# Tissue Characterization with Surgical Smart Grasper and Hybrid CNN–GRU Model

Yana Sosnovskaya

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

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Blake Hannaford, Chair

Eli Shlizerman

Joshua Smith

Program Authorized to Offer Degree:

Electrical & Computer Engineering

University of Washington

**Abstract**

Tissue Characterization with Surgical Smart Grasper and Hybrid CNN–GRU Model

Yana Sosnovskaya

Chair of the Supervisory Committee:
Professor Blake Hannaford
Electrical and Computer Engineering

Minimally Invasive Surgery (MIS) is now standard in modern medicine and involves operating through small incisions using laparoscopic instruments (graspers) for manipulation and endoscopic cameras for visual feedback. Advantages of MIS include faster recovery, less blood loss, and lower risk of complications. Alongside its benefits, MIS brings new challenges, such as the lack of tactile feedback for surgeons because, in open surgery, surgeons can palpate the tissue to gain information about its non-visible structure and abnormalities, such as tumors, blood vessels, and foreign bodies (lost surgical instruments or shrapnel). Adding miniaturizing sensors to laparoscopic graspers can solve the problem of palpating and sensing for surgeons. Such devices paired with artificial intelligence (AI) algorithms for noise removal and data processing can provide valuable new information to surgeons and help with early diagnosis and treatment. The thesis aims to extend the work done by Philip Roan on the motorized Smart Grasper Robot. The thesis involves re-designing all electronics involved in Roan's design to accommodate integrating a new sensor - a 1D A-mode ultrasound sensor- into one of the grasper's tips. Ultrasound transducers were not used in bidirectional mode on surgical graspers before, which gives a set of problems solved in the thesis. One issue involves finding the methods for time series ultrasound signals to clear them from the ringing artifact and the noise that interferes with the echo. Another problem is estimating the Time of Flight on filtered ultrasound signals. Another problem the thesis aims to solve is

evaluating and developing a multimodal AI algorithm that can process all sensor modalities simultaneously. Multimodal AI algorithms with physiological data are rarely used in research nowadays, but having the data from different sensor modalities that can measure and process simultaneously a lot of vital parameters from the patient could give valuable information for surgeons of minimal changes in the patient's vitals during the surgery. As a result of the thesis, we collected *in-vitro* and *in-vivo* multimodal time series data from animals and released it openly to use it for the community to train more sophisticated Deep Learning models.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# GLOSSARY

ABLATION: removal of the tissue or body part. Could be applicable to removal of the sensor's modality as well.

ABNORMALITY: deviation from the normal state.

CLASSIFICATION: the categorization of a population into two or more sub-populations, according to shared qualities or characteristics.

CONVOLUTIONAL NEURAL NETWORKS (CNN): a deep learning method that learns features by itself by filter optimization.

DATA FUSION: a combination of data.

DEEP LEARNING: the subset of methods of machine learning based on artificial neural networks.

FEATURE: distinctive component or characteristics of something, e.g. signal.

GATED RECURRENT UNIT (GRU): a type of RNN model, has advantages in terms of memory and it's more accurate for long sequences.

HEMOGLOBIN: a protein found in red-blood cells which binds to oxygen to transport it through the blood.

HYBRID MODEL: a combination of models.

IMPEDANCE: the resistance to the flow of electrons. It is typically used to refer to materials which are exposed to alternating currents and incorporates the magnitude and phase.

LONG SHORT TERM MEMORY (LSTM): a type of RNN model, similar to GRU, can deal with vanishing gradient problem.

MODALITY: a particular form of perception.

MULTIMODAL DEEP LEARNING: building or expanding models that process and relate information from multiple modalities.

PHOTOPLETHYSMOGRAPHY (PPG): a signal that uses the infrared light to measure the volumetric blood variations. Commonly known as heart rate.

RECURRENT NEURAL NETWORKS (RNN): a deep learning method chacterized by direction of information between layers. Used for time series signals.

RESISTANCE: an opposition to the flow of electrical current.

RINGDOWN ARTIFACT: an oscillatory artifact

SPO2: A measure of the dissolve oxygen in the blood.

# ACKNOWLEDGMENTS

I would like to thank my advisor, Blake Hannaford, for his guidance, support, and patience. Additionally, I would like to thank all members of the UW Biorobotics Laboratory. In particular, Astrini Sie, Andrew Pace, Melody Su, Nivii Kalavakonda, Joey Sullivan, Andrew Lewis, and Haonan Peng, who helped me with technical advice on this project, encouraged me and inspired me. I'm deeply going to miss our discussions.

I want to acknowledge my collaborators Eli Shlizerman, Mika N. Sinanan and thank for their provided expertise on this project.

This project would not have been possible without the tremendous help of dedicated undergraduate researchers. Their role in debugging hardware and software, running experiments, and calibrating sensors was crucial. They also played a significant part in my personal growth as a mentor and leader. I would like to extend my gratitude to two exceptional MS students, Jack Kaplan and Shaheryar Hasnain, who managed to revive the old motor of the Smart Grasper with limited information and access due to the COVID-19 pandemic.

I would like to thank UW Harborview Medical Center and Dr. Nathan White's Emergency Medicine Research group for helping us conduct animal experiments and answering many of our questions about medical devices and procedures. Big thanks also to MS student Eric Chang and undergraduate student Benito Elmer for helping with animal experiments and giving their inputs for design decisions for improving Smart Grasper based on surgical room experience.

I want to thank my parents, Natalia Gavrilova and Dmitriy Gavrilov, and my parents-in-law, Olga and Raymond Hayes, for their support, love, patience, and encouragement. Also, thanks to my brother, Alexey Gavrilov, for his encouragement and for checking on me almost

every day despite being in a different country.

In particular, I'd like to thank my husband, Eugeny Sosnovsky, who supported me with his love and patience through this hard and long journey and gave me valuable technical advice. And my pug, Kefir, for being the best labmate during the COVID-19 lockdown.

Lastly, I'd like to thank my dear best friends, Mayuree Binjolkar, Ekaterina Nazarova, Julia Starceva, Elena Vasilieva, and Elena Smirnova and many others, who supported me throughout this journey.

This material is based upon work supported by the National Science Foundation under Grant #2036255. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# DEDICATION

to my father, Dmitriy Gavrilov,

who had the tremendous patience to explain mathematical concepts to me over and over again.

# Chapter 1

# INTRODUCTION

## *1.1 Medical Need*

Minimally-Invasive Surgery (MIS) is now standard in modern medicine and involves operating through small incisions using laparoscopic graspers for tissue manipulation and endoscopic cameras for visual feedback. MIS has a lot of advantages to compare with open surgeries: faster recovery, less blood loss, and lower risk of complications [1, 2]. However, MIS has new challenges, such as the lack of sense of touch for surgeons because in open surgery, surgeons can palpate tissue to gain sense information about its non-visible structure, abnormalities, such as tumors and blood vessels [1, 3]. Adding miniaturized sensors to laparoscopic graspers can solve the problem of lost sense of touch for surgeons, and provide valuable information about vitals and tissue change during the surgery.

One of the solution could be developing the surgical grasper that can sense different types of properties of tissue and solve a variety of surgeon's problems, such as sense abnormalities in tissue, classify tissue to help with cancer diagnosis, tissue slippery detection. Adding motor to the surgical grasper can help to have applied uniformed pressure to the tissue without fear to damage the tissue.

