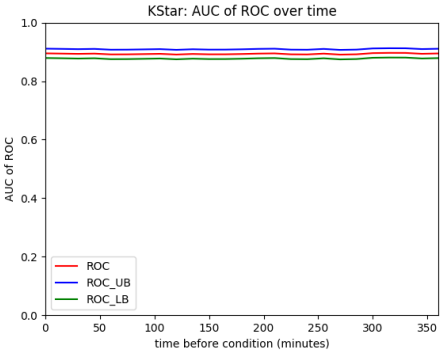
(b) Decision Table



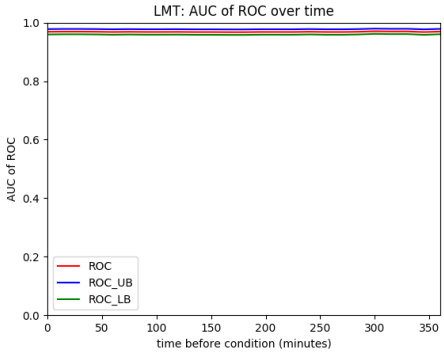
(c) J48



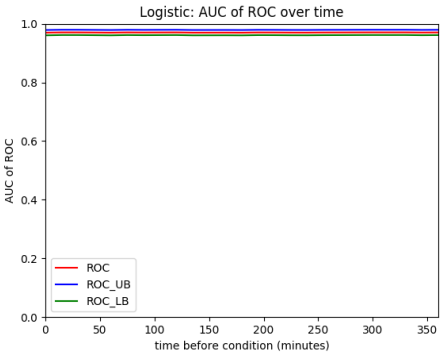
(d) JRip



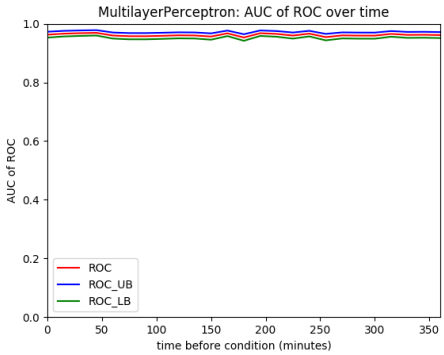
(e) KStar



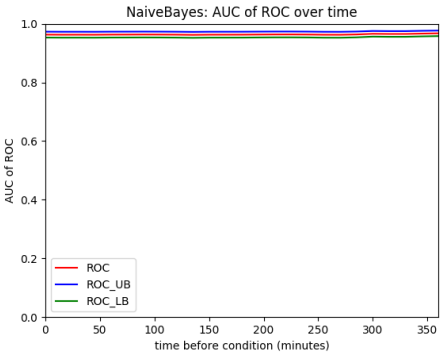
(f) LMT



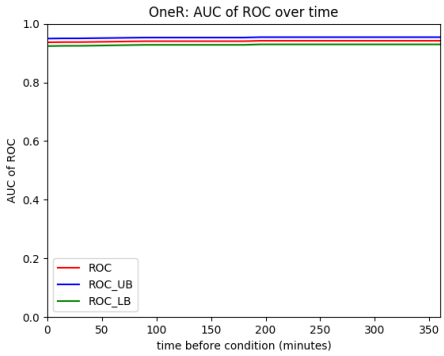
(g) Logistic Regression



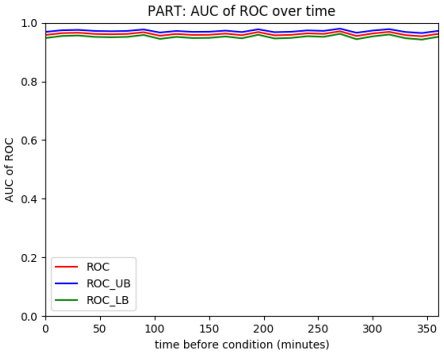
(h) Multilayer Perceptron



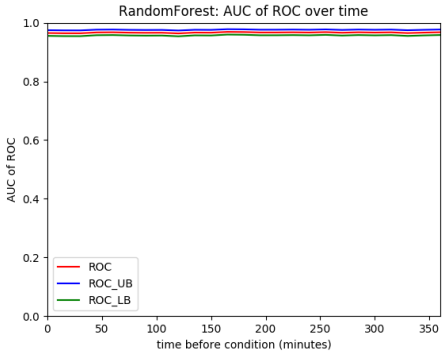
(i) Naive Bayes



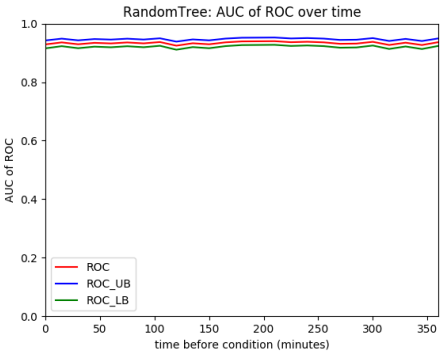
(j) One R



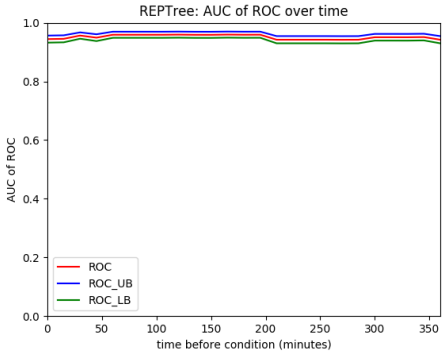
(k) PART

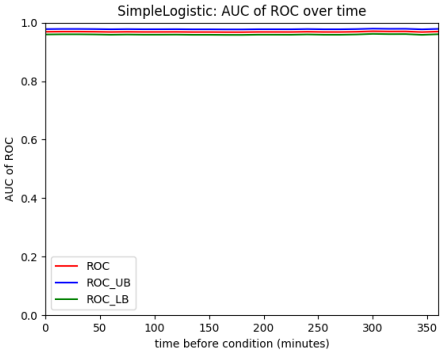


(l) Random Forest

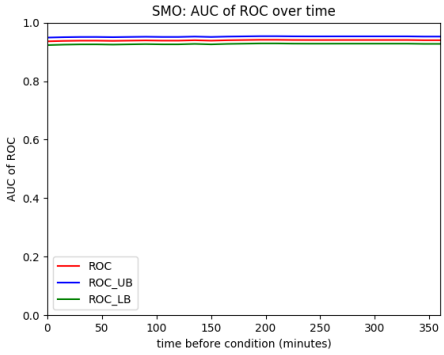


(m) Random Tree

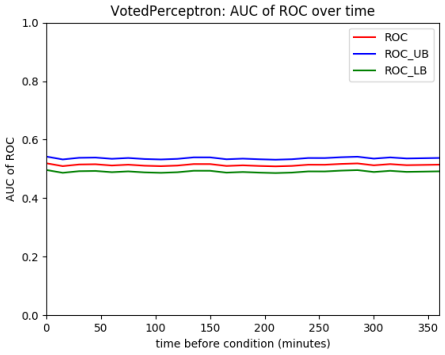




(n) Simple Logistic Regression



(o) SMO



(p) Voted Perceptron

Figure E.39: Algorithm's performance over time on the AKI Stage 3 selection criteria.

E.3.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Temperature	36.4	36.9	37.6	36.4	36.9	37.2	<0.0001
PO Intake	50	60	120	30	50	100	<0.0001
Hyaline Casts	1	1	2	1	1	1	<0.0001
Sodium, Urine	12	18	37	12	12	17	<0.0001
Creatinine	2.2	2.5	2.9	0.6	0.7	0.9	<0.0001
GT Flush	20	20	30	20	20	30	0.1998
Oxygen, blood gas	40	45	60	2	20	40	<0.0001
Urea Nitrogen	32.5	47	66	10	14	19	<0.0001
Albumin	2.3	2.7	3.3	2.6	3.2	3.7	<0.0001
NTproBNP	130	306	1005	51	140	295	<0.0001
Gastric Meds	30	30	60	10	30	60	<0.0001
Urea Nitrogen, Urine	85	148	343	60	85	148	<0.0001

Table E.13: AKI Stage 3 model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Temperature	36.4	37	37.6	36.4	36.9	37.2	36.4	36.9	37.3	36.375	36.8	37.225	<0.0001
PO Intake	50	60	120	30	50	100	30	50	100	30	40	60	<0.0001
Hyaline Casts	1	1	2	1	1	1	1	1	1	1	1	1	<0.0001
Sodium, Urine	12	21	40	12	12	17	12	13	20	11	13	17	<0.0001
Creatinine	2.2	2.6	3	0.6	0.7	0.9	0.7	0.9	1.2	0.6	0.9	1	<0.0001
GT Flush	20	20	30	20	20	30	20	20	30	20	20	30	0.0780
Oxygen, blood gas	40	45	60	2	20	40	2	40	50	6.75	45	52.75	<0.0001
Urea Nitrogen	36	50	68	10	14	19	13.75	21	31.25	13	18	26	<0.0001
Albumin	2.2	2.7	3.3	2.6	3.2	3.7	2.2	2.85	3.6	2.6	2.8	3.4	<0.0001
NTproBNP	136	357	1183.5	51	140	295	65	220.5	387	59	223	306	<0.0001
Gastric Meds	30	30	60	10	30	60	10	30	32.5	10	30	30	<0.0001
Urea Nitrogen, Urine	85	166	344	60	85	146	42	134	242.5	60	135	160.5	<0.0001

Table E.14: AKI Stage 3 model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	7531	683	6848	
Age, median	61	70	60	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.7389
Weight, median (Q1,Q3)	77.00 (65.00, 90.40)	78.00 (65.00, 94.80)	77.00 (65.00, 90.00)	0.1212
BMI, median (Q1,Q3)	27.78 (24.22, 31.85)	27.78 (24.24, 33.56)	27.78 (24.22, 31.61)	0.2078
Male	4154 (0.55)	394 (0.58)	3760 (0.55)	0.3509
In-Hospital Mortality	958 (0.13)	324 (0.47)	634 (0.09)	<0.0001
30 Day Mortality	1156 (0.15)	347 (0.51)	809 (0.12)	<0.0001
ICU LOS, median(Q1,Q3)	1.90 (1.13, 3.27)	5.82 (2.42, 12.41)	1.81 (1.09, 2.99)	<0.0001
ICU				
CCU	987 (0.13)	143 (0.21)	844 (0.12)	<0.0001
CSRU	1126 (0.15)	100 (0.15)	1026 (0.15)	0.8260
MICU	2898 (0.38)	290 (0.42)	2608 (0.38)	0.0788
SICU	1400 (0.19)	111 (0.16)	1289 (0.19)	0.1372
TSICU	1120 (0.15)	39 (0.06)	1081 (0.16)	<0.0001
ethnicity				
Asian	188 (0.02)	14 (0.02)	174 (0.03)	0.4386
Black	611 (0.08)	82 (0.12)	529 (0.08)	0.0002
Hispanic	293 (0.04)	23 (0.03)	270 (0.04)	0.4673
White	5426 (0.72)	452 (0.66)	4974 (0.73)	0.0580
insurance				
Government	266 (0.04)	7 (0.01)	259 (0.04)	0.0003
Medicaid	746 (0.10)	51 (0.07)	695 (0.10)	0.0337
Medicare	3531 (0.47)	457 (0.67)	3074 (0.45)	<0.0001
Private	2862 (0.38)	166 (0.24)	2696 (0.39)	<0.0001
Self Pay	126 (0.02)	2 (0.00)	124 (0.02)	0.0034

Table E.15: AKI Stage 3 demographics of validation set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	7531	79	9	6839	604	
Age, median	61	66	73	60	70	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 69.00)	0.00 (0.00, 0.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.1483
Weight, median (Q1,Q3)	77.00 (65.00, 90.40)	72.05 (63.00, 88.25)	73.20 (73.20, 73.20)	77.00 (65.00, 90.00)	80.00 (65.50, 95.00)	0.0874
BMI, median (Q1,Q3)	27.78 (24.22, 31.85)	27.64 (24.64, 31.55)	0.00 (0.00, 0.00)	27.78 (24.22, 31.61)	27.78 (24.24, 34.34)	0.1665
Male	4154 (0.55)	36 (0.46)	7 (0.78)	3753 (0.55)	358 (0.59)	0.2506
In-Hospital Mortality	958 (0.13)	30 (0.38)	2 (0.22)	632 (0.09)	294 (0.49)	<0.0001
30 Day Mortality	1156 (0.15)	32 (0.41)	2 (0.22)	807 (0.12)	315 (0.52)	<0.0001
ICU LOS, median(Q1,Q3)	1.90 (1.13, 3.27)	4.08 (2.04, 8.15)	1.79 (1.20, 2.09)	1.81 (1.09, 3.00)	6.20 (2.52, 12.91)	<0.0001
ICU						
CCU	987 (0.13)	16 (0.20)	1 (0.11)	843 (0.12)	127 (0.21)	<0.0001
CSRU	1126 (0.15)	14 (0.18)	1 (0.11)	1025 (0.15)	86 (0.14)	0.8719
MICU	2898 (0.38)	33 (0.42)	5 (0.56)	2603 (0.38)	257 (0.43)	0.2820
SICU	1400 (0.19)	13 (0.16)	1 (0.11)	1288 (0.19)	98 (0.16)	0.4755
TSICU	1120 (0.15)	3 (0.04)	1 (0.11)	1080 (0.16)	36 (0.06)	<0.0001
ethnicity						
Asian	188 (0.02)	1 (0.01)	0 (0.00)	174 (0.03)	13 (0.02)	0.7884
Black	611 (0.08)	9 (0.11)	4 (0.44)	525 (0.08)	73 (0.12)	<0.0001
Hispanic	293 (0.04)	2 (0.03)	0 (0.00)	270 (0.04)	21 (0.03)	0.7895
White	5426 (0.72)	47 (0.59)	5 (0.56)	4969 (0.73)	405 (0.67)	0.2113
insurance						
Government	266 (0.04)	1 (0.01)	0 (0.00)	259 (0.04)	6 (0.01)	0.0033
Medicaid	746 (0.10)	8 (0.10)	0 (0.00)	695 (0.10)	43 (0.07)	0.1076
Medicare	3531 (0.47)	51 (0.65)	6 (0.67)	3068 (0.45)	406 (0.67)	<0.0001
Private	2862 (0.38)	19 (0.24)	2 (0.22)	2694 (0.39)	147 (0.24)	<0.0001
Self Pay	126 (0.02)	0 (0.00)	1 (0.11)	123 (0.02)	2 (0.00)	0.0041

Table E.16: AKI Stage 3 demographics of validation set patients in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	7531	683	6848	
Comorbidity				
Congestive heart failure	1437 (0.19)	342 (0.50)	1095 (0.16)	<0.0001
Cardiac arrhythmias	2219 (0.29)	314 (0.46)	1905 (0.28)	<0.0001
Valvular disease	915 (0.12)	124 (0.18)	791 (0.12)	<0.0001
Pulmonary circulation disorders	282 (0.04)	53 (0.08)	229 (0.03)	<0.0001
Peripheral vascular disorders	706 (0.09)	121 (0.18)	585 (0.09)	<0.0001
Hypertension, uncomplicated	3258 (0.43)	188 (0.28)	3070 (0.45)	<0.0001
Hypertension, complicated	351 (0.05)	195 (0.29)	156 (0.02)	<0.0001
Paralysis	246 (0.03)	15 (0.02)	231 (0.03)	0.1046
Other neurological disorders	577 (0.08)	38 (0.06)	539 (0.08)	0.0378
Chronic pulmonary disease	1642 (0.22)	169 (0.25)	1473 (0.22)	0.0844
Diabetes, uncomplicated	1352 (0.18)	169 (0.25)	1183 (0.17)	<0.0001
Diabetes, complicated	304 (0.04)	84 (0.12)	220 (0.03)	<0.0001
Hypothyroidism	752 (0.10)	80 (0.12)	672 (0.10)	0.1340
Renal failure	417 (0.06)	223 (0.33)	194 (0.03)	<0.0001
Liver disease	687 (0.09)	168 (0.25)	519 (0.08)	<0.0001
Peptic ulcer disease excluding bleeding	50 (0.01)	3 (0.00)	47 (0.01)	0.4498
AIDS/HIV	63 (0.01)	10 (0.01)	53 (0.01)	0.0600
Lymphoma	102 (0.01)	11 (0.02)	91 (0.01)	0.5464
Metastatic cancer	91 (0.01)	1 (0.00)	90 (0.01)	0.0081
Solid tumor without metastasis	636 (0.08)	56 (0.08)	580 (0.08)	0.8166
Rheumatoid arthritis/collagen vascular diseases	246 (0.03)	35 (0.05)	211 (0.03)	0.0048
Coagulopathy	561 (0.07)	150 (0.22)	411 (0.06)	<0.0001
Weight loss	280 (0.04)	67 (0.10)	213 (0.03)	<0.0001
Fluid and electrolyte disorders	1765 (0.23)	337 (0.49)	1428 (0.21)	<0.0001
Blood loss anemia	152 (0.02)	26 (0.04)	126 (0.02)	0.0006
Deficiency anemia	233 (0.03)	21 (0.03)	212 (0.03)	0.9761
Alcohol abuse	353 (0.05)	42 (0.06)	311 (0.05)	0.0642
Drug abuse	354 (0.05)	21 (0.03)	333 (0.05)	0.0399
Psychoses	136 (0.02)	10 (0.01)	126 (0.02)	0.4858
Depression	769 (0.10)	51 (0.07)	718 (0.10)	0.0186
Trauma	1176 (0.16)	72 (0.11)	1104 (0.16)	0.0004

Table E.17: AKI Stage 3 comorbidities of validation set patients by condition positive or negative. P-values for patient counts calculated by chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	7531	79	9	6839	604	
Comorbidity						
Congestive heart failure	1437 (0.19)	26 (0.33)	2 (0.22)	1093 (0.16)	316 (0.52)	<0.0001
Cardiac arrhythmias	2219 (0.29)	34 (0.43)	1 (0.11)	1904 (0.28)	280 (0.46)	<0.0001
Valvular disease	915 (0.12)	13 (0.16)	2 (0.22)	789 (0.12)	111 (0.18)	<0.0001
Pulmonary circulation disorders	282 (0.04)	4 (0.05)	0 (0.00)	229 (0.03)	49 (0.08)	<0.0001
Peripheral vascular disorders	706 (0.09)	11 (0.14)	1 (0.11)	584 (0.09)	110 (0.18)	<0.0001
Hypertension, uncomplicated	3258 (0.43)	34 (0.43)	5 (0.56)	3065 (0.45)	154 (0.25)	<0.0001
Hypertension, complicated	351 (0.05)	8 (0.10)	1 (0.11)	155 (0.02)	187 (0.31)	<0.0001
Paralysis	246 (0.03)	1 (0.01)	1 (0.11)	230 (0.03)	14 (0.02)	0.2103
Other neurological disorders	577 (0.08)	5 (0.06)	0 (0.00)	539 (0.08)	33 (0.05)	0.1638
Chronic pulmonary disease	1642 (0.22)	17 (0.22)	2 (0.22)	1471 (0.22)	152 (0.25)	0.3330
Diabetes, uncomplicated	1352 (0.18)	15 (0.19)	2 (0.22)	1181 (0.17)	154 (0.25)	0.0001
Diabetes, complicated	304 (0.04)	4 (0.05)	2 (0.22)	218 (0.03)	80 (0.13)	<0.0001
Hypothyroidism	752 (0.10)	6 (0.08)	1 (0.11)	671 (0.10)	74 (0.12)	0.2865
Renal failure	417 (0.06)	9 (0.11)	1 (0.11)	193 (0.03)	214 (0.35)	<0.0001
Liver disease	687 (0.09)	11 (0.14)	0 (0.00)	519 (0.08)	157 (0.26)	<0.0001
Peptic ulcer disease excluding bleeding	50 (0.01)	0 (0.00)	0 (0.00)	47 (0.01)	3 (0.00)	0.8267
AIDS/HIV	63 (0.01)	2 (0.03)	0 (0.00)	53 (0.01)	8 (0.01)	0.1856
Lymphoma	102 (0.01)	0 (0.00)	0 (0.00)	91 (0.01)	11 (0.02)	0.5335
Metastatic cancer	91 (0.01)	0 (0.00)	0 (0.00)	90 (0.01)	1 (0.00)	0.0671
Solid tumor without metastasis	636 (0.08)	10 (0.13)	0 (0.00)	580 (0.08)	46 (0.08)	0.4037
Rheumatoid arthritis/collagen vascular diseases	246 (0.03)	4 (0.05)	0 (0.00)	211 (0.03)	31 (0.05)	0.0420
Coagulopathy	561 (0.07)	13 (0.16)	0 (0.00)	411 (0.06)	137 (0.23)	<0.0001
Weight loss	280 (0.04)	5 (0.06)	0 (0.00)	213 (0.03)	62 (0.10)	<0.0001
Fluid and electrolyte disorders	1765 (0.23)	32 (0.41)	1 (0.11)	1427 (0.21)	305 (0.50)	<0.0001
Blood loss anemia	152 (0.02)	1 (0.01)	0 (0.00)	126 (0.02)	25 (0.04)	0.0019
Deficiency anemia	233 (0.03)	4 (0.05)	0 (0.00)	212 (0.03)	17 (0.03)	0.7004
Alcohol abuse	353 (0.05)	4 (0.05)	0 (0.00)	311 (0.05)	38 (0.06)	0.2565
Drug abuse	354 (0.05)	4 (0.05)	0 (0.00)	333 (0.05)	17 (0.03)	0.1429
Psychoses	136 (0.02)	5 (0.06)	0 (0.00)	126 (0.02)	5 (0.01)	0.0062
Depression	769 (0.10)	6 (0.08)	1 (0.11)	717 (0.10)	45 (0.07)	0.1360
Trauma	1176 (0.16)	6 (0.08)	1 (0.11)	1103 (0.16)	66 (0.11)	0.0046

Table E.18: AKI Stage 3 comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by chi-squared test.

E.4 AKI progression from Stage 1 to Stage 2

E.4.1 Training and Testing

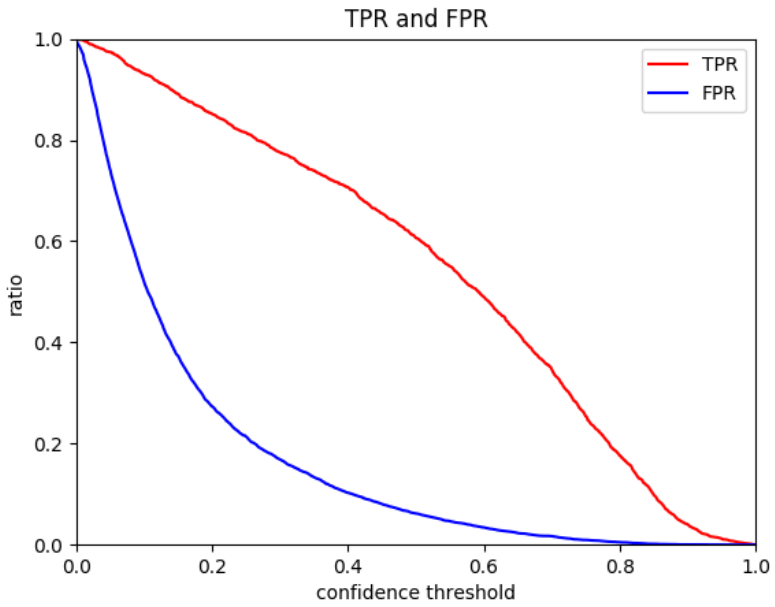


Figure E.40: AKI progression from Stage 1 to Stage 2 training/testing true positive and false positive rates.

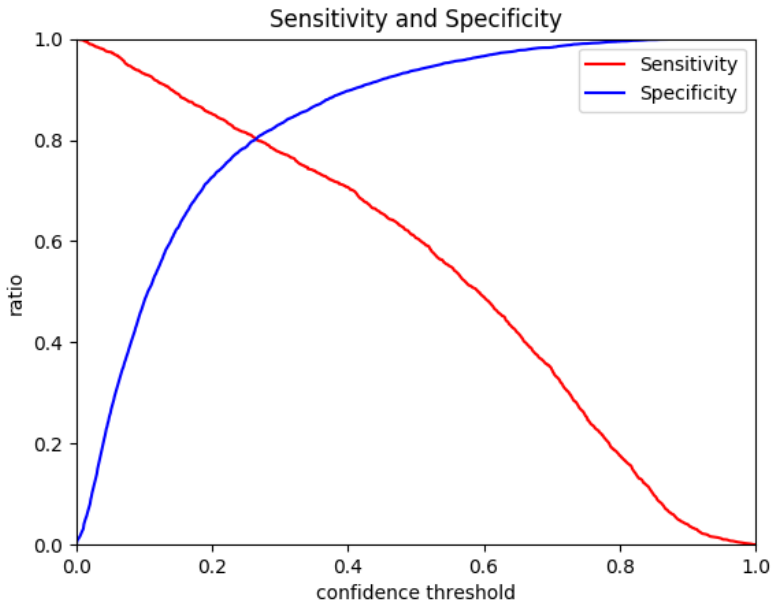


Figure E.41: AKI progression from Stage 1 to Stage 2 training/testing sensitivity and specificity.

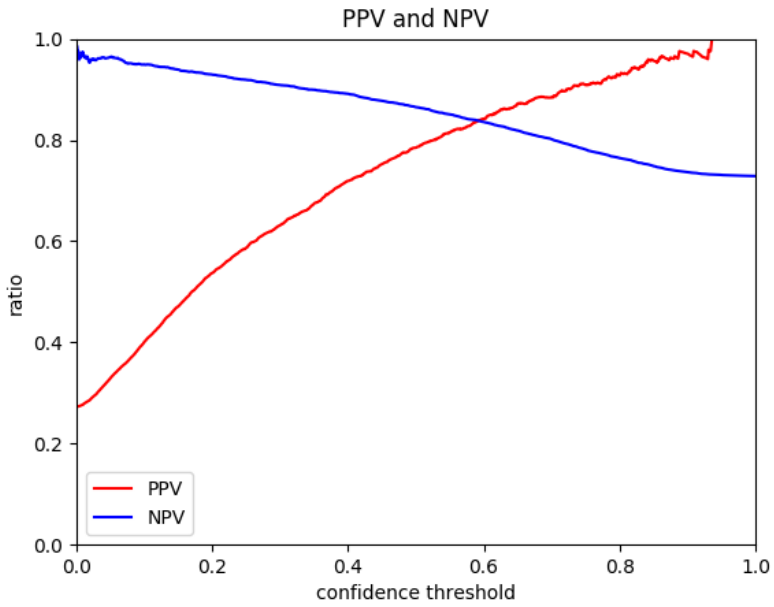


Figure E.42: AKI progression from Stage 1 to Stage 2 training/testing positive and negative predictive value.

