Machine Learning Approach to Predict Life-threatening Outcomes with Admit Electrocardiograms of Hospitalized Patients with COVID-19

Zih-Hua Chen

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Bioengineering

University of Washington

2021

Committee:

Patrick M. Boyle

Arun R. Sridhar

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Machine Learning Approach to Predict Life-threatening Outcomes with Admit Electrocardiograms of Hospitalized Patients with COVID-19

Zih-Hua Chen

Chair of the Supervisory Committee:
Patrick M. Boyle
Department of Bioengineering

COVID-19 has been straining the health care systems worldwide due to its high fatality rate and high infection rate. Recent studies found that COVID-19 can cause life-threatening cardiovascular complications in patients. Therefore, there is a need for early risk-stratified tools. In this thesis, I proposed two machine learning algorithms for predicting the probabilities of mortality or developing cardiovascular complications for hospitalized patients with confirmed COVID-19. Machine learning is a technique that combines statistics and algorithms which enable the machine to extract underlying features and draw inferences from a huge amount of data. The models in this work only use a standard 10-second 12-lead intake electrocardiogram (ECG) as input data. The models were trained and evaluated with a database containing 1270 intake ECGs which were split into a training set, a validation set, and a testing set in a ratio of 8:1:1. The prediction results yield an average sensitivity of 0.65 and an average specificity of 0.52 for all-cause mortality. For predicting life-threatening cardiovascular outcomes, it reaches an average sensitivity of 0.63 and an average specificity of 0.44.

Acknowledgment

I would like to express my special thankfulness of gratitude to my advisor, Dr. Patrick Boyle, my primary collaborator Dr. Arun Sridhar, as well as the Department of Bioengineering at the University of Washington who gave me this precious opportunity to conduct this project. Dr. Patrick Boyle and Dr. Arun Sridhar have provided me with lots of supportiveness and professional insights when I was doing this research.

In addition, this research could not be done without the University of Washington Population Health Initiative's COVID-19 Rapid Response Grant.

I want to express my special thanks to Dr. Gregory John, Dr. Alison Fohner, and Dr. Jake Mayfield for providing constructive feedback and suggestions when I was stuck somewhere. In addition, I want to thank Dr. Gregory John for helping me scale up the machine learning framework. Thank Dr. Alison Fohner for helping me building the conventional model for internal evaluation. Thank Dr. Jake Mayfield for helping me adjudicated all of the ECGs.

I want to thank Sarah, Chris, and all of the collaborators for their significant work on building the first few COVID-19 patient databases in the world. The project could not be done without any part mentioned above.

As a trainee in the Cardiac System Simulation (CardSS) Lab, Dr. Patrick Boyle is a great mentor who teaches me not only scientific research but also some important values that I have never thought of until I came to the United States. I also want to thank everyone in the CardSS Lab, especially Chelsea, Alex, Savannah, and Dr. Patrick Boyle. Thank you all for supporting me along the way and giving me lots of constructive feedback.

Last, I want to thank my family in Taiwan for supporting me to study abroad. Words are not enough to describe my thankfulness to you.

I have never expected I would pursue my master's degree under a pandemic. This two-year experience is challenging and memorable. It is a bit chaotic due to COVID-19, but thanks for all of your kind help and support, I feel it's an extremely meaningful stage in my life.

Table of Contents

Chapter 1. Introduction

Section 1.1 Background and motivation

Currently, a pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is affecting the world seriously. World Health Organization (WHO) named this disease COVID-19 and declared a global health emergency. [1] By April 2021, there have been more than 145,000,000 confirmed cases of COVID-19, including more than 3,000,000 death worldwide, reported to WHO. [2] In a systematic review, Ana Macedo et al. conducted a meta-analysis of the COVID-19 mortality rate among 33 studies with a total of 13,398 hospitalized patients.[3] They reported that the global mortality rate was 17.1%. For generally admitted patients, the COVID-19 mortality rate was 11.5%. As regards critically ill patients, their study found a 40.5% COVID-19 mortality.[3] Due to this high fatality rate and high infection rate, COVID-19 has been straining the health care systems worldwide.

COVID-19 is known for severe respiratory syndromes, such as shortness of breath, difficulty of breathing, cough, or sore throat.[1] In addition, many recent studies have pointed out that COVID-19 can affect the heart and lead to life-threatening cardiovascular comorbidities.[4-7] This study aimed at predicting three common cardiovascular comorbidities which were arrhythmic events, thromboembolism, and heart failure.

Heart arrhythmias is a condition in which the heart beats with irregular rhythms or abnormal rates. The abnormal heart beasts are due to cardiac tissue damage caused by diseases, injury, or genetics. [8] In a retrospective study from Wuhan, China, among a case series of 138 patients, 16.7% of patients were reported with arrhythmias complications after COIVD-19 infection. [9] Another study characterized the first 393 consecutive patients who were admitted to two hospitals in New York City. 29 (7.4%) patients were observed with arrhythmia during

their hospital stay. [10] There is a wide range of potential mechanisms might develop arrhythmia after COVID-19 infection, including hypoxia, myocardial strain, abnormal host immune response, and drug side effect. [6] Since arrhythmias are often asymptomatic or not causing obvious signs, there is a need to develop an early warning system for hospitals to predict patient potential outcomes and give adequate medical treatments.

In addition to arrhythmia, COVID-19 infection may lead to thromboembolism due to hypoxia, excessive inflammation, platelet activation, and diffuse intravascular coagulation. The incidence of thromboembolic complications was first reported in a study of 183 critically ill ICU patients admitted to three Dutch hospitals. [11] The cumulative incidence was 31%, including 27% of venous thromboembolism (VTE) and 3.7% of atrial thromboembolic events. Due to the hypercoagulable state of COVID-19 patients, drug interactions are an important issue of COVID-19 treatments.[12] Given that, early risk stratification is needed for COVID-19 patients.

Heart failure often develops after other severe diseases which damage the heart. A study enrolled 3080 consecutive patients with confirmed COVID-19 in Spain and followed them up for at least 30 days. The authors concluded that patients with COVID-19 have a higher incidence of acute heart failure and are associated with a high fatality rate. [13]

In clinical practices, biomarkers, physiologic variables, and comorbidity have been used as risk indicators. Acute Physiology and Chronic Health Evaluation II (APACHE II)[14], Sequential Organ Failure Assessment scores (SOFA)[15], and the Padua Prediction Score[16] are commonly used in predicting hospitalized patient outcomes. In a retrospective study of 178 patients in Wuhan, China, the sensitivity of using APACHE II to predict hospital mortality was 96.15% and the specificity was 86.27%. [17] Despite the high effectiveness, the APACHE II score is derived from 12 physiologic variables which require corresponding equipment and

specialist. The efficiency could be challenging when the number of patients was extremely high during a pandemic. Therefore, there is a need to develop an accessible assessment tool for a hospital to perform an early risk assessment of mortality and/or cardiovascular complications on patients with confirmed COVID-19.

Section 1.2 The advantage of electrocardiogram

Standard 12-lead electrocardiograms (ECG or EKG) records the surface potential activity of the heart through non-invasive electrodes. By interpreting ECG, cardiologists can diagnosis underlying heart conditions. ECG is considered the primary diagnosis tool of arrhythmias. The acquisition of ECG is a common non-invasive procedure in all hospitals which is informative and efficient. Accordingly, ECG is an accessible tool for cardiovascular studies in clinical practice.

Section 1.3 Previous research

Machine learning (ML) has demonstrated promising abilities in investigating cardiac electrophysiology (EP), such as heartbeat classification[18] and heart disease detection.[19, 20] Early studies heavily rely on peak detection and feature extraction, including RR-interval, QRS complex, and P wave. [21] In recent years, researchers have used deep learning (DL) algorithms to replaced predefined feature extraction and thus accelerated the application of machine learning in EP. In 2019, Zachi Attia et al. proposed a convolution neural network (CNN) architecture for the identification of patients with atrial fibrillation (AFib) during sinus rhythm.[22] The model was built with ten residual blocks. They fed a standard 12-lead 10-second ECG as input without requiring any manual feature extraction. The model was trained with 454789 ECGs from 126526

patients and evaluated with a hold-out test set which contained 130802 ECGs from 36280 patients, yielding an AUC of 0.85, the sensitivity of 79%, specificity of 79.5%, and F1-score of 39.2%. In the same year, Arjun Gupta et al. proposed a fine-tuned version of the ConvNetQuake [23] neural network to detect myocardial infection in ECGs. [19] The model is an eight-layer one-dimensional CNN architecture. It takes a 10-second long raw digital ECG from lead v6 and leads vz as input, yielding a 99.43% accuracy.

CNN is a common architecture of the DL neural network. Unlike conventional regression models, DL models are non-linear models which are more capable of fitting high dimensional data. In other words, DL models have more potential to solve complex problems or discover unrevealed patterns. By taking the advantage of modern fast computing resources, deep learning has been applied in many aspects in the real world, such as natural language processing, object detection, facial recognition, etc. The clinical application of ML in cardiology has also been discussed over the past decade.[24] Although there are some intrinsic challenges, such as the need for a huge amount of labeled training data, deep learning models could still be an early risk assessment tool for disease screening.

As to the prediction of COVID-19 prognosis, this is the first study that proposed a single ECG-based machine learning approach to my knowledge. Previous machine learning approaches used demographics, biomarkers information, physiological values, and/or health records as model inputs. [25-28]

Section 1.4 Specific goals

The objective of this study is to demonstrate the potential of the ECG-based machine learning approach in predicting life-threatening outcomes for hospitalized patients with COVID-

19. The research outcomes include three parts. First, developed a binary ML model for predicting all-cause in-hospital mortality of hospitalized COVID-19 patients. Second, build a multi-label ML model for predicting life-threatening cardiovascular outcomes of hospitalized COVID-19 patients. The only input data of both ML models is the intake ECG of each patient. Last, characterized the study cohort with statistical analysis.

Chapter 2. Methodology

Machine learning (ML) is the scientific study of statistics and algorithms where computer systems perform given tasks without specific commands or instructions but rely on learning features or drawing inferences from a big amount of data. A common application for machine learning techniques is targeted advertising. Digital marketers train machine learning models with browsing history, purchasing history, seasonal trends, etc. Given that information, the models can predict the purchasing likelihood and then deliver the advertisement to potential customers.

Herein, I hypothesized that there were underlying patterns in the intake ECG waveforms of COVID-19 patients who were at a high risk of mortality or cardiovascular complications due to damaged cardiomyocytes. I trained an ML model to learn the unidentified features and their difference between the admit ECGs from patients who survived and the admit ECGs from patients who died. If the model could discover these underlying patterns in the intake ECGs, the model could serve as a predictor to evaluate the risk of mortality when a new patient was admitted with COVID-19.