In the dissertation, I upgraded such surgical device, Smart Grasper, that has a multisensory set to measure different tissue properties (mechanical, electrical, acoustic, thermal properties) simultaneously, and can provide useful information to the surgeons. Smart Grasper is based on the original work of Philip Roan [4], but with modifications and updates in hardware to accommodate a new sensor – a 1D bi-directional ultrasound sensor. To our best knowledge, it's the first motorized surgical grasper with bi-directional ultrasound on its tip. In addition to ultrasound, Smart Grasper also has pulse-oximeter, bioimpedance, thermistor,

indirect force sensing, and an encoder to measure the position of grasper's jaws. All of the sensor modalities require its own hardware and signal processing development in order to receive reliable signals for further processing, as well as calibration procedures. Some of the calibration procedures were developed during the thesis work in order to increase precision of the device. This kind of instrument can perform different experiments, both *in-vitro* (butcher shop materials) or *in-vivo* (animal experiments).

These new multimodal data streams demand exploration of new multimodal deep learning models, based on time series data obtained from the sensors. A multimodal deep learning model will be able to learn hundredth features from the time series signals to compare with classical statistical methods.

The surgical device paired with AI algorithms for sensor's filtering techniques and data fusion will help for real-time diagnostics procedures, abnormality detection and tissue identification in operating room where the time is limited.

### *1.2 Broader Medical Impact*

The broader impacts of the dissertation range from improved surgical care to creating a new surgical multimodal robotic solutions. The multimodal sensor system concept arose from a well-established collaboration between engineers and clinicians on smart medical devices and algorithms for improved surgical care. This deep collaboration ensures significant and valid potential of this forward looking research to impact medicine.

The clinical benefits are creating proven technology, preliminary *ex-vivo*, *in-situ* and *in-vivo* multimodal data, and new ideas for future clinical-focused research such as the detection of cancerous tissue, abnormalities and tissue identification in real time.

**Impact 1: smart surgical instruments for manual or robotic surgery.** From a therapeutic standpoint, surgery requires that tissue planes be defined and then divided, often with diathermy or cautery for hemostasis. Being able to interrogate tissues just before this division would help confirm that the tissue could be safely cauterized and cut, that there were no unexpected structures such as vessels too large for electrocautery to control,

or nerves and other vital structures in the divided tissue plane. Dissection around the common bile duct during biliary surgery or around the ureter in colorectal, retroperitoneal, or gynecologic surgery, often for cancer, are specific examples where this information would improve safety. With further investigation, it's possible to define acoustic and other multimodal tissue characteristics around different types of cancer or abnormalities that indicate more precisely the minimum safe margin of resection. This would help minimize the risk for local cancer recurrence and augment the surgeon's "art" with real-time tissue evaluation.

**Impact 2: new scientific data about living organs in surgically realistic contexts.** Knowledge of how interactions such as grasping with controlled contact pressure affect tissue function (including deformation, oxygen supply, damage and tissue repair responses) is fundamental to new research in semi-autonomous surgical robotics.

Companies such as SynDaver (Tampa, FL) and Simulab (Seattle, WA) produce physical models attempting to mimic human tissue properties for teaching anatomy and surgery. However, medical students claimed that the current state of the art in synthetic cadavers is a useful adjunct, but still lacks realistic representation of the human body's tissue and organs and should not serve as a replacement for cadavers in a medical school curriculum [5]. The research described in the dissertation (particularly the biomechanical and ultrasound measurements) provides a scientific foundation for medical education and automated intervention devices which interact with living tissues.

**Impact 3: novel multimodal and times series datasets released to the community for research on advanced healthcare.** This project released its multimodal chicken tissue dataset on IEEE DataPort repositories [6], along with identified ground truth labels, and Python scripts for reading the data. Ultrasound dataset for denoising time series acoustic signals on short distances from the ringdown artifact also released on IEEE DataPort repositories [7]. The multimodal porcine *in-vivo* and *in-situ* datasets will be released after publishing the article. The datasets will help develop better multimodal fusion techniques and novel multimodal deep learning models to combine data from different sensors, denoise the signals, and learn new features from the signals.

### *1.3 Dissertation Overview*

The following chapters will describe literature review (Chapter 2), Smart Grasper design and its sensor's modalities (Chapter 3), calibration procedures (Chapter 4), an introduction to an ultrasound sensor, artifact removal algorithms from acoustic signals (Chapter 5), animal *ex-vivo*, *in-situ*, and *in-vivo* experiments (Chapter 6), new datasets, a deep learning CNN–GRU introduction, the application of the model to the collected data (Chapter 7), discussion of results (Chapter 8), conclusion (Chapter 9) and future work (Chapter 10). The appendix contains schematic designs and layouts for the developed sensor sets and peripheral electronics.

# Chapter 2

# LITERATURE REVIEW AND BACKGROUND

## *2.1 Tissue Properties*

Tissue characterization can be done by studying the tissue properties: thermal, acoustic, optical, mechanical and electrical properties. Knowledge of different tissue properties could help with:

- providing real-time tactile feedback to the surgeons;
- sensing abnormalities on early stage of a disease;
- modeling the tissue for medical simulators and the robotic surgeries;
- understanding tissue behavior for different types of exposure;
- substitution of tissue by similar properties materials.

Tissue properties have high dependence on different natural factors, such as age of the subject, the amount of protein (collagen) and fat in tissue. Death is also one of the crucial factors that changes the tissue properties. It's much easier to study tissue in a laboratory *in-vitro* than on living animals or human subjects. However, it may be that *in-vivo* and *in-vitro* tissue properties measurements have little correlation [8]. One of the goals of this dissertation is to evaluate the tissue characterization *in-vitro* and *in-vivo* by analyzing the different properties of the tissue. Thermal, acoustic, optical, mechanical and electrical properties are integrated into a proposed new surgical instrument.

**Thermal Properties** To a first approximation, the internal body temperature is constant throughout the body. However, differential tissue blood flow [9], water content [10], and local metabolic activity [11] likely determine regional temperature variations within an organ, which are measured by our developed grasper and have clinical significance to organ treatment and healing. Local temperature variations can be helpful in surgical intervention. For example, [12] showed that local temperature indicated viable tissue for bowel anastomosis. Some of the Smart Grasper's other modalities, such as acoustic and electrical, have been shown to depend on temperature, so contact-point temperature sensing needs to be tested to see the influence of temperature on acoustic and electrical signals.

**Acoustic Properties** Ultrasound is widely used for non-invasive medical imaging. One of the useful characteristics in tissue classification is velocity of sound through tissue. Several previous studies show thermal-induced changes in tissue velocity of sound. When temperature increases, acoustic waves propagate faster in water-based tissue, and slower in fatty-tissue [13, 14, 15]. But there is a need for further research on temperature-dependent velocity of sound propagation through actual tissue *in-vitro* and specially, *in-vivo*, which current work addresses. Another acoustic property that used for tissue identification is attenuation. One application used a transmitter-receiver transducer pair on Martin forceps to detect soft tissue cancers by evaluating the echo attenuation [16].

**Optical Properties** Optical Spectroscopy is increasingly used for cancer detection diagnostics [17]. We can measure scattered or transmitted optical properties with light-emitting diodes (LED) at selected wavelengths and a photodiode to capture the scattered or transmitted signal after it interacts with the tissue. Correlation of optical and other parameters may enhance the accuracy and discrimination of pathological from healthy tissue, cancer from non-cancer tissue, vascularized from nonviable tissue. Smart Grasper optically measures hemoglobin oxygen saturation, locally under the grasper jaws, via relative absorption of two LED emissions at 660 nm and 940 nm (the well known values used in common Pulse

Oximeters).