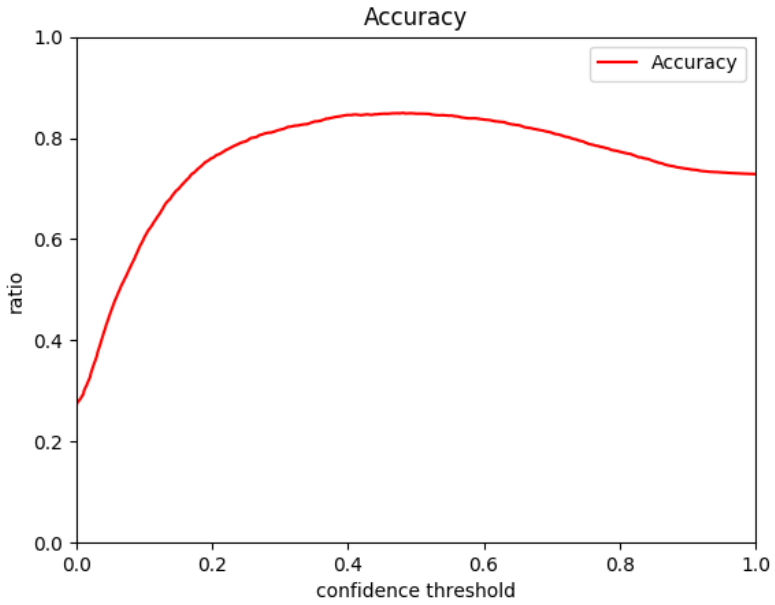


Figure E.43: AKI progression from Stage 1 to Stage 2 training/testing accuracy.

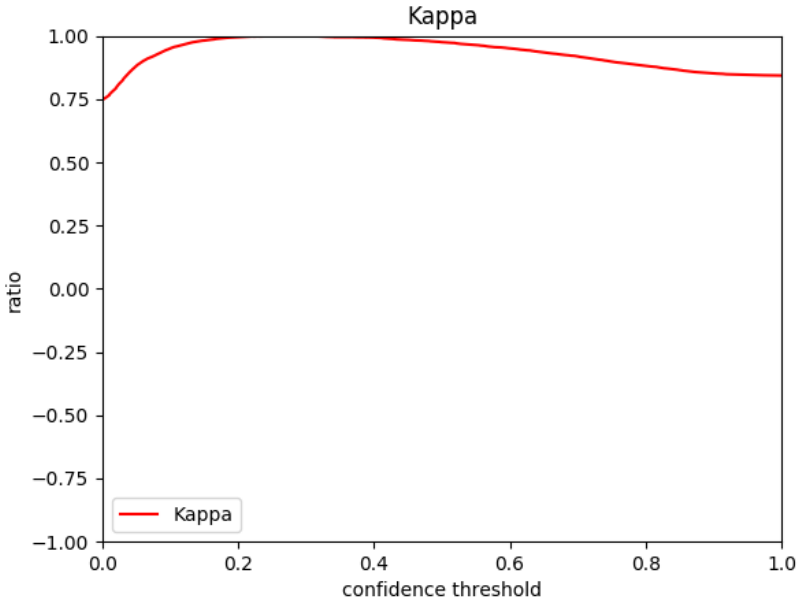


Figure E.44: AKI progression from Stage 1 to Stage 2 training/testing kappa.

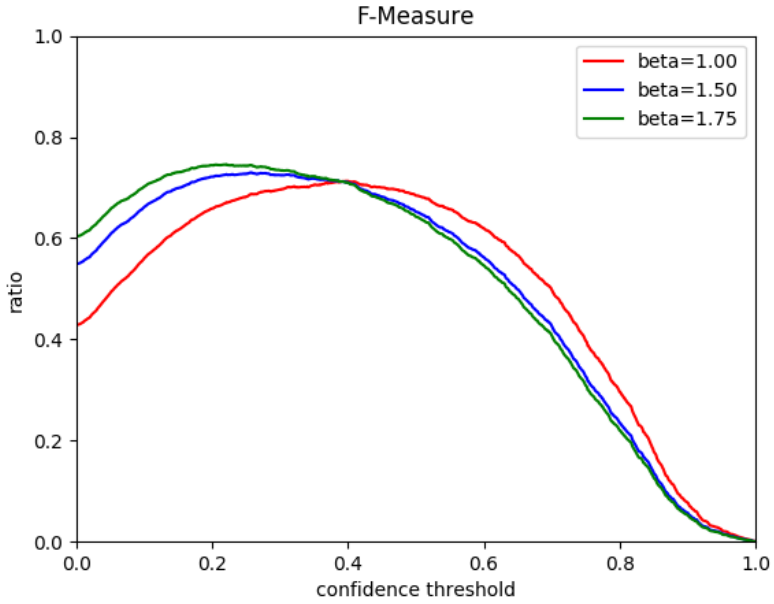


Figure E.45: AKI progression from Stage 1 to Stage 2 training/testing F-measures.

E.4.2 Validation

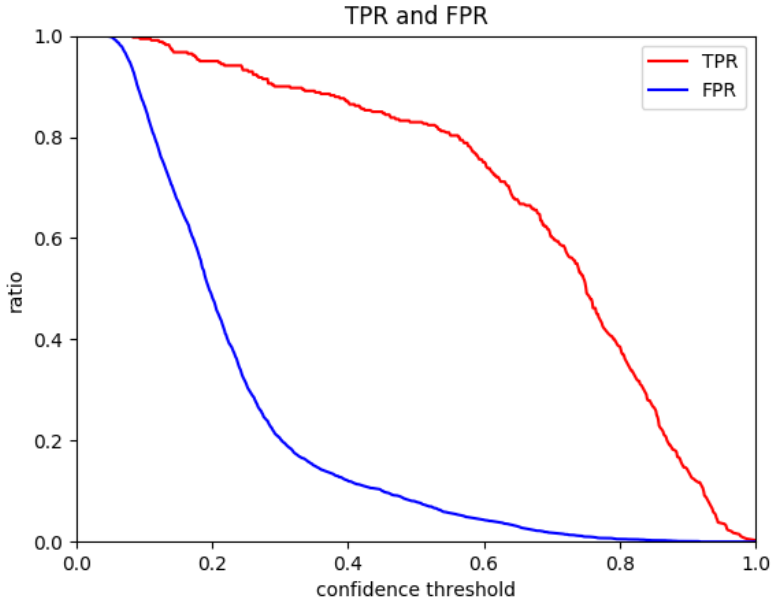


Figure E.46: AKI progression from Stage 1 to Stage 2 validation true and false positive rates.

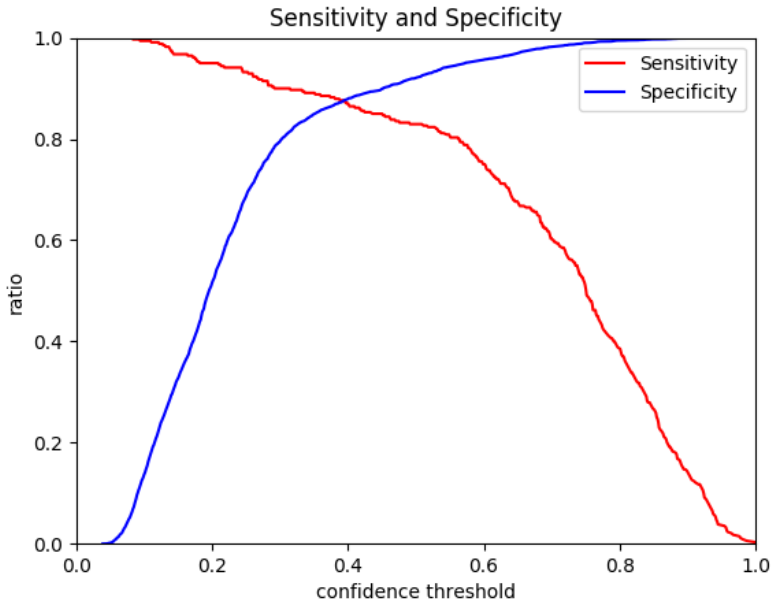


Figure E.47: AKI progression from Stage 1 to Stage 2 validation sensitivity and specificity.

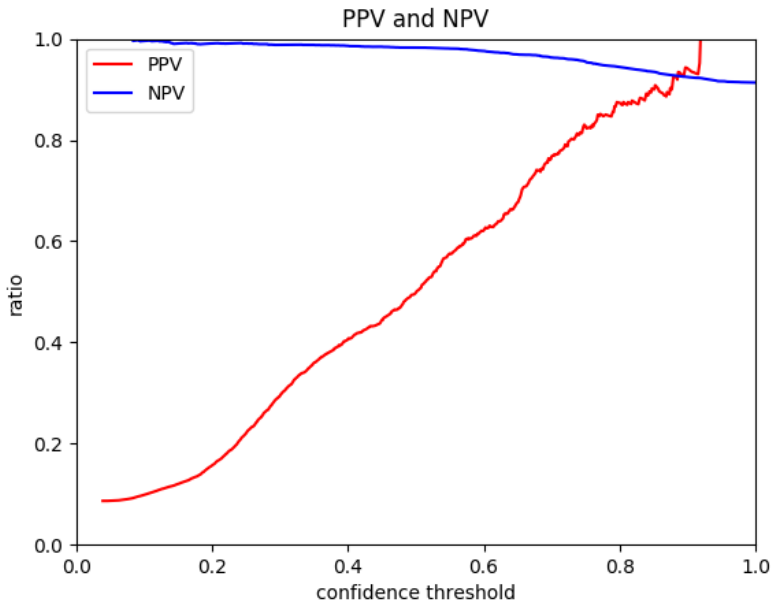


Figure E.48: AKI progression from Stage 1 to Stage 2 validation positive and negative predictive values.

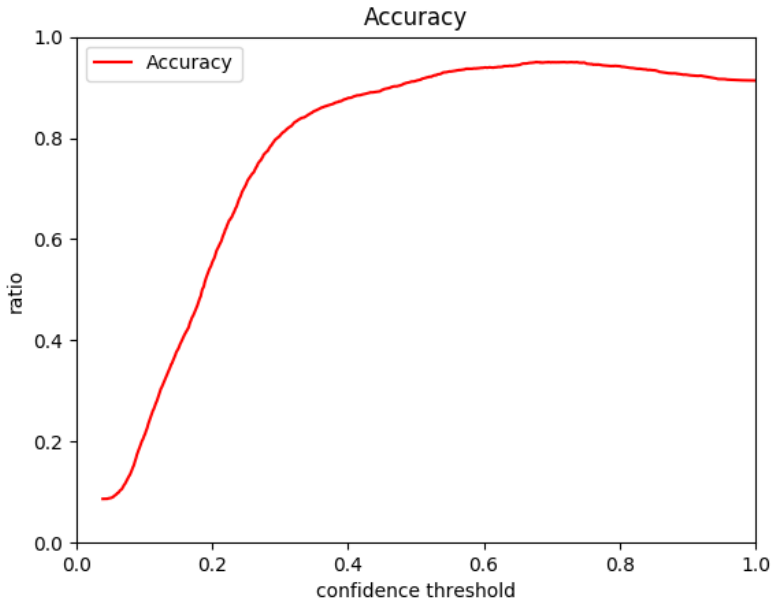


Figure E.49: AKI progression from Stage 1 to Stage 2 validation accuracy.

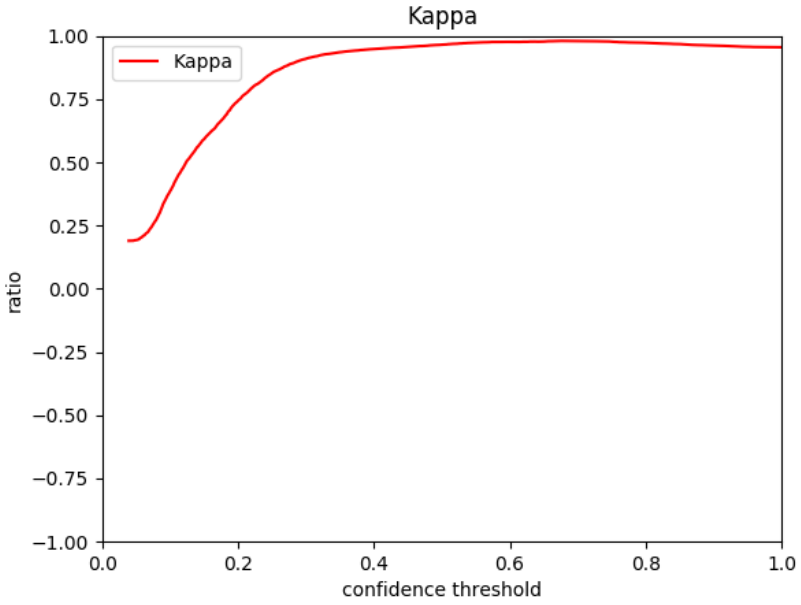


Figure E.50: AKI progression from Stage 1 to Stage 2 validation kappa.

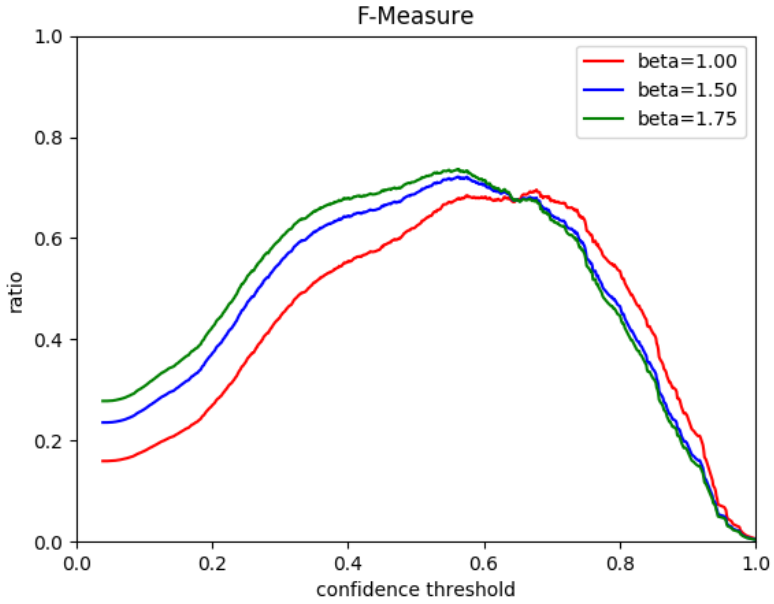
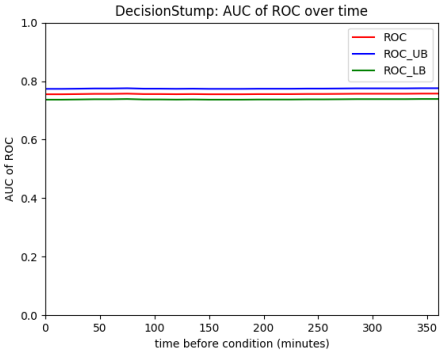
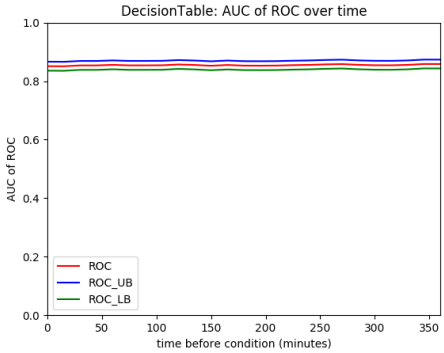


Figure E.51: AKI progression from Stage 1 to Stage 2 validation F-measures.

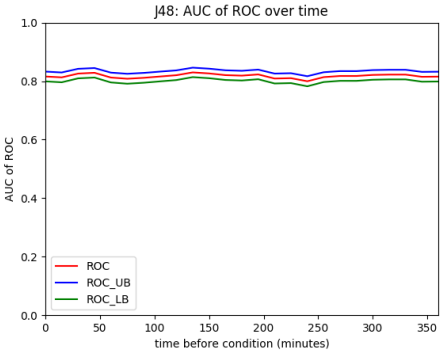
E.4.3 Machine Learning Algorithms over time



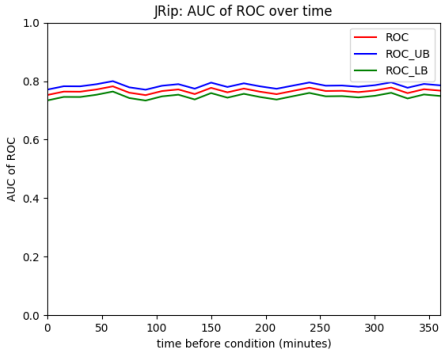
(a) Decision Stump



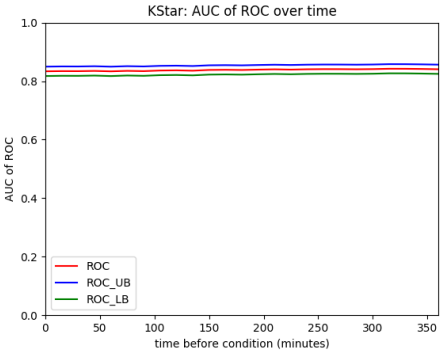
(b) Decision Table



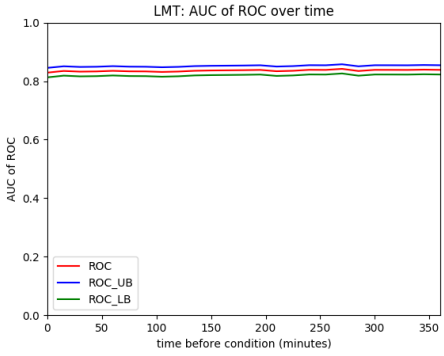
(c) J48



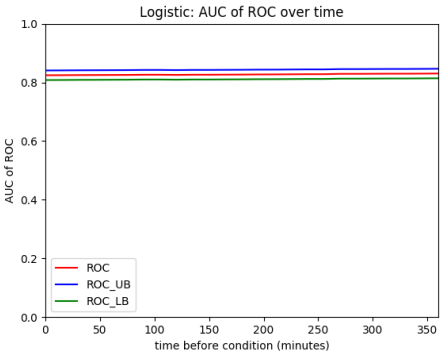
(d) JRip



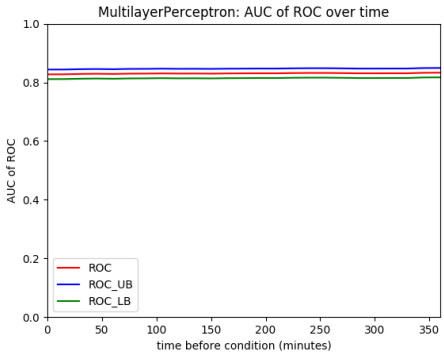
(e) KStar



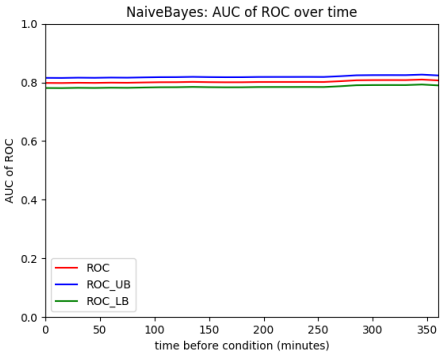
(f) LMT



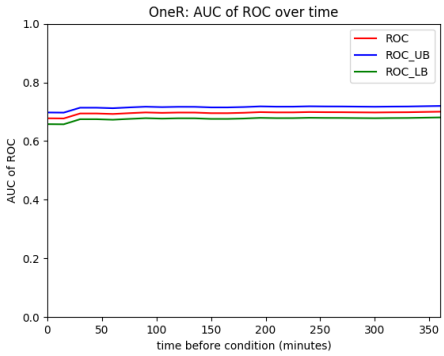
(g) Logistic Regression



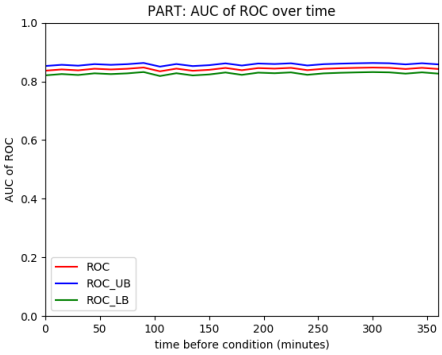
(h) Multilayer Perceptron



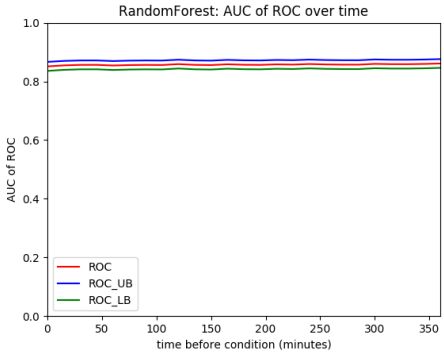
(i) Naive Bayes



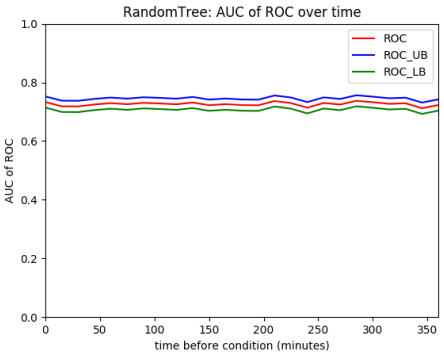
(j) One R



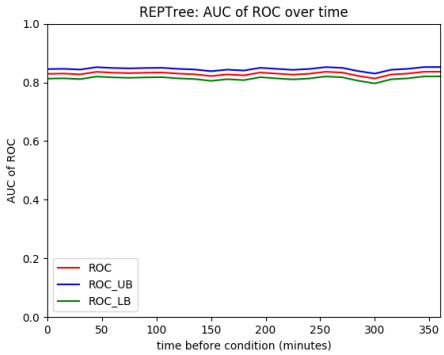
(k) PART



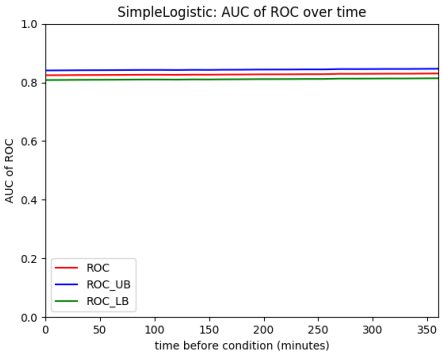
(l) Random Forest



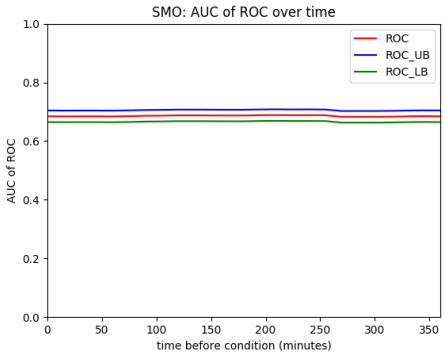
(m) Random Tree



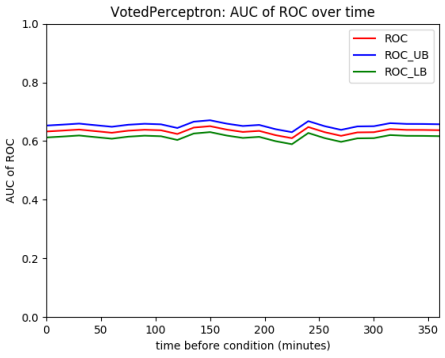
(n) REP Tree



(o) Simple Logistic Regression



(p) SMO



(q) Voted Perceptron

Figure E.52: Algorithm's performance over time on the AKI progression from Stage 1 to Stage 2 selection criteria.