Under the same hypothesis, I further trained another ML model to identify the unrevealed ECG patterns for patients who develop arrhythmia, thromboembolism, or heart failure after COVID-19 infection. When a new patient was admitted to the hospital, this model could predict the probabilities of developing the above three cardiovascular complications individually.

In this chapter, I introduced the detailed methods and experimental design in my research.

Section 2.1 Deep learning and supervised learning

Deep learning (DL) is a sub-area of ML where the model is constructed with at least one hidden layer, in addition to an input layer and an output layer. The basic unit in each hidden layer is a "neuron". A deep learning network is also known as an artificial neural network (ANN). Each neuron would take multiple input values, conduct a mathematical operation, and then generate an output value for the next layer. For example, **figure 1** demonstrates a simple three-input neuron. Input x_1 x_2, and x_3 would be multiplied by a corresponding weight, w_1, w_2, and w_3 , and then add up together. A bias might be added to this summation before mapping with an activation function.

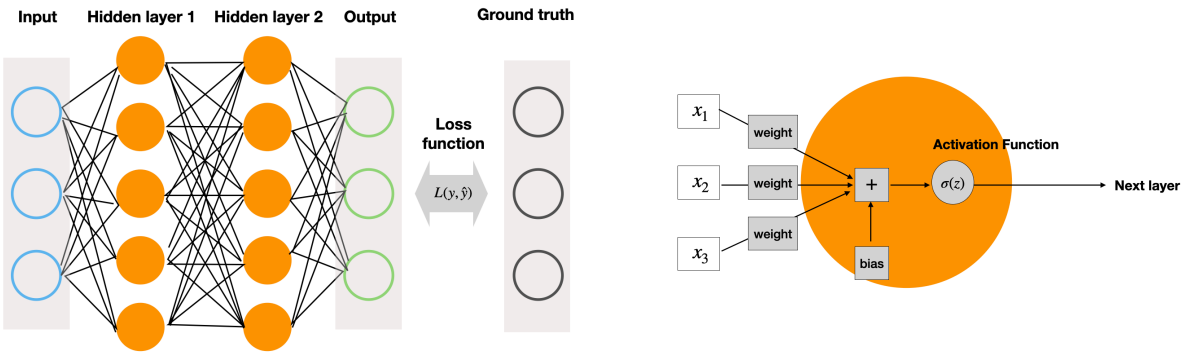


Figure 1. An example scheme of (left) an ANN and (right) a neuron. (left) All of the data must be transfer into multi-dimensional vectors before feeding into an ANN. In this example, ANN, both of the input and output are 1-d arrays with three elements. Hidden Layers 1 and 2 both have five neurons which are the key computing units of the ANN. The loss function is used to evaluate the error between ground truth and output. (right) In a three-input neuron, each input is multiplied by a weight and then add up with a bias. The sum is mapped with an activation function and then pass to the next layer.

Supervised learning refers to a type of learning scenario where the model was given both training data and corresponding labels. The learning progress usually involves hundreds of iteration. Each learning iteration is called an epoch. A loss function is used to represent the goodness of an ANN during the training progress. For example, calculate the mean square error (MSE) between the output vector and the ground truth vector. Meanwhile, the training goal is to

gradually lower the loss value per epoch. To reach this goal, an algorithm called backpropagation [29], which is short for backward propagation of errors, and an optimizing algorithm called stochastic gradient descent[30] are commonly used together in the supervised learning of ANN. During each epoch, the gradient of the loss function regarding current weights and biases of each neuron was first calculated. Next, update the existing weights and biases based on the gradient. These two steps can be repeated until the model reaches the optimal loss or the desired number of epochs.

Section 2.2 Model development and architecture

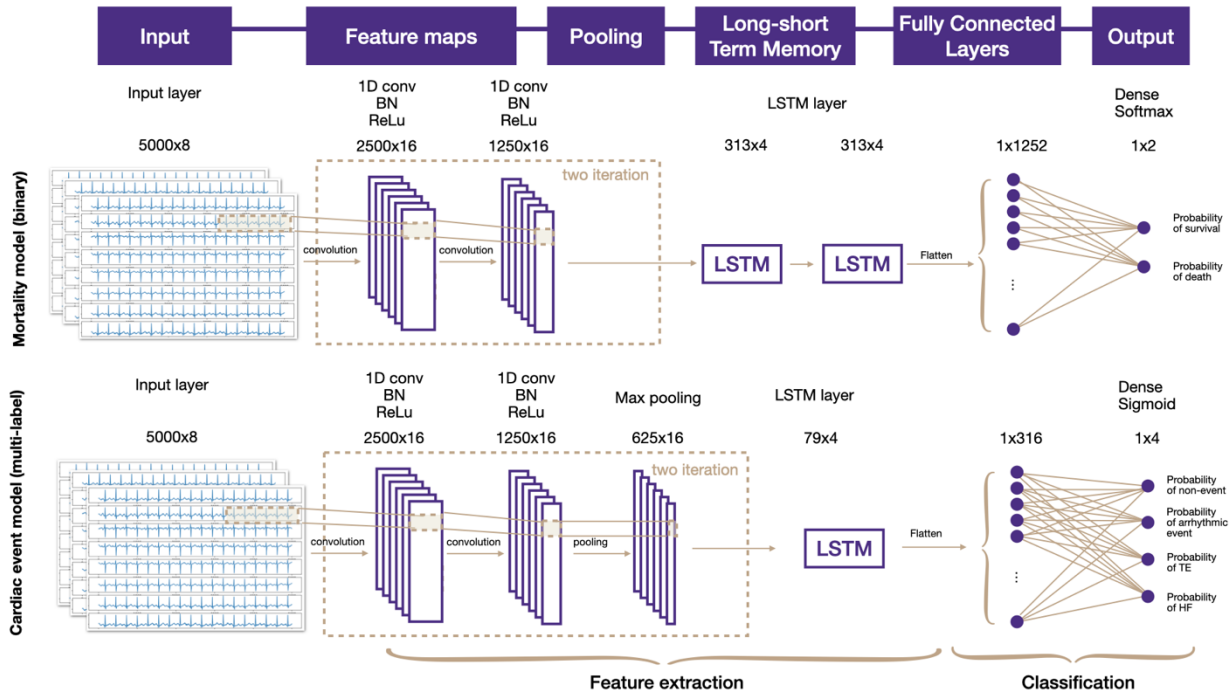
In this study, both of the ML models are trained under a supervised learning scenario. The models were developed using the Keras deep learning API [31] with the Tensorflow backend [32] and the Scikit-learn API [33, 34]. The deep learning models were written with Python version 3.7 [35].

I formulated two distinct Convolutional Neural Networks with Long Short-Term Memory (CNN-LSTM); these were designed to predict the incidence during COVID hospitalization of, respectively, mortality (Mortality model) and life-threatening cardiovascular events as described in the next subsection (MACE model). **Figure 2** shows schematic representations for both model architectures, both of which had three main sections. In section I, a series of 1D convolution layers was used to extract temporal features from each channel separately; each convolution layer was followed by a batch normalization layer to centralize data distribution and a rectified linear unit (ReLU) to weight the output of past layers. The number of feature maps, convolution kernel size, and feature size were selected empirically. Section II was a recurrent neural network (RNN), consisting of two or one LSTM layer(s) [36] to process spatial features across all eight channels;

each LSTM layer contained four units with feedback connections. Section III was a fully connected layer and activation function used to output a confidence value describing the model's confidence that a particular ECG belonged to each class. The mortality model was a binary model (i.e., predicted likelihood of survival vs. death) where the last layer contained a softmax function; in contrast, the MACE model was a multi-label model (independent likelihoods for all four event types: arrhythmic, heart failure, thromboembolic, or none) where the last layer was a sigmoid function.

Figure 2. Model Architectures for the mortality model and MACE model. Both models take digital ECG signals as input and then pass them through convolution layers, long-short term memory layers, and fully connected layers sequentially. The numbers refer to the output shape of each layer.

Section 2.3 Dataset and outcome labels



The dataset is a collaborative project between the University of Washington Healthcare System (**UW**; Seattle, WA), the Karolinska Institute University Hospital (**KI**; Stockholm, Sweden), the Uppsala University Hospital (**UU**; Uppsala, Sweden), and the Copenhagen

University Hospital (**UC**; Copenhagen, Denmark). The construction of the dataset was approved by the Institutional Review Board (IRB) of each center.

I investigated 1270 hospitalized COVID-19 patients (60.4% male, 63.7±17.2 years old) admitted to the four centers listed above between 3/2/20 and 6/8/20. Standard 10-second 12-lead ECGs and outcomes were collected for the study cohort, along with patient demographics, comorbidities, hospitalization variables, and laboratory values. Missing information was allowed. ECGs were acquired at a sampling rate of 500 Hz or 250 Hz using GE ECGs systems (GE, USA) and raw data were managed using the GE MUSE Cardiologist Information system (GE Healthcare, USA). All intake ECGs were adjudicated to exclude ECGs containing pacing artifacts caused by a cardiovascular implantable electronic device (CIED). 1207 of 1270 (95.0%) intake ECGs were acquired within 24h of the hospital admission date. For patients who had multiple ECGs in the MUSE system from the date of intake, only the latest one was exported to the dataset. All data were collected and managed using Research Electronic Data Capture (REDCap) electronic data capture tools hosted at the Institute of Translational Health Sciences, University of Washington.[37, 38]

Figure 3 presents a data flow diagram for this study, including the elimination of records due to exclusion criteria. To eliminate data leakage and enforce the assumption that all data points were independent and identically distributed random variables, I included only one intake ECG per patient.