**Mechanical Properties** The existing literature on *in-vivo* biomechanics of internal abdominal organs is very limited [18, 19]. Rosen [20] and Ottensmeyer [19] work has documented very substantial differences in terms of biomechanical responses (e.g. compressive strain vs applied grasp pressure) between *in-vivo* and *in-vitro* tissue samples, differences that are thought to be due to blood perfusion and the effects of metabolic activity in living tissues. When applied to estimation of tissue health status and to calibration of surgical simulator responses, these differences are highly significant.

**Electrical Properties** Electrical impedance (bioimpedance) has long been studied for various biomedical applications, such as to measure BMI [21], blood cholesterol [22], blood glucose analysis [23], and malignant tumor detection [24]. Bioimpedance has characteristics that depend on tissue structure, fat, composition and frequency [22]. Buono et al. measured bioimpedance change with tissue temperature [25]. Developed multimodal surgical grasper is able to measure complex bio-impedance (magnitude $|V(j\omega)| \,/\, |I(j\omega)|$ and phase angle $\angle\,(V(j\omega)/I(j\omega))$ at range of frequencies from 30 kHz to 70 kHz.

### *2.2 Related Previous UW Biorobotics Lab Research*

Earlier work in UW Biorobotics Lab produced an initial version of a multimodal grasper [4, 20, 26]. Roan's prototype device added optical, thermal and electrical impedance sensing to the $7 \times 7$ mm tip of the commercial surgical grasper. A 10 W brushed DC motor (Maxon, Fall River, MA, USA) drove the grasper via a capstan, a 19:1 planetary gearhead and a partial pulley with a 0.024 in uncoated stainless steel cable (#2024, Sava Industries, NJ, USA). Rotating motion of the partial pulley drove the pushrod of the grasper head via a parallel plate load cell and strain gages. A 500 count-per-revolution rotary differential encoder ($19 \times 500 = 9500$ effective counts per capstan revolution) measured jaw opening and resulting tissue displacement. A special-purpose Linux embedded PC with custom processing

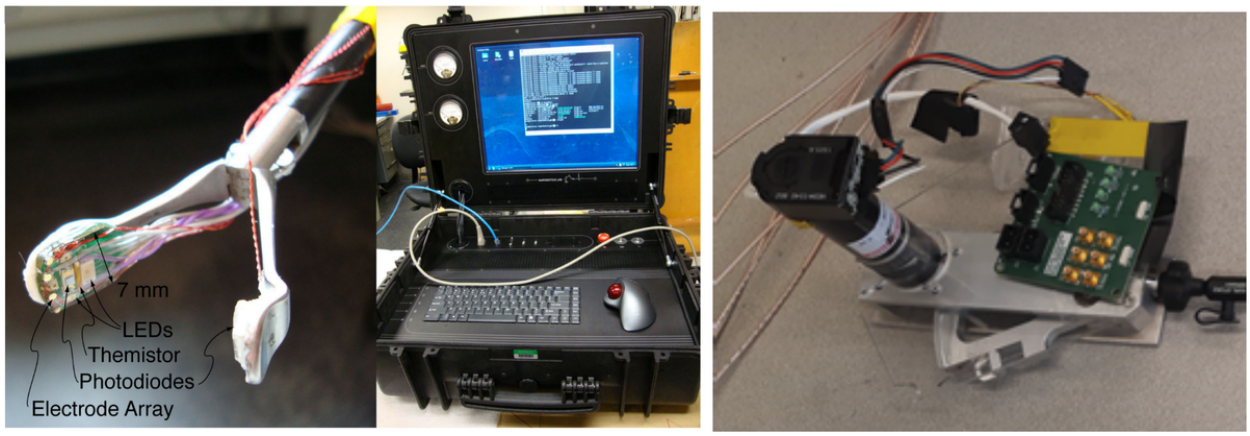



Figure 2.1: Instrumented grasping prototype from previous work. $7 \times 7$ mm instrument tips contain optical, temperature, electrical impedance sensors connected to electronics case via fine wires. Custom embedded Linux PC collected data in animal lab. Close up of drive mechanism (DC motor with optical encoder and alpha-wrapped pulley) and new signal interface board.

board amplified and filtered the analog signals from the sensors, converted them to digital form, and connected them via a USB interface. Qt-based graphical interface controlled the motor, and visualized the real-time signals from the sensors such as pulse, oxygen level and bioimpedance. Sensors were surface mount soldered to a $7 \times 7$ mm rigid PCB with small fine gauge wires connecting from the grasper tip to the analog board (on Figure 2.1).

The main goal of *in-vivo* experiments in [26] was to study responses of female porcine small bowel to ischemia by occluding the blood flow to specific tissues (approved UW IACUC protocol 4201-01). A laparotomy exposed the small bowel and well-vascularized small bowel tissue was squeezed between the jaws of the grasper for 10 s at several constant pressures up to and including 40 kPa (300 mmHg). Previous work [26] showed that 95 % of grasps during minimally invasive procedures were held for less than $8.86 \pm 7.06$ s. Earlier study showed 40 kPa grasp will cause moderate internal tissue damage in the liver. For each grasp, a new section of tissue was used to prevent reperfusion injury from affecting the data. After reading

sensor data at 20 kPa (safe grasp condition) with normal blood flow, some of the mesenteric vessels were clamped for three to five minutes, creating a mild level of ischemia in the tissue. The grasper again acquired data in the low blood flow condition for 10 s. Finally, the grasper measured responses of the watershed (dual-arterial supply) region of the tissue, intermediate between low and healthy blood flow, for 10 s. After these measurements, the artery clamp was removed.

In these early experiments, only the temperature, grasper jaw position, grasper jaw force and pulse oximetry were recorded [26]. In 2014, Sie *et al.* [27, 28] performed additional experiments with the same multimodal grasper at University of Minnesota. Sie's modified design included the double-layered pressure sensing films (4LW, Fujifilm Holdings Corp, Tokyo, Japan) placed on a flat glass base. An Extended Kalman Filter (EKF) was developed to estimate tissue parameters such that stress-strain relationships among all tissue types are distinct. Their *in-vivo* tests were performed with liver, bladder, and gallbladder to validate tissue parameter identification. Two porcine models were used for data collection on two different days.

### *2.3 Other Research with Sensorized Surgical Graspers*

Most of the other researchers, who focused on sensorized surgical graspers, worked on improving multi-axes force measurements. For example, Soakhanvar *et al.* [29] used three uniaxial polyvinylidene fluoride (PVDF) films based on the piezoelectric effect to measure force, the location of concentrated load (e.g. a hidden lump) and the softness of the grasped object. Recent research developed by [30] presents a novel clamping force sensor made from fiber Bragg grating integrated in a manual laparoscopic instrument and tested on both *ex-vivo* tissue and *in-vivo* porcine liver. Other surgical instruments use strain gauges for direct [31] and indirect [32] force measurement. Other researchers focused on environment influence of force sensing. For example, Seok et al. [33] worked on a multimodal approach to build sensorized-forceps. Their work mostly focused on compensating for environmental influences (e.g. temperature and humidity change) that affect force measurements.