E.4.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Fentanyl	0.098	0.256	0.750	0.04	0.167	0.918	<0.0001
Oxygen, blood gas	4	40	54	3	21	50	<0.0001
Alanine Aminotransferase (ALT)	12	22	46	10	17	31	<0.0001
Cholesterol, HDL	4	8	9	4	8	8	0.6124
Urea Nitrogen	23	32	45	12	17	22	<0.0001
Gastric Meds	4	10	20	0	10	20	<0.0001
Phosphate, blood chemistry	3	3.7	4.6	2.4	3.2	3.9	<0.0001

Table E.19: AKI progression from Stage 1 to Stage 2 model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Fentanyl	0.100	0.292	0.663	0.040	0.167	0.918	0.050	0.225	0.893	0.098	0.256	0.756	<0.0001
Oxygen, blood gas	4	40	60	3	21	50	4	31	50	4	40	50	<0.0001
Alanine Aminotransferase (ALT)	10	17	39.5	10	17	31	10	15	30	12	22	47	<0.0001
Cholesterol, HDL	4	8	8	4	8	8	4	8	8	4	8	9	0.6479
Urea Nitrogen	21	28	42	12	17	22	13	17	22	23	33	45	<0.0001
Gastric Meds	4	10	20	0	10	20	0	10	20	4	10	20	<0.0001
Phosphate, blood chemistry	2.8	3.3	4.3	2.4	3.2	3.875	2.4	3.1	3.9	3	3.7	4.7	<0.0001

Table E.20: AKI progression from Stage 1 to Stage 2 model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	3924	340	3584	
Age, median	67	68	67	*
Height, median (Q1,Q3)	68.00 (64.00, 70.00)	67.00 (64.00, 70.00)	68.00 (64.00, 70.00)	0.6832
Weight, median (Q1,Q3)	81.00 (68.10, 95.00)	81.60 (67.30, 95.40)	81.00 (68.30, 95.00)	0.8234
BMI, median (Q1,Q3)	28.63 (25.00, 33.07)	28.57 (24.49, 34.82)	28.64 (25.04, 33.01)	0.7523
Male	2462 (0.63)	201 (0.59)	2261 (0.63)	0.3773
In-Hospital Mortality	814 (0.21)	63 (0.19)	751 (0.21)	0.3481
30 Day Mortality	899 (0.23)	68 (0.20)	831 (0.23)	0.2407
ICU LOS, median(Q1,Q3)	3.13 (1.71, 6.97)	3.30 (1.79, 7.79)	3.11 (1.70, 6.92)	0.3340
ICU				
CCU	600 (0.15)	41 (0.12)	559 (0.16)	0.1108
CSRU	1135 (0.29)	108 (0.32)	1027 (0.29)	0.3083
MICU	1204 (0.31)	104 (0.31)	1100 (0.31)	0.9737
SICU	569 (0.15)	44 (0.13)	525 (0.15)	0.4295
TSICU	416 (0.11)	43 (0.13)	373 (0.10)	0.2254
ethnicity				
Asian	88 (0.02)	11 (0.03)	77 (0.02)	0.2009
Black	277 (0.07)	31 (0.09)	246 (0.07)	0.1350
Hispanic	116 (0.03)	10 (0.03)	106 (0.03)	0.9866
White	2810 (0.72)	242 (0.71)	2568 (0.72)	0.9212
insurance				
Government	68 (0.02)	8 (0.02)	60 (0.02)	0.3635
Medicaid	321 (0.08)	27 (0.08)	294 (0.08)	0.8718
Medicare	2285 (0.58)	197 (0.58)	2088 (0.58)	0.9415
Private	1215 (0.31)	104 (0.31)	1111 (0.31)	0.8965
Self Pay	35 (0.01)	4 (0.01)	31 (0.01)	0.5611

Table E.21: AKI progression from Stage 1 to Stage 2 demographics of validation set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	3924	79	166	3418	261	
Age, median	67	68	68	67	68	*
Height, median (Q1,Q3)	68.00 (64.00, 70.00)	68.00 (65.75, 70.00)	66.00 (63.88, 70.00)	68.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.4069
Weight, median (Q1,Q3)	81.00 (68.10, 95.00)	80.00 (67.30, 91.25)	80.00 (63.18, 98.00)	81.00 (68.70, 95.00)	83.10 (67.33, 99.95)	0.9017
BMI, median (Q1,Q3)	28.63 (25.00, 33.07)	26.77 (24.25, 33.84)	28.06 (24.54, 33.32)	28.69 (25.13, 32.99)	30.10 (24.86, 36.29)	0.5632
Male	2462 (0.63)	54 (0.68)	102 (0.61)	2159 (0.63)	147 (0.56)	0.5215
In-Hospital Mortality	814 (0.21)	11 (0.14)	33 (0.20)	718 (0.21)	52 (0.20)	0.5663
30 Day Mortality	899 (0.23)	11 (0.14)	38 (0.23)	793 (0.23)	57 (0.22)	0.3853
ICU LOS, median(Q1,Q3)	3.13 (1.71, 6.97)	3.32 (1.34, 8.01)	3.14 (1.77, 7.29)	3.11 (1.69, 6.90)	3.29 (1.90, 7.69)	0.6657
ICU						
CCU	600 (0.15)	11 (0.14)	22 (0.13)	537 (0.16)	30 (0.11)	0.3337
CSRU	1135 (0.29)	37 (0.47)	37 (0.22)	990 (0.29)	71 (0.27)	0.0091
MICU	1204 (0.31)	15 (0.19)	66 (0.40)	1034 (0.30)	89 (0.34)	0.0270
SICU	569 (0.15)	8 (0.10)	20 (0.12)	505 (0.15)	36 (0.14)	0.5728
TSICU	416 (0.11)	8 (0.10)	21 (0.13)	352 (0.10)	35 (0.13)	0.4054
ethnicity						
Asian	88 (0.02)	4 (0.05)	3 (0.02)	74 (0.02)	7 (0.03)	0.3533
Black	277 (0.07)	7 (0.09)	9 (0.05)	237 (0.07)	24 (0.09)	0.4306
Hispanic	116 (0.03)	2 (0.03)	3 (0.02)	103 (0.03)	8 (0.03)	0.8404
White	2810 (0.72)	57 (0.72)	124 (0.75)	2444 (0.72)	185 (0.71)	0.9693
insurance						
Government	68 (0.02)	4 (0.05)	0 (0.00)	60 (0.02)	4 (0.02)	0.0459
Medicaid	321 (0.08)	6 (0.08)	13 (0.08)	281 (0.08)	21 (0.08)	0.9951
Medicare	2285 (0.58)	43 (0.54)	103 (0.62)	1985 (0.58)	154 (0.59)	0.8843
Private	1215 (0.31)	26 (0.33)	48 (0.29)	1063 (0.31)	78 (0.30)	0.9318
Self Pay	35 (0.01)	0 (0.00)	2 (0.01)	29 (0.01)	4 (0.02)	0.5398

Table E.22: AKI progression from Stage 1 to Stage 2 demographics of validation set patients in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	3924	340	3584	
Comorbidity				
Congestive heart failure	1347 (0.34)	120 (0.35)	1227 (0.34)	0.7502
Cardiac arrhythmias	1709 (0.44)	166 (0.49)	1543 (0.43)	0.1233
Valvular disease	823 (0.21)	88 (0.26)	735 (0.21)	0.0386
Pulmonary circulation disorders	267 (0.07)	16 (0.05)	251 (0.07)	0.1206
Peripheral vascular disorders	555 (0.14)	46 (0.14)	509 (0.14)	0.7526
Hypertension, uncomplicated	1927 (0.49)	184 (0.54)	1743 (0.49)	0.1678
Hypertension, complicated	280 (0.07)	21 (0.06)	259 (0.07)	0.4885
Paralysis	105 (0.03)	12 (0.04)	93 (0.03)	0.3140
Other neurological disorders	244 (0.06)	18 (0.05)	226 (0.06)	0.4746
Chronic pulmonary disease	968 (0.25)	87 (0.26)	881 (0.25)	0.7209
Diabetes, uncomplicated	898 (0.23)	87 (0.26)	811 (0.23)	0.2756
Diabetes, complicated	209 (0.05)	16 (0.05)	193 (0.05)	0.6040
Hypothyroidism	389 (0.10)	38 (0.11)	351 (0.10)	0.4389
Renal failure	351 (0.09)	27 (0.08)	324 (0.09)	0.5173
Liver disease	472 (0.12)	25 (0.07)	447 (0.12)	0.0093
Peptic ulcer disease excluding bleeding	24 (0.01)	2 (0.01)	22 (0.01)	0.9540
AIDS/HIV	46 (0.01)	6 (0.02)	40 (0.01)	0.2911
Lymphoma	71 (0.02)	3 (0.01)	68 (0.02)	0.1836
Metastatic cancer	38 (0.01)	5 (0.01)	33 (0.01)	0.3248
Solid tumor without metastasis	334 (0.09)	21 (0.06)	313 (0.09)	0.1225
Rheumatoid arthritis/collagen vascular diseases	120 (0.03)	9 (0.03)	111 (0.03)	0.6502
Coagulopathy	525 (0.13)	43 (0.13)	482 (0.13)	0.6994
Weight loss	203 (0.05)	13 (0.04)	190 (0.05)	0.2522
Fluid and electrolyte disorders	1156 (0.29)	103 (0.30)	1053 (0.29)	0.7668
Blood loss anemia	77 (0.02)	4 (0.01)	73 (0.02)	0.2791
Deficiency anemia	96 (0.02)	12 (0.04)	84 (0.02)	0.1816
Alcohol abuse	195 (0.05)	14 (0.04)	181 (0.05)	0.4610
Drug abuse	142 (0.04)	9 (0.03)	133 (0.04)	0.3244
Psychoses	63 (0.02)	4 (0.01)	59 (0.02)	0.5136
Depression	314 (0.08)	27 (0.08)	287 (0.08)	0.9669
Trauma	491 (0.13)	50 (0.15)	441 (0.12)	0.2316

Table E.23: AKI progression from Stage 1 to Stage 2 comorbidities of validation set patients by condition positive or negative. P-values for patient counts calculated by chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	3924	79	166	3418	261	
Comorbidity						
Congestive heart failure	1347 (0.34)	27 (0.34)	55 (0.33)	1172 (0.34)	93 (0.36)	0.9775
Cardiac arrhythmias	1709 (0.44)	45 (0.57)	67 (0.40)	1476 (0.43)	121 (0.46)	0.2377
Valvular disease	823 (0.21)	23 (0.29)	20 (0.12)	715 (0.21)	65 (0.25)	0.0133
Pulmonary circulation disorders	267 (0.07)	1 (0.01)	8 (0.05)	243 (0.07)	15 (0.06)	0.1436
Peripheral vascular disorders	555 (0.14)	11 (0.14)	16 (0.10)	493 (0.14)	35 (0.13)	0.4447
Hypertension, uncomplicated	1927 (0.49)	41 (0.52)	79 (0.48)	1664 (0.49)	143 (0.55)	0.5633
Hypertension, complicated	280 (0.07)	3 (0.04)	10 (0.06)	249 (0.07)	18 (0.07)	0.6484
Paralysis	105 (0.03)	0 (0.00)	0 (0.00)	93 (0.03)	12 (0.05)	0.0171
Other neurological disorders	244 (0.06)	4 (0.05)	12 (0.07)	214 (0.06)	14 (0.05)	0.8594
Chronic pulmonary disease	968 (0.25)	11 (0.14)	41 (0.25)	840 (0.25)	76 (0.29)	0.1215
Diabetes, uncomplicated	898 (0.23)	15 (0.19)	41 (0.25)	770 (0.23)	72 (0.28)	0.3241
Diabetes, complicated	209 (0.05)	2 (0.03)	8 (0.05)	185 (0.05)	14 (0.05)	0.7322
Hypothyroidism	389 (0.10)	3 (0.04)	16 (0.10)	335 (0.10)	35 (0.13)	0.0998
Renal failure	351 (0.09)	6 (0.08)	13 (0.08)	311 (0.09)	21 (0.08)	0.8691
Liver disease	472 (0.12)	6 (0.08)	23 (0.14)	424 (0.12)	19 (0.07)	0.0704
Peptic ulcer disease excluding bleeding	24 (0.01)	0 (0.00)	0 (0.00)	22 (0.01)	2 (0.01)	0.6463
AIDS/HIV	46 (0.01)	1 (0.01)	0 (0.00)	40 (0.01)	5 (0.02)	0.3643
Lymphoma	71 (0.02)	0 (0.00)	2 (0.01)	66 (0.02)	3 (0.01)	0.4450
Metastatic cancer	38 (0.01)	0 (0.00)	4 (0.02)	29 (0.01)	5 (0.02)	0.0643
Solid tumor without metastasis	334 (0.09)	2 (0.03)	18 (0.11)	295 (0.09)	19 (0.07)	0.1791
Rheumatoid arthritis/collagen vascular diseases	120 (0.03)	3 (0.04)	2 (0.01)	109 (0.03)	6 (0.02)	0.4421
Coagulopathy	525 (0.13)	8 (0.10)	22 (0.13)	460 (0.13)	35 (0.13)	0.8866
Weight loss	203 (0.05)	4 (0.05)	14 (0.08)	176 (0.05)	9 (0.03)	0.1779
Fluid and electrolyte disorders	1156 (0.29)	16 (0.20)	43 (0.26)	1010 (0.30)	87 (0.33)	0.2285
Blood loss anemia	77 (0.02)	1 (0.01)	2 (0.01)	71 (0.02)	3 (0.01)	0.6172
Deficiency anemia	96 (0.02)	3 (0.04)	6 (0.04)	78 (0.02)	9 (0.03)	0.3973
Alcohol abuse	195 (0.05)	5 (0.06)	9 (0.05)	172 (0.05)	9 (0.03)	0.6583
Drug abuse	142 (0.04)	1 (0.01)	7 (0.04)	126 (0.04)	8 (0.03)	0.6510
Psychoses	63 (0.02)	1 (0.01)	2 (0.01)	57 (0.02)	3 (0.01)	0.8865
Depression	314 (0.08)	7 (0.09)	18 (0.11)	269 (0.08)	20 (0.08)	0.6021
Trauma	491 (0.13)	11 (0.14)	28 (0.17)	413 (0.12)	39 (0.15)	0.2235

Table E.24: AKI progression from Stage 1 to Stage 2 comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by chi-squared test.

E.5 AKI progression from Stage 2 to Stage 3

E.5.1 Training and Testing

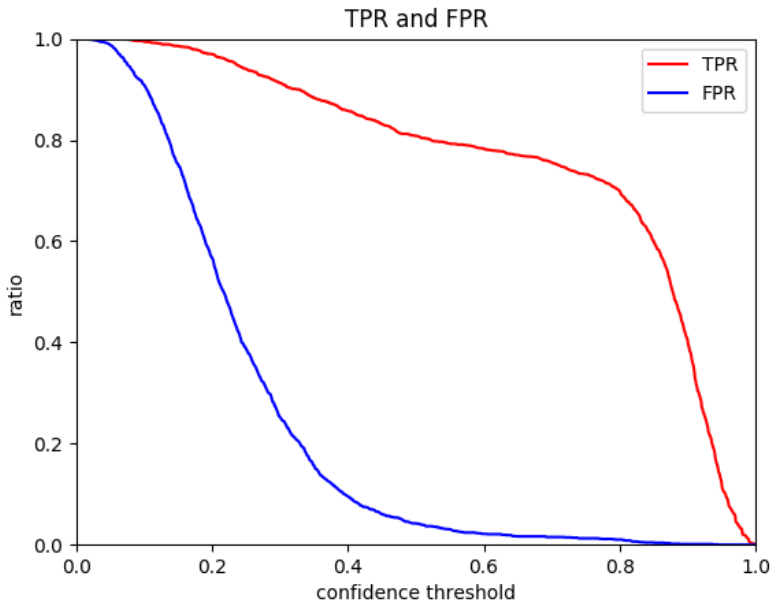


Figure E.53: AKI progression from Stage 2 to Stage 3 training/testing true positive and false positive rates.

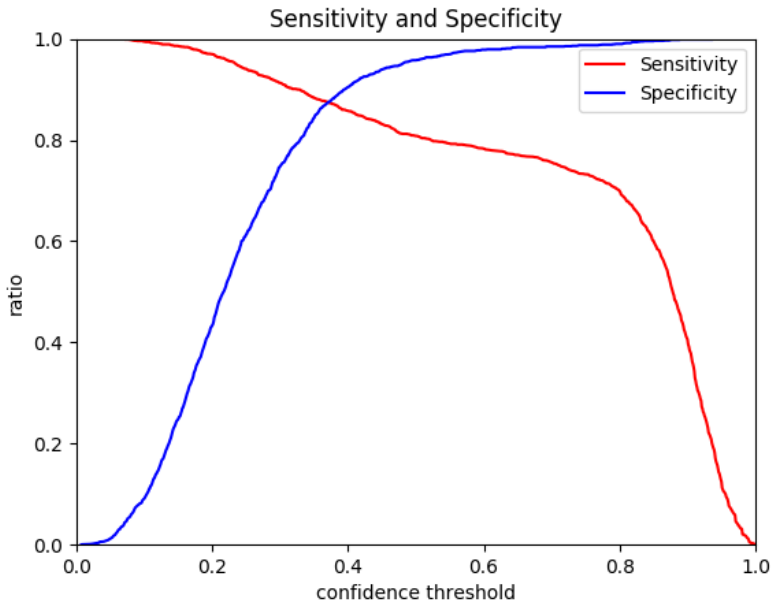


Figure E.54: AKI progression from Stage 2 to Stage 3 training/testing sensitivity and specificity.

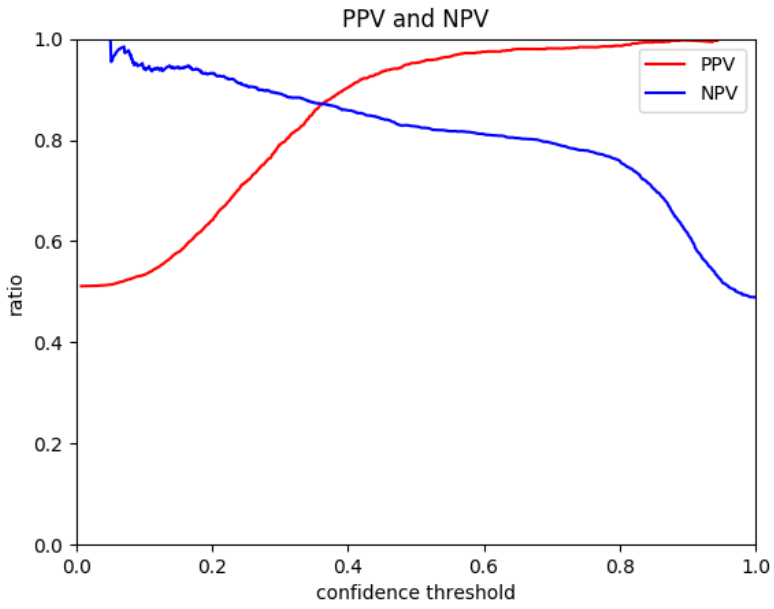


Figure E.55: AKI progression from Stage 2 to Stage 3 training/testing positive and negative predictive value.

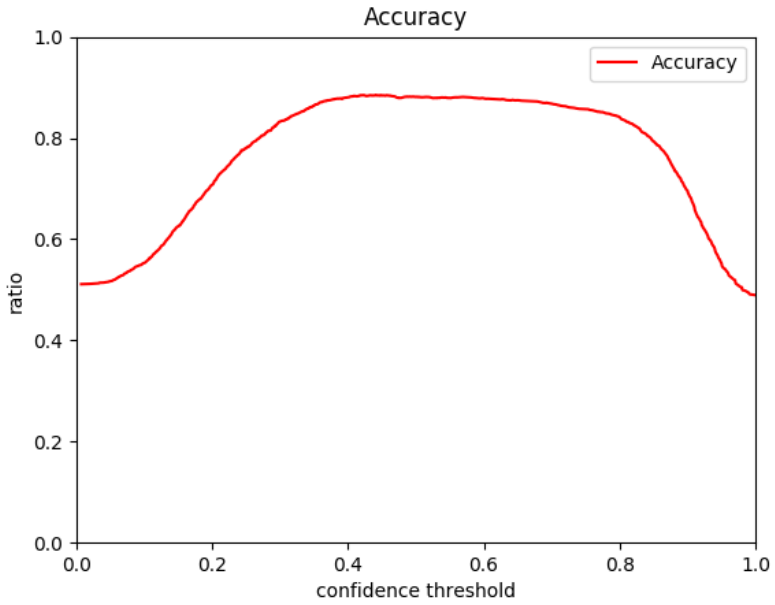


Figure E.56: AKI progression from Stage 2 to Stage 3 training/testing accuracy.

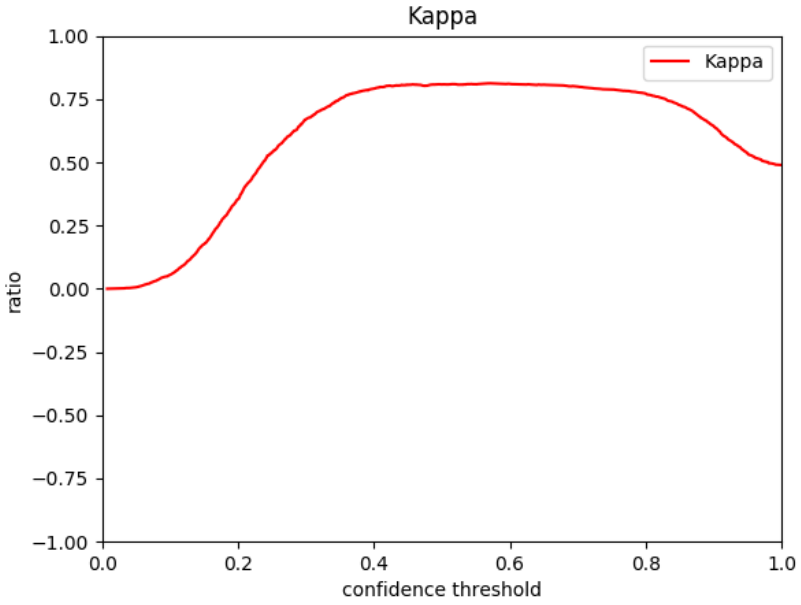


Figure E.57: AKI progression from Stage 2 to Stage 3 training/testing kappa.

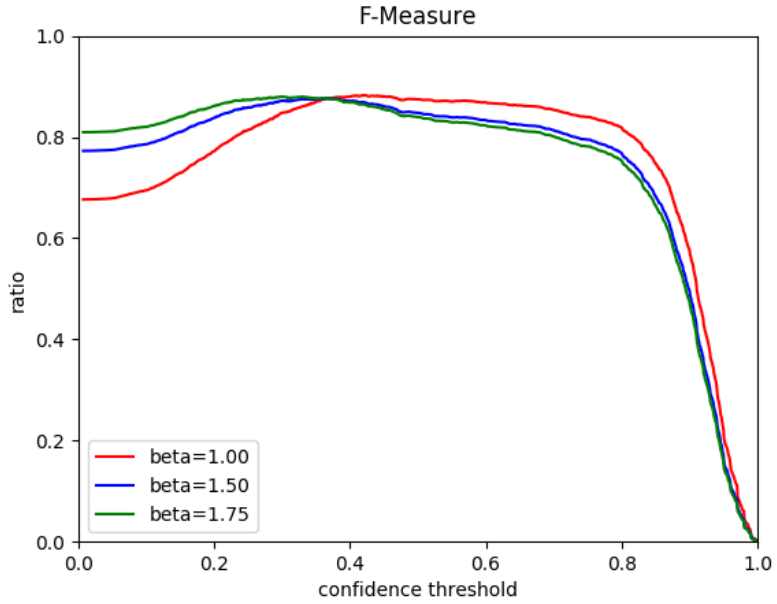


Figure E.58: AKI progression from Stage 2 to Stage 3 training/testing F-measures.

E.5.2 Validation

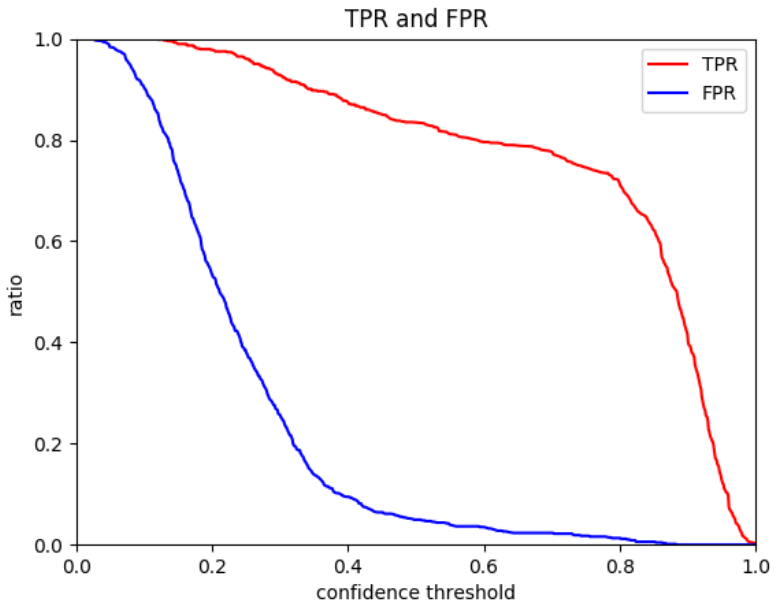


Figure E.59: AKI progression from Stage 2 to Stage 3 validation true and false positive rates.

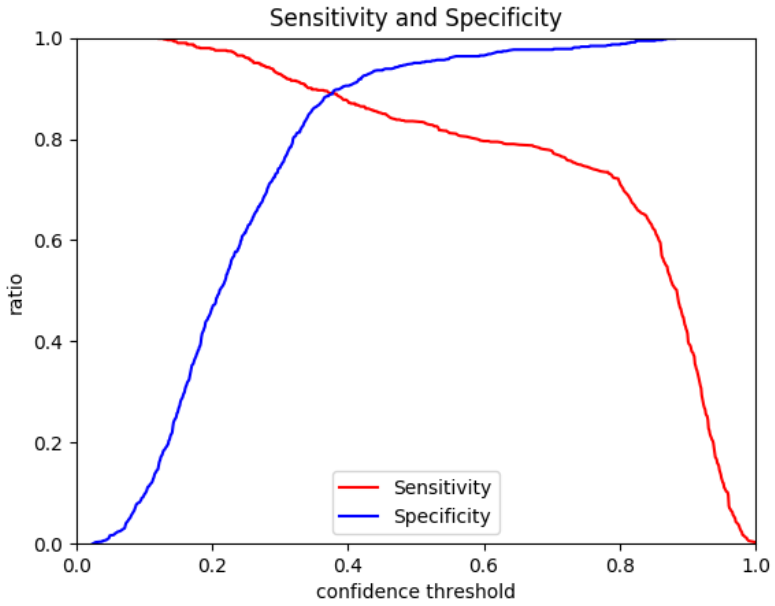


Figure E.60: AKI progression from Stage 2 to Stage 3 validation sensitivity and specificity.

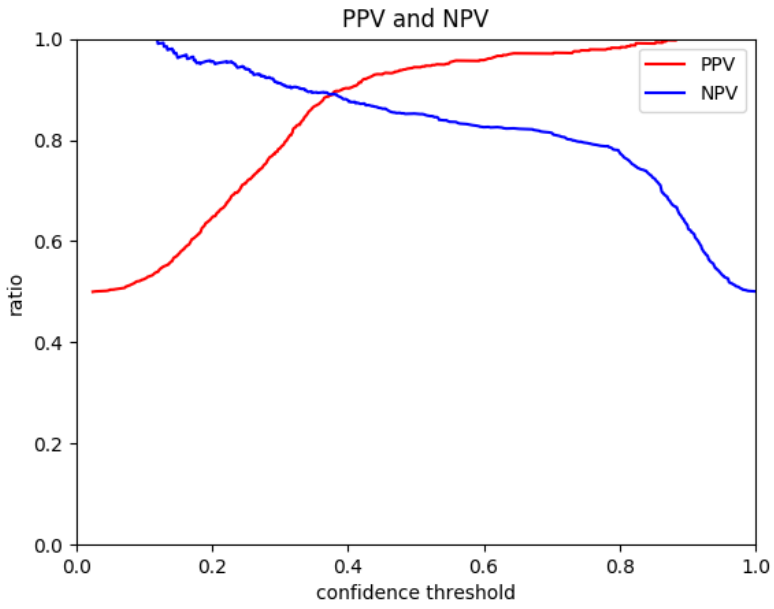


Figure E.61: AKI progression from Stage 2 to Stage 3 validation positive and negative predictive values.

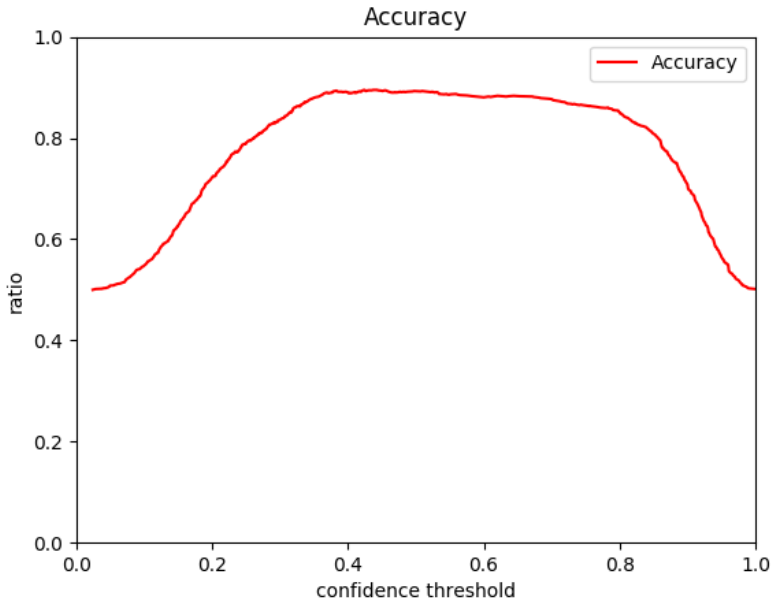


Figure E.62: AKI progression from Stage 2 to Stage 3 validation accuracy.