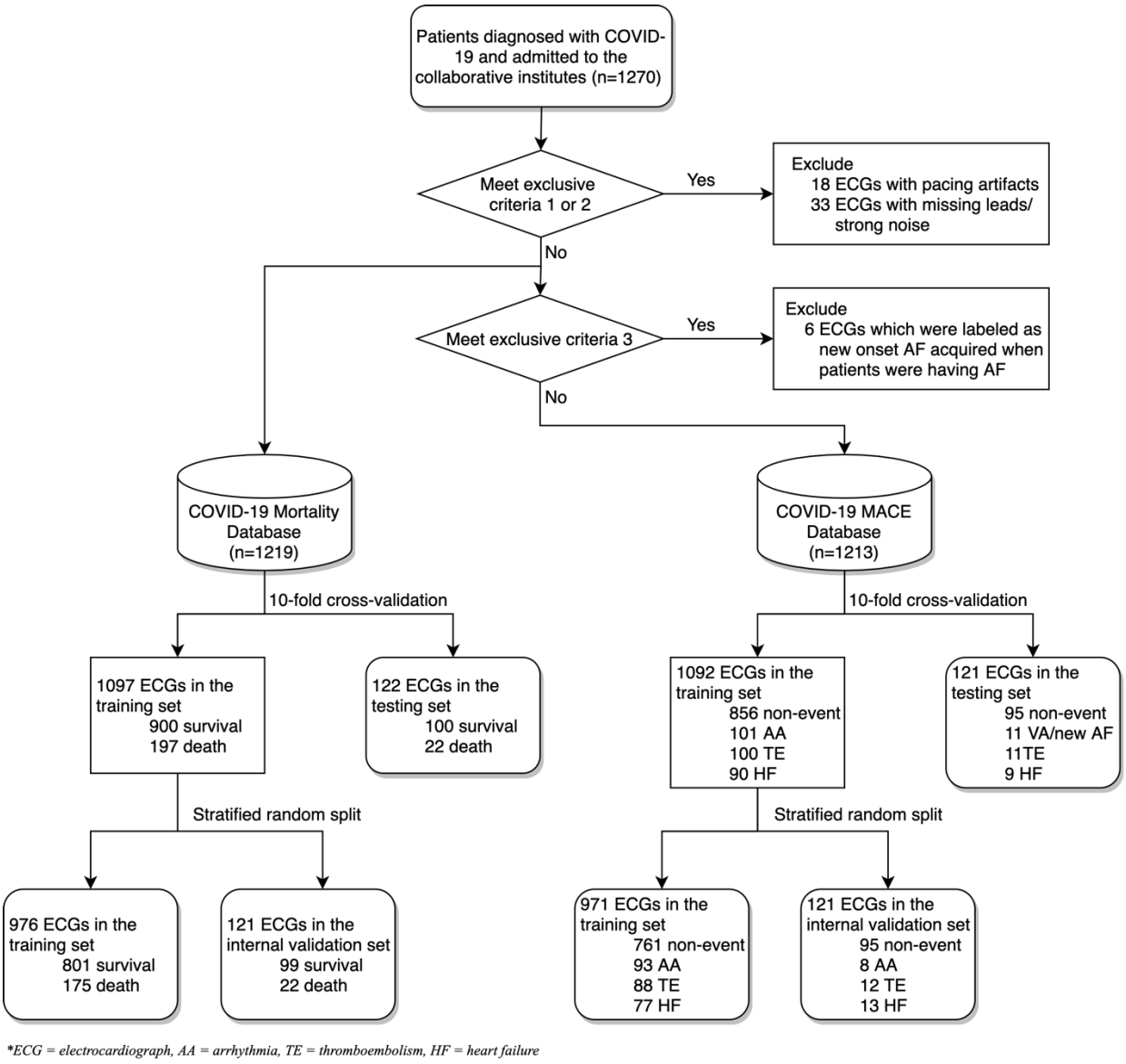


Figure 3. Patient Diagram. The admit ECGs of 1270 patients were used in this study. Exclusive criteria were applied to create the COVID-19 mortality database and the COVID-19 MACE database. The two databases were used to train the mortality model and the MACE model respectively.

All ECG records were labeled in terms of (1) patient all-cause mortality vs. survival (i.e., discharge to home or other care) and (2) incidence of MACE during COVID hospitalization (background vs. arrhythmic events [AA] vs. thromboembolic complication [TE] vs. heart failure [HF] events). Patients who remained hospitalized for COVID-19 illness were not included in this

cohort. The arrhythmic endpoint was defined as a composite of new-onset atrial fibrillation (AFib), high burden PVCs, sustained ventricular tachycardia (VT), ventricular fibrillation (VF), tachycardia arrest, and bradycardia arrest. Patients with recurrent AFib or persistent AFib were noted, but these were not classified as events under the arrhythmic endpoint since the study aim was to predict complications that arose as a result of COVID-19. HF events included new-onset HF and cardiogenic shock. The background class refers to non-event cases where patients did not develop any above of the above MACE during their COVID-19 hospitalization.

Section 2.4 Preprocessing

ECGs and associated labeling information were split into training and testing sets using a 10-fold stratified cross-validation scheme [39]. This allowed us to estimate model performance and generality with lower bias compared to a single hold-out testing set paradigm. Stratified data partitioning ensured that the training and testing sets always contained percentages of every target class consistent with proportions in the entire dataset. To avoid over-fitting and under-fitting, an internal validation set was randomly split from the training set and used to evaluate the model over successive iterations. The ratio of records in the training, validation, and testing sets was 8:1:1.

The deep learning model input data was a stack of standard 10-second ECG records sampled at either 500 (n=1112) or 250 Hz (n=107); the latter records were up-sampled to 500 Hz via interpolation. Standard 12-lead ECG records were recorded, but four leads (III, aVR, aVL, aVF) were omitted from the analysis since they can be derived from linear combinations of other included leads. Thus, for each clinical record used the model input was an 8×5000 matrix.

Data preprocessing is crucial in machine learning as it can help ensure useful features in raw signals are appropriately prioritized during network training. Various techniques have been

proposed for ECG-based algorithms [18], but most rely on expert-defined ECG features (e.g., RR interval, QRS complex duration, etc.) To facilitate the comparison of ECGs with amplitude differences due to a variety of factors potentially unrelated to the underlying cardiac state (e.g., body shape), all signals were normalized to the range [0,1].

Section 2.5 Training

During the learning process, I used ADAM optimization algorithm [40] to schedule gradient descent and categorical cross-entropy as the loss function for iterative updating weights and biases. Given the fact that the database was small by deep learning standards and imbalanced, containing a much lower number of patients compared to without events, I added an early stopping callback, which halted the training process after the loss of the validation set had not been improved for ten epochs, to prevent overfitting.

Training and testing of CNN-LSTM models were carried out using advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system of the University of Washington. All jobs were run on one standard compute node (32 cores, Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz, 128 GB RAM).

Section 2.6 Statistical Assessment of Model Performance

Model performance was evaluated by calculating the area under the receiver operator characteristic curves (AUROC) as well as sensitivity, specificity, and f1 score values, and confusion matrices for hold-out testing sets in a 10-fold cross-validation scheme, as described above. Optimal probability thresholds were determined based on the ROC curves of internal validation sets for each output class:

$$D = {fpr_{val}}^2 + 2 \times (1 - tpr_{val})^2$$

where D is a weighted distance from the origin of the coordinate to a specific point in the ROC curve of the validation set, fpr is a false positive rate, and tpr is a true positive rate. This formulation deliberately weighted sensitivity (i.e. $1 - tpr$) over specificity (i.e. fpr), since the importance of minimizing the number of false negatives was deemed a much higher priority than reducing the false positive rate. Each fpr, tpr pair maps to a unique threshold determined by minimizing D.

The ROC curves for each class in each model independently consider the comparison of model output to ground-truth labels. Thus, assessing the overall performance of each model necessitated the use of ROC aggregation techniques. For the MACE model, which was a multi-label model, we calculated a *macro-average* ROC curve (i.e., each class's ROC curve contributed equally to the average) and a *micro-average* ROC curve (i.e., values of all classes were aggregated, weighting by the number of entries in each class).

The false positive (type I error) and the false negative (type II error) possess different meaning in this scenario. I investigated the false positive rate (FPR) and the false negative rate (FNR) with a moving threshold. FPR and FNR were derived from the formula below:

$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP}$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN}$$

In the above formula, FN means false negative, FP means false positive, TN means true negative, and TP means true positive. FNR is the miss rate of the event cases (i.e., patients who died or developed AA/TE/HF complications but the model did not predict correctly on their

outcomes). FPR is the false alarm ratio of the event cases (i.e. patients who survived or did not develop life-threatening outcomes, but the model predicted that they were at high risk).

Principle component analysis (PCA) was implemented with the sklearn.decomposition module from Scikit-learn 0.24.2 version. I extracted the input of the last classification layer for PCA to visualize the variance in the data. Each sample was transformed into a set of values of its first two principal components.

The Precision-Recall curve for each class in each model measures the model's ability to predict a specific positive class. The equations for calculating precision and recall are listed below:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Finally, when aggregating results of different model performances from all different train/test iterations carried out as part of the 10-fold cross-validation process, two-sided 95% confidence intervals were used to estimate the overall performance matrix for the system.

Chapter 3. Results

Section 3.1 Study Cohort

In this study cohort of 1270 COVID-19 patients, the mean age was 63.5 years (SD 17.0), 60.4% (n = 736) were men. **Table 1** shows the statistical characteristic of overall patients and patients from four centers. The overall all-cause mortality rate was 17.9% (n=219). For major adverse cardiovascular events (MACE), the percentage were 9.7% (n = 118), 9.3% (n = 113), and 8.3% (n = 101), respectively for developing arrhythmic event (AA), thromboembolism (TE), and heart failure (HF).

It is noteworthy that the mean BMI values of overall patients and patients from individual centers exceeded 27 with a standard deviation (SD) greater than 5. According to the Centers for Disease Control and Prevention (CDC) guideline, overweight refers to a person's BMI value is greater than 25, and lower than 30. People whose BMI > 30 are considered as obesity. More the half of the patients in this database were overweight.

One-way ANOVA was run on the continuous variables of the four centers to test whether there was a statistically significant difference between the mean of the four groups of patients from different centers. The Chi-square test was applied for categorical variables. P-values and the percentage of missing values were shown in the last two columns of the table.

Age, BMI, and race distribution were at variance among the four centers ($P<0.01$). Beyond that, I observed a significantly low mortality rate at the University of Copenhagen ($P<0.0001$).