More recent research [34] developed and tested a sensor-equipped surgical grasper with a slip sensor on the tip. The slip sensor includes four thermistors surround a heater and can maintain at constant temperature, generating a thermal gradient in an object contacting the sensor. When the object slips, the thermal gradient shifts over the thermistors. Their sensor set could be integrated with a *da Vinci* surgical system. This group experimented with 5 types of porcine *ex-vivo* tissue to test moisture related grasp slip. They were able to detect small (2.1 mm) slips at a 92 % true positive rate.

### *2.4 Multiple Sensor Data Fusion*

Sensor data could be represented in different ways, such as an RGB image, spectrum image, time series signal, or windowed signal. When multiple different sensor modalities are used, the data could be fused by time stamp, some event (e.g., grasp time), and action. There are different ways to fuse the data. One of them is the classical way – when the distinct features are extracted manually based on physics and known correlations. For example, the time of flight of an acoustic echo could be combined with the distance that the echo travels, and we can obtain the velocity of sound as a feature. Other features are the signal's maximum amplitude, frequency components, phase, slope angle, etc. After extracting the features, they could be concatenated and used in machine learning or probabilistic algorithms for classification tasks. Another way to fuse the data is to use deep learning methods. Deep learning can learn hundredth features based on patterns of many signals. Both approaches, classical and deep learning, have advantages and disadvantages. The classical method's advantage is that it does not require an extensive dataset to perform classification tasks. However, the classical method suffers from low accuracy for multi-class problems due to the low number of potential features we can extract. The deep learning approach has the advantage of learning the hundredth features. Still, it requires an extensive dataset with various examples and can suffer from measurement biases and overfitting solutions.

### *2.5 Deep Learning for Sensor's Fusion*

Deep Learning (DL) is currently used for classification, noise reduction, and trend prediction tasks. Combining the data from different sources of information could be beneficial. Multimodal deep learning is still a new field, and work is mainly done to combine images with audio, text with audio, and text with both images and audio. The combination of multimodal time series data and DL algorithms is still an ongoing research area due to needing more available datasets for training. Time series training based on physiological signals is even more valuable because of the potential health benefits for the community, such as faster diagnosis, rehabilitation, and treatment. Data fusion from different modalities aims to understand what features are complementary to each other and which ones are redundant or cooperative in different modalities [35].

The most common neural networks for time series multimodal data learning are Recurrent Neural Networks (RNN), Fully Connected Neural Networks (FCNN), Convolutional Neural Networks (CNN), Graph Neural Networks (GNN), Autoencoders, Transformers, and their variations. Different challenges are raised when multimodal biomedical data is used, such as small dataset size, missing data from modalities, an imbalance of dimensionality between modalities, and what data fusion technique to use. Early layers of these models learn simple abstractions of the time series signal when the deeper layers are combined into abstract representation [35].

Several different fusion techniques exist, such as early fusion, intermediate fusion and late fusion of data. At early fusion, the data is fused before training, and the model is trained already on fused data. At intermediate fusion, the data is trained separately for different modalities and then fused in the middle of the model to continue training on fused data. Late fusion is when the modal is trained separately on different modalities, and the data is fused at the end of the training for decision-making [35]. These strategies have advantages and disadvantages, and the choice of which fusion strategy to use depends on the signals. For example, some signals are combined before the training, while others make sense to train

separately and then combine. Intermediate and late fusion strategies could be used when using different models for different modalities, e.g., CNN is better work for images, and RNN is better for audio.

# Chapter 3

# DESIGN

## *3.1 Requirements*

Smart Grasper is a surgical instrument, and it must work in an operating room for *in-vivo*, *in-situ*, and *ex-vivo* abdominal surgeries. Because of that, there are requirements the instrument should satisfy.

- Sensor set should be compact in order to fit on Grasper's jaws and protected from moisture for *in-vivo* experiments.
- The full instrument's electronics should be compact and robust to work in the operating room, and to fit into a surgical cart.
- The instrument should be powered from the wall socket connection or battery to accommodate the operating room requirements and work robustly.
- The data acquisition needs to work fast to capture the data from ultrasound, optical, and bioimpedance sensing during the 10 s of grasp time.
- The sensor set should be cheap enough to be a disposable after each surgery.

## *3.2 Grasper's Hardware Architecture*

### *3.2.1 Comparison with Original Design*

The original design of Smart Grasper that Roan developed in his Ph.D. Dissertation [4] had the following sensors: thermistor, 4-electrodes bioimpedance, 4-wavelength optical sensor, in-direct force sensing, and jaw position. However, Roan's sensor set and Analog Front

Table 3.1: Comparison of a power and motor control with Roan's design.

| **Subsystem** | **Roan's Design** | **New Design** |
|---|---|---|
| Power | ±15 V for all subsystems | +15 V for motor; 3.3/5 V for sensors |
| Motor Control | Fuzzy Controller | PID Controller, force sensor in feedback loop |
| Motor | RE25-10W, Maxon | RE25-10W, Maxon |
| Motor Hardware Controller | LSC 30/2 Servo Controller, Maxon | ESCON 50/5 Servo Controller, Maxon |

End for processing the signals were damaged during the travel and storage. The Smart Grasper requires significant system rework. In the current research, I extended the work by adding a 1-D piezo element (ultrasound sensor) to one of the Grasper's jaws. The ultrasound transducer occupied one of the Grasper's jaws, and due to restrictive space, the other sensor's modalities were reworked to accommodate this change. The Bioimpedance sensor's capability was reduced to 2 electrodes; optical sensing was decreased to 2 wavelengths (RED and IR). Also, I decided to pivot from pure analog design to mixed-signal design to simplify the development and integration of the sensors with the motor control. The changes for power, motor control, and sensor set were summarized in Table 3.1, Table 3.2 and Table 3.3.

The instrument was powered from ±15 V in Roan Grasper's design, including the sensor set, peripheral electronics, filters, operational amplifiers, and MOSFETs. It required costly components, such as operational amplifiers OPA128 (obsolete in 2024), to power optical sensing and amplify the small signals from photodiodes. However, we decided to use different power rails to simply the development and take an opportunity to use mixed signals circuitry for development: one +15 V for powering the motor, motor controller, and force sensor's amplifier, and a second +5 V for the encoder's buffer, bioimpedance, optical sensing, and

+3.3 V for the thermistor.

Also, Roan's design had an in-built Linux-based computer for the Grasper to display and save the signals. In Roan's design, the user is supposed to push the button to open and close the Grasper without predefined grasp time. In contrast, we used a regular Windows laptop to control microcontrollers and systems on a chip to streamline the design and speed up the development. We utilized the predefined 10 s grasped time to ensure we collected the data for a specific time window. Timely framed data is also easier to input into deep learning models.

### *3.2.2 New Smart Grasper Design*

Smart Grasper was significantly reworked based on the previous section 3.2.1, and several different solutions were tried for each modality before the final prototype and integration of sensors in hardware and software. Each modality is required to calibrate and use analog and digital signal processing in order to obtain good signal integrity for small biomedical signals. The Grasper's architecture is presented in Fig. 3.1.

Grasper hardware is presented on Fig. 3.2, where Grasper shown (a) on test bench; (b) Grasper is in use; (c) and (d) are sensor sets developed for Grasper's jaws.