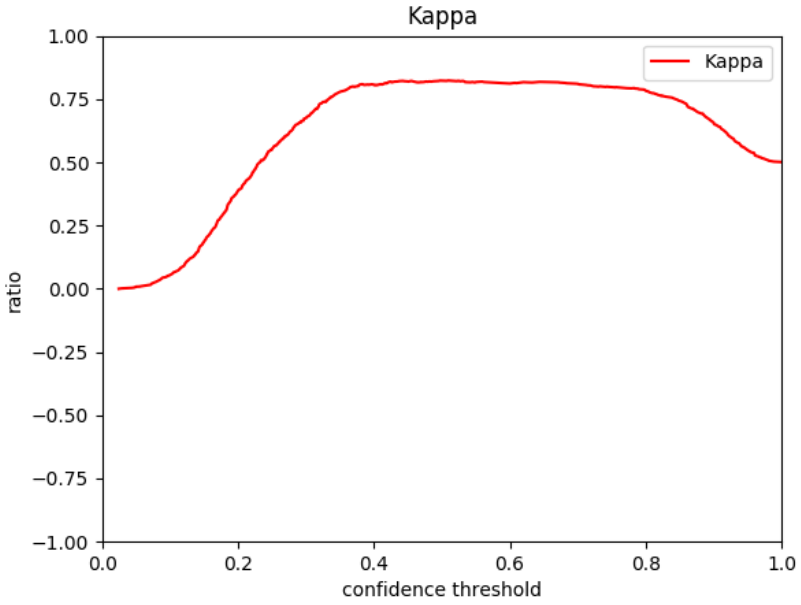


Figure E.63: AKI progression from Stage 2 to Stage 3 validation kappa.

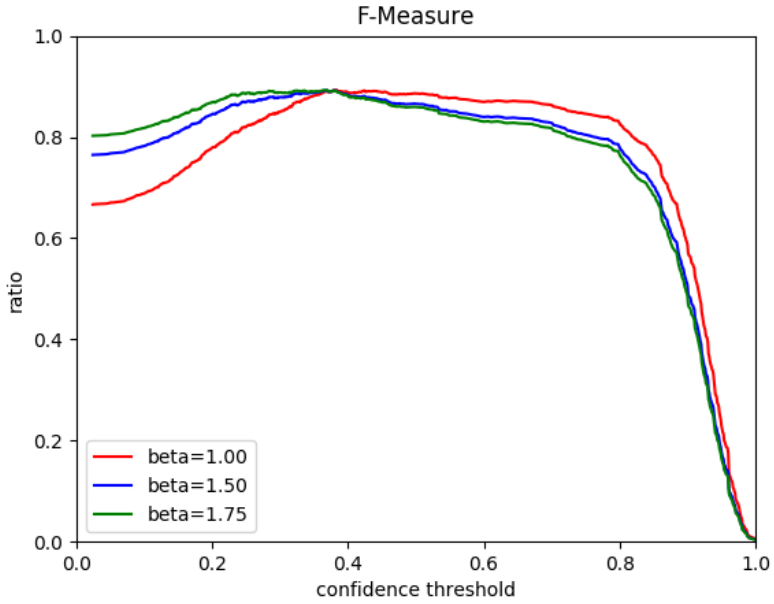
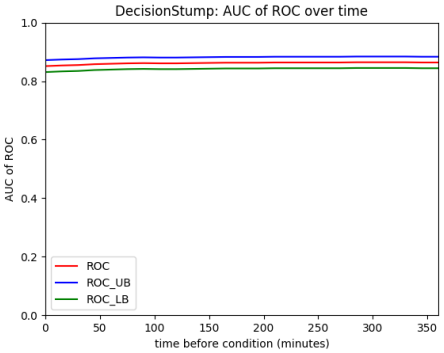
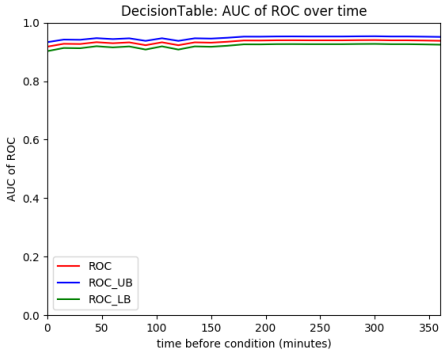


Figure E.64: AKI progression from Stage 2 to Stage 3 validation F-measures.

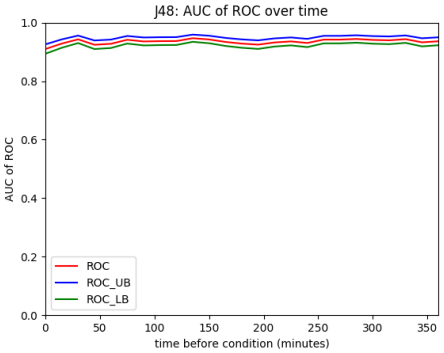
E.5.3 Machine Learning Algorithms over time



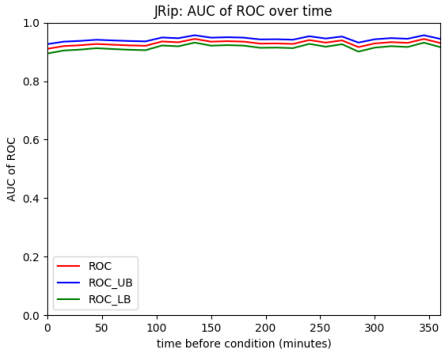
(a) Decision Stump



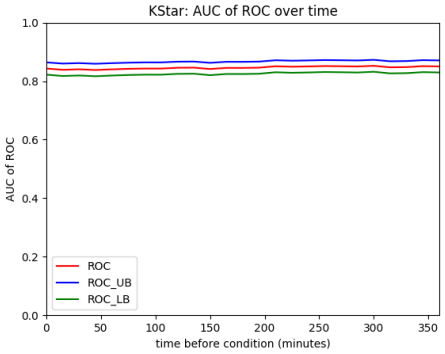
(b) Decision Table



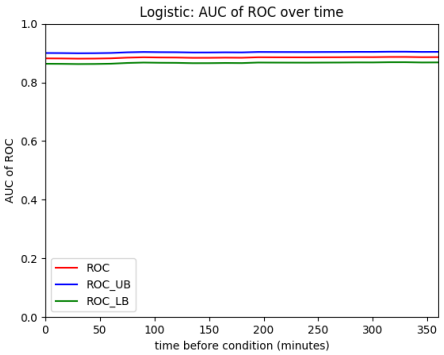
(c) J48



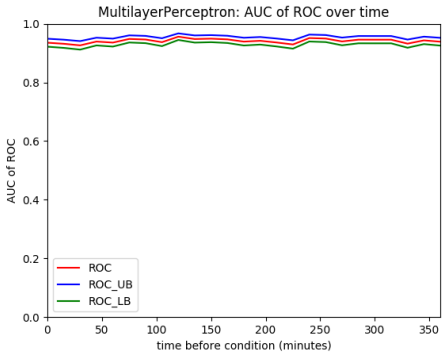
(d) JRip



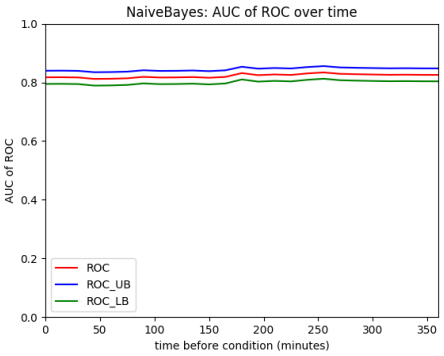
(e) KStar



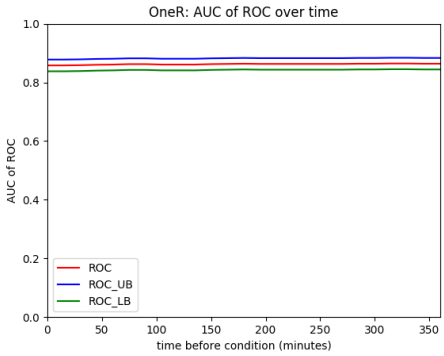
(f) Logistic Regression



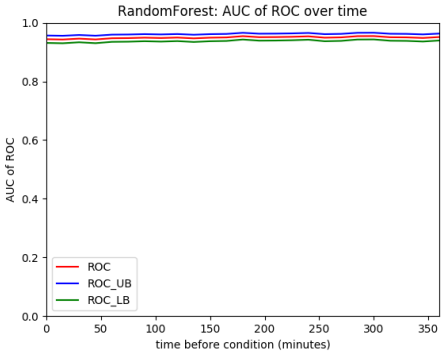
(g) Multilayer Perceptron



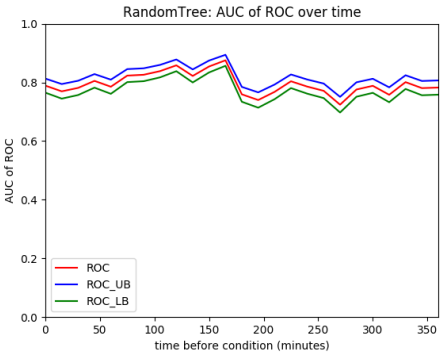
(h) Naive Bayes



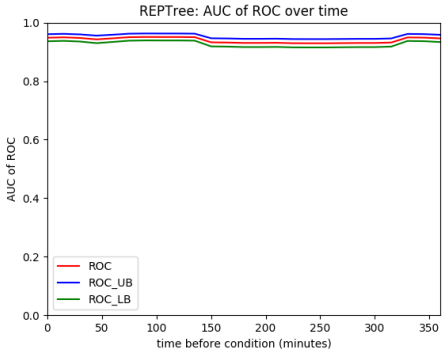
(i) One R

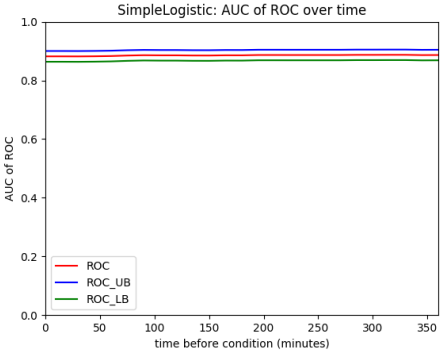


(j) Random Forest

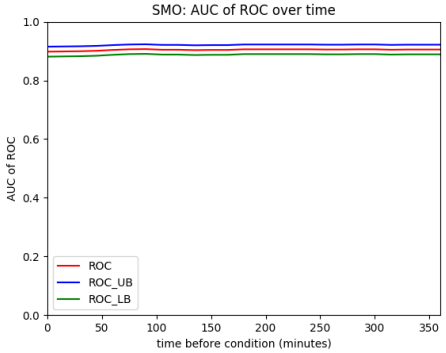


(k) Random Tree

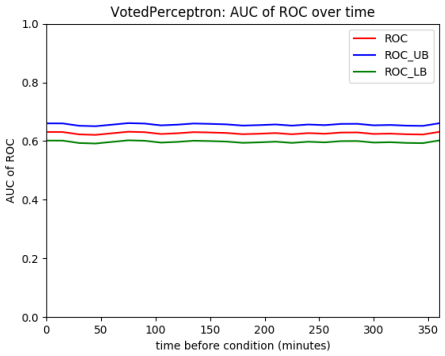




(l) Simple Logistic Regression



(m) SMO



(n) Voted Perceptron

Figure E.65: Algorithm's performance over time on the progression from AKI Stage 2 to AKI Stage 3 selection criteria.

E.5.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
pCO_2	33	38	44	32	39	45	0.3746
Fentanyl	0.064	0.100	1.658	0.064	0.076	0.803	0.0102
Osmolality	298	298	298	298	298	298	0.6732
Lactate Dehydrogenase (LD)	166.25	244	401.5	153	199	309.75	<0.0001
Creatinine	2.2	2.5	2.9	1.3	1.5	1.9	<0.0001
D5/.45NS	0	0	0	0	0	0	0.2565
Sodium (whole blood)	132.25	134	138	133	134	138	0.5256
Urea Nitrogen	32	47	66	23.25	33	45	<0.0001
Midazolam (Versed)	0.5	1	1	0.1	0.809	1	<0.0001
Alanine Aminotransferase (ALT)	15	26	62.75	15	23	45	0.0084
Calcium, Total	7.7	8.3	8.9	7.9	8.4	8.9	0.3516
Hemoglobin	8.2	8.8	10.5	8.2	8.7	10.9	0.1017
Cefazolin	1	1	1	1	1	1	1.0000

Table E.25: AKI progression from Stage 2 to Stage 3 model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
pCO_2	33	38	44	33	39	45	27	38	43	34	38	44	0.4071
Fentanyl	0.064	0.100	2.145	0.064	0.076	0.803	0.064	0.160	25.000	0.064	0.076	0.803	0.0007
Osmolality	298	298	298	298	298	298	280	298	298	298	298	298	0.9312
Lactate Dehydrogenase (LD)	172.5	251	416	153	198	298	196	415	993	157	234	330	<0.0001
Creatinine	2.3	2.7	3	1.3	1.5	1.9	2	2.2	2.3	0.8	1.1	1.8	<0.0001
D5/.45NS	0	0	0	0	0	0	0	0	0	0	0	0	0.0252
Sodium (whole blood)	132	134	138	133	134	138	134	137	140	134	134	137	0.0257
Urea Nitrogen	37	51	70	23	32	44	29	44	52	17.5	26	37	<0.0001
Midazolam (Versed)	0.5	1	1	0.1	0.717	1	0.5	1	2	0.1	0.5	1	<0.0001
Alanine Aminotransferase (ALT)	15	26	66.5	15	22	43	30	54	613	16	23	47	<0.0001
Calcium, Total	7.7	8.3	8.9	7.9	8.4	8.9	7.4	8.4	9	7.8	8.3	9.1	0.7244
Hemoglobin	8.1	8.7	10.4	8.2	8.6	10.9	8.6	10	11.1	8.2	8.9	10.85	0.0177
Cefazolin	1	1	1	1	1	1	1	1	1	1	1	1	0.1704

Table E.26: AKI progression from Stage 2 to Stage 3 model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	1372	686	686	
Age, median	70	69.5	72.5	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 69.00)	0.1949
Weight, median (Q1,Q3)	80.00 (66.30, 95.60)	78.00 (64.75, 94.90)	81.00 (68.00, 96.30)	0.1860
BMI, median (Q1,Q3)	28.51 (24.69, 34.04)	27.78 (24.23, 33.59)	29.40 (25.34, 34.26)	0.1059
Male	773 (0.56)	396 (0.58)	377 (0.55)	0.4944
In-Hospital Mortality	563 (0.41)	325 (0.47)	238 (0.35)	0.0002
30 Day Mortality	611 (0.45)	348 (0.51)	263 (0.38)	0.0006
ICU LOS, median(Q1,Q3)	4.66 (2.12, 10.04)	5.82 (2.39, 12.42)	3.75 (1.97, 7.52)	<0.0001
ICU				
CCU	283 (0.21)	143 (0.21)	140 (0.20)	0.8585
CSRU	279 (0.20)	101 (0.15)	178 (0.26)	<0.0001
MICU	513 (0.37)	292 (0.43)	221 (0.32)	0.0017
SICU	201 (0.15)	111 (0.16)	90 (0.13)	0.1385
TSICU	96 (0.07)	39 (0.06)	57 (0.08)	0.0662
ethnicity				
Asian	24 (0.02)	14 (0.02)	10 (0.01)	0.4142
Black	134 (0.10)	83 (0.12)	51 (0.07)	0.0057
Hispanic	40 (0.03)	23 (0.03)	17 (0.02)	0.3428
White	951 (0.69)	454 (0.66)	497 (0.72)	0.1632
insurance				
Government	11 (0.01)	7 (0.01)	4 (0.01)	0.3657
Medicaid	98 (0.07)	51 (0.07)	47 (0.07)	0.6862
Medicare	939 (0.68)	460 (0.67)	479 (0.70)	0.5352
Private	319 (0.23)	166 (0.24)	153 (0.22)	0.4667
Self Pay	5 (0.00)	2 (0.00)	3 (0.00)	0.6547

Table E.27: AKI progression from Stage 2 to Stage 3 demographics of validation set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	1372	572	652	34	114	
Age, median	70	69.5	73	51	69.5	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	68.00 (64.00, 70.50)	67.00 (64.00, 69.00)	68.00 (64.75, 72.00)	67.00 (64.00, 68.00)	0.0792
Weight, median (Q1,Q3)	80.00 (66.30, 95.60)	79.95 (65.76, 95.00)	80.70 (68.00, 96.00)	87.95 (79.35, 100.70)	75.00 (63.15, 92.80)	0.1424
BMI, median (Q1,Q3)	28.51 (24.69, 34.04)	27.79 (24.39, 34.08)	29.40 (25.32, 34.27)	29.99 (26.37, 32.10)	27.33 (23.98, 33.28)	0.3864
Male	773 (0.56)	350 (0.61)	351 (0.54)	26 (0.76)	46 (0.40)	0.0133
In-Hospital Mortality	563 (0.41)	282 (0.49)	220 (0.34)	18 (0.53)	43 (0.38)	0.0002
30 Day Mortality	611 (0.45)	303 (0.53)	245 (0.38)	18 (0.53)	45 (0.39)	0.0006
ICU LOS, median(Q1,Q3)	4.66 (2.12, 10.04)	6.20 (2.62, 12.88)	3.79 (1.97, 7.59)	3.43 (1.98, 7.22)	4.00 (2.04, 9.80)	<0.0001
ICU						
CCU	283 (0.21)	118 (0.21)	135 (0.21)	5 (0.15)	25 (0.22)	0.8794
CSRU	279 (0.20)	76 (0.13)	171 (0.26)	7 (0.21)	25 (0.22)	<0.0001
MICU	513 (0.37)	247 (0.43)	209 (0.32)	12 (0.35)	45 (0.39)	0.0164
SICU	201 (0.15)	96 (0.17)	83 (0.13)	7 (0.21)	15 (0.13)	0.2206
TSICU	96 (0.07)	35 (0.06)	54 (0.08)	3 (0.09)	4 (0.04)	0.2295
ethnicity						
Asian	24 (0.02)	13 (0.02)	10 (0.02)	0 (0.00)	1 (0.01)	0.5400
Black	134 (0.10)	74 (0.13)	39 (0.06)	12 (0.35)	9 (0.08)	<0.0001
Hispanic	40 (0.03)	20 (0.03)	15 (0.02)	2 (0.06)	3 (0.03)	0.4635
White	951 (0.69)	375 (0.66)	479 (0.73)	18 (0.53)	79 (0.69)	0.2509
insurance						
Government	11 (0.01)	6 (0.01)	4 (0.01)	0 (0.00)	1 (0.01)	0.8001
Medicaid	98 (0.07)	42 (0.07)	43 (0.07)	4 (0.12)	9 (0.08)	0.7025
Medicare	939 (0.68)	385 (0.67)	464 (0.71)	15 (0.44)	75 (0.66)	0.2758
Private	319 (0.23)	137 (0.24)	139 (0.21)	14 (0.41)	29 (0.25)	0.1068
Self Pay	5 (0.00)	2 (0.00)	2 (0.00)	1 (0.03)	0 (0.00)	0.0831

Table E.28: AKI progression from Stage 2 to Stage 3 demographics of validation set patients in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and P-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's chi-squared test.

Variable	Total	Condition Positive	condition Negative	P-value
N	1372	686	686	
Comorbidity				
Congestive heart failure	663 (0.48)	342 (0.50)	321 (0.47)	0.4147
Cardiac arrhythmias	658 (0.48)	314 (0.46)	344 (0.50)	0.2422
Valvular disease	270 (0.20)	124 (0.18)	146 (0.21)	0.1806
Pulmonary circulation disorders	124 (0.09)	54 (0.08)	70 (0.10)	0.1508
Peripheral vascular disorders	236 (0.17)	121 (0.18)	115 (0.17)	0.6961
Hypertension, uncomplicated	486 (0.35)	188 (0.27)	298 (0.43)	<0.0001
Hypertension, complicated	323 (0.24)	197 (0.29)	126 (0.18)	<0.0001
Paralysis	28 (0.02)	15 (0.02)	13 (0.02)	0.7055
Other neurological disorders	76 (0.06)	39 (0.06)	37 (0.05)	0.8185
Chronic pulmonary disease	365 (0.27)	170 (0.25)	195 (0.28)	0.1907
Diabetes, uncomplicated	351 (0.26)	170 (0.25)	181 (0.26)	0.5571
Diabetes, complicated	150 (0.11)	84 (0.12)	66 (0.10)	0.1416
Hypothyroidism	172 (0.13)	81 (0.12)	91 (0.13)	0.4458
Renal failure	388 (0.28)	225 (0.33)	163 (0.24)	0.0016
Liver disease	283 (0.21)	169 (0.25)	114 (0.17)	0.0011
Peptic ulcer disease excluding bleeding	9 (0.01)	3 (0.00)	6 (0.01)	0.3173
AIDS/HIV	14 (0.01)	10 (0.01)	4 (0.01)	0.1088
Lymphoma	21 (0.02)	11 (0.02)	10 (0.01)	0.8273
Metastatic cancer	9 (0.01)	1 (0.00)	8 (0.01)	0.0196
Solid tumor without metastasis	123 (0.09)	56 (0.08)	67 (0.10)	0.3213
Rheumatoid arthritis/collagen vascular diseases	61 (0.04)	35 (0.05)	26 (0.04)	0.2492
Coagulopathy	292 (0.21)	150 (0.22)	142 (0.21)	0.6397
Weight loss	101 (0.07)	67 (0.10)	34 (0.05)	0.0010
Fluid and electrolyte disorders	568 (0.41)	337 (0.49)	231 (0.34)	<0.0001
Blood loss anemia	41 (0.03)	26 (0.04)	15 (0.02)	0.0858
Deficiency anemia	44 (0.03)	21 (0.03)	23 (0.03)	0.7630
Alcohol abuse	82 (0.06)	42 (0.06)	40 (0.06)	0.8252
Drug abuse	39 (0.03)	21 (0.03)	18 (0.03)	0.6310
Psychoses	18 (0.01)	10 (0.01)	8 (0.01)	0.6374
Depression	105 (0.08)	51 (0.07)	54 (0.08)	0.7697
Trauma	136 (0.10)	72 (0.10)	64 (0.09)	0.4927

Table E.29: AKI progression from Stage 2 to Stage 3 comorbidities of validation set patients by condition positive or negative. P-values for patient counts calculated by chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	1372	572	652	34	114	
Comorbidity						
Congestive heart failure	663 (0.48)	293 (0.51)	312 (0.48)	9 (0.26)	49 (0.43)	0.1676
Cardiac arrhythmias	658 (0.48)	265 (0.46)	333 (0.51)	11 (0.32)	49 (0.43)	0.2668
Valvular disease	270 (0.20)	102 (0.18)	142 (0.22)	4 (0.12)	22 (0.19)	0.3152
Pulmonary circulation disorders	124 (0.09)	46 (0.08)	66 (0.10)	4 (0.12)	8 (0.07)	0.5180
Peripheral vascular disorders	236 (0.17)	102 (0.18)	108 (0.17)	7 (0.21)	19 (0.17)	0.9119
Hypertension, uncomplicated	486 (0.35)	138 (0.24)	287 (0.44)	11 (0.32)	50 (0.44)	<0.0001
Hypertension, complicated	323 (0.24)	184 (0.32)	120 (0.18)	6 (0.18)	13 (0.11)	<0.0001
Paralysis	28 (0.02)	13 (0.02)	13 (0.02)	0 (0.00)	2 (0.02)	0.8260
Other neurological disorders	76 (0.06)	31 (0.05)	36 (0.06)	1 (0.03)	8 (0.07)	0.8304
Chronic pulmonary disease	365 (0.27)	141 (0.25)	188 (0.29)	7 (0.21)	29 (0.25)	0.4645
Diabetes, uncomplicated	351 (0.26)	149 (0.26)	171 (0.26)	10 (0.29)	21 (0.18)	0.4514
Diabetes, complicated	150 (0.11)	75 (0.13)	63 (0.10)	3 (0.09)	9 (0.08)	0.2081
Hypothyroidism	172 (0.13)	69 (0.12)	91 (0.14)	0 (0.00)	12 (0.11)	0.1227
Renal failure	388 (0.28)	207 (0.36)	156 (0.24)	7 (0.21)	18 (0.16)	<0.0001
Liver disease	283 (0.21)	155 (0.27)	97 (0.15)	17 (0.50)	14 (0.12)	<0.0001
Peptic ulcer disease excluding bleeding	9 (0.01)	3 (0.01)	6 (0.01)	0 (0.00)	0 (0.00)	0.6115
AIDS/HIV	14 (0.01)	8 (0.01)	4 (0.01)	0 (0.00)	2 (0.02)	0.4221
Lymphoma	21 (0.02)	10 (0.02)	10 (0.02)	0 (0.00)	1 (0.01)	0.7975
Metastatic cancer	9 (0.01)	1 (0.00)	7 (0.01)	1 (0.03)	0 (0.00)	0.0656
Solid tumor without metastasis	123 (0.09)	44 (0.08)	65 (0.10)	2 (0.06)	12 (0.11)	0.4867
Rheumatoid arthritis/collagen vascular diseases	61 (0.04)	31 (0.05)	21 (0.03)	5 (0.15)	4 (0.04)	0.0085
Coagulopathy	292 (0.21)	131 (0.23)	131 (0.20)	11 (0.32)	19 (0.17)	0.2369
Weight loss	101 (0.07)	61 (0.11)	31 (0.05)	3 (0.09)	6 (0.05)	0.0016
Fluid and electrolyte disorders	568 (0.41)	293 (0.51)	217 (0.33)	14 (0.41)	44 (0.39)	<0.0001
Blood loss anemia	41 (0.03)	24 (0.04)	14 (0.02)	1 (0.03)	2 (0.02)	0.1781
Deficiency anemia	44 (0.03)	19 (0.03)	23 (0.04)	0 (0.00)	2 (0.02)	0.5574
Alcohol abuse	82 (0.06)	36 (0.06)	38 (0.06)	2 (0.06)	6 (0.05)	0.9747
Drug abuse	39 (0.03)	18 (0.03)	16 (0.02)	2 (0.06)	3 (0.03)	0.6468
Psychoses	18 (0.01)	4 (0.01)	7 (0.01)	1 (0.03)	6 (0.05)	0.0010
Depression	105 (0.08)	45 (0.08)	53 (0.08)	1 (0.03)	6 (0.05)	0.5592
Trauma	136 (0.10)	64 (0.11)	58 (0.09)	6 (0.18)	8 (0.07)	0.2005

Table E.30: AKI progression from Stage 2 to Stage 3 comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by chi-squared test.