	Overall	UW	KI	UU	UC	*P*-value	Missing(%)
n	1218	275	459	307	177		
Demographics							
Age (years) (mean (SD))	63.50 (16.95)	63.00 (16.83)	60.04 (16.80)	65.89 (17.99)	69.15 (13.24)	<0.001	0.0
BMI (kg/m^2) (mean (SD))	28.34 (6.63)	29.89 (8.53)	28.65 (5.90)	27.85 (6.90)	27.06 (5.79)	0.001	15.8
Sex at birth = male (%)	736 (60.4)	166 (60.4)	303 (66.0)	173 (56.4)	94 (53.1)	0.007	0.0
Ethnicity (%)						-	14.9
Hisp/Latinx	57 (5.5)	56 (20.7)	0 (0.0)	1 (0.3)	-		
Non-Hisp/Latinx	496 (47.8)	214 (79.0)	13 (2.8)	269 (87.6)	-		
Unkn/Unavail	484 (46.7)	1 (0.4)	446 (97.2)	37 (12.1)	-		
Race(%)							
FN/AK Native	5 (0.4)	5 (1.8)	0 (0.0)	0 (0.0)	-	0.001	0.0
Asian	86 (7.1)	39 (14.2)	7 (1.5)	40 (13.0)	-	<0.001	0.0
Black/Afr Amer	49 (4.0)	33 (12.0)	10 (2.2)	6 (2.0)	-	<0.001	0.0
HI FN/Pac Isl	4 (0.3)	4 (1.5)	0 (0.0)	0 (0.0)	-	0.003	0.0
White	608 (49.9)	186 (67.6)	221 (48.1)	201 (65.5)	-	<0.001	0.0
Other	57 (4.7)	3 (1.1)	50 (10.9)	4 (1.3)	-	<0.001	0.0
Unkn/Unavail	234 (19.2)	5 (1.8)	173 (37.7)	56 (18.2)	-	<0.001	0.0
Comorbidity							
Hypertension (%)	657 (54.8)	143 (55.4)	227 (49.7)	184 (59.9)	103 (58.2)	0.030	1.6
CAD (%)	163 (13.7)	40 (15.7)	60 (13.3)	48 (15.6)	15 (8.5)	0.110	2.3
CIED (%)						-	15.1
None	1009 (97.6)	256 (95.5)	451 (98.3)	302 (98.4)	-		
Pacemaker	20 (1.9)	8 (3.0)	7 (1.5)	5 (1.6)	-		
ICD	5 (0.5)	4 (1.5)	1 (0.2)	0 (0.0)	-		
Outcomes							
Arrhythmic event (%)	118 (9.7)	22 (8.0)	47 (10.2)	36 (11.7)	13 (7.3)	0.305	0.0
TE (%)	113 (9.3)	19 (6.9)	50 (10.9)	29 (9.4)	15 (8.5)	0.334	0.0
HF (%)	101 (8.3)	16 (5.8)	53 (11.5)	23 (7.5)	9 (5.1)	0.010	0.0
mortality (%)	219 (18.0)	66 (24.0)	69 (15.0)	69 (22.5)	15 (8.5)	<0.001	0.0

Table 1. Characteristics of overall patients and patients admitted to individual centers.

	survival	death	*P*-value	Missing (%)
n	999	219		
Demographics				
Age (years) (mean (SD))	60.91 (16.42)	75.34 (14.09)	<0.001	0.0
BMI (kg/m^2) (mean (SD))	28.54 (6.55)	27.39 (6.94)	0.037	15.8
Sex at birth = male (%)	591 (59.2)	145 (66.2)	0.063	0.0
Ethnicity (%)			0.002	14.9
Hisp/Latinx	48 (5.8)	9 (4.4)		
Non-Hisp/Latinx	376 (45.1)	120 (58.8)		
Unkn/Unavail	409 (49.1)	75 (36.8)		
Race(%)				
FN/AK Native	4 (0.4)	1 (0.5)	1.000	0.0
Asian	71 (7.1)	15 (6.8)	1.000	0.0
Black/Afr Amer	41 (4.1)	8 (3.7)	0.906	0.0
HI FN/Pac Isl	3 (0.3)	1 (0.5)	1.000	0.0
White	481 (48.1)	127 (58.0)	0.010	0.0
Other	46 (4.6)	11 (5.0)	0.929	0.0
Unkn/Unavail	193 (19.3)	41 (18.7)	0.913	0.0
Comorbidity				
Hypertension (%)	503 (51.2)	154 (71.3)	<0.001	1.6
CAD (%)	116 (11.9)	47 (22.1)	<0.001	2.3
CIED (%)			0.999	15.1
None	810 (97.6)	199 (97.5)		
Pacemaker	16 (1.9)	4 (2.0)		
ICD	4 (0.5)	1 (0.5)		

Table 2. Characteristics of patients who survived and patients who died.

Table 2 shows similar characteristic information as Table 1, but they are grouped by outcomes (survival vs. death). The survival cases consist of 63.7% of patients (n = 776) who were discharged home and 19.3% of patients (n = 235) who were discharged to other care units after COVID-19 stabilization. The same statistical tests, one-way ANOVA and Chi-square, were applied as described above.

Herein, I observed a significant difference between the mean age of the survival group and the death group (P<0.001). In addition, there were significant differences between the prevalence rate of hypertension and coronary artery disease (CAD) of the two groups (P<0.001). Therefore, we could reject the null hypothesis of independence between mortality and risk factors such as age, hypertension, and CAD.

	None	Arrhythmic	TE	HF	P-value	Missing (%)
n	950	112	111	99		
Demographics						
Age (years) (mean (SD))	62.04 (17.10)	69.59 (15.49)	63.66 (14.32)	72.71 (14.66)	<0.001	0.0
BMI (kg/m^2) (mean (SD))	28.22 (6.48)	28.33 (6.77)	28.20 (6.46)	29.21 (7.57)	0.606	15.7
Sex at birth = male (%)	556 (58.5)	75 (67.0)	78 (70.3)	68 (68.7)	0.014	0.0
Ethnicity (%)					0.003	14.6
Hisp/Latinx	51 (6.4)	1 (1.0)	6 (6.2)	1 (1.1)		
Non-Hisp/Latinx	393 (49.1)	52 (52.5)	38 (39.6)	32 (35.6)		
Unkn/Unavail	357 (44.6)	46 (46.5)	52 (54.2)	57 (63.3)		
Race(%)						
FN/AK Native	4 (0.4)	1 (0.9)	0 (0.0)	1 (1.0)	0.657	0.0
Asian	71 (7.5)	5 (4.5)	5 (4.5)	5 (5.1)	0.383	0.0
Black/Afr Amer	42 (4.4)	3 (2.7)	4 (3.6)	2 (2.0)	0.573	0.0
HI FN/Pac Isl	4 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0.715	0.0
White	456 (48.0)	63 (56.2)	65 (58.6)	57 (57.6)	0.033	0.0
Other	38 (4.0)	9 (8.0)	6 (5.4)	6 (6.1)	0.222	0.0
Unkn/Unavail	191 (20.1)	19 (17.0)	16 (14.4)	19 (19.2)	0.477	0.0
Comorbidity						
Hypertension (%)	474 (50.7)	77 (69.4)	57 (52.3)	83 (84.7)	<0.001	1.6
CAD (%)	110 (11.9)	24 (21.8)	10 (9.2)	31 (31.6)	<0.001	2.3
CIED (%)					0.316	14.9
None	780 (97.6)	97 (98.0)	95 (100.0)	85 (95.5)		
Pacemaker	14 (1.8)	2 (2.0)	0 (0.0)	4 (4.5)		
ICD	5 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)		

Table 3. Characteristics of patients who developed cardiovascular complications.

Table 3 shows the characteristic of subgroups stratified by arrhythmic events (9.7%), thromboembolic complication (TE) (9.3%), and heart failure event (HF)(8.3%). The 'None' group refers to patients who did not develop AA, TE, and HF. The same statistical tests, one-way ANOVA and Chi-square, were applied as described above.

There were significant differences between the mean of age, the prevalence of hypertension, and the prevalence of CAD of each subgroup. The mean age of the HF group was 72 years old, higher than the mean age of the AA group (69 years), TE group (63 years), and None group (62 years). The prevalence of hypertension was 50.7% and 52.3% for the None group and TE group respectively, while the values were much higher in the AA group (69.4%) and HF group (84.7%).

Section 3.2 Mortality Model

(a) Optimal model architecture

After applying the exclusive criteria, I trained a binary model for predicting in-hospital all-cause mortality after COVID-19 infection using 976 standard 10-second 12-lead intake ECGs and test on the testing set containing 122 intake ECGs. (**Fig.1**)

The optimal model was selected through a process called hyperparameter tuning or hyperparameter optimization where the model was trained under a different combination of a set of hyperparameters. To find the optimal model, the hyperparameters that I adjusted include the number of convolution layers, the number of LSTM layers, approaches of data augmentation, and approaches of data preprocessing.

The optimal model architecture consists of four convolution layers, followed by two LSTM layers. (**Fig.2)** The numbers of filters were 16, 16, 32, and 32 for each convolution layer, respectively. The kernel sizes were 7,7,5, and 5. The stride value of the kernel is 2 for all convolution layers. Both LSTM layers contain four LSTM units. **Table 4** summarizes the output shape and the detailed parameters of each model layer.

Mortality model parameters of each layer					
Layer	Output Shape	Kernal size	Number of filters	strides	Number of LSTM units
Input	5000x8	-	-	-	-
Conv1d_1	2500x16	7	16	2	-
Conv1d_2	1250x16	7	16	2	-
Conv1d_3	625x32	5	32	2	-
Conv1d_4	313x32	5	32	2	-
LSTM_1	313x4	-	-	-	4
LSTM_2	313x4	-	-	-	4
Flatten	1252	-	-	-	-
Dense	2	-	-	-	-

Table. 4 Architecture and parameters of the mortality model.

(b) Mortality prediction results

All of the performance metrics were derived from the average of individual performance metrics on a 10-fold cross-validation scheme and calculated 95% confidence interval. Average AUROC, sensitivity, and specificity were summarized in **Table 5**. For internal evaluation, the same training sets were used to train another multivariable logistic regression model with R version 3.6.1 and then examined the conventional model on the same hold-out testing sets. The baseline logistic regression model used age, sex, race/ethnicity, BMI, history of hypertension, history of CAD, and presence of CIED as input variables. Four measurements extracted from ECG were also used to adjust the baseline model, including Ventricular Rate, PR Interval, QRS Duration, and QT Internal. The logistic regression model used a threshold = 0.5, while the ML model used an optimal threshold determined through the ROC curve of the validation set.

[95% CI]	AUROC	Sensitivity	Specificity	PPV	NPV
Survival vs Death					
Regression model	0.68 [0.68-0.68]	0.97 [0.96-0.98]	0.13 [0.10-0.17]	0.83 [0.83-0.84]	0.52 [0.41-0.62]
ML model	0.83 [0.79-0.87]	0.65 [0.59-0.70]	0.52 [0.46-0.58]	0.23 [0.21-0.25]	0.87 [0.85-0.88]

Table 5. Performance metrics of the mortality model and baseline model.