**Force and Distance Measurements** Force sensing is crucial for avoiding tissue trauma during grasping in abdominal surgeries. Excessive forces will cause tissue to be traumatized, and insufficient forces can lead to instrument slippage reducing operation efficiency and endangering patient safety [36, 20]. Conventional laparoscopic surgical instruments still lack force sensing, making diagnostic quality tactile feedback infeasible [2, 3, 36].
The mechanical hardware was mostly unchanged from Brown's design [37]. Fig. 3.3 demonstrates how the motor (RE25-10W, Maxon) with attached 19:1 planetary gearbox (GP26, Maxon) and encoder (HEDL55, Maxon) opened and closed the Grasper tip's jaws. While the Grasper's mechanical hardware remained untouched, a few changes were made to the sensing and controls. An updated motor controller (2018 ESCON 50/5 4-Q Servocontroller,

Table 3.2: Sensors comparison with Roan's design.

| **Subsystem** | **Roan's Design** | **New Design** |
|---|---|---|
| Force | FR1010, 40 lb, FUTEK | FR1010, 40 lb, FUTEK (LCL-020, 20 lb, OMEGA) |
| Position | HEDL55, Maxon Encoder | HEDL55, Maxon Encoder |
| Temperature | QT06002-532 Quality Thermistor | NTC Thermistor 10 kΩ (NCP15XH103J03RC) |
| Bioimpedance | Mill-Max Manufacturing Corp. Gold Electrodes | Mill-Max Manufacturing Corp. Gold Electrodes |
| Optical | Photodiode PDB-C152SM, Adanced Photonix; Epitex SMC810, 810 nm Infrared LED; LTST-C190KGKT RED LED, Lite-on Electronics; LTST-C190KGKT, Lite-On Electronics Green LED; LSTS-C190TBKT, Lite-on Electronics Blue LED | Würth Elektronik Red 630nm LED (155060SS75300), IR Emitter 850 nm (15412085A9000); ams-OSRAM USA INC. Photodiode 850 nm (BPW 34 S-Z) |
| Ultrasound | Not implemented | Steminc SMD063T07R111 3 MHz piezo-element |

Maxon) is controlled via microcontroller (Teensy 4.0) by Pulse-Width Modulation (PWM) in current control mode. The relationship between the PWM input and the command current was linear but varied based on controller setup parameters. For all the data gathered in this dissertation, the motor controller was set to output anywhere from 0 to 400 mA of current. A quadrature encoder buffer breakout board (LS7366R, SuperDroid Robots) keeps track of encoder counts at 40 MHz.

Table 3.3: Comparison of peripheral electronics with Roan's design.

| **Subsystem** | **Roan's Design** | **New Design** |
|---|---|---|
| Force DAQ | CSG110, FUTEK amplifier | CSG110, FUTEK amplifier |
| Temperature Driver | Wheatstone Bridge; INA128U, Burr-Brown op-amp buffer | Wheatstone Bridge ($R = 10\,\text{k}\Omega, 10\,\text{k}\Omega, 470\,\text{k}\Omega$) |
| Bioimpedance Driver | Oscillator AD9832, Analog Devices; LM318D, National Semiconductor high speed amplifier | AD5933 SoC driver, Analog Devices |
| Optical Driver | AD549, Analog Devices; OPA 128, Burr-Brown; LM324, National Semiconductor | MSP430FR2533 Smart Combo, MOSFET drivers for LEDs |
| Ultrasound Driver | Not implemented | TDC1000 as a driver, Analog Discover 2.0 as data acquisition and 6 MHz clock |

Only one of the internal force sensors (FR1010, 40 lb, FUTEK) was used because of damage to the second's wiring during storage, but no need could be found to have two load cells besides redundancy. After amplification (CSG110, FUTEK), a 10-bit analog-to-digital converter (ADC) mapped the 0 to 5 V force sensor output to discretized values (DV) ranging from 0 to 1023.

During the data collection and experiments at UW Harborview Medical Center in 2023, force sensor FR1010 got broken, and substituted with OMEGA LCL-020 Full Bridge, Thin Beam Load Cell, 20 lb. The FUTEK amplifier was used the same, and mapping of DV values was repeated for a new force sensor.

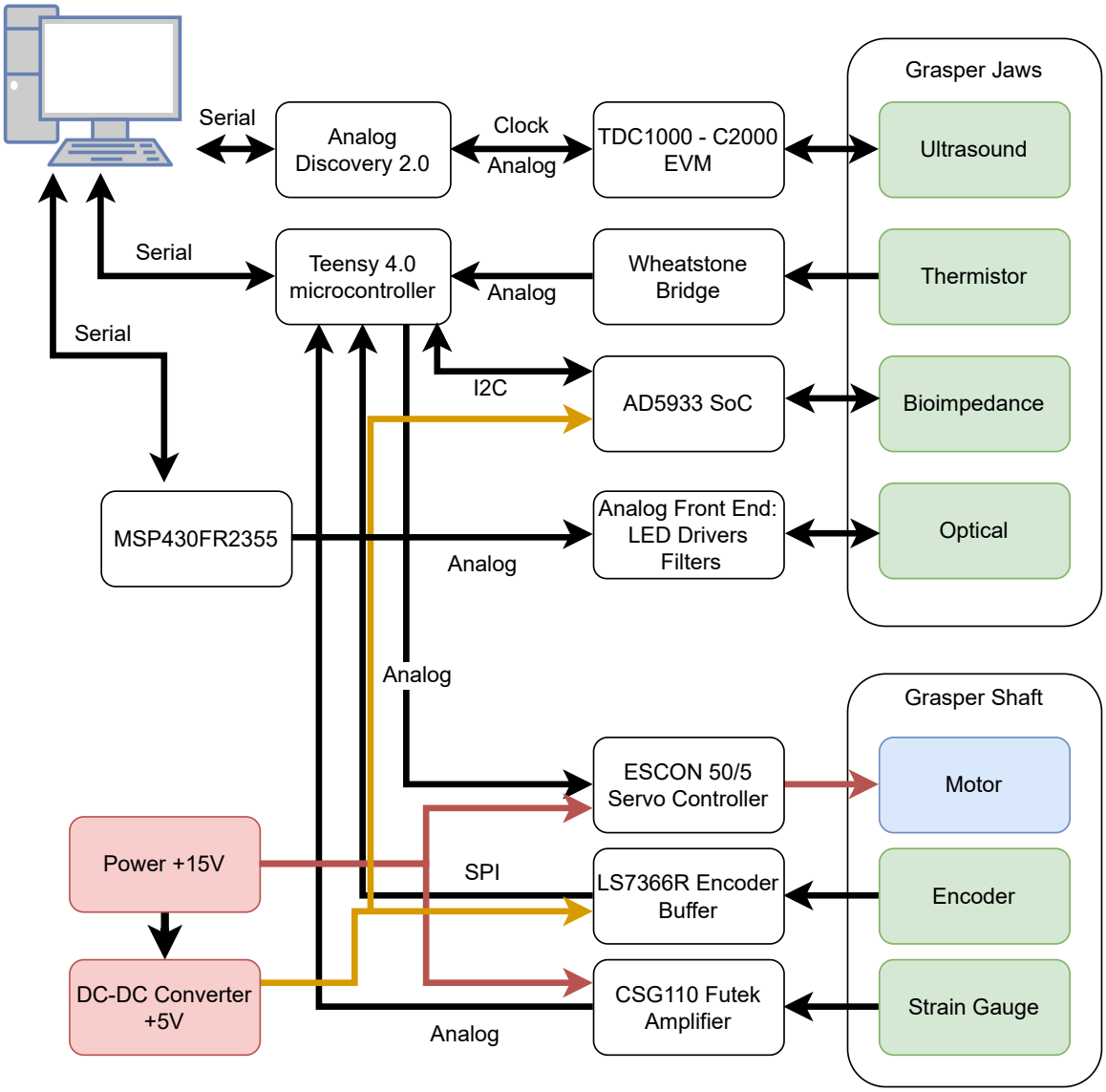



Figure 3.1: New Smart Grasper architecture.