Appendix F

SEPSIS

F.1 Sepsis

F.1.1 Training and Testing

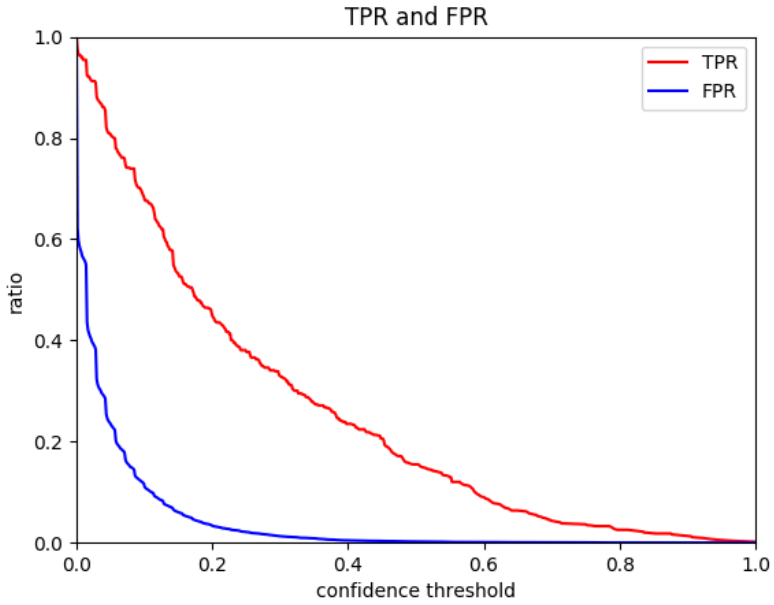


Figure F.1: Sepsis training/testing true positive and false positive rates.

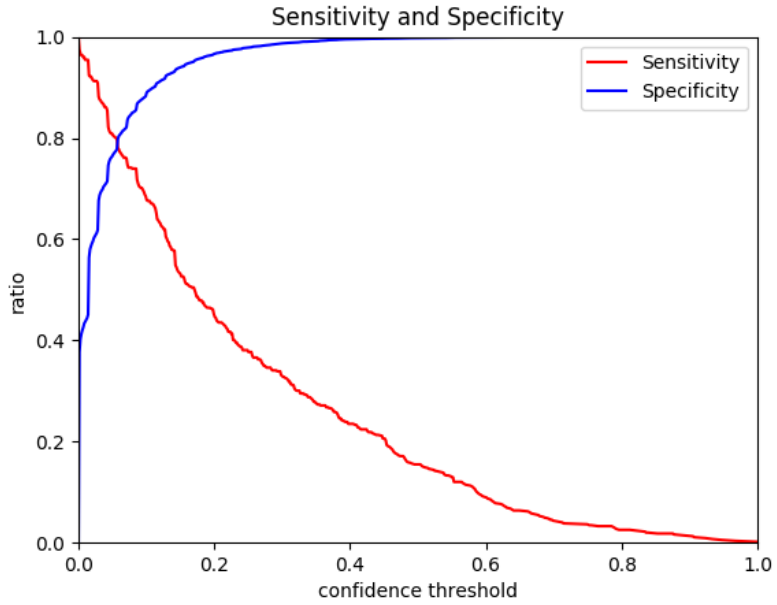


Figure F.2: Sepsis training/testing sensitivity and specificity.

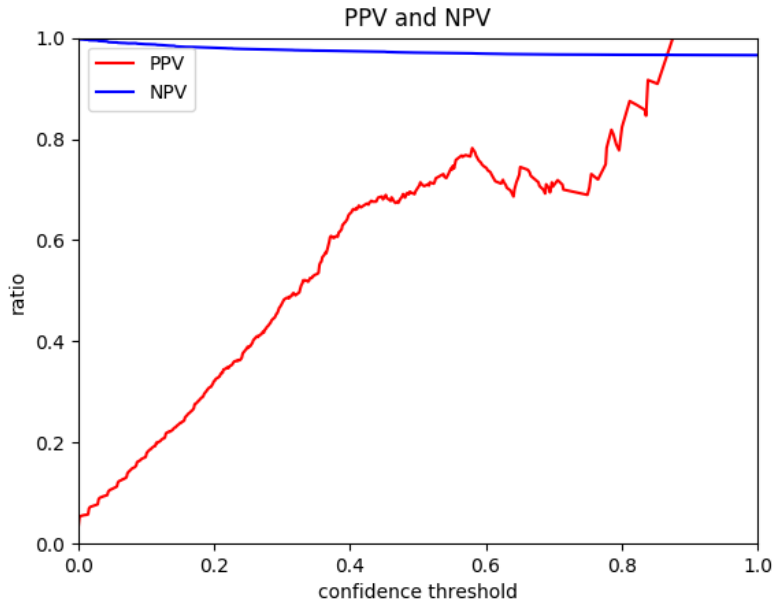


Figure F.3: Sepsis training/testing positive and negative predictive value.

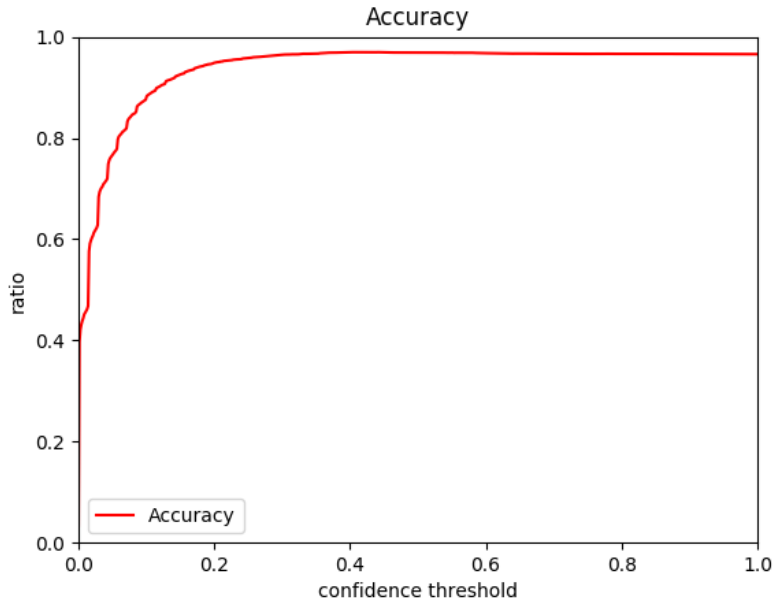


Figure F.4: Sepsis training/testing accuracy.

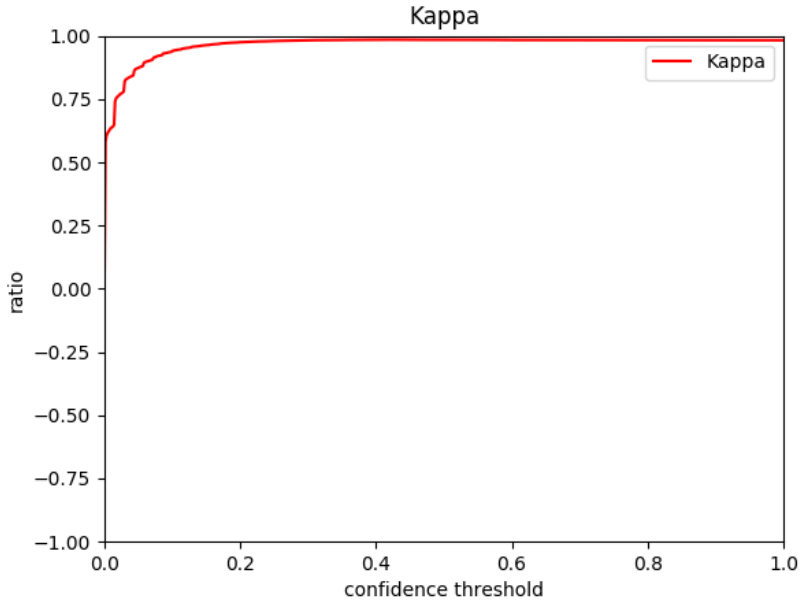


Figure F.5: Sepsis training/testing kappa.

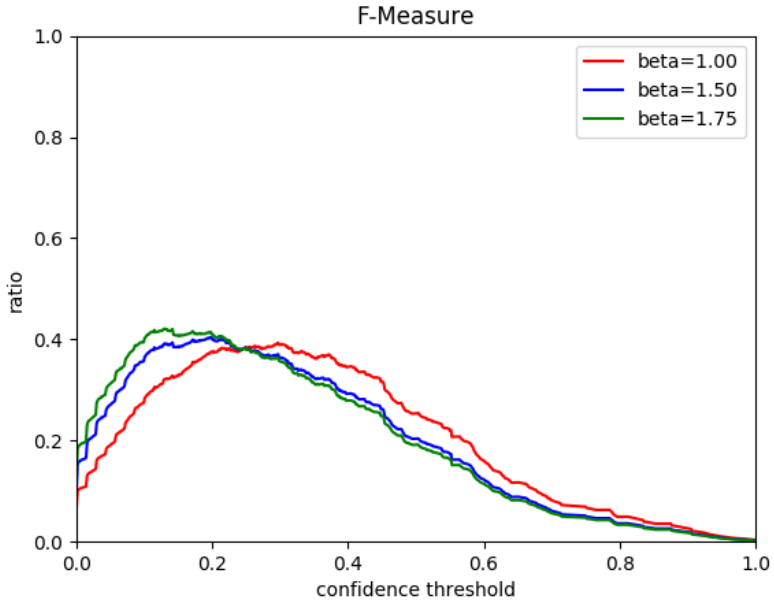


Figure F.6: Sepsis training/testing F-measures.

F.1.2 Validation

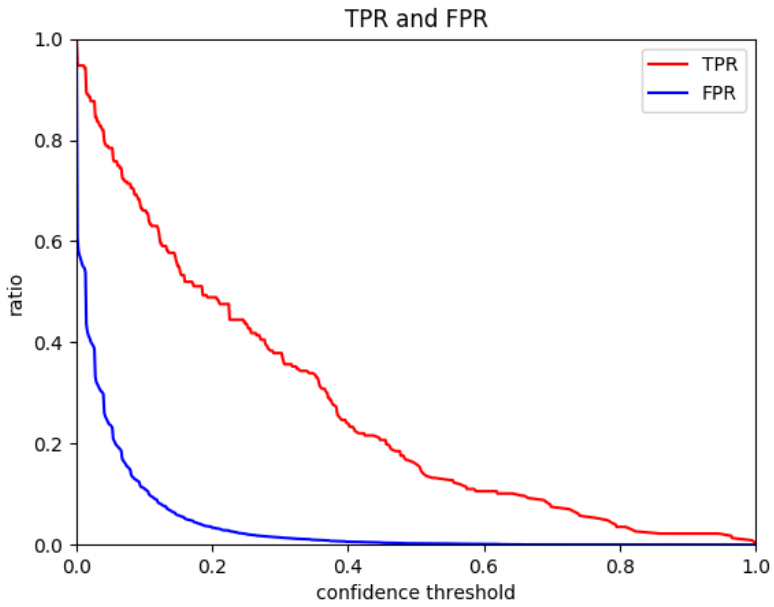


Figure F.7: Sepsis validation true and false positive rates.

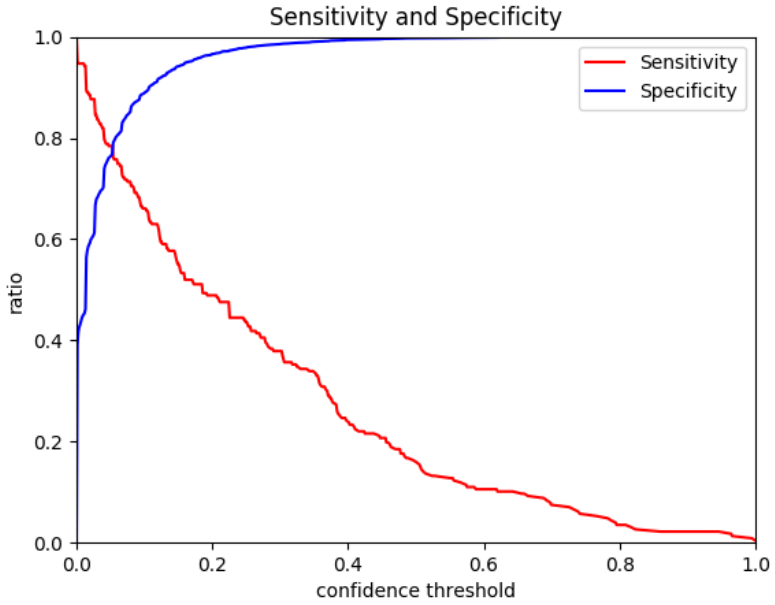


Figure F.8: Sepsis validation sensitivity and specificity.

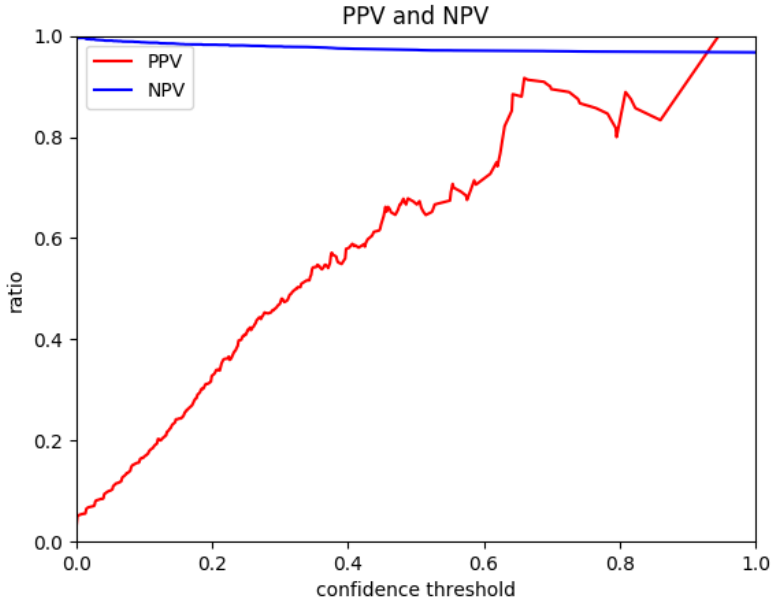


Figure F.9: Sepsis validation positive and negative predictive values.

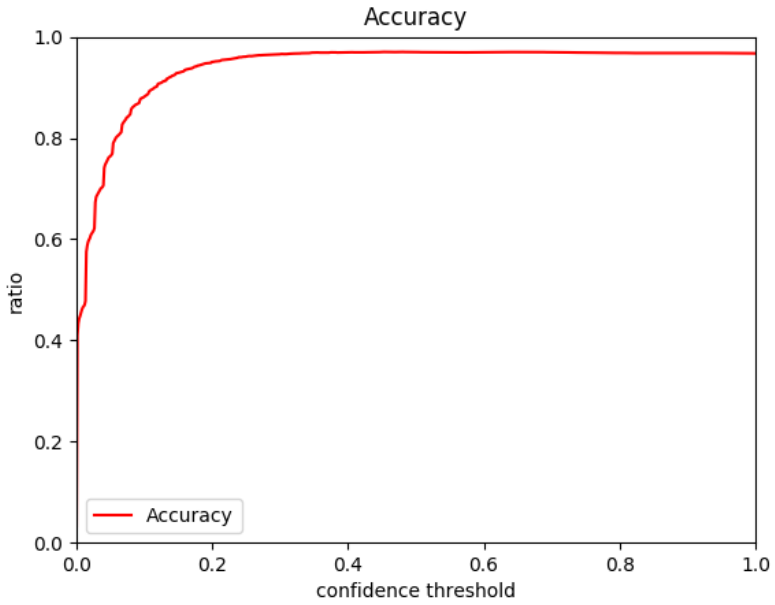


Figure F.10: Sepsis validation accuracy.

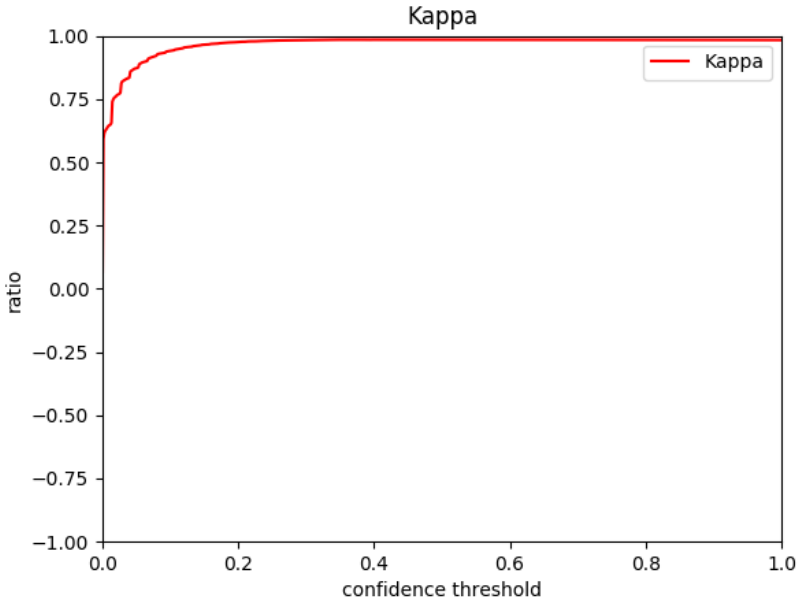


Figure F.11: Sepsis validation kappa.

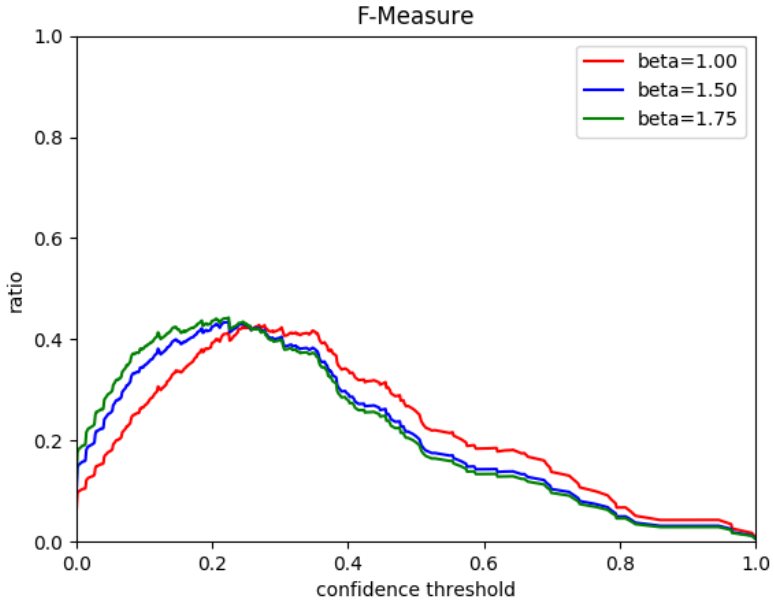
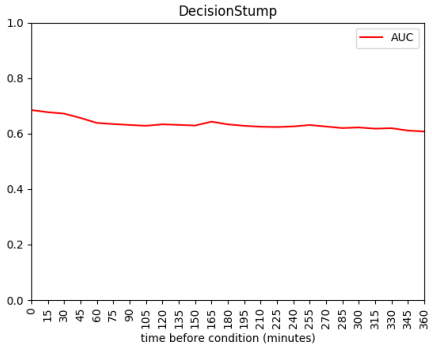
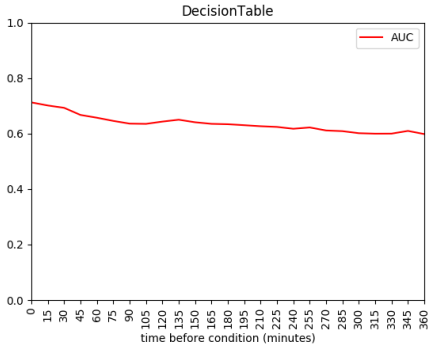


Figure F.12: Sepsis validation F-measures.

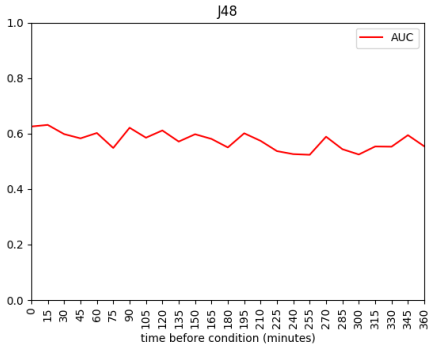
F.1.3 Machine Learning Algorithms over time



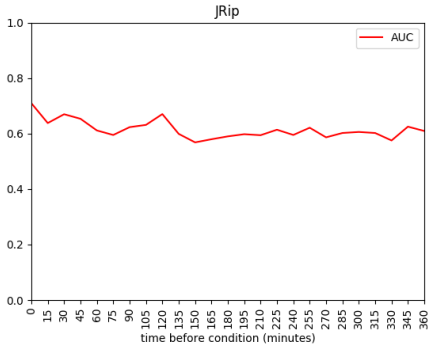
(a) Decision Stump



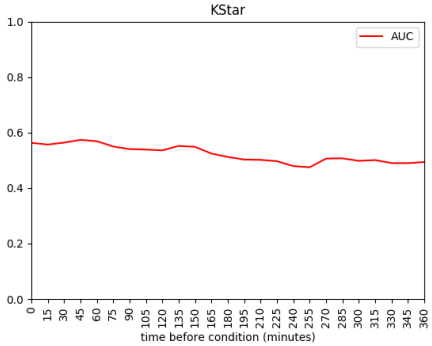
(b) Decision Table



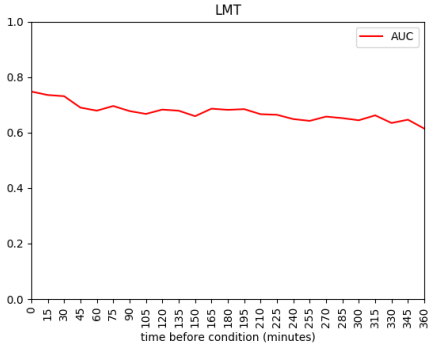
(c) J48



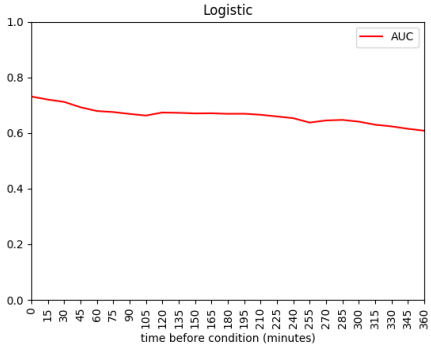
(d) JRip



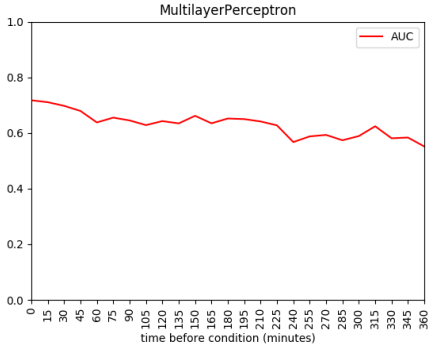
(e) KStar



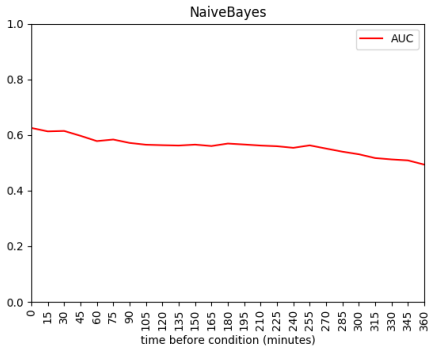
(f) LMT



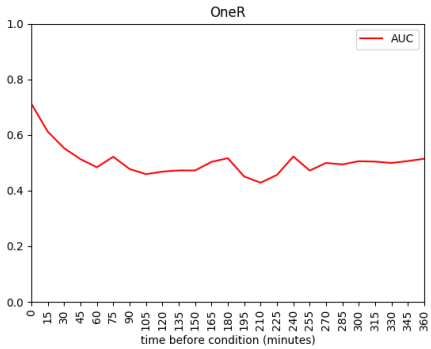
(g) Logistic Regression



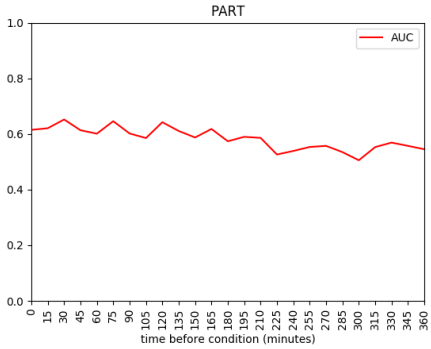
(h) Multilayer Perceptron



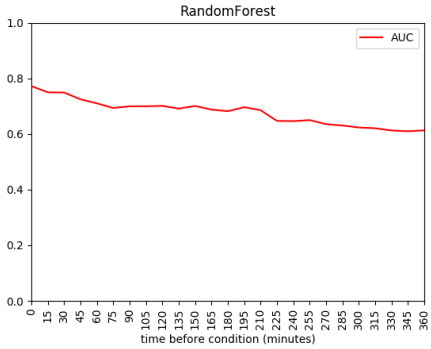
(i) Naive Bayes



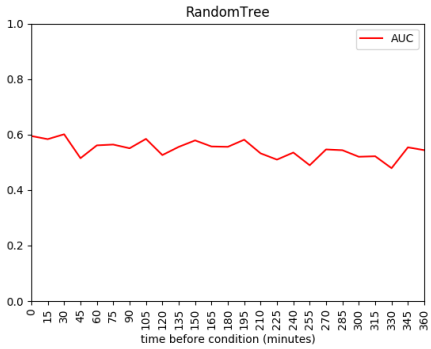
(j) One R



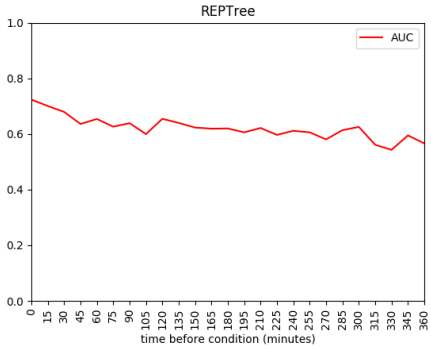
(k) PART

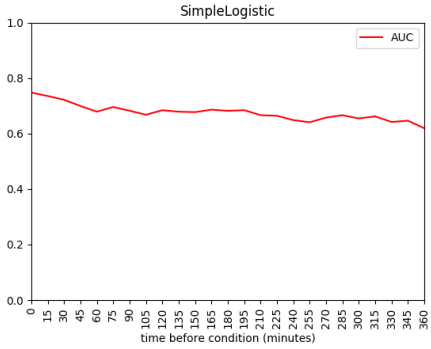


(l) Random Forest

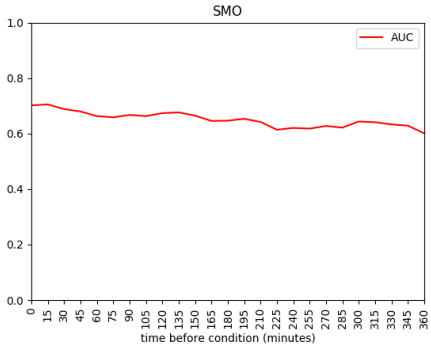


(m) Random Tree

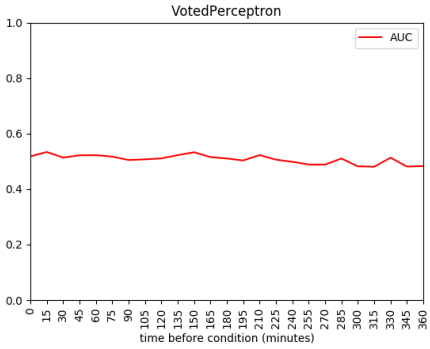




(n) Simple Logistic Regression



(o) SMO



(p) Voted Perceptron

Figure F.13: Algorithm's performance over time on the sepsis selection criteria.