The deep learning technique raised the average AUROC to 0.61 (95% CI 0.54 - 0.68), compared to the average AUROC of the logistic regression model was 0.68 (95% CI 0.72-0.64). The ML model for predicting in-hospital all-cause mortality with COVID-19 patients reaches a sensitivity of 0.65, a specificity of 0.52, a positive predictive value (PPV) of 0.23, and a negative predictive value (NPV) of 0.87. (**Table 5**)

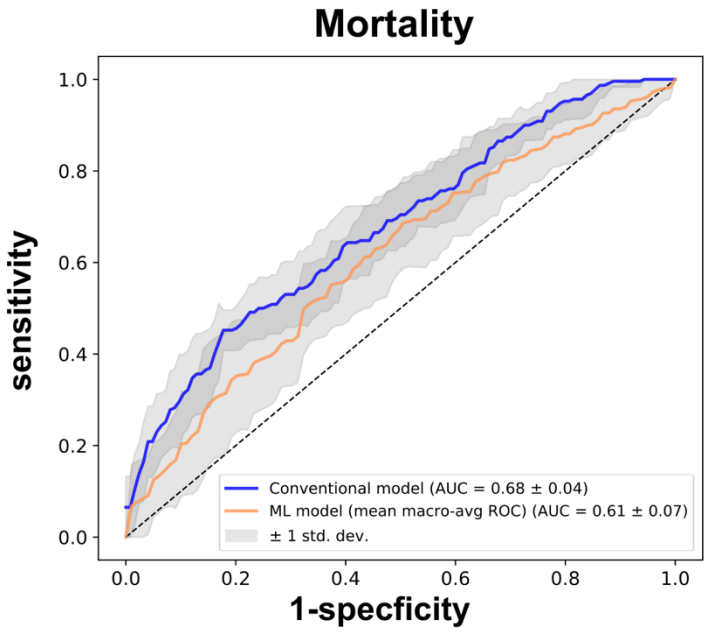


Figure 4. Overall receiver operating characteristic (ROC) curve of the mortality ML model and baseline model. The ROC curves show the average ROC curves and +/- 1 standard deviation across the 10-fold cross-validation scheme.

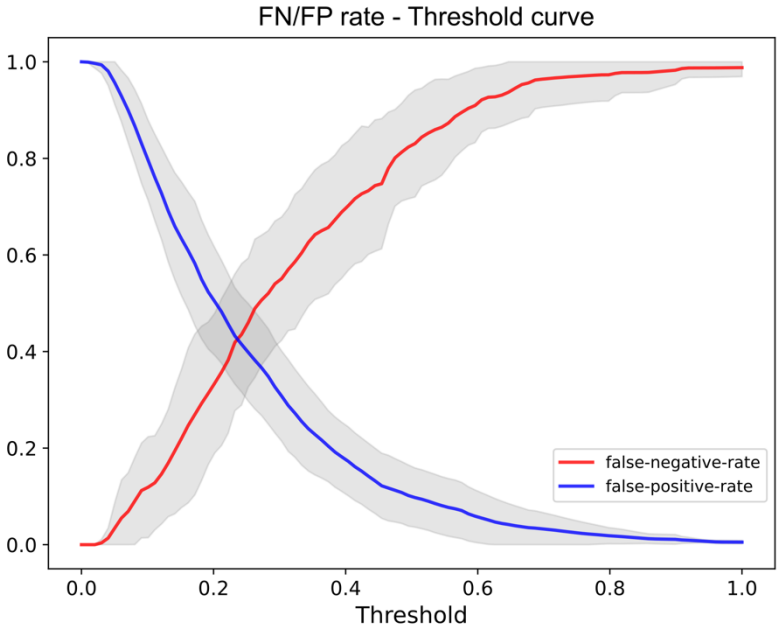


Figure 5. FNR curve and FPR curve with a moving threshold.

Figure 5 plots the false-positive rate (FPR) and the false-negative rate (FNR) with a moving threshold. There was an obvious trade-off between FPR and FNR. However, keeping a

low FNR is more important than a low FPR. Overall, one should carefully consider the resource in all aspects when deciding the acceptable values for FNR, FPR, and optimal threshold.

(c) Optimal train-test-split

This subsection shows the raw model output from one of the ten-fold cross-validations which yielded the greatest AUROC. **Figure 6** depicts the distribution of the model raw output for the two groups, patients who survived and patients who died, in a boxplot and a violin plot. The y-axis means the predicted probability of mortality. The x-axis is the ground truth outcome. The dashed line refers to the optimal threshold determined from the ROC curve of the internal validation set. For people who died, about 75% of them had a predicted probability greater than the threshold. These cases were considered as true-positive (TP). For people who survived, however, there were some false-positive (FP). **Figure 7** gives the example ECGs for true-negative, false-positive, false-negative, and true-positive.

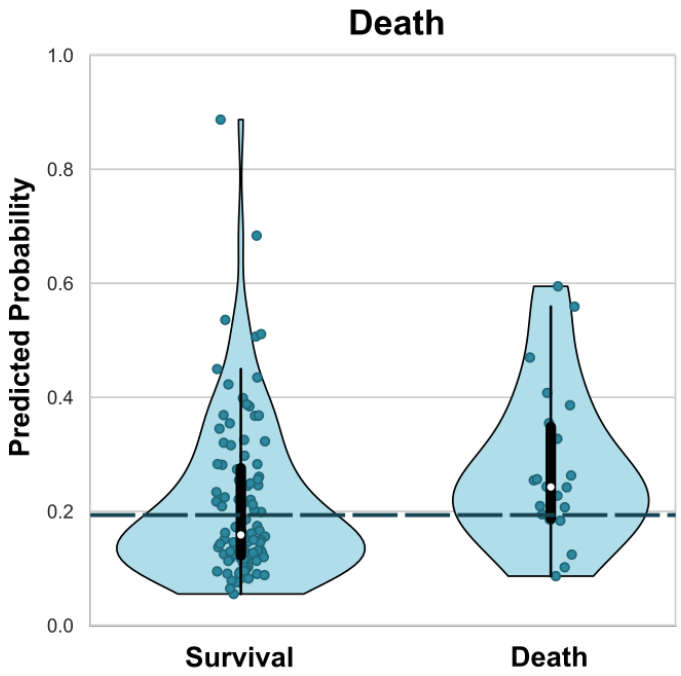


Figure 6. Distribution of the mortality model raw output.

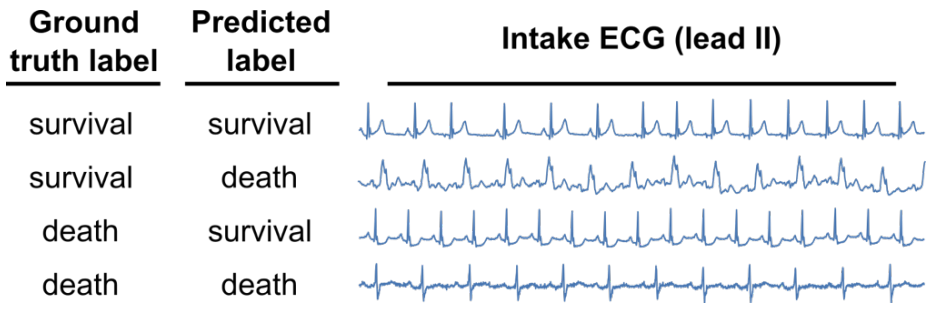


Figure 7. Examples of ECGs and corresponding labels from the test set.

Since the database was imbalanced, a precision-recall curve gives more information about the ability to predict a specific class. In **Figure 8**, class 0 refers to the survival label and class 1 refers to the death label. The precision-recall (PR) curves remain a similar trend among the three sets (training/validation/testing). For a random classifier, the baseline PR curve for class 1 was a horizontal line that had a y-axis value of 0.18 which equaled to the mortality rate in the database. Thus, in this train-test-split, the ML model yielded a 61% relative improvement at the area under the PR curve compared to a random classifier.

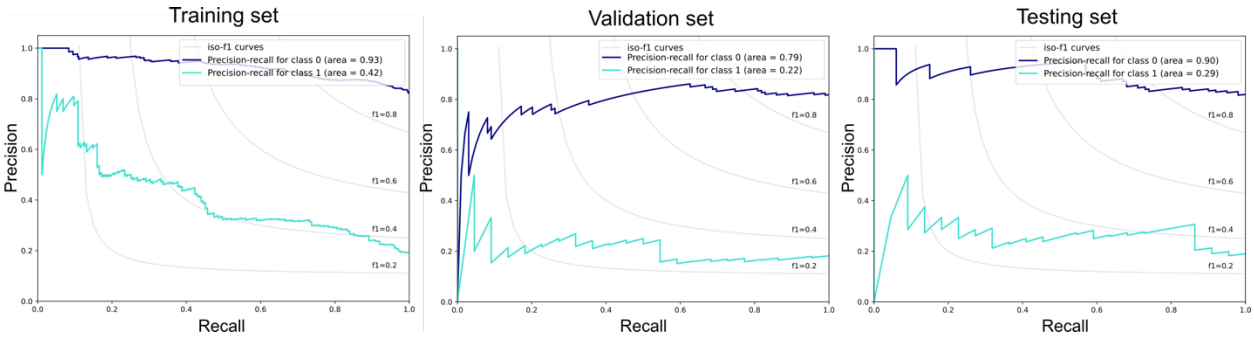


Figure 8. Precision-Recall curves of each class for the training/validation/testing subset.

In **Figure 9**, the results of PCA for the three sets were shown for data visualization. Each intake ECG was transformed into a set of two values representing its first two principal

components and marked on the scatter plot. The samples in the overlapping area between the non-event group and event group means that they were the challenging cases for classification.

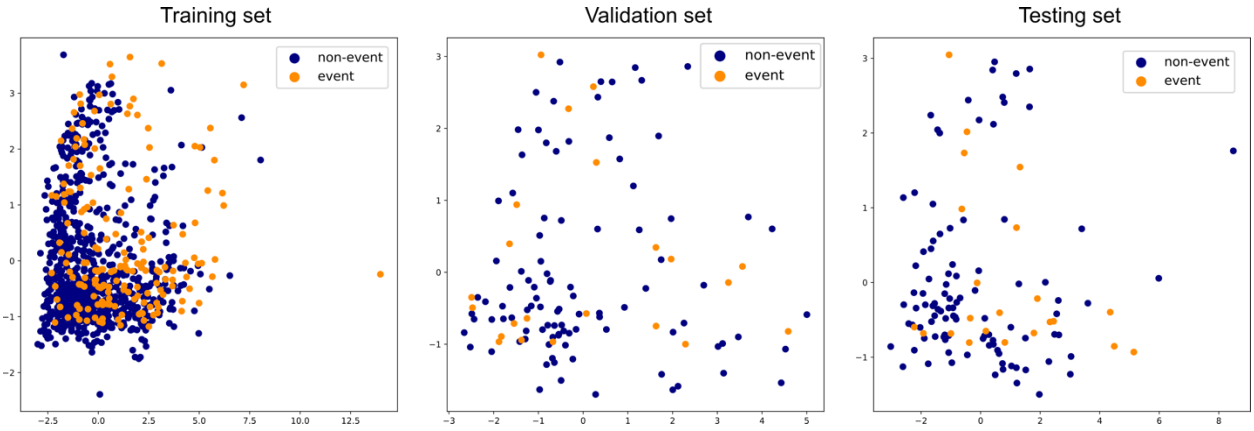


Figure 9. Principle component analysis for the training/validation/testing subset.