**Temperature** is measured by a thermistor – a temperature-dependable resistor. After experimenting with both a voltage divider and a Wheatstone bridge for temperature-dependable resistor – NTC Thermistor 10 kΩ (NCP15XH103J03RC), the favor was given to

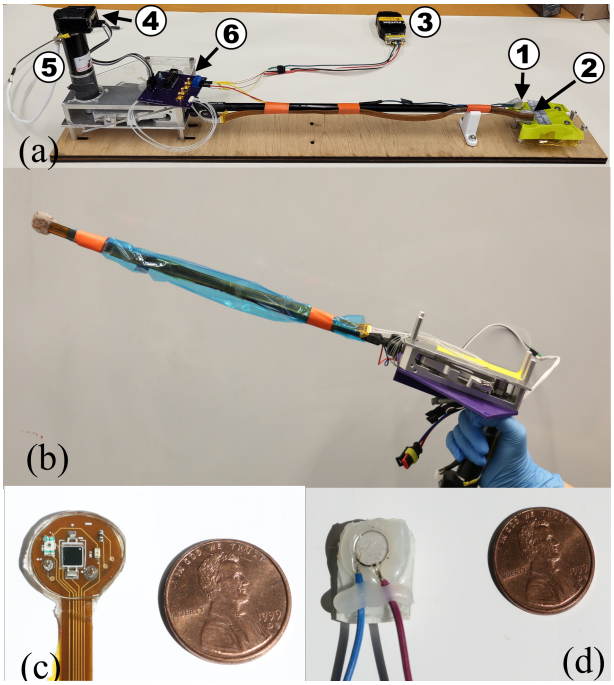



Figure 3.2: (a) Smart Grasper instrument inverted on a test bench: (1) 3 MHz Steminc SMD063T07R111 1D ultrasound transducer, (2) NCP15XH103J03RC 10 kΩ thermistor, (3) Futek amplifier for FT1010 force sensor, (4) Maxon HEDL-5540 encoder, (5) Maxon RE25-10W motor, (6) Futek FT1010 40 lb force sensor. Motor body serves as a handle for the user. (b) The Smart Grasper Instrument in use. (c) Sensor set on flex pcb with epoxy cover to protect sensors from moisture. It includes: thermistor, IR and RED LEDs, two electrodes and photodiode. (d) 3 MHz Steminc SMD063T07R111 1D ultrasound transducer was epoxied to protect it from moisture.

a Wheatstone bridge configuration due to the simplicity to use and precision; the schematic is presented on Figure 3.4. The calibration procedure for the thermistor is presented in Section 4.4. A Wheatstone bridge circuit has three resistors (10 kΩ, 10 kΩ, 470 kΩ), and a 10 kΩ thermistor is the fourth temperature-dependent resistor in the bridge. Two points (ADC1 and ADC2) on schematic, Figure 3.4, are where the microcontroller's ADC is plugged in to get a differential signal and convert it to temperature in Celsius based on calibration.

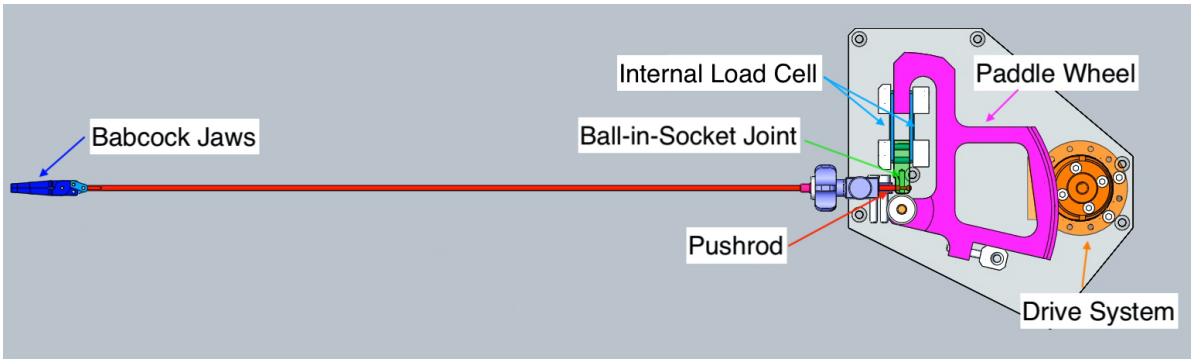



Figure 3.3: The Smart Grasper (formerly MEG) without its top cover. The drive system turns a paddle wheel using a steel cable (not shown). The paddle wheel it attached to a ball-in-socket joint that moves the pushrod back in forth, which opens and closes the Babcock jaws. Readings from the internal load cell record the force on the pushrod.

**Ultrasound** measures acoustic wave propagation through media. In our case, media are different types of soft tissue or fat. We used a 1D bidirectional Steminc SMD063T07R111 piezo-element as a sensor for transmitting and receiving the acoustic waves, shown on Fig. 3.2. To drive the piezo-element with a 3 MHz carrier frequency, we used TDC1000 System-on-Chip (SoC) with evaluation board C2000EVM, shown on Fig. 3.5. An Analog Discovery 2.0 portable oscilloscope (Fig. 3.5) is used in Smart Grasper to obtain raw ultrasound waveforms and provide an external 6 MHz clock to the C2000EVM. 6 MHz clock was provided due to internal clock divider in TDC1000 SoC. We used three test points at TDC1000-C2000 EVM: COMPIN, START, and EXT CLCK. COMPIN collects the buffered raw echo ultrasound data. START pin is to collect the start impulse of the TDC1000 to know the time when exactly the transducer was excited. Lastly, EXT CLCK is the external clock, 6 MHz square wave generated by Analog Discovery 2.0 as a clock signal. Internal clock wasn't utilized because it was only 8 MHz in frequency, and we had 3 MHz transducer attached to the Grasper's jaws. Analog Discovery 2.0 has API available and could be programmable in Python, making the integration procedure simpler. The calibration procedure and methods

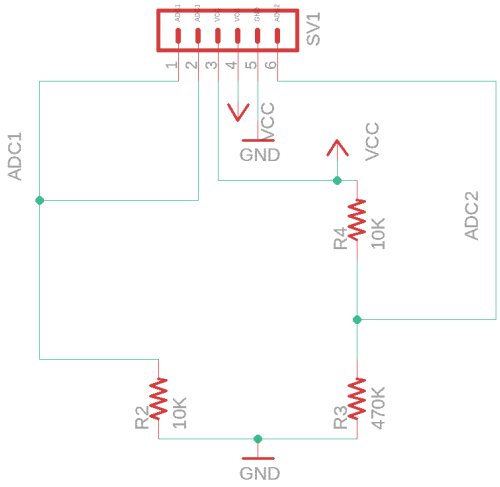



Figure 3.4: Wheatstone bridge for thermistor measurements.

for denoising ultrasound and TOF estimation are presented in Chapter 5.