F.1.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Magnesium Sulfate (Bolus)	50	50	50	50	50	50	0.001
Bands	0	0	5.5	0	0	0	<0.0001
Arterial Line Zero/Calibrate	0	1	1	0	0	1	0.006
Calculated Bicarbonate, Whole Blood	17	20	23	5	20	22	<0.0001
Nitroglycerin	0.275	0.578	1.601	0.126	0.454	1.390	0.026
Low Exhaled Min Vol	3	3.5	4	1	3.5	4	0.001
Non-Invasive Blood Pressure Alarm - High	160	160	160	160	160	160	0.981
Total PEEP Level	5	5	6	2	5	5.5	<0.0001
pCO_2	34	40	47	31	38	45	0.000
GCS Total	15	15	15	15	15	15	<0.0001
Stroke Volume	31.404	41.250	44.634	31.404	41.250	44.634	0.176
ALT	10	16	25	9	15	19	<0.0001
LVSW	31.62	34.014	39.168	31.62	34.014	39.168	0.102
Arterial Base Excess	-4	0	0	0	0	0	<0.0001
D5/.45NS	0	0	5	0	0	0	0.001
Digoxin	0.3	0.6	0.6	0.3	0.6	0.6	0.047
Lactate	1.2	1.8	3	0.9	1.3	1.8	<0.0001
Ventilator Mode	11	11	47	11	11	11	<0.0001
Creatine Kinase (CK)	38	79	148	37	73	144	0.800
Potassium	3.7	4.1	4.5	3.7	4	4.3	0.081
Asparate Aminotransferase (AST)	19	29	50	15	21	33	<0.0001
Cholesterol, Total	115	139	156	115	142	162	0.395
GT Flush	10	30	30	10	30	30	0.365

Table F.1: Sepsis model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Magnesium Sulfate (Bolus)	50	50	50	50	50	50	50	50	50	50	50	50	<0.0001
Bands	0	0	8	0	0	0	0	0	3	0	0	0	<0.0001
Arterial Line Zero/Calibrate	0	1	1	0	0	1	0	1	1	0	1	1	<0.0001
Calculated Bicarbonate, Whole Blood	19	21	23	5	20	22	16	21	24	6	20	22	<0.0001
Nitroglycerin	0.401	0.657	2.225	0.107	0.454	1.390	0.401	1.015	3.029	0.143	0.454	1.390	<0.0001
Low Exhaled Min Vol	3	4	4	0	3.5	4	2.875	3.5	4	1.5	3.5	4	<0.0001
Non-Invasive Blood Pressure Alarm - High	160	160	160	160	160	160	160	160	160	160	160	160	0.919
Total PEEP Level	5	5	8	0	5	5.4	5	5	6	5	5	6	<0.0001
pCO_2	35	40	49	31	38	45	32	39	47	33	39	46	<0.0001
GCS Total	14	15	15	15	15	15	15	15	15	15	15	15	<0.0001
Stroke Volume	31.404	41.250	44.634	31.404	41.250	44.634	31.404	41.250	44.634	31.404	41.250	44.634	0.427
ALT	12	18	28	9	15	19	11	16	22	10	14	19	<0.0001
LVSW	31.620	34.014	39.168	31.620	34.014	39.168	31.620	34.014	39.168	31.620	34.014	39.168	0.278
Arterial Base Excess	-5	-2	0	0	0	0	-2	0	1	0	0	1	<0.0001
D5/.45NS	0	0	10	0	0	0	0	0	10	0	0	0	<0.0001
Digoxin	0.3	0.6	0.6	0.3	0.6	0.6	0.3	0.6	0.6	0.3	0.6	0.6	0.199
Lactate	1.408	2.4	3.525	0.9	1.2	1.8	1.1	1.6	2.4	1	1.3	1.8	<0.0001
Ventilator Mode	11	11	49	11	11	11	11	11	30	11	11	22	<0.0001
Creatine Kinase (CK)	37.75	87	177.5	37	71	138	45.75	97.5	220	38.5	66	126.5	<0.0001
Potassium	3.7	4.1	4.6	3.7	4	4.3	3.7	4.1	4.5	3.7	4	4.4	0.006
Asparate Aminotransferase (AST)	21	33	59.5	14	21	31	20	29	53	17	23	37	<0.0001
Cholesterol, Total	115	139.5	155.75	115	142	164	114	139	149.25	115	139	153	0.025
GT Flush	20	30	30	10	30	30	20	30	30	10	30	30	<0.0001

Table F.2: Sepsis model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	6914	227	6687	
Age, median	62	70	62	*
Height, median (Q1,Q3)	67.00 (63.00, 70.00)	65.00 (62.00, 69.00)	67.00 (64.00, 70.00)	0.0625
Weight, median (Q1,Q3)	77.55 (64.12, 92.00)	75.60 (63.00, 90.00)	78.00 (64.40, 92.50)	0.3206
BMI, median (Q1,Q3)	28.00 (24.44, 31.86)	27.84 (24.14, 33.49)	28.06 (24.45, 31.80)	0.8515
Male	3694 (0.53)	123 (0.54)	3571 (0.53)	0.8739
In-Hospital Mortality	566 (0.08)	99 (0.44)	467 (0.07)	<0.0001
30 Day Mortality	733 (0.11)	106 (0.47)	627 (0.09)	<0.0001
ICU LOS, median(Q1,Q3)	1.63 (1.01, 2.84)	7.05 (3.33, 13.80)	1.57 (1.00, 2.70)	<0.0001
ICU				
CCU	1062 (0.15)	21 (0.09)	1041 (0.16)	0.0169
CSRU	877 (0.13)	16 (0.07)	861 (0.13)	0.0153
MICU	2482 (0.36)	139 (0.61)	2343 (0.35)	<0.0001
SICU	1208 (0.17)	37 (0.16)	1171 (0.18)	0.6675
TSICU	950 (0.14)	14 (0.06)	936 (0.14)	0.0017
ethnicity				
Asian	181 (0.03)	10 (0.04)	171 (0.03)	0.0906
Black	579 (0.08)	24 (0.11)	555 (0.08)	0.2445
Hispanic	268 (0.04)	10 (0.04)	258 (0.04)	0.6806
White	5072 (0.73)	157 (0.69)	4915 (0.74)	0.4530
insurance				
Government	232 (0.03)	0 (0.00)	232 (0.03)	0.0050
Medicaid	678 (0.10)	15 (0.07)	663 (0.10)	0.1177
Medicare	3368 (0.49)	154 (0.68)	3214 (0.48)	<0.0001
Private	2544 (0.37)	56 (0.25)	2488 (0.37)	0.0022
Self Pay	92 (0.01)	2 (0.01)	90 (0.01)	0.5505

Table F.3: Sepsis demographics of validation set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's Chi-Squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	6914	140	6137	550	87	
Age, median	62	71.5	61	68	68	*
Height, median (Q1,Q3)	67.00 (63.00, 70.00)	65.00 (62.00, 69.00)	67.00 (64.00, 70.00)	66.00 (62.75, 70.00)	66.50 (63.00, 69.25)	0.0745
Weight, median (Q1,Q3)	77.55 (64.12, 92.00)	75.45 (60.00, 90.10)	78.60 (65.00, 93.00)	69.00 (61.23, 86.32)	76.60 (65.10, 86.60)	0.0348
BMI, median (Q1,Q3)	28.00 (24.44, 31.86)	27.04 (24.00, 33.49)	28.24 (24.56, 31.94)	26.13 (23.77, 29.62)	28.54 (25.92, 32.76)	0.2072
Male	3694 (0.53)	72 (0.51)	3274 (0.53)	297 (0.54)	51 (0.59)	0.8999
In-Hospital Mortality	566 (0.08)	66 (0.47)	381 (0.06)	86 (0.16)	33 (0.38)	<0.0001
30 Day Mortality	733 (0.11)	71 (0.51)	521 (0.08)	106 (0.19)	35 (0.40)	<0.0001
ICU LOS, median(Q1,Q3)	1.63 (1.01, 2.84)	7.98 (3.84, 14.94)	1.51 (0.99, 2.61)	2.08 (1.16, 3.87)	5.69 (2.84, 11.69)	<0.0001
ICU						
CCU	1062 (0.15)	16 (0.11)	967 (0.16)	74 (0.13)	5 (0.06)	0.0356
CSRU	877 (0.13)	8 (0.06)	769 (0.13)	92 (0.17)	8 (0.09)	0.0038
MICU	2482 (0.36)	83 (0.59)	2084 (0.34)	259 (0.47)	56 (0.64)	<0.0001
SICU	1208 (0.17)	21 (0.15)	1097 (0.18)	74 (0.13)	16 (0.18)	0.1030
TSICU	950 (0.14)	12 (0.09)	887 (0.14)	49 (0.09)	2 (0.02)	<0.0001
ethnicity						
Asian	181 (0.03)	5 (0.04)	165 (0.03)	6 (0.01)	5 (0.06)	0.0327
Black	579 (0.08)	14 (0.10)	505 (0.08)	50 (0.09)	10 (0.11)	0.5838
Hispanic	268 (0.04)	6 (0.04)	234 (0.04)	24 (0.04)	4 (0.05)	0.9015
White	5072 (0.73)	94 (0.67)	4512 (0.74)	403 (0.73)	63 (0.72)	0.8565
insurance						
Government	232 (0.03)	0 (0.00)	218 (0.04)	14 (0.03)	0 (0.00)	0.0244
Medicaid	678 (0.10)	8 (0.06)	619 (0.10)	44 (0.08)	7 (0.08)	0.1728
Medicare	3368 (0.49)	97 (0.69)	2883 (0.47)	331 (0.60)	57 (0.66)	<0.0001
Private	2544 (0.37)	33 (0.24)	2332 (0.38)	156 (0.28)	23 (0.26)	<0.0001
Self Pay	92 (0.01)	2 (0.01)	85 (0.01)	5 (0.01)	0 (0.00)	0.5644

Table F.4: Sepsis demographics of validation set patients in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's Chi-Squared test.

Variable	Total	Condition Positive	condition Negative	P-value
N	6914	227	6687	
Comorbidity				
Congestive heart failure	1442 (0.21)	97 (0.43)	1345 (0.20)	<0.0001
Cardiac arrhythmias	2128 (0.31)	97 (0.43)	2031 (0.30)	0.0010
Valvular disease	901 (0.13)	31 (0.14)	870 (0.13)	0.7909
Pulmonary circulation disorders	339 (0.05)	10 (0.04)	329 (0.05)	0.7305
Peripheral vascular disorders	670 (0.10)	32 (0.14)	638 (0.10)	0.0301
Hypertension, uncomplicated	3140 (0.45)	98 (0.43)	3042 (0.45)	0.6101
Hypertension, complicated	362 (0.05)	13 (0.06)	349 (0.05)	0.7423
Paralysis	224 (0.03)	1 (0.00)	223 (0.03)	0.0172
Other neurological disorders	493 (0.07)	17 (0.07)	476 (0.07)	0.8370
Chronic pulmonary disease	1637 (0.24)	68 (0.30)	1569 (0.23)	0.0480
Diabetes, uncomplicated	1326 (0.19)	63 (0.28)	1263 (0.19)	0.0027
Diabetes, complicated	294 (0.04)	7 (0.03)	287 (0.04)	0.3853
Hypothyroidism	755 (0.11)	35 (0.15)	720 (0.11)	0.0370
Renal failure	412 (0.06)	23 (0.10)	389 (0.06)	0.0088
Liver disease	345 (0.05)	30 (0.13)	315 (0.05)	<0.0001
Peptic ulcer disease excluding bleeding	49 (0.01)	2 (0.01)	47 (0.01)	0.7538
AIDS/HIV	51 (0.01)	3 (0.01)	48 (0.01)	0.2976
Lymphoma	76 (0.01)	6 (0.03)	70 (0.01)	0.0241
Metastatic cancer	78 (0.01)	2 (0.01)	76 (0.01)	0.7215
Solid tumor without metastasis	540 (0.08)	27 (0.12)	513 (0.08)	0.0252
Rheumatoid arthritis/collagen vascular diseases	224 (0.03)	14 (0.06)	210 (0.03)	0.0127
Coagulopathy	308 (0.04)	57 (0.25)	251 (0.04)	<0.0001
Weight loss	228 (0.03)	28 (0.12)	200 (0.03)	<0.0001
Fluid and electrolyte disorders	1589 (0.23)	119 (0.52)	1470 (0.22)	<0.0001
Blood loss anemia	121 (0.02)	6 (0.03)	115 (0.02)	0.3010
Deficiency anemia	210 (0.03)	5 (0.02)	205 (0.03)	0.4631
Alcohol abuse	218 (0.03)	6 (0.03)	212 (0.03)	0.6600
Drug abuse	303 (0.04)	3 (0.01)	300 (0.04)	0.0251
Psychoses	132 (0.02)	9 (0.04)	123 (0.02)	0.0227
Depression	739 (0.11)	17 (0.07)	722 (0.11)	0.1338
Trauma	1002 (0.14)	25 (0.11)	977 (0.15)	0.1615

Table F.5: Sepsis Comorbidities of validation set patients by condition positive or negative. P-values for patient counts calculated by chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	6914	140	6137	550	87	
Comorbidity						
Congestive heart failure	1442 (0.21)	69 (0.49)	1123 (0.18)	222 (0.40)	28 (0.32)	<0.0001
Cardiac arrhythmias	2128 (0.31)	63 (0.45)	1768 (0.29)	263 (0.48)	34 (0.39)	<0.0001
Valvular disease	901 (0.13)	23 (0.16)	760 (0.12)	110 (0.20)	8 (0.09)	<0.0001
Pulmonary circulation disorders	339 (0.05)	8 (0.06)	281 (0.05)	48 (0.09)	2 (0.02)	0.0003
Peripheral vascular disorders	670 (0.10)	22 (0.16)	567 (0.09)	71 (0.13)	10 (0.11)	0.0053
Hypertension, uncomplicated	3140 (0.45)	59 (0.42)	2794 (0.46)	248 (0.45)	39 (0.45)	0.9471
Hypertension, complicated	362 (0.05)	10 (0.07)	276 (0.04)	73 (0.13)	3 (0.03)	<0.0001
Paralysis	224 (0.03)	0 (0.00)	214 (0.03)	9 (0.02)	1 (0.01)	0.0105
Other neurological disorders	493 (0.07)	11 (0.08)	444 (0.07)	32 (0.06)	6 (0.07)	0.6748
Chronic pulmonary disease	1637 (0.24)	46 (0.33)	1392 (0.23)	177 (0.32)	22 (0.25)	<0.0001
Diabetes, uncomplicated	1326 (0.19)	41 (0.29)	1124 (0.18)	139 (0.25)	22 (0.25)	<0.0001
Diabetes, complicated	294 (0.04)	4 (0.03)	243 (0.04)	44 (0.08)	3 (0.03)	0.0002
Hypothyroidism	755 (0.11)	25 (0.18)	642 (0.10)	78 (0.14)	10 (0.11)	0.0052
Renal failure	412 (0.06)	15 (0.11)	307 (0.05)	82 (0.15)	8 (0.09)	<0.0001
Liver disease	345 (0.05)	25 (0.18)	244 (0.04)	71 (0.13)	5 (0.06)	<0.0001
Peptic ulcer disease excluding bleeding	49 (0.01)	1 (0.01)	38 (0.01)	9 (0.02)	1 (0.01)	0.0548
AIDS/HIV	51 (0.01)	0 (0.00)	38 (0.01)	10 (0.02)	3 (0.03)	0.0002
Lymphoma	76 (0.01)	2 (0.01)	59 (0.01)	11 (0.02)	4 (0.05)	0.0019
Metastatic cancer	78 (0.01)	1 (0.01)	64 (0.01)	12 (0.02)	1 (0.01)	0.1106
Solid tumor without metastasis	540 (0.08)	19 (0.14)	460 (0.07)	53 (0.10)	8 (0.09)	0.0257
Rheumatoid arthritis/collagen vascular diseases	224 (0.03)	12 (0.09)	186 (0.03)	24 (0.04)	2 (0.02)	0.0014
Coagulopathy	308 (0.04)	53 (0.38)	131 (0.02)	120 (0.22)	4 (0.05)	<0.0001
Weight loss	228 (0.03)	20 (0.14)	154 (0.03)	46 (0.08)	8 (0.09)	<0.0001
Fluid and electrolyte disorders	1589 (0.23)	86 (0.61)	1177 (0.19)	293 (0.53)	33 (0.38)	<0.0001
Blood loss anemia	121 (0.02)	5 (0.04)	104 (0.02)	11 (0.02)	1 (0.01)	0.3710
Deficiency anemia	210 (0.03)	4 (0.03)	170 (0.03)	35 (0.06)	1 (0.01)	<0.0001
Alcohol abuse	218 (0.03)	4 (0.03)	188 (0.03)	24 (0.04)	2 (0.02)	0.3989
Drug abuse	303 (0.04)	2 (0.01)	268 (0.04)	32 (0.06)	1 (0.01)	0.0588
Psychoses	132 (0.02)	7 (0.05)	108 (0.02)	15 (0.03)	2 (0.02)	0.0211
Depression	739 (0.11)	12 (0.09)	658 (0.11)	64 (0.12)	5 (0.06)	0.3850
Trauma	1002 (0.14)	18 (0.13)	905 (0.15)	72 (0.13)	7 (0.08)	0.2871

Table F.6: Sepsis comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by chi-squared test.

Appendix G

DISSEMINATED INTRAVASCULAR COAGULATION

G.1 DIC

G.1.1 Training and Testing

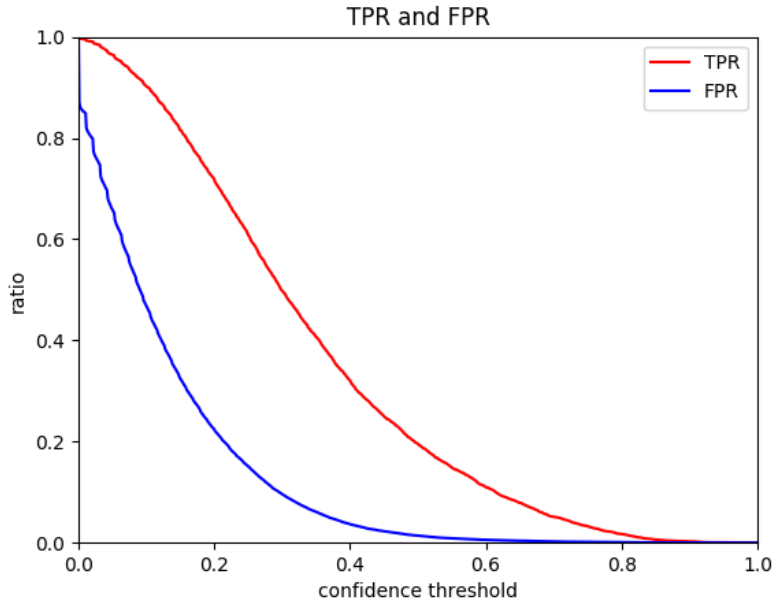


Figure G.1: DIC training/testing true positive and false positive rates.

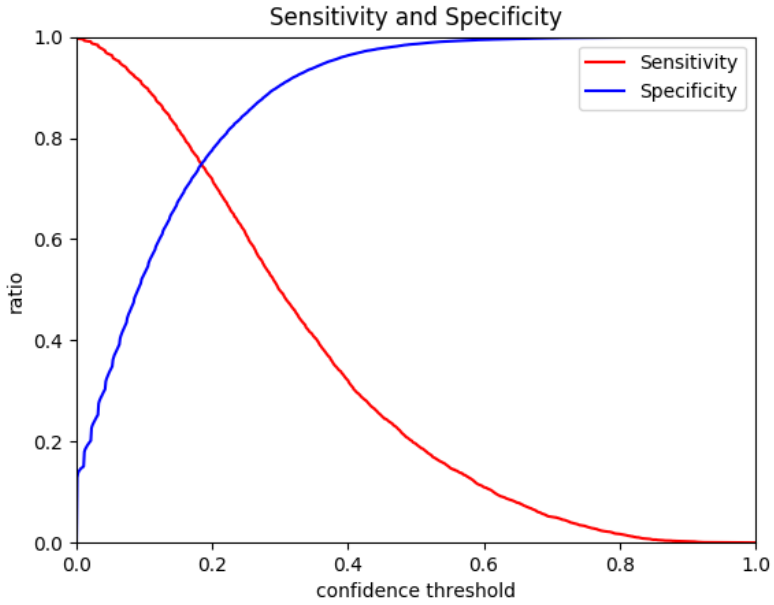


Figure G.2: DIC training/testing sensitivity and specificity.

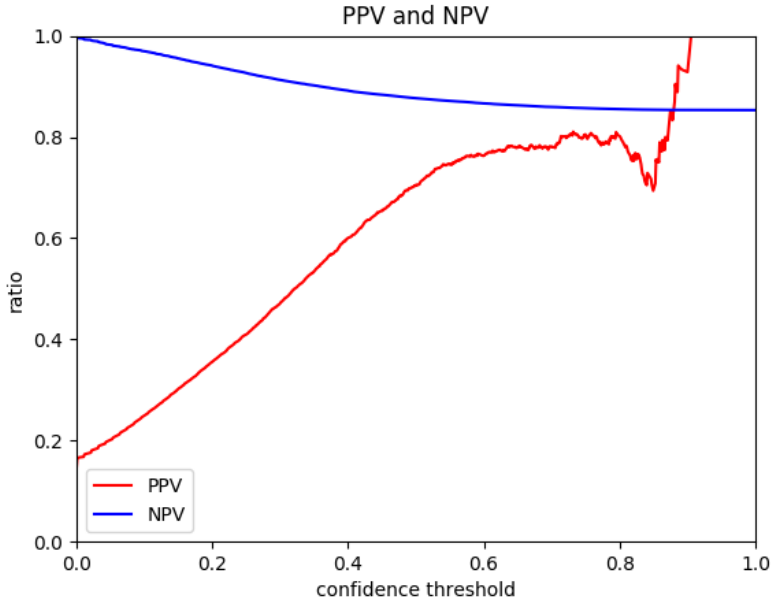


Figure G.3: DIC training/testing positive and negative predictive value.

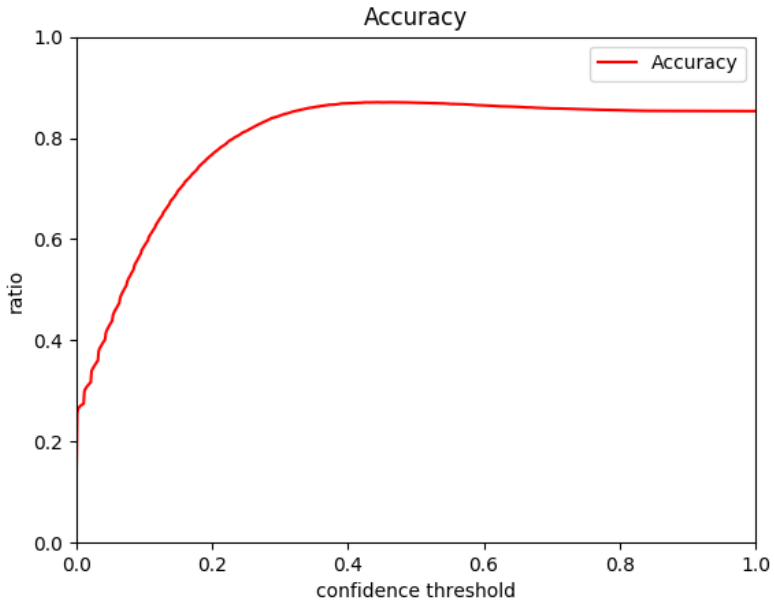


Figure G.4: DIC training/testing accuracy.

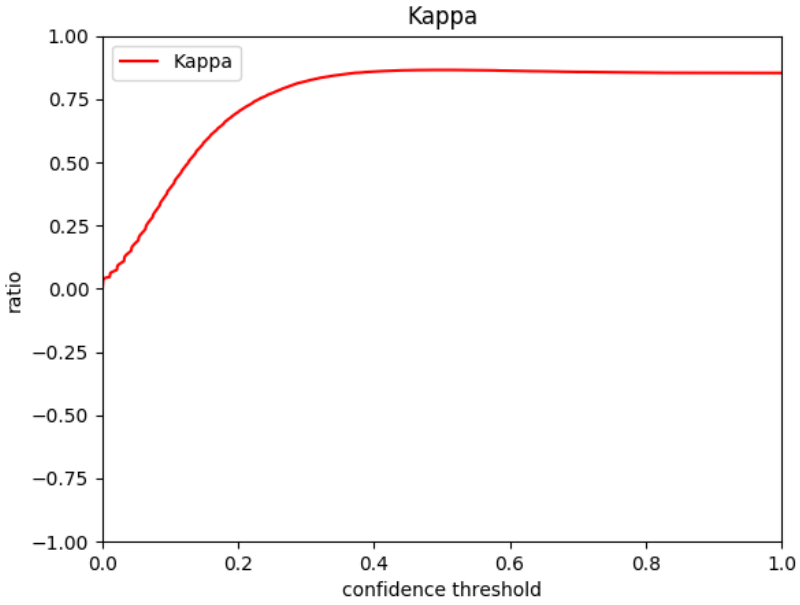


Figure G.5: DIC training/testing kappa.