Section 3.3 MACE model

(a) Optimal model architecture

I further trained a multi-label model for predicting independent risk scores of developing arrhythmic events (9.3%), thromboembolic complication (TE) (9.3%), and heart failure event (HF)(8.3%) using 971 standard 10-second 12-lead intake ECGs and test on the hold-out testing set containing 127 intake ECGs. (**Fig.1**)

The optimal model architecture consists of four convolution layers, two max-pooling layers, and followed by one LSTM layer. (**Fig.2)** The numbers of filters were 16, 16, 32, and 32 for each convolution layer, respectively. The kernel sizes were 7,7,5, and 5. The stride value of the kernel is 2 for all convolution layers. The LSTM layer contains four LSTM units. **Table 6** summarizes the output shape and the detailed parameters of each layer in this model.

Cardiovascular model parameters of each layer						
Layer	Output Shape	Pool size	Kernal size	Number of filters	strides	Number of LSTM units
Input	5000x8	-	-	-	-	-
Conv1d_1	2500x16	-	7	16	2	-
Conv1d_2	1250x16	-	7	16	2	-
MaxPooling	625x16	2	-	-	2	
Conv1d_3	313x32	-	5	32	2	-
Conv1d_4	157x32	-	5	32	2	-
MaxPooling	79x32	2	-	-	-	-
LSTM_1	79x4	-	-	-	-	4
Flatten	316	-	-	-	-	-
Dense	4	-	-	-	-	-

Table. 6 Architecture and parameters of the MACE model.

(b) MACE prediction results

All of the performance metrics were derived from the average of individual performance metrics on a 10-fold cross-validation scheme and calculated 95% confidence interval. Micro-average AUROC, sensitivity, and specificity were summarized in **Table 7**. For internal evaluation, the same training sets were used to train another multivariable logistic regression model with R version 3.6.1 and then examined the conventional model on the same hold-out testing sets. The baseline logistic regression model used the same variables as the baseline mortality mode. Noted that the baseline model here was a binary classifier which had a class of none-event cases and the other class of all arrhythmic event, thromboembolic event, and heart failure cases.

The ML model yield sensitivities of 0.72, 0.59, 0.57, 0.62 for predicting none-event, arrhythmic events, thromboembolic events, and heart failure, respectively, while the logistic regression model yields a sensitivity of 0.97 and specificity of 0.08. (**Table 7**)

[95% CI]	AUROC	Sensitivity	Specificity	PPV	NPV
Regression model	0.68 [0.65-0.61]	0.97 [0.97-0.98]	0.08 [0.06-0.10]	0.77 [0.75-0.79]	0.52 [0.37-0.66]
ML model					
None-event	0.79 [0.77-0.81]	0.72 [0.65-0.79]	0.36 [0.26-0.46]	0.81 [0.80-0.82]	0.26 [0.24-0.28]
Arrhythmic		0.59 [0.44-0.74]	0.45 [0.37-0.53]	0.10 [0.08-0.12]	0.92 [0.90-0.94]
Thromboembolism		0.57 [0.45-0.69]	0.47 [0.47-0.53]	0.10 [0.08-0.12]	0.92 [0.90-0.93]
Heart failure		0.62 [0.50-0.74]	0.46 [0.34-0.57]	0.09 [0.08-0.10]	0.93 [0.92-0.94]

Table 7. Performance metrics of the MACE model and baseline model.

The deep learning technique yield a the micro-average AUROC of predicting life-threatening cardiovascular complications of 0.79 (95% CI 0.77 - 0.81) and macro-average AUROC of 0.68 (95% CI 0.65 - 0.61), compared to the average AUROC of the logistic regression model was 0.68 (95% CI 0.72-0.64). (**Figure 10**) The macro-average ROC added the y-axis value (sensitivities) of the four classes that matched the same x-axis value (specificities) and then divided by four. Therefore, the macro-average ROC gives an overview of the average ability of predicting the four labels. However, since the thresholds were not calibrated across the four classes, the macro-average ROC curve should not be used for selecting thresholds. On the other hand, the micro-average ROC curve first aggregated the samples of all four classes and then computed the ROC curve. At each point of the micro-average ROC curve, the exact same threshold was applied to all classes. Since it merged all four classes into one class, the micro-average ROC curve indicated the goodness of correctly predicting at least one label.

To select viable thresholds for each individual class, **Figure 11** shows the FPR and the FNR on a moving threshold of each individual class. There was an trade-off between FPR and FNR. However, keeping a low FNR is more important than a low FPR when predicting arrhythmia, thromboembolism, and heart failure. This means the model is sensitive to these positive cases and it is unlikely to miss out any patients that would develop these cardiovascular

events. Though, the con is an increasing number of false-positive. Oppositely, for the non-event class, keeping a low false-positive rate is prior to keeping a low false-negative rate. Therefore, though the model is less sensitive to the non-event cases, if the model predicts that someone has a high probability of being a non-event, we could confidently trust the prediction, no matter what the other three probabilities are.

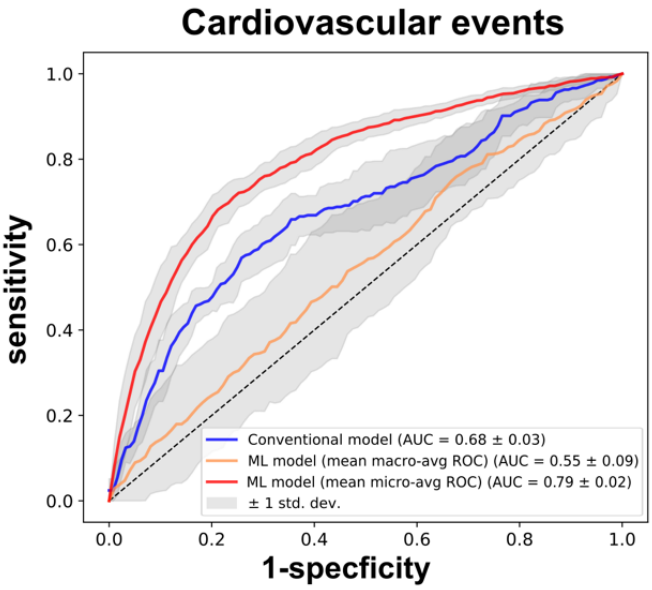


Figure 10. Overall receiver operating characteristic (ROC) curve of the MACE model and baseline model. The ROC curves show the average ROC curves and +/- 1 standard deviation across the 10-fold cross-validation scheme.

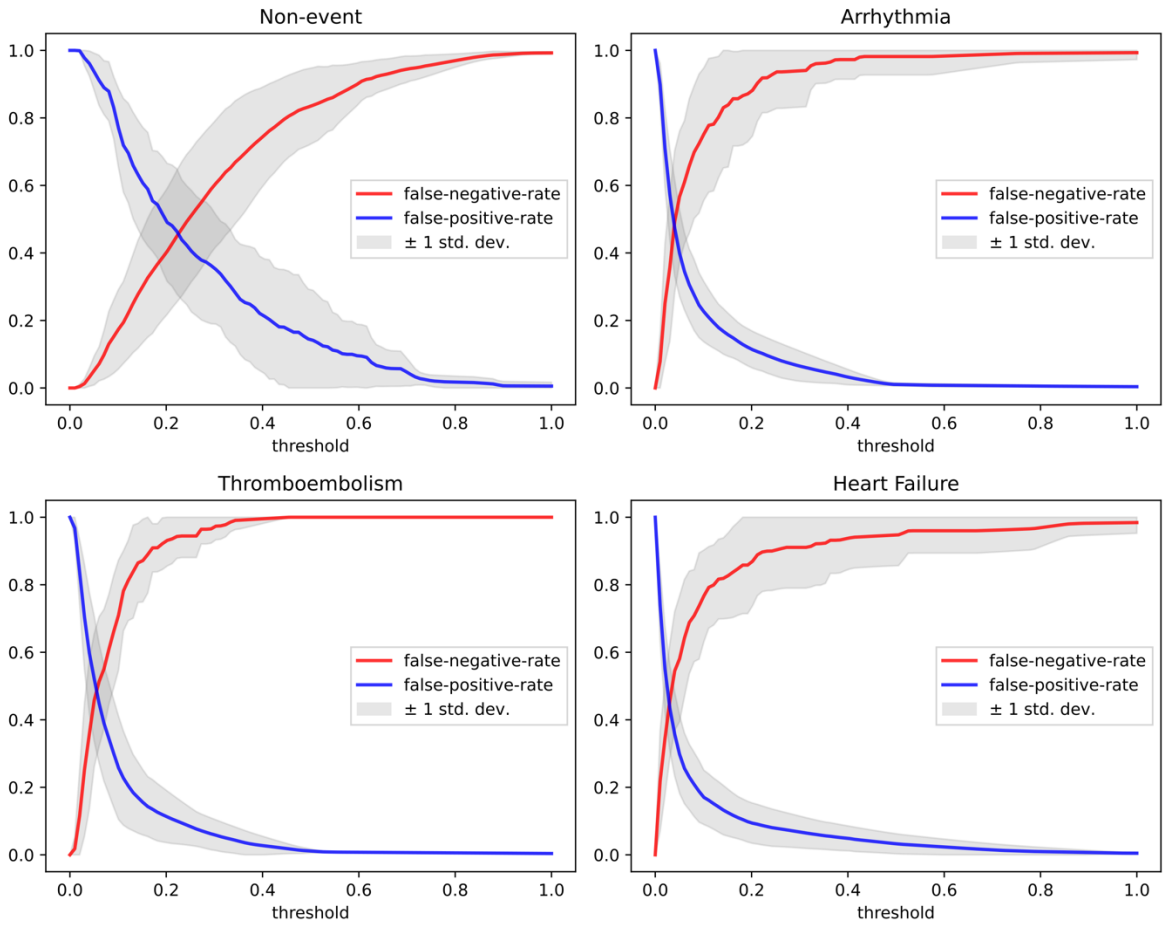


Figure 11. Average FNR curve and average FPR curve of the MACE model with a moving threshold.

(c) Optimal train-test-split

This subsection shows the raw model output from one of the ten-fold cross-validation which yielded the greatest AUROC. **Figure 12** depicts the distribution of the model raw output for the four groups in a boxplot and a violin plot. The y-axis means the predicted probability of the event recorded in the title. The x-axis is the ground truth outcome. The dashed line refers to the optimal threshold determined from the ROC curve of the internal validation set. If a patient developed both an arrhythmic event and heart failure, the predicted probabilities were marked in both ground truth labels. **Figure 13** gives the example ECGs of each type of ground truth label.