**Bioimpedance** is a measurement of the ability of a material to create an opposition to the flow of electric current through the tissues and the impedance. Impedance can be presented mathematically as a complex magnitude in Cartesian coordinates comprising two components – Resistance and Reactance [38]. In the current design of Smart Grasper, we use only two electrodes, driving it by Analog Devices AD5933 SoC, embedded in Digilent PIMODIA breakout impedance board (Fig. 3.6). Both electrodes are used to inject the electrical current and measure the tissue's response. In Roan's original design, four electrodes were used instead, where two electrodes were used to inject electrical current, and the impedance measured by a second pair of electrodes.

**Optical Sensing** is a sensing technique based on the optical response of absorption and reflection of the wavelength of tissue. The optical sensing circuit consists of two LEDs (red and infrared) and a photodiode, two stages of operational amplifiers for the photodiode, and a driver MOSFET circuit for LEDs. In the original work, Roan used four different wavelengths (red, infrared, green, and blue) to measure pulse and oxygen level, and two photodiodes, one

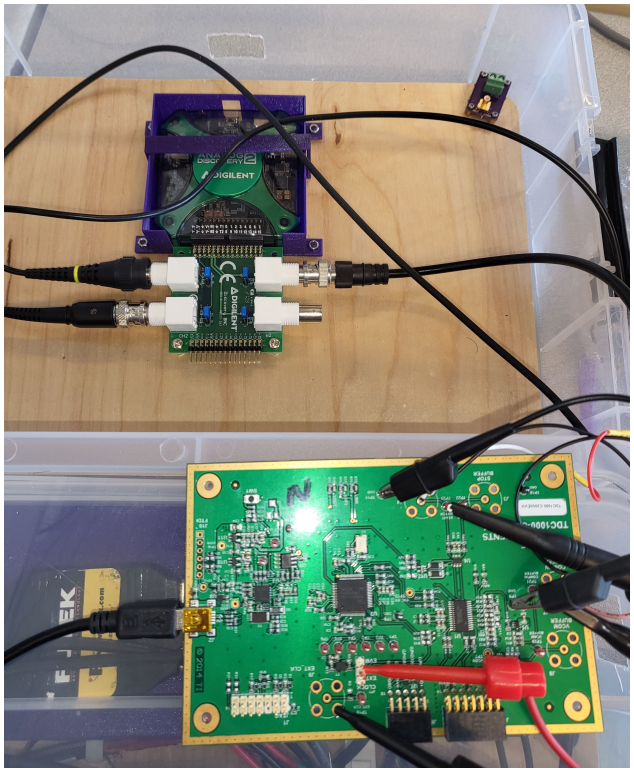


Figure 3.5: Ultrasound data acquisition setup. TDC1000-C2000 EVM (on the bottom of the Figure) hooked up with probes to the Analog Discovery 2.0 portable oscilloscope (on the top of the Figure).

for infrared wavelength only. In a new Smart Grasper, we decided to simplify the development and save space using only two wavelengths - red and infrared (IR), and one photodiode with a wide range to obtain scattered optical signals from different wavelengths. Roan developed the whole signal processing circuitry for pulse oximetry from scratch, utilizing only analog components. However, in the current design, we decided to pivot from the original design and use the advantage of Texas Instruments MSP430FR2355 Smart Combo SoC. MSP430FR2355 allows the use of operational amplifiers, digital-analog converters (DAC), and analog-digital converters (ADC) in a convenient way to collect data and amplify from the photodiode. We used a combination of developed PCB (Analog Front End PCB is presented in Appendix


Figure 3.6: AD5933-based PIMODIA breakout board. Pins on the left are connected to Teensy 4.0, and coax cables on the right connected to the electrodes.

D) with MOSFETs, filters and DAC, operational amplifiers from MSP430FR2355 to drive LEDs and obtain signals from the photodiode.

### *3.3 New Grasper's Software Architecture*

The Grasper software architecture for control and sensing has two levels. However, this architecture doesn't include post-processing signal processing and deep learning part, which described in the following Section 7.1. First level is low level, which is programmed in C and C++ for the Teensy 4.0 microcontroller and MSP430FR2355 LaunchPad. According to the Fig. 3.1, Teensy 4.0 is responsible for the most functionality of the Grasper, such as motor control, thermistor, force, position and bioimpedance data collection.

Teensy 4.0 was programmed with using Arduino 1.8 IDE with Teensy 4.0 plugin. It gives an advantage to use Arduino libraries which are easy to use, and combine. And Teensy 4.0 has a better functionality (clock speed and memory) then Arduino microcontrollers.

Texas Instruments MSP430FR2355 (MSP430) LaunchPad has in-built microcontroller to program and control MSP430. It was programmed on C using TI Code Composer Studio 12.6.0 IDE, where MSP430 library needs to be installed.

The second level of software has high level code, written on Python 3.11 in Anaconda

IDE on the Windows computer. High level software combines the data acquisitions from multiple different devices with time stamps, and save it in appropriate format and folders on the computer. High level code combines the following functionality: Analog Discovery 2.0 portable oscilloscope to drive and obtain data from ultrasound, Teensy 4.0 board data collection from sensors and motor control, MSP430FR2355 Smart Combo LaunchPad to obtain data from photodiode and to drive LEDs.

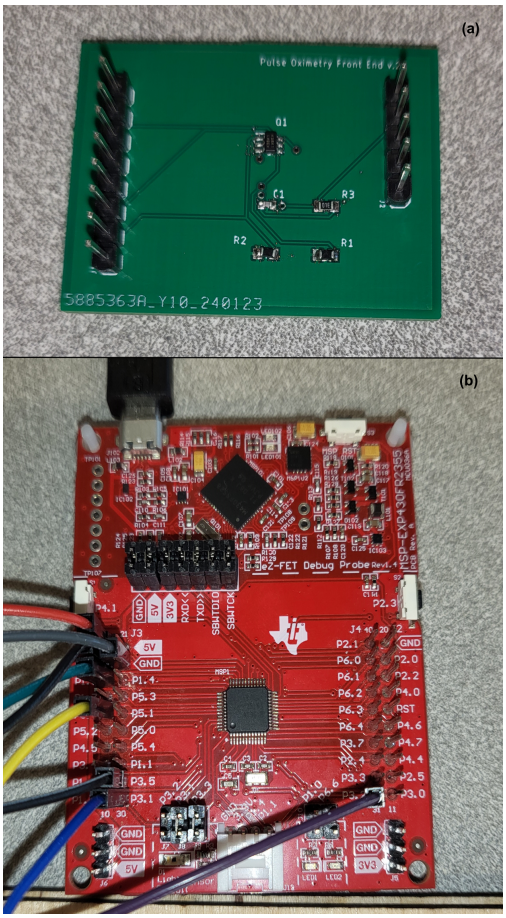



Figure 3.7: Hardware Setup for LEDs and a Photodiode: (a) analog front end with LED drivers and filters for a photodiode, (b) MSP430FR2355 Smart Combo LaunchPad with operational amplifiers, analog-to-digital convers, and digital-to-analog converters.

# Chapter 4

# CALIBRATION AND SIGNAL PRE-PROCESSING

This chapter describes the calibration procedures and signal pre-processing for the force sensor, encoder (position sensor), thermistor (temperature sensor), bioimpedance, and optical sensing. Ultrasound sensor is placed in separate Chapter 5 due to development of new algorithms for filtering 1-D ultrasound from a ringdown artifact.

The work in Sections 4.1, 4.2, 4.3 was done together with MS ME students Jack Kaplan and Sheryar Hassain and published in the proceedings at ISMR Conference [39].