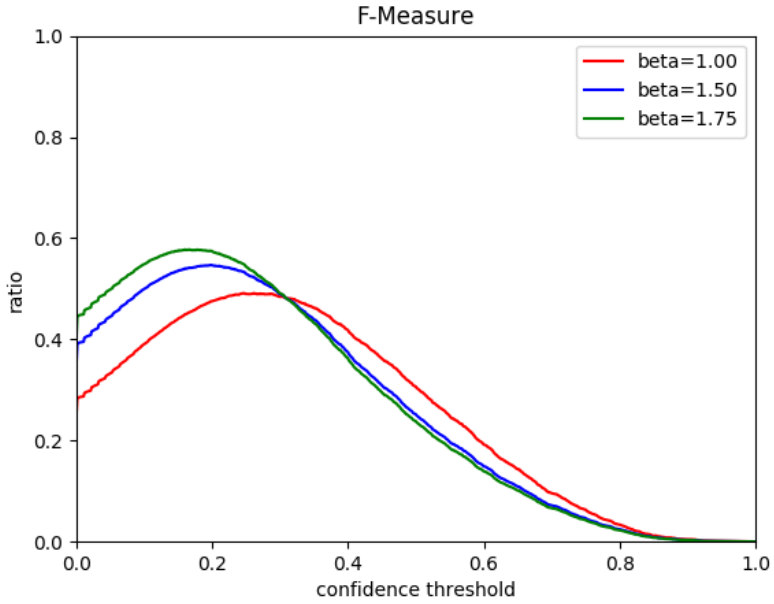


Figure G.6: DIC training/testing F-measures.

G.1.2 Validation

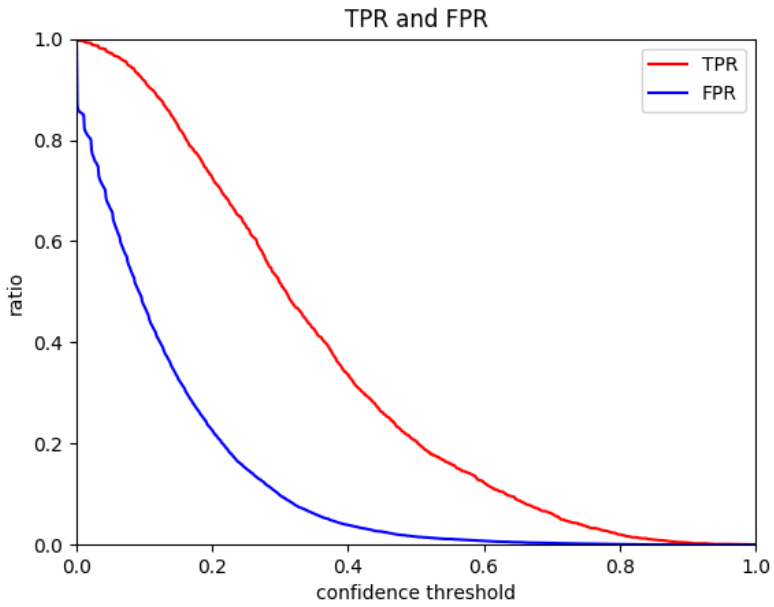


Figure G.7: DIC validation true and false positive rates.

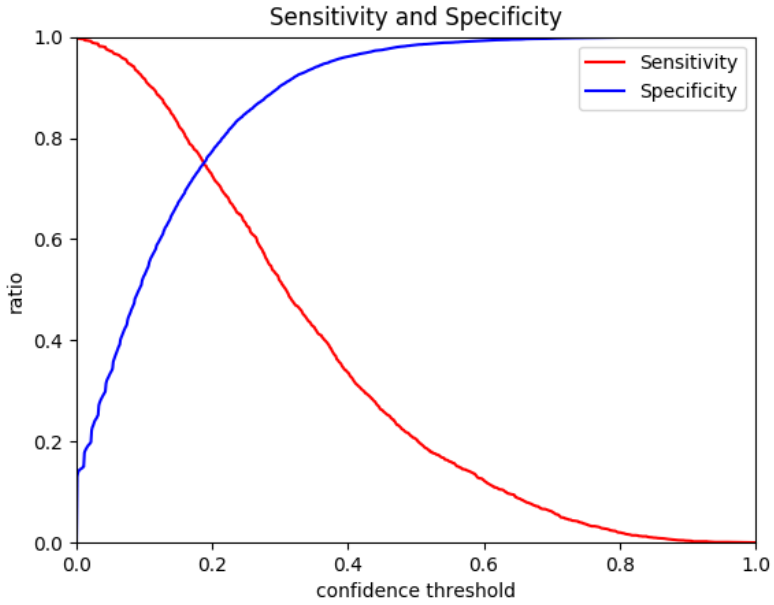


Figure G.8: DIC validation sensitivity and specificity.

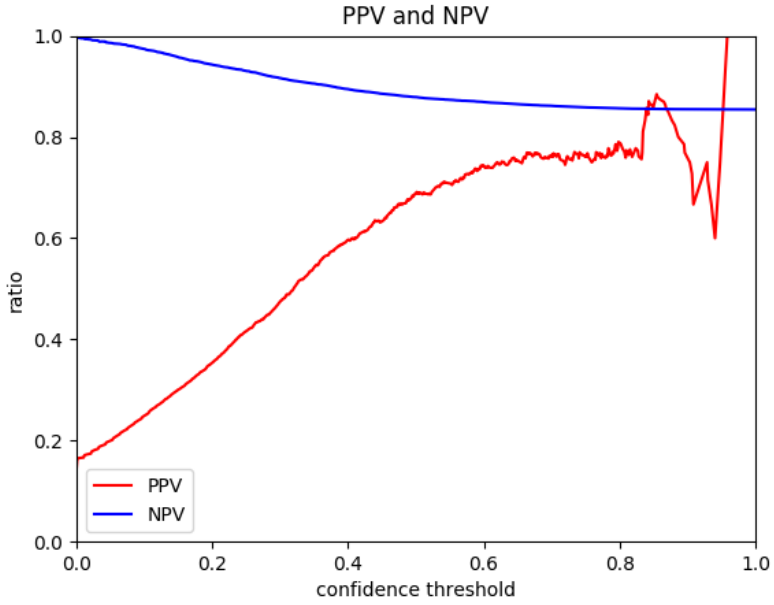


Figure G.9: DIC validation positive and negative predictive values.

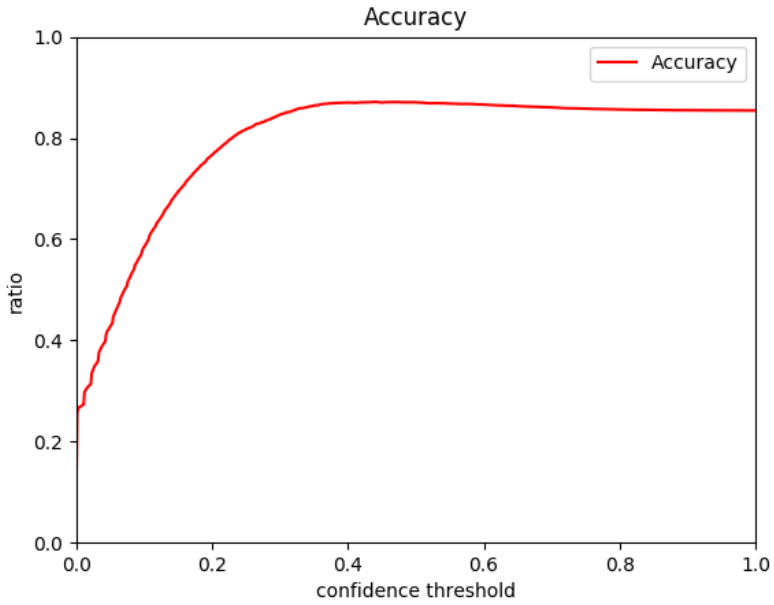


Figure G.10: DIC validation accuracy.

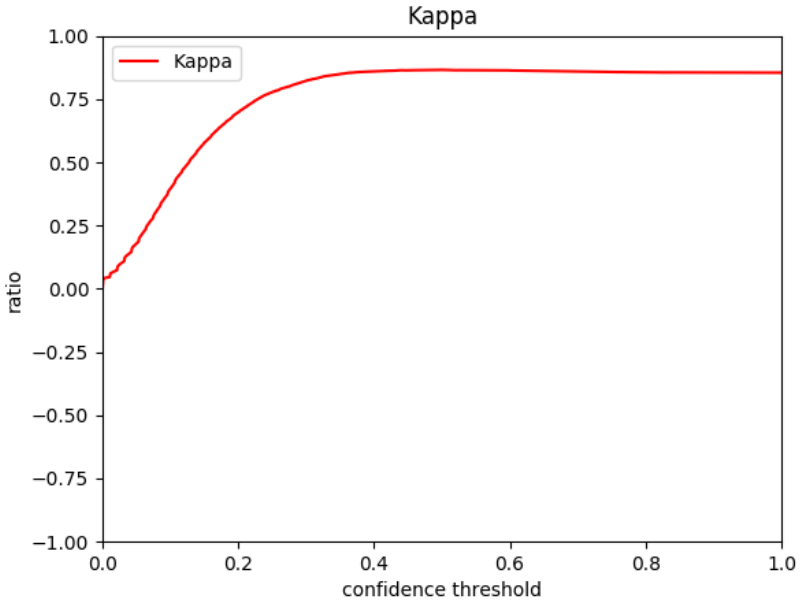


Figure G.11: DIC validation kappa.

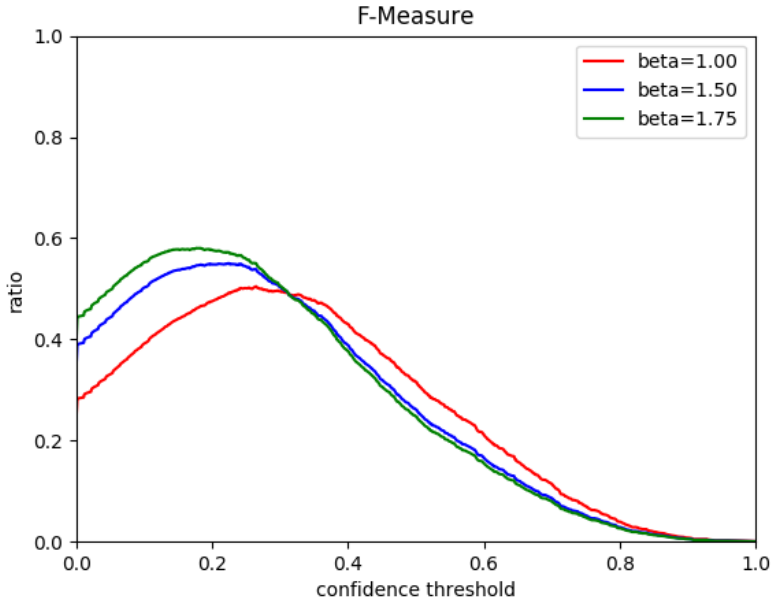
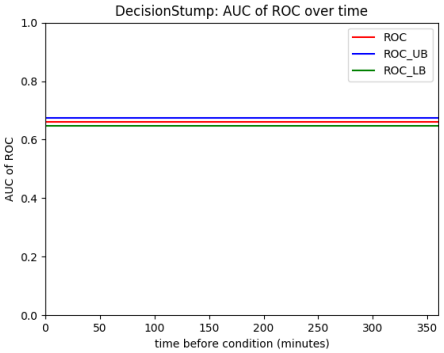
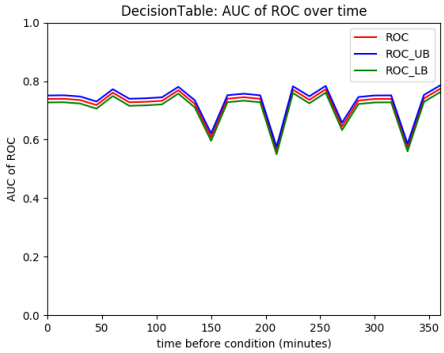


Figure G.12: DIC validation F-measures.

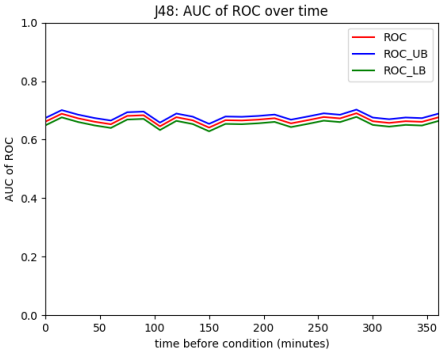
G.1.3 Machine Learning Algorithms over time



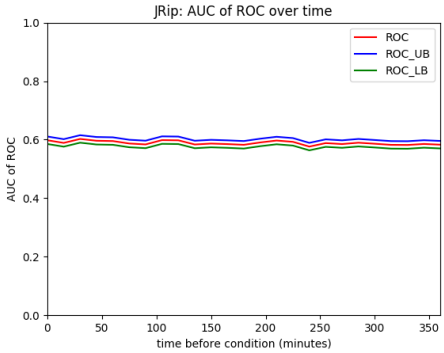
(a) Decision Stump



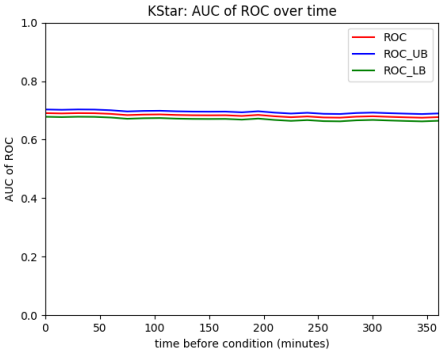
(b) Decision Table



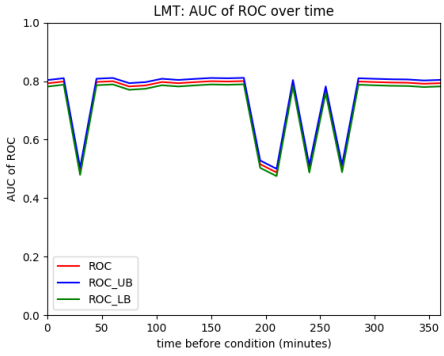
(c) J48



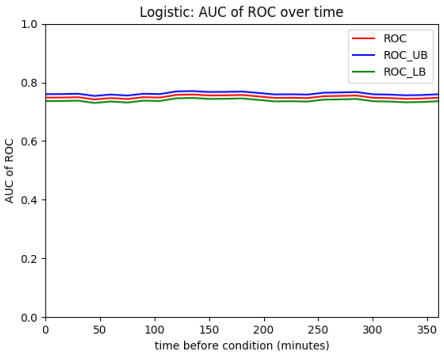
(d) JRip



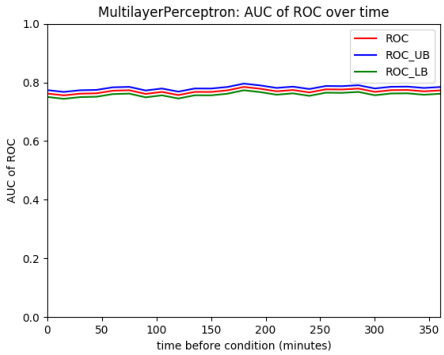
(e) KStar



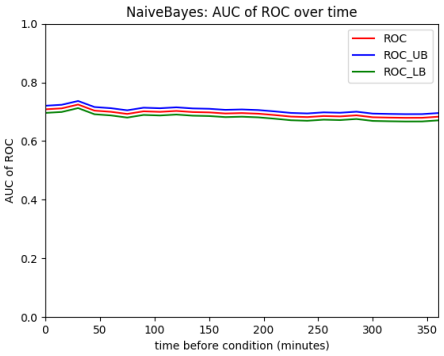
(f) LMT



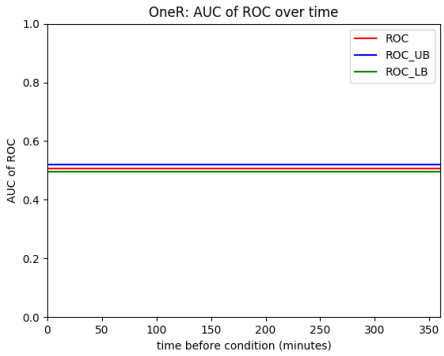
(g) Logistic Regression



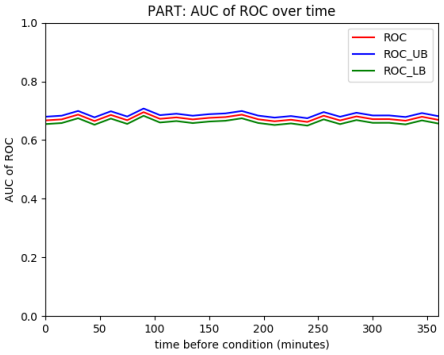
(h) Multilayer Perceptron



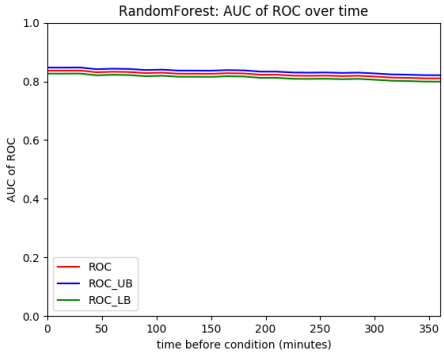
(i) Naive Bayes



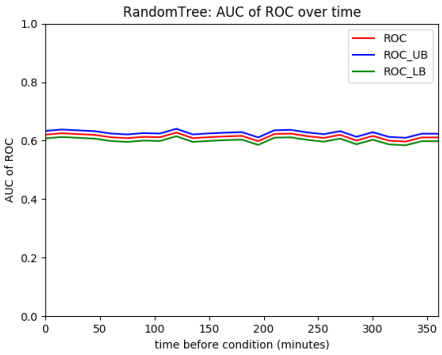
(j) One R



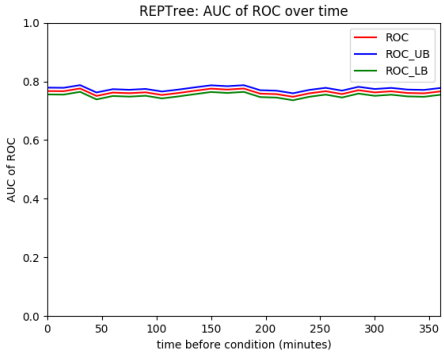
(k) PART

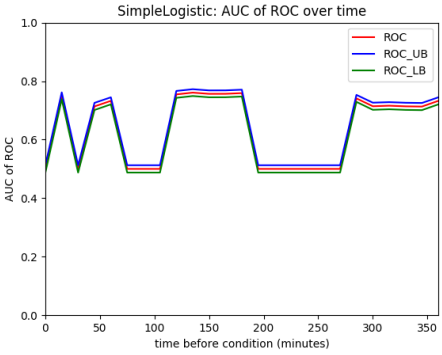


(l) Random Forest

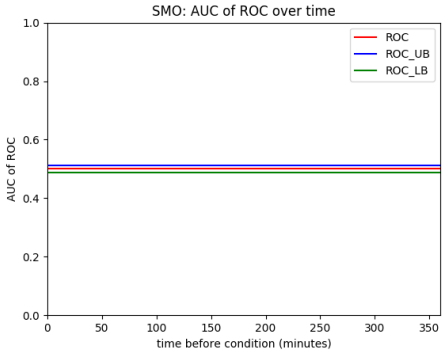


(m) Random Tree

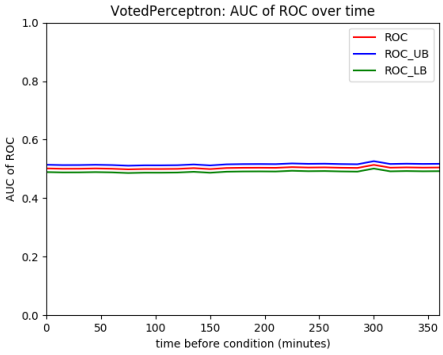




(n) Simple Logistic Regression



(o) SMO



(p) Voted Perceptron

Figure G.13: Algorithm's performance over time on the DIC selection criteria.

G.1.4 Comparative Demographics, Variables, and Comorbidities

Label	Condition Positive			Condition Negative			P-value
	Q1	median	Q3	Q1	median	Q3	
Admit Wt	58.3	66.6	82	58.3	63.6	79.6	<0.0001
Previous Weight	57.5	69	84.7	58	68.8	80.275	0.6502
Calcium Gluconate	2	2	2	1	2	2	<0.0001
CPK	31	59	121	30	65	120	0.1556
Digoxin	0.2	0.3	0.4	0.2	0.3	0.4	0.1109
Eosinophils	0	0.3	1.1	0.1	0.5	1.5	<0.0001
Temperature C (calc)	35.889	36.611	37.389	35.944	36.667	37.278	0.4849
Total Bilirubin	0.3	0.5	0.7	0.3	0.5	0.8	0.1789
Creatinine	0.7	1	1.5	0.6	0.8	1.1	<0.0001
CK-MB	2	3	4	2	3	4	0.7481
Pre-Admission Intake	0	110	300	0	110	350	0.0740
Differential-Basos	0	0	0	0	0	0.1	0.0205
Braden Friction/Shear	2	2	3	2	2	3	0.2073
Braden Nutrition	2	2	3	2	2	3	0.0834
Difficulty swallowing	0	0	0	0	0	0	<0.0001
Arterial Blood Pressure Alarm - High	140	160	160	140	160	160	<0.0001
Minute Volume Alarm - Low	3	3.5	4	3	3.5	4	<0.0001
Minute Volume(Obser)	6.4	7.2	9	6.5	7.2	8.49	0.1491
INR(PT)	13.3	14.1	15.9	12.5	13.3	14.6	<0.0001
Discharge needs	0	0	0	0	0	0	<0.0001
Central Venous Pressure Alarm - High	20	20	20	20	20	20	<0.0001
Non-Invasive Blood Pressure Alarm - High	160	160	160	160	160	160	<0.0001
Minute Volume	6.9	8	10.3	6.3	7.5	10.2	<0.0001
Resp Rate (Total)	10	12	16	10	12	15	<0.0001
Tube Feeding Residual	0	0	0	0	0	0	0.9169

Table G.1: DIC model variable composition of the validation data set. P-values calculated by Kruskal-Wallis test.

Label	True Positive			True Negative			False Positive			False Negative			P-value
	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	Q1	median	Q3	
Admit Wt	58.3	67	82	57.7	63.6	78	58.3	68	82.35	58.3	65	79.6	<0.0001
Previous Weight	56.9	68.3	85.1	58	68.3	78.7	56.8	69.1	85	58	70	83.075	0.0883
Calcium Gluconate	2	2	2	1	2	2	2	2	2	1	2	2	<0.0001
CPK	31	58	124	29	67	115	34	63	127	29	61	116	0.0189
Digoxin	0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4	<0.0001
Eosinophils	0	0.3	1.1	0.1	0.6	1.7	0	0.4	1.3	0.075	0.4	1.2	<0.0001
Temperature C (calc)	35.833	36.611	37.389	35.889	36.667	37.278	36.111	36.722	37.333	36.217	36.722	37.333	<0.0001
Total Bilirubin	0.3	0.5	0.7	0.3	0.4	0.7	0.3	0.5	0.8	0.3	0.5	0.7	0.0259
Creatinine	0.8	1.1	1.7	0.6	0.7	1	0.7	1	1.5	0.6	0.8	1	<0.0001
CK-MB	2	3	4	2	3	4	2	3	5	2	4	5	<0.0001
Pre-Admission Intake	0	120	300	0	110	350	0	110	350	0	0	350	<0.0001
Differential-Basos	0	0	0	0	0	0.1	0	0	0.1	0	0	0.1	0.0008
Braden Friction/Shear	2	2	3	2	2	3	2	2	3	2	2	3	0.5561
Braden Nutrition	2	2	3	2	2	3	2	2	3	2	2	3	0.3128
Difficulty swallowing	0	0	0	0	0	0	0	0	0	0	0	0	<0.0001
Arterial Blood Pressure Alarm - High	140	160	160	130	160	160	140	160	160	130	160	160	<0.0001
Minute Volume Alarm - Low	3	3.5	4	3	3.5	4	3	3.5	4	3	3.5	4	<0.0001
Minute Volume(Obser)	6.3	7.2	9	6.5	7.2	8.11	6.5	7.32	9.1	6.575	7.5	8.8	<0.0001
INR(PT)	13.5	14.5	16.4	12.2	13	14.2	13.2	14.1	15.8	12.5	13.2	14.125	<0.0001
Discharge needs	0	0	0	0	0	0	0	0	0	0	0	0	<0.0001
Central Venous Pressure Alarm - High	20	20	20	16	20	20	20	20	20	20	20	20	<0.0001
Non-Invasive Blood Pressure Alarm - High	160	160	160	160	160	160	160	160	160	160	160	160	<0.0001
Minute Volume	6.9	7.5	10.3	6.3	7.5	9.6	6.9	8.3	10.3	6.9	8.3	10.3	<0.0001
Resp Rate (Total)	10	13	17	10	12	14	10	12	18	10	12	16	<0.0001
Tube Feeding Residual	0	0	0	0	0	0	0	0	0	0	0	0	0.2138

Table G.2: DIC model variable performance of the validation data set. P-values calculated by Kruskal-Wallis test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	16399	2389	14010	
Age, median	62	67	61	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	0.2646
Weight, median (Q1,Q3)	78.00 (65.00, 92.00)	77.00 (64.00, 91.00)	78.50 (65.50, 92.50)	0.0470
BMI, median (Q1,Q3)	28.07 (24.56, 32.60)	27.99 (24.24, 32.33)	28.12 (24.64, 32.63)	0.2625
Male	9174 (0.56)	1410 (0.59)	7764 (0.55)	0.0295
In-Hospital Mortality	2300 (0.14)	675 (0.28)	1625 (0.12)	<0.0001
30 Day Mortality	2699 (0.16)	765 (0.32)	1934 (0.14)	<0.0001
ICU LOS, median(Q1,Q3)	2.09 (1.13, 4.40)	3.23 (1.86, 7.00)	1.98 (1.09, 4.02)	<0.0001
ICU				
CCU	2124 (0.13)	315 (0.13)	1809 (0.13)	0.7316
CSRU	2595 (0.16)	462 (0.19)	2133 (0.15)	<0.0001
MICU	5636 (0.34)	909 (0.38)	4727 (0.34)	0.0009
SICU	2361 (0.14)	403 (0.17)	1958 (0.14)	0.0006
TSICU	1773 (0.11)	289 (0.12)	1484 (0.11)	0.0387
ethnicity				
Asian	527 (0.03)	63 (0.03)	464 (0.03)	0.0890
Black	1559 (0.10)	207 (0.09)	1352 (0.10)	0.1487
Hispanic	582 (0.04)	60 (0.03)	522 (0.04)	0.0036
White	11502 (0.70)	1711 (0.72)	9791 (0.70)	0.3496
insurance				
Government	471 (0.03)	43 (0.02)	428 (0.03)	0.0008
Medicaid	1532 (0.09)	192 (0.08)	1340 (0.10)	0.0239
Medicare	8154 (0.50)	1419 (0.59)	6735 (0.48)	<0.0001
Private	6047 (0.37)	709 (0.30)	5338 (0.38)	<0.0001
Self Pay	195 (0.01)	26 (0.01)	169 (0.01)	0.6251