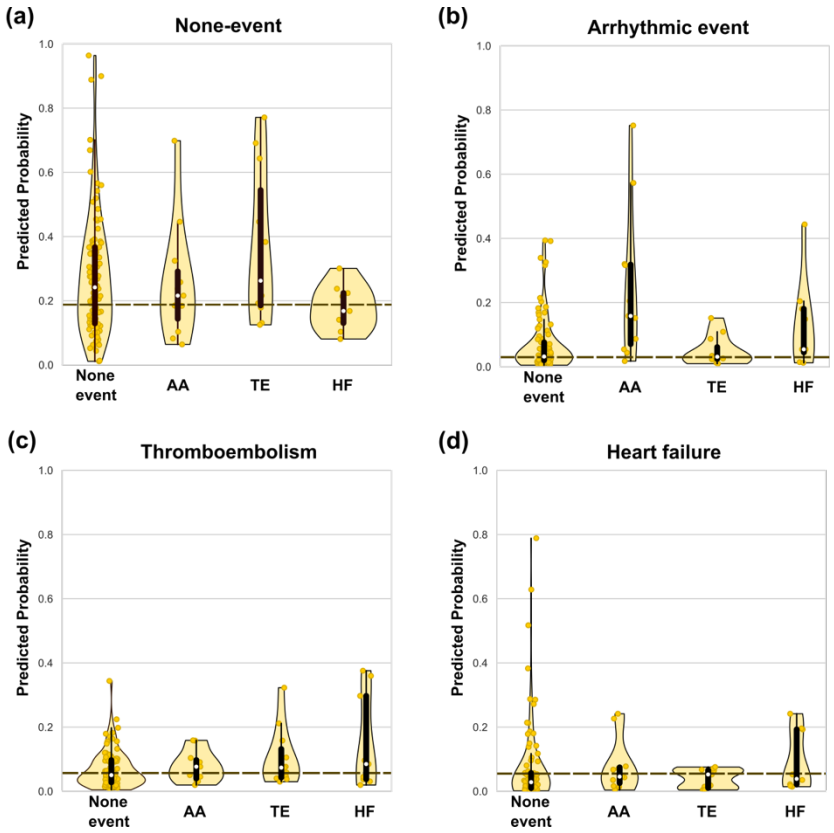


Figure 12. Distribution of the MACE model raw output.

Ground truth label	Predicted label	Intake ECG (lead II)
none	none	
none	AA	
none	TE	
none	AA, HF	
none	none, AA, TE, HF	
none	----	
AA	AA	
AA	none, AA, TE	
AA	AA, TE, HF	
AA	----	
TE	none, TE	
TE	HF	
TE	none	
TE	AA, TE, HF	
HF	none, AA, TE, HF	
HF	TE	
AA, HF	none, AA, HF	
AA, HF	AA	
AA, TE	none, AA, TE, HF	
TE, HF	----	

Figure 13. Examples of ECGs and corresponding labels from the test set.

Since the database was imbalanced, a precision-recall curve gives more information about the ability to predict a specific class. In **Figure 14**, class 0 refers to the non-event label,

class 1 refers to the arrhythmic label, class 2 refers to the thromboembolic label, and class 3 refers to the heart failure label. For a random classifier, the baseline PR curves were horizontal lines with their y-axis values of 0.093, 0.093, and 0.083 for AA, TE, and HF respectively. Thus, in this train-test-split, the ML model relatively improved the area under PR curve by 351%, 72%, and 56% compared to a random classifier. The area under PR curve for class 1 (AA) was much lower in the validation set than in the testing set. This means that the representatives of AA in the validation set and the testing set were distinct from each other. The data of none-evet and AA were visualized by PCA in **Figure 15**. We can observe that, in the training set and the testing set, most samples labeled with AA event have a x-axis value greater than zero; however, only one AA sample has a x-axis value greater than zero.

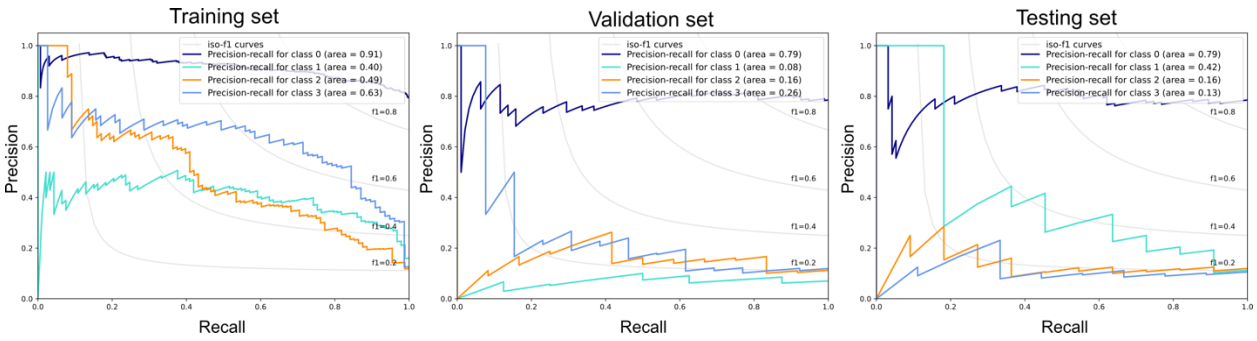


Figure 14. Precision-Recall curves of each class for the training/validation/testing subset.

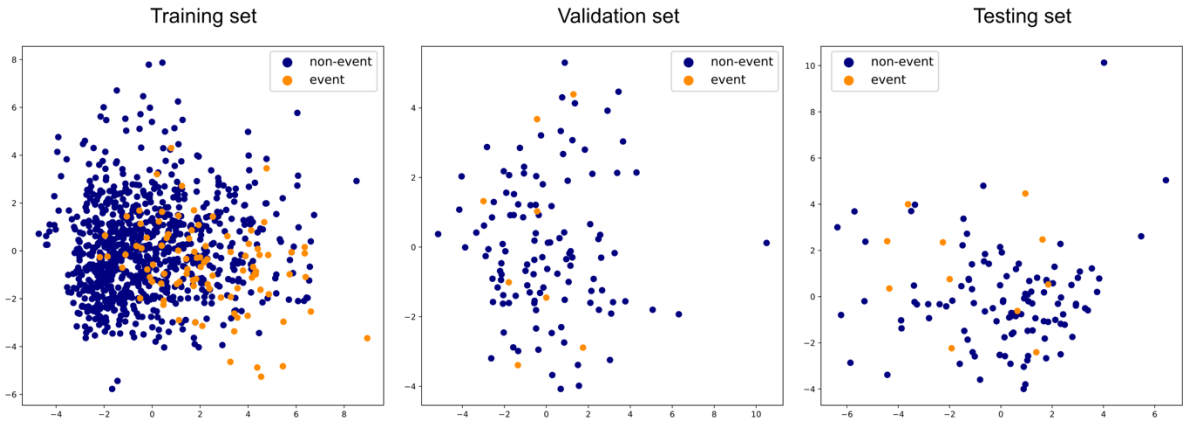


Figure 15. Principle component analysis for non-event class and arrhythmic class in the training/validation/testing subset.

Chapter 4. Discussion and Conclusion

Section 4.1 Discussion

This study demonstrates the potential of using an ECG-based machine learning approach to predict all-cause mortality and cardiovascular outcomes of hospitalized COVID-19 patients. My models can provide an early risk assessment of life-threatening outcomes only with a 10-second intake ECG. Intake ECGs are accessible data that is usually acquired 24 h after a patient was admitted to the hospital.

On the other hand, the length of hospital stay ranges from 0 to 200 days in the database. This implies that the model prediction results are effective in more than six months regardless of comorbidities, treatments, and patient demography.

Though intake ECGs are convenient data for hospitals, comparing to other ML models in predicting COVID-19 outcomes, my models' sensitivities and specificities are relatively lower. These ML models were built on demographics, biomarkers information, physiological values, APACHE-II scores, SOFA scores, and/or health records as model inputs. [25-28] Julie Shade et al developed COVID-HEART predictor, a multivariable continuous-updating risk predictor for hospitalized COVID-19 patients. It predicts imaging-confirmed thromboembolism with a median of 72 hours before the event and yields an AUROC of 0.70.[25] Chuanyu Hu et al trained an ML model with patient demographic, clinical, and first laboratory findings for predicting mortality. While they only built the model with 183 patients, it reaches an AUROC of 0.881, a sensitivity of 0.839, and a specificity of 0.794 on an external validation set.[26] Akil Vaid et al reported their ML model for predicting in-hospital mortality. Their study analyzed the electronic health records of 4098 patients with confirmed COVID-19 to predict the mortality at time windows of 3, 5, 7, and 10 days from admission. The results show an AUROC for mortality of

0.89 at 3 days, 0.85 at 5 and 7 days, and 0.84 at 10 days. The latter two research both conclude that age and C-reactive protein level are critical risk factors for predicting the survival rate of COVID-19 patients.

In sum, my model could act as a primary screening tool for hospitalized patients before they could receive a further thorough examination.

Section 4.2 Limitations

There are some limitations to this work. First, the model outputs are not intuitive. Since the optimal threshold was selected independently for every outcome class, the MACE model might classify a patient under 'None-event' and 'Arrhythmic event' at the same time. It requires a further investigation on how to interpret this kind of model row output. Next, this work was evaluated on a relatively small case population and the label in the database is highly imbalanced. Therefore, there is still room for improving the generality and accuracy of these models.

Section 4.3 Future goals

To advance this work as a primary risk assessment tool for hospitals, my future goals include building a multi-stage model that could provide more intuitive outputs and increasing the accuracy and generality. For example, the multi-stage model could be a series of binary classification processes where the first stage would classify the data under event case or non-event case. If it is classified as a non-event case, the data will not be passed down the latter stages for predicting the probabilities of each cardiovascular complication. The current models serve as good baseline models for future works. To increase the generality, one simple way is to

train the model with a larger dataset. However, building any database, especially for medical data, requires lots of effort and time.

Section 4.4 Conclusion

In conclusion, this study proposed two accessible tools which were built with a deep learning algorithm for hospitals to conduct early risk stratified on hospitalized patients with COVID-19. One model focuses on in-hospital all-cause mortality, while the other model targets life-threatening cardiovascular events, including arrhythmic events thromboembolism, and heart failure. A 10-second 12-lead intake ECG is everything the model needs to predict the outcomes. This work demonstrates the potential of using an intake ECG to predict in-hospital mortality and cardiovascular complications for patients with confirmed COVID-19. In addition, the characteristics of this study cohort were also described in this thesis. There is a trend that patients who died or developed cardiovascular complications are older and processing hypertension. Overall, using the deep learning technique ECGs is a great approach to investigate cardiovascular complications. Deep learning has the potential of discovering underlying ECG patterns and might be able to provide insights into ECG interpretation.

Reference

[1] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery,* vol. 76, pp. 71-76, 2020/04/01/, 2020.