Table G.3: DIC demographics of validation set patients by condition positive or negative. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's Chi-Squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	16399	1881	9975	4035	508	
Age, median	62	68	58	68	64.5	*
Height, median (Q1,Q3)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (64.00, 70.00)	67.00 (63.00, 70.00)	67.00 (64.00, 70.00)	0.0006
Weight, median (Q1,Q3)	78.00 (65.00, 92.00)	77.00 (64.33, 90.98)	79.00 (66.00, 92.00)	78.00 (65.00, 93.00)	76.70 (63.80, 90.95)	0.2141
BMI, median (Q1,Q3)	28.07 (24.56, 32.60)	27.66 (24.24, 32.36)	28.24 (24.86, 32.26)	28.00 (24.49, 32.79)	28.95 (24.54, 32.11)	0.6297
Male	9174 (0.56)	1119 (0.59)	5461 (0.55)	2303 (0.57)	291 (0.57)	0.0488
In-Hospital Mortality	2300 (0.14)	580 (0.31)	829 (0.08)	796 (0.20)	95 (0.19)	<0.0001
30 Day Mortality	2699 (0.16)	655 (0.35)	1020 (0.10)	914 (0.23)	110 (0.22)	<0.0001
ICU LOS, median(Q1,Q3)	2.09 (1.13, 4.40)	3.36 (1.88, 7.26)	1.88 (1.04, 3.74)	2.22 (1.21, 4.64)	2.98 (1.77, 5.85)	<0.0001
ICU						
CCU	2124 (0.13)	246 (0.13)	1188 (0.12)	621 (0.15)	69 (0.14)	<0.0001
CSRU	2595 (0.16)	349 (0.19)	1325 (0.13)	808 (0.20)	113 (0.22)	<0.0001
MICU	5636 (0.34)	746 (0.40)	3058 (0.31)	1669 (0.41)	163 (0.32)	<0.0001
SICU	2361 (0.14)	314 (0.17)	1440 (0.14)	518 (0.13)	89 (0.18)	0.0007
TSICU	1773 (0.11)	222 (0.12)	1096 (0.11)	388 (0.10)	67 (0.13)	0.0187
ethnicity						
Asian	527 (0.03)	48 (0.03)	384 (0.04)	80 (0.02)	15 (0.03)	<0.0001
Black	1559 (0.10)	161 (0.09)	959 (0.10)	393 (0.10)	46 (0.09)	0.5248
Hispanic	582 (0.04)	42 (0.02)	394 (0.04)	128 (0.03)	18 (0.04)	0.0016
White	11502 (0.70)	1364 (0.73)	7003 (0.70)	2788 (0.69)	347 (0.68)	0.4956
insurance						
Government	471 (0.03)	29 (0.02)	335 (0.03)	93 (0.02)	14 (0.03)	<0.0001
Medicaid	1532 (0.09)	143 (0.08)	1016 (0.10)	324 (0.08)	49 (0.10)	<0.0001
Medicare	8154 (0.50)	1158 (0.62)	4231 (0.42)	2504 (0.62)	261 (0.51)	<0.0001
Private	6047 (0.37)	532 (0.28)	4273 (0.43)	1065 (0.26)	177 (0.35)	<0.0001
Self Pay	195 (0.01)	19 (0.01)	120 (0.01)	49 (0.01)	7 (0.01)	0.8739

Table G.4: DIC demographics of validation set patients in context of model performance. *Age above 89 is obfuscated by MIMIC for privacy protection, making distributions and p-value calculations invalid. P-values for distributions of continuous variables calculated by Kruskal-Wallis test. P-values for patient counts calculated by Pearson's Chi-Squared test.

Variable	Total	Condition Positive	condition Negative	P-value
N	16399	2389	14010	
Comorbidity				
Congestive heart failure	4259 (0.26)	939 (0.39)	3320 (0.24)	<0.0001
Cardiac arrhythmias	5167 (0.32)	1003 (0.42)	4164 (0.30)	<0.0001
Valvular disease	2307 (0.14)	429 (0.18)	1878 (0.13)	<0.0001
Pulmonary circulation disorders	798 (0.05)	140 (0.06)	658 (0.05)	0.0172
Peripheral vascular disorders	1731 (0.11)	339 (0.14)	1392 (0.10)	<0.0001
Hypertension, uncomplicated	6167 (0.38)	868 (0.36)	5299 (0.38)	0.2724
Hypertension, complicated	1958 (0.12)	329 (0.14)	1629 (0.12)	0.0051
Paralysis	399 (0.02)	48 (0.02)	351 (0.03)	0.1507
Other neurological disorders	1015 (0.06)	139 (0.06)	876 (0.06)	0.4303
Chronic pulmonary disease	3427 (0.21)	499 (0.21)	2928 (0.21)	0.9906
Diabetes, uncomplicated	3101 (0.19)	485 (0.20)	2616 (0.19)	0.0906
Diabetes, complicated	1091 (0.07)	187 (0.08)	904 (0.06)	0.0160
Hypothyroidism	1610 (0.10)	225 (0.09)	1385 (0.10)	0.5002
Renal failure	2296 (0.14)	395 (0.17)	1901 (0.14)	0.0003
Liver disease	1259 (0.08)	335 (0.14)	924 (0.07)	<0.0001
Peptic ulcer disease excluding bleeding	108 (0.01)	18 (0.01)	90 (0.01)	0.5364
AIDS/HIV	132 (0.01)	28 (0.01)	104 (0.01)	0.0305
Lymphoma	180 (0.01)	54 (0.02)	126 (0.01)	<0.0001
Metastatic cancer	141 (0.01)	20 (0.01)	121 (0.01)	0.8973
Solid tumor without metastasis	1116 (0.07)	190 (0.08)	926 (0.07)	0.0200
Rheumatoid arthritis/collagen vascular diseases	514 (0.03)	82 (0.03)	432 (0.03)	0.3733
Coagulopathy	1226 (0.07)	430 (0.18)	796 (0.06)	<0.0001
Weight loss	595 (0.04)	146 (0.06)	449 (0.03)	<0.0001
Fluid and electrolyte disorders	4133 (0.25)	782 (0.33)	3351 (0.24)	<0.0001
Blood loss anemia	300 (0.02)	67 (0.03)	233 (0.02)	0.0001
Deficiency anemia	445 (0.03)	67 (0.03)	378 (0.03)	0.7703
Alcohol abuse	577 (0.04)	123 (0.05)	454 (0.03)	<0.0001
Drug abuse	601 (0.04)	76 (0.03)	525 (0.04)	0.1816
Psychoses	254 (0.02)	27 (0.01)	227 (0.02)	0.0752
Depression	1405 (0.09)	124 (0.05)	1281 (0.09)	<0.0001
Trauma	1900 (0.12)	330 (0.14)	1570 (0.11)	0.0005

Table G.5: DIC Comorbidities of validation set patients by condition positive or negative. P-values for patient counts calculated by chi-squared test.

Variable	Total	True Positive	True Negative	False Positive	False Negative	P-value
N	16399	1881	9975	4035	508	
Comorbidity						
Congestive heart failure	4259 (0.26)	809 (0.43)	1784 (0.18)	1536 (0.38)	130 (0.26)	<0.0001
Cardiac arrhythmias	5167 (0.32)	827 (0.44)	2428 (0.24)	1736 (0.43)	176 (0.35)	<0.0001
Valvular disease	2307 (0.14)	343 (0.18)	1050 (0.11)	828 (0.21)	86 (0.17)	<0.0001
Pulmonary circulation disorders	798 (0.05)	116 (0.06)	395 (0.04)	263 (0.07)	24 (0.05)	<0.0001
Peripheral vascular disorders	1731 (0.11)	263 (0.14)	855 (0.09)	537 (0.13)	76 (0.15)	<0.0001
Hypertension, uncomplicated	6167 (0.38)	643 (0.34)	3699 (0.37)	1600 (0.40)	225 (0.44)	0.0007
Hypertension, complicated	1958 (0.12)	279 (0.15)	998 (0.10)	631 (0.16)	50 (0.10)	<0.0001
Paralysis	399 (0.02)	40 (0.02)	263 (0.03)	88 (0.02)	8 (0.02)	0.1706
Other neurological disorders	1015 (0.06)	112 (0.06)	642 (0.06)	234 (0.06)	27 (0.05)	0.4287
Chronic pulmonary disease	3427 (0.21)	400 (0.21)	1932 (0.19)	996 (0.25)	99 (0.19)	<0.0001
Diabetes, uncomplicated	3101 (0.19)	386 (0.21)	1755 (0.18)	861 (0.21)	99 (0.19)	<0.0001
Diabetes, complicated	1091 (0.07)	160 (0.09)	569 (0.06)	335 (0.08)	27 (0.05)	<0.0001
Hypothyroidism	1610 (0.10)	168 (0.09)	916 (0.09)	469 (0.12)	57 (0.11)	0.0002
Renal failure	2296 (0.14)	335 (0.18)	1132 (0.11)	769 (0.19)	60 (0.12)	<0.0001
Liver disease	1259 (0.08)	294 (0.16)	386 (0.04)	538 (0.13)	41 (0.08)	<0.0001
Peptic ulcer disease excluding bleeding	108 (0.01)	12 (0.01)	62 (0.01)	28 (0.01)	6 (0.01)	0.4932
AIDS/HIV	132 (0.01)	24 (0.01)	55 (0.01)	49 (0.01)	4 (0.01)	<0.0001
Lymphoma	180 (0.01)	40 (0.02)	76 (0.01)	50 (0.01)	14 (0.03)	<0.0001
Metastatic cancer	141 (0.01)	16 (0.01)	80 (0.01)	41 (0.01)	4 (0.01)	0.6669
Solid tumor without metastasis	1116 (0.07)	142 (0.08)	616 (0.06)	310 (0.08)	48 (0.09)	0.0007
Rheumatoid arthritis/collagen vascular diseases	514 (0.03)	70 (0.04)	262 (0.03)	170 (0.04)	12 (0.02)	<0.0001
Coagulopathy	1226 (0.07)	409 (0.22)	81 (0.01)	715 (0.18)	21 (0.04)	<0.0001
Weight loss	595 (0.04)	127 (0.07)	239 (0.02)	210 (0.05)	19 (0.04)	<0.0001
Fluid and electrolyte disorders	4133 (0.25)	647 (0.34)	2091 (0.21)	1260 (0.31)	135 (0.27)	<0.0001
Blood loss anemia	300 (0.02)	53 (0.03)	113 (0.01)	120 (0.03)	14 (0.03)	<0.0001
Deficiency anemia	445 (0.03)	54 (0.03)	245 (0.02)	133 (0.03)	13 (0.03)	0.0527
Alcohol abuse	577 (0.04)	106 (0.06)	240 (0.02)	214 (0.05)	17 (0.03)	<0.0001
Drug abuse	601 (0.04)	58 (0.03)	349 (0.03)	176 (0.04)	18 (0.04)	0.0491
Psychoses	254 (0.02)	20 (0.01)	156 (0.02)	71 (0.02)	7 (0.01)	0.2477
Depression	1405 (0.09)	87 (0.05)	972 (0.10)	309 (0.08)	37 (0.07)	<0.0001
Trauma	1900 (0.12)	257 (0.14)	1100 (0.11)	470 (0.12)	73 (0.14)	0.0044

Table G.6: DIC comorbidities of validation set patients in context of model performance. P-values for patient counts calculated by chi-squared test.

Appendix H

MISCLASSIFICATIONS

Variable	Total	Condition Positive	condition Negative	P-value
N	18074	7523	10551	
Comorbidity				
Congestive heart failure	4661 (0.26)	2414 (0.32)	2247 (0.21)	<0.0001
Cardiac arrhythmias	5651 (0.31)	2866 (0.38)	2785 (0.26)	<0.0001
Valvular disease	2460 (0.14)	1359 (0.18)	1101 (0.10)	<0.0001
Pulmonary circulation disorders	865 (0.05)	408 (0.05)	457 (0.04)	0.0009
Peripheral vascular disorders	1878 (0.10)	914 (0.12)	964 (0.09)	<0.0001
Hypertension, uncomplicated	6537 (0.36)	3124 (0.42)	3413 (0.32)	<0.0001
Hypertension, complicated	2123 (0.12)	815 (0.11)	1308 (0.12)	0.0025
Paralysis	419 (0.02)	175 (0.02)	244 (0.02)	0.9527
Other neurological disorders	1083 (0.06)	446 (0.06)	637 (0.06)	0.7682
Chronic pulmonary disease	3663 (0.20)	1791 (0.24)	1872 (0.18)	<0.0001
Diabetes, uncomplicated	3330 (0.18)	1569 (0.21)	1761 (0.17)	<0.0001
Diabetes, complicated	1175 (0.07)	478 (0.06)	697 (0.07)	0.5122
Hypothyroidism	1723 (0.10)	791 (0.11)	932 (0.09)	0.0003
Renal failure	2528 (0.14)	997 (0.13)	1531 (0.15)	0.0258
Liver disease	1713 (0.09)	849 (0.11)	864 (0.08)	<0.0001
Peptic ulcer disease excluding bleeding	113 (0.01)	50 (0.01)	63 (0.01)	0.5714
AIDS/HIV	163 (0.01)	81 (0.01)	82 (0.01)	0.0366
Lymphoma	240 (0.01)	108 (0.01)	132 (0.01)	0.2886
Metastatic cancer	155 (0.01)	80 (0.01)	75 (0.01)	0.0116
Solid tumor without metastasis	1264 (0.07)	604 (0.08)	660 (0.06)	<0.0001
Rheumatoid arthritis/collagen vascular diseases	574 (0.03)	272 (0.04)	302 (0.03)	0.0051
Coagulopathy	1684 (0.09)	949 (0.13)	735 (0.07)	<0.0001
Weight loss	741 (0.04)	346 (0.05)	395 (0.04)	0.0051
Fluid and electrolyte disorders	4673 (0.26)	2068 (0.27)	2605 (0.25)	0.0003
Blood loss anemia	344 (0.02)	188 (0.02)	156 (0.01)	<0.0001
Deficiency anemia	476 (0.03)	219 (0.03)	257 (0.02)	0.0523
Alcohol abuse	750 (0.04)	361 (0.05)	389 (0.04)	0.0003
Drug abuse	659 (0.04)	310 (0.04)	349 (0.03)	0.0048
Psychoses	275 (0.02)	132 (0.02)	143 (0.01)	0.0319
Depression	1490 (0.08)	551 (0.07)	939 (0.09)	0.0003
Trauma	2081 (0.12)	986 (0.13)	1095 (0.10)	<0.0001

Table H.1: Comorbidities for false positive misclassificatied patients of ARDS, AKI, sepsis, and DIC. P-values for patient counts calculated by chi-squared test.

Variable	Total	Condition Positive	Condition Negative	P-value
N	18074	673	17401	
Comorbidity				
Congestive heart failure	4661 (0.26)	183 (0.27)	4478 (0.26)	0.4650
Cardiac arrhythmias	5651 (0.31)	252 (0.37)	5399 (0.31)	0.0035
Valvular disease	2460 (0.14)	94 (0.14)	2366 (0.14)	0.7983
Pulmonary circulation disorders	865 (0.05)	35 (0.05)	830 (0.05)	0.6162
Peripheral vascular disorders	1878 (0.10)	93 (0.14)	1785 (0.10)	0.0049
Hypertension, uncomplicated	6537 (0.36)	321 (0.48)	6216 (0.36)	<0.0001
Hypertension, complicated	2123 (0.12)	51 (0.08)	2072 (0.12)	0.0013
Paralysis	419 (0.02)	36 (0.05)	383 (0.02)	<0.0001
Other neurological disorders	1083 (0.06)	51 (0.08)	1032 (0.06)	0.0867
Chronic pulmonary disease	3663 (0.20)	143 (0.21)	3520 (0.20)	0.5643
Diabetes, uncomplicated	3330 (0.18)	131 (0.19)	3199 (0.18)	0.5215
Diabetes, complicated	1175 (0.07)	33 (0.05)	1142 (0.07)	0.0976
Hypothyroidism	1723 (0.10)	74 (0.11)	1649 (0.09)	0.2104
Renal failure	2528 (0.14)	57 (0.08)	2471 (0.14)	<0.0001
Liver disease	1713 (0.09)	72 (0.11)	1641 (0.09)	0.2945
Peptic ulcer disease excluding bleeding	113 (0.01)	6 (0.01)	107 (0.01)	0.3732
AIDS/HIV	163 (0.01)	6 (0.01)	157 (0.01)	0.9771
Lymphoma	240 (0.01)	7 (0.01)	233 (0.01)	0.5091
Metastatic cancer	155 (0.01)	8 (0.01)	147 (0.01)	0.3445
Solid tumor without metastasis	1264 (0.07)	66 (0.10)	1198 (0.07)	0.0049
Rheumatoid arthritis/collagen vascular diseases	574 (0.03)	22 (0.03)	552 (0.03)	0.8901
Coagulopathy	1684 (0.09)	73 (0.11)	1611 (0.09)	0.1852
Weight loss	741 (0.04)	44 (0.07)	697 (0.04)	0.0015
Fluid and electrolyte disorders	4673 (0.26)	219 (0.33)	4454 (0.26)	0.0005
Blood loss anemia	344 (0.02)	20 (0.03)	324 (0.02)	0.0406
Deficiency anemia	476 (0.03)	19 (0.03)	457 (0.03)	0.7574
Alcohol abuse	750 (0.04)	29 (0.04)	721 (0.04)	0.8360
Drug abuse	659 (0.04)	26 (0.04)	633 (0.04)	0.7636
Psychoses	275 (0.02)	10 (0.01)	265 (0.02)	0.9391
Depression	1490 (0.08)	60 (0.09)	1430 (0.08)	0.5364
Trauma	2081 (0.12)	113 (0.17)	1968 (0.11)	<0.0001

Table H.2: Comorbidities for false negative misclassificatied patients of ARDS, AKI, sepsis, and DIC. P-values for patient counts calculated by chi-squared test.

VITA

Current Positions:

- Doctoral Student (PhD Candidate), Department of Biomedical Informatics Medical Education, University of Washington
- Program Operations Analyst, Department of Emergency Medicine, School of Medicine, University of Washington
- Director of Analytics, T Exponents
- Board member of a 501(c)(3) international activism organization.
- Board member of a 501(c)(3) child care.

Degrees:

- Masters of Science, University of Chicago, Computer Science, 2004
- Bachelors of Science, University of Wisconsin, Astrophysics, Physics, 2001

Professional Experience:

- 2014-Present: Program Operations Analyst, Emergency Medicine, University of Washington
- 2013-Present: Principal Investigator, Health Informatics Group, Open Science Grid
- 2012-Present: PhysioNet MIMIC Reviewer

- 2004-2009: Bioinformatics Software Developer, Ben-May Department for Cancer Research, University of Chicago
- 2001-2002: Department of Energy, Lawrence Livermore National Laboratory
- 1999-2001: Atomic Collisions Group, University of Wisconsin

Fellowships:

- 2009-2012: National Library of Medicine, Informatics Research Fellow, University of Washington, School of Medicine
- 2001-2002: Graduate Research Fellowship, Department of Energy, Lawrence Livermore National Laboratory, University of California

Awards:

- 2014 1st Place Prize at MITs Critical Data Marathon competition
- 2004 Letter of Commendation, Department of Computer Science, University of Chicago
- 2000 Research Symposium, University of Wisconsin
- 1998 Phi Eta Sigma National Honors Society
- 1998 Golden Key National Honors Society

Book Chapters:

- Jablonowski, Karl. Hidden Markov Models for Protein Domain Homology Identification and Analysis. SH2 Domains. Ed. Kazuya Machida, Ed. Bernard Liu. Humana Press, 2017. Print.

- Jablonowski, Karl. Proteomic Clustering Analysis of SH2 domain datasets. SH2 Domains. Ed. Kazuya Machida, Ed. Bernard Liu. Humana Press, 2017. Print.

Journal Articles:

- Sachita P Shah, MD; Kevin Penn, MD; Stephen J Kaplan, MD, MPH; Michael Vrablik, DO; Karl Jablonowski, MS; Tam N Pham,MD; May J. Reed, MD. Comparison of Bedside Screening Methods for Frailty Assessment in Older Adult Trauma Patients in the Emergency Department. The American Journal of Emergency Medicine (in press).

- Mitchell Kim, Taketo Watase, Karl D Jablonowski, Medley O Gatewood, Daniel J Henning. A Sepsis-related Diagnosis Impacts Interventions and Predicts Outcomes for Emergency Patients with Severe Sepsis. Western Journal of Emergency Medicine, 2017 Oct; 18(6): 10981107.

- Bernard A Liu, Brett W Engelmann, Karl Jablonowski, Katherine Higginbotham, Andrew S Stergachis and Piers D Nash. Src Homology 2 Domain Binding Sites in Insulin, IGF-1 and FGF Receptor Mediated Signaling Networks Reveal an Extensive Potential Interactome. Cell Communication and Signaling2012, 10:27 (14 September 2012)

- Liu BA, Shah E, Jablonowski K, Stergachis A, Engelmann B, Nash PD. The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signalingin Eukaryotes.Science SignalingVol. 4,Issue 202, p. ra83.December 6, 2011

- Liu BA, Jablonowski K, Shah E, Engelmann B, Jones RB, Nash PD. SH2 domains recognize contextual peptide sequence information to determine selectivity. Molecular and Cellular Proteomics mcp.M110.001586First Published on July 13, 2010, doi:10.1074/mcp.M110.001586

- Machida K, Thompson CM, Dierck K, Jablonowski K, Krkkinen S, Liu B, Zhang H, Nash PD, Newman DK, Nollau P, Pawson T, Renkema GH, Saksela K, Shin DG, Mayer BJ. High-Throughput Phosphotyrosine Profiling Using SH2 Domains. Molecular Cell26: 899-915. June 22, 2007

- Liu BA, Jablonowski K, Raina M, Arce M, Pawson T, Nash PD. The Human and Mouse Complement of SH2 Domain Proteins establishing the boundaries of phosphotyrosine signaling. Molecular Cell22: 851-868. June 23, 2006

Acknowledgements:

- Johnson et al. Insights from atomic-resolution X-ray structures of chemically synthesized HIV-1 protease in complex with inhibitors. Journal of Molecular Biology(2007)

- Boffard et al. Measurement of electron-impact excitation cross sections out of the neon $^3P_2$ metastable level. Physical Review A (2001)

Presented Posters:

- Marie Vrablik, MD MCR; Liza Rosenman, MD; Karl Jablonowski, MS; Annie Chipman, MD MS. Violence in the Emergency Department: An Observational Study. Society for Emergency Medicine, 2018.

- Ben Friendman, MD; Dan Henning, MD MPH; Arvin Akhavan, MD; Karl Jablonowski, MS; Kennedy Hall, MD MHS; Nick Johnson MD. Trends in Emergency Department Lactate Utilization and Prognostic Accuracy. Society for Emergency Medicine, 2018.

- Arvin Akhavan, MD; NickJohnson,MD;Karl Jablonowski, MS; Kennedy Hall, MD MHS; Dan Henning, MD MPH; Ben Friedman MD. Assessing the Prognostic Value of Lactate in the Presence of Ethanol. Society for Emergency Medicine, 2018.

- David Cheever, MD MS; Lauren Whiteside, MD MS FACEP; Karl Jablonowski, MS; Marie Vrablik, MD MCR. Improving Transitions Between the Emergency Department and Primary Care: A Health Information Exchange Pilot Project. Society for Emergency Medicine, 2018.

- Marie Vrablik, MD MCR; Karl Jablonowski, MS; Steve Mitchell, MD; Tak Watase, MD MBA; Ken Linnau, MD. Time and Workflow Barriers to Obtaining a Spine MRI from the Emergency Department. UW Patient Safety Improvement Program, 2016.

- Karl Jablonowski, M.S., Joseph Paonessa, M.D., Aaron Mittel, M.D., Tom Schaus, M.D., Ph.D, Brian Malley, B.S., Wendy Chen, Pharm.D. "Mean Arterial Pressure in the ICU, Vasopressors,and Mortality." Critical Data Conference: Secondary Use of Big Data from Critical Care. Cambridge, MA January 7, 2014.

- Liu BA, Jablonowski K, Engelmann B, Nash PD. "Determinants of Physiological LigandSpecificity of SH2 Domains using Peptide Arrays." Keystone Symposia: B2 -Omics Meets Cell Biology. Breckenridge, Colorado., January 25-30, 2009.

- Liu BA, Jablonowski K, Engelmann B, Nash PD. "Phosphotyrosine Interaction Network for the Insulin/IGF-1 Signaling." 3rdAnnual Diabetes Day Symposium. Chicago, IL., May 17, 2008.

- Liu BA, Jablonowski K, Shah E, Stergachis A, Nash PD. "Defining SH2 Specificity and Diversity using PeptideArrays." Keystone Symposia: Systems Biology and Regulatory Networks. Steamboat Springs, CO., March 22-27, 2007.

- Liu BA, Jablonowski K, Shah E, Stergachis A, Pawson T, Nash PD. "Directing phosphotyrosine signals: an integrated bioinformatics and proteomics examination of SH2 domains." 2006 Growth Factor Signaling: Gordon Research Conference. New London, CT., July 16-21, 2006.

- Liu BA, Jablonowski K, Nalluri R, Djordjevic D, Shah E, Stergachis A, Pawson T, Nash PD. "Directing phosphotyrosine signals: an integrated bioinformatics and proteomics examination of SH2 domains." American Association for Cancer Research Annual Meeting. Washington D.C., April 1-5, 2006.

- Jablonowski K. Thorne-Zytkow Objects University of Wisconsin Research Symposium. April 11, 2000.

Funding Support Agencies:

- Air Force Office of Scientific Research

- Department of Energy

- King County Medic One

- National Aeronautics and Space Administration

- National Institute of Health

- National Science Foundation

- University of Chicago Cancer Research Foundation Womens Board