[2] World Health Organization, "Coronavirus disease 2019 (COVID-19): situation reports.," 2021. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed April 23, 2021.

[3] A. Macedo, N. Gonçalves, and C. Febra, "COVID-19 fatality rates in hospitalized patients: systematic review and meta-analysis," *Annals of epidemiology,* vol. 57, pp. 14-21, 2021.

[4] A. Akhmerov, and E. Marbán, "COVID-19 and the Heart," *Circulation Research,* vol. 126, no. 10, pp. 1443-1455, 2020/05/08, 2020.

[5] R. B. Azevedo, B. G. Botelho, J. V. G. d. Hollanda, L. V. L. Ferreira, L. Z. Junqueira de Andrade, S. S. M. L. Oei, T. d. S. Mello, and E. S. Muxfeldt, "Covid-19 and the cardiovascular system: a comprehensive review," *Journal of Human Hypertension,* vol. 35, no. 1, pp. 4-11, 2021/01/01, 2021.

[6] P. Dherange, J. Lang, P. Qian, B. Oberfeld, W. H. Sauer, B. Koplan, and U. Tedrow, "Arrhythmias and COVID-19: A Review," *JACC. Clinical electrophysiology,* vol. 6, no. 9, pp. 1193-1204, 2020.

[7] E. J. Topol, "COVID-19 can affect the heart," no. 1095-9203 (Electronic).

[8] National Heart, Lung and Blood Institute, "Arrhythmia," Available at https://www.nhlbi.nih.gov/health-topics/arrhythmia. Accessed April 23, 2021.

[9] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, and Z. Peng, "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," *JAMA,* vol. 323, no. 11, pp. 1061-1069, 2020.

[10] P. Goyal, J. J. Choi, L. C. Pinheiro, E. J. Schenck, R. Chen, A. Jabri, M. J. Satlin, T. R. Campion, M. Nahid, J. B. Ringel, K. L. Hoffman, M. N. Alshak, H. A. Li, G. T. Wehmeyer, M. Rajan, E. Reshetnyak, N. Hupert, E. M. Horn, F. J. Martinez, R. M. Gulick, and M. M. Safford, "Clinical Characteristics of Covid-19 in New York City," *New England Journal of Medicine,* vol. 382, no. 24, pp. 2372-2374, 2020/06/11, 2020.

[11] F. A. Klok, M. J. H. A. Kruip, N. J. M. van der Meer, M. S. Arbous, D. A. M. P. J. Gommers, K. M. Kant, F. H. J. Kaptein, J. van Paassen, M. A. M. Stals, M. V. Huisman, and H. Endeman, "Incidence of thrombotic complications in critically ill ICU patients with COVID-19," *Thrombosis Research,* vol. 191, pp. 145-147, 2020/07/01/, 2020.

[12] I. H. Khan, S. Savarimuthu, M. S. T. Leung, and A. Harky, "The need to manage the risk of thromboembolism in COVID-19 patients," *Journal of Vascular Surgery,* vol. 72, no. 3, pp. 799-804, 2020/09/01/, 2020.

[13] J. R. Rey, J. Caro-Codón, S. O. Rosillo, M. Iniesta Á, S. Castrejón-Castrejón, I. Marco-Clement, L. Martín-Polo, C. Merino-Argos, L. Rodríguez-Sotelo, J. M. García-Veas, L. A. Martínez-Marín, M. Martínez-Cossiani, A. Buño, L. Gonzalez-Valle, A. Herrero, J. L. López-Sendón, and J. L. Merino, "Heart failure in COVID-19 patients: prevalence, incidence and prognostic implications," no. 1879-0844 (Electronic).

[14] D. P. Wagner, and E. A. Draper, "Acute physiology and chronic health evaluation (APACHE II) and Medicare reimbursement," *Health care financing review,* vol. Suppl, no. Suppl, pp. 91-105, 1984.
[15] A. E. Jones, S. Trzeciak, and J. A. Kline, "The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation," *Critical care medicine,* vol. 37, no. 5, pp. 1649-1654, 2009.
[16] S. Barbar, V. Noventa F Fau - Rossetto, A. Rossetto V Fau - Ferrari, B. Ferrari A Fau - Brandolin, M. Brandolin B Fau - Perlati, E. Perlati M Fau - De Bon, D. De Bon E Fau - Tormene, A. Tormene D Fau - Pagnan, P. Pagnan A Fau - Prandoni, and P. Prandoni, "A risk assessment model for the identification of hospitalized medical patients at risk for venous thromboembolism: the Padua Prediction Score," no. 1538-7836 (Electronic).
[17] X. Zou, S. Li, M. Fang, M. Hu, Y. Bian, J. Ling, S. Yu, L. Jing, D. Li, and J. Huang, "Acute Physiology and Chronic Health Evaluation II Score as a Predictor of Hospital Mortality in Patients of Coronavirus Disease 2019," *Critical care medicine,* vol. 48, no. 8, pp. e657-e665, 2020.
[18] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ECG signals using machine learning techniques: A survey." pp. 714-721.
[19] A. Gupta, E. Huerta, Z. Zhao, and I. Moussa, *Deep Learning for Cardiologist-level Myocardial Infarction Detection in Electrocardiograms*, 2019.
[20] A. K. Feeny, M. K. Chung, A. Madabhushi, Z. I. Attia, M. Cikes, M. Firouznia, P. A. Friedman, M. M. Kalscheur, S. Kapa, S. M. Narayan, P. A. Noseworthy, R. S. Passman, M. V. Perez, N. S. Peters, J. P. Piccini, K. G. Tarakji, S. A. Thomas, N. A. Trayanova, M. P. Turakhia, and P. J. Wang, "Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology," *Circulation. Arrhythmia and electrophysiology,* vol. 13, no. 8, pp. e007952-e007952, 2020.
[21] S. Sahoo, M. Dash, S. Behera, and S. Sabut, "Machine Learning Approach to Detect Cardiac Arrhythmias in ECG Signals: A Survey," *IRBM,* vol. 41, no. 4, pp. 185-194, 2020/08/01/, 2020.
[22] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, S. J. Asirvatham, A. J. Deshmukh, B. J. Gersh, R. E. Carter, X. Yao, A. A. Rabinstein, B. J. Erickson, S. Kapa, and P. A. Friedman, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet,* vol. 394, no. 10201, pp. 861-867, 2019.
[23] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Science Advances,* vol. 4, no. 2, pp. e1700578, 2018.
[24] C. Siontis Konstantinos, X. Yao, P. Pirruccello James, A. Philippakis Anthony, and A. Noseworthy Peter, "How Will Machine Learning Inform the Clinical Care of Atrial Fibrillation?," *Circulation Research,* vol. 127, no. 1, pp. 155-169, 2020/06/19, 2020.
[25] J. K. Shade, A. N. Doshi, E. Sung, D. M. Popescu, A. S. Minhas, N. A. Gilotra, K. N. Aronis, A. G. Hays, and N. A. Trayanova, "COVID-HEART: Development and Validation of a Multi-Variable Model for Real-Time Prediction of Cardiovascular Complications in Hospitalized Patients with COVID-19," *medRxiv*, pp. 2021.01.03.21249182, 2021.
[26] C. Hu, Z. Liu, Y. Jiang, O. Shi, X. Zhang, K. Xu, C. Suo, Q. Wang, Y. Song, K. Yu, X. Mao, X. Wu, M. Wu, T. Shi, W. Jiang, L. Mu, D. C. Tully, L. Xu, L. Jin, S. Li, X. Tao,

T. Zhang, and X. Chen, “Early prediction of mortality risk among patients with severe COVID-19, using machine learning,” *International Journal of Epidemiology,* vol. 49, no. 6, pp. 1918-1929, 2020.

[27] M. Mehta, J. Julaiti, P. Griffin, and S. Kumara, “Early Stage Machine Learning–Based Prediction of US County Vulnerability to the COVID-19 Pandemic: Machine Learning Approach,” *JMIR Public Health Surveill,* vol. 6, no. 3, pp. e19446, 2020/9/11, 2020.

[28] A. Vaid, S. Somani, A. J. Russak, J. K. De Freitas, F. F. Chaudhry, I. Paranjpe, K. W. Johnson, S. J. Lee, R. Miotto, F. Richter, S. Zhao, N. D. Beckmann, N. Naik, A. Kia, P. Timsina, A. Lala, M. Paranjpe, E. Golden, M. Danieletto, M. Singh, D. Meyer, P. F. O'Reilly, L. Huckins, P. Kovatch, J. Finkelstein, R. M. Freeman, E. Argulian, A. Kasarskis, B. Percha, J. A. Aberg, E. Bagiella, C. R. Horowitz, B. Murphy, E. J. Nestler, E. E. Schadt, J. H. Cho, C. Cordon-Cardo, V. Fuster, D. S. Charney, D. L. Reich, E. P. Bottinger, M. A. Levin, J. Narula, Z. A. Fayad, A. C. Just, A. W. Charney, G. N. Nadkarni, and B. S. Glicksberg, “Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation,” *J Med Internet Res,* vol. 22, no. 11, pp. e24018, 2020/11/6, 2020.

[29] R. Rojas, "The Backpropagation Algorithm," *Neural Networks: A Systematic Introduction*, R. Rojas, ed., pp. 149-182, Berlin, Heidelberg: Springer Berlin Heidelberg, 1996.

[30] H. Robbins, and S. Monro, “A Stochastic Approximation Method,” *The Annals of Mathematical Statistics,* vol. 22, pp. pp. 400-407, 1951.

[31] F. Chollet, and others, "Keras," GitHub, 2015.

[32] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. I. , Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research,* vol. 12, pp. 2825--2830, 2011.

[34] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, “API design for machine learning software: experiences from the scikit-learn project,” *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108--122, 2013.

[35] G. v. Rossum, and J. d. Boer, “Interactively Testing Remote Servers Using the Python Programming Language,” *CWI Quarterly,* vol. 4, no. 4, 1991.

[36] S. Hochreiter, and J. Schmidhuber, “Long short-term memory,” *Neural Comput,* vol. 9, no. 8, pp. 1735-80, Nov 15, 1997.

[37] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, “Research electronic data capture (REDCap)—A metadata-driven methodology and workflow

process for providing translational research informatics support," *Journal of Biomedical Informatics,* vol. 42, no. 2, pp. 377-381, 2009/04/01/, 2009.
[38] P. A. Harris, R. Taylor, B. L. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, and S. N. Duda, "The REDCap consortium: Building an international community of software platform partners," *Journal of Biomedical Informatics,* vol. 95, pp. 103208, 2019/07/01/, 2019.
[39] R. Koha, "A Study of CrossValidation and Bootstrap for Accuracy Estimation and Model Selecti," *International Joint Conference on Articial Intelligence*, 1995.
[40] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.