## *4.1 Force Calibration*

The test fixture consisted of a wooden base with screw points to attach the Grasper and the smaller calibration subassemblies for force and position calibration.

The force calibration fixture (Figure 4.1) was attached to the base and consisted of an aluminum plate connected to the external load cell (TAL220B, 5 kg). The upper Grasper jaw engaged with the plate and the lower jaw engaged with a 3D printed base, as seen in Figure 4.1. Raising the external load cell and lowering the 3D printed piece increased the jaw distance. Output from the external load cell was routed through a breakout board (HX711, Sparkfun) with an amplifier and a 24-bit ADC before going to the microcontroller. Calibrating the relationship between the external load cell digital values (DV) and forces measured in Newtons required applying known forces to the load cell and recording the measurement. We hung masses from a 3D printed fixture. The equation 4.1 for correlation between external load cell and Grasper's strain gauge:

$$F(z) = -5.63 \times 10^{-3} + 4.46 \times 10^{-5} \times z \tag{4.1}$$

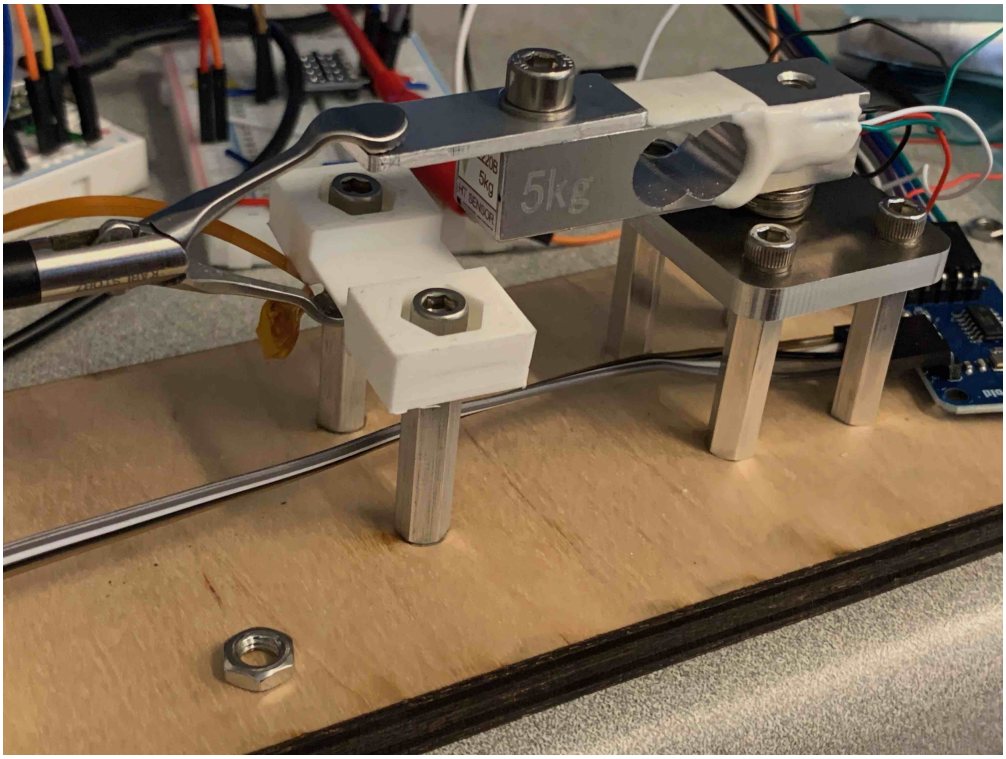


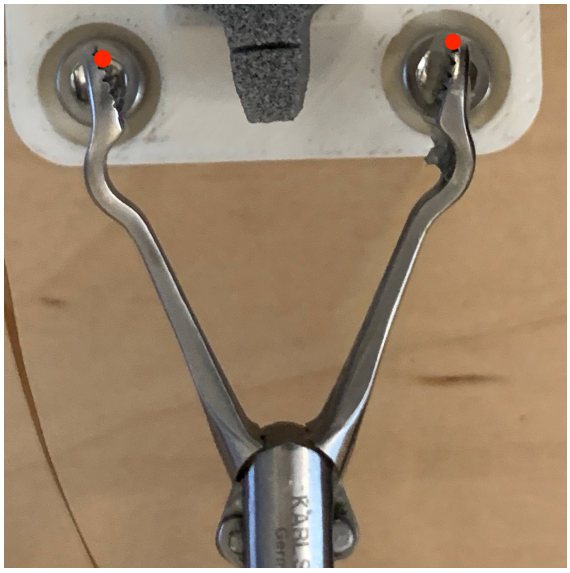

Figure 4.1: Force calibration fixture with external load cell. HX711 breakout board for load cell interfacing is not shown. The external load cell and the 3D printed piece were raised and lowered by adding and removing washers.

where $z$ is the force reading from the external load cell measured in DV and $F$ is the force measurement in Newtons.

### *4.2 Position Calibration*

Being able to accurately measure the distance between the Grasper's jaws is crucial to combining data from other planned sensor modalities into a cohesive picture of the tissue being grasped. In an infinitely stiff grasper, there would be a direct relationship between the motor rotation and jaw distance. However, stretching and bending of the mechanical components add compliance between the motor angle (measured by an encoder) and the jaw

(a)



(b)

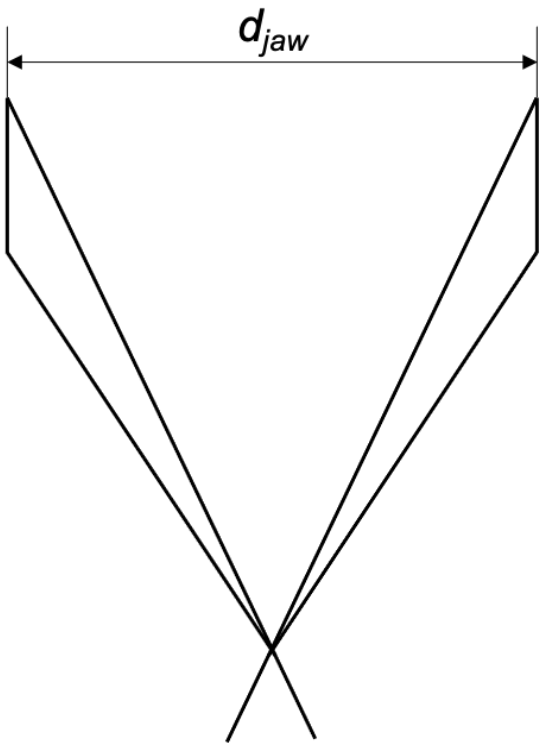



Figure 4.2: Jaw distance. Note that the distance between the red dots in (a) corresponds to $d_{\text{jaw}}$ in (b).

distance ($d_{\text{jaw}}$ in Fig. 4.2). Therefore, the jaw distance is a function of both motor angle and applied force (Eq. (4.3)). Deformation and slack in the system will vary with force and the state of the cable. The position calibration procedure began by putting a block with known width in the fixture (Fig. 4.3). The Grasper was then fed a current causing it grasp the block. Calipers were used to measure the actual distance between the jaws to the nearest 0.1 mm. The current to the Grasper was increased, and measurements from the internal force sensor and the motor encoder during this process were recorded. The process was then repeated for every block. Each run of the calibration process can record data for multiple distances. The encoder was zeroed by feeding the Grasper a current that causes the jaws to open until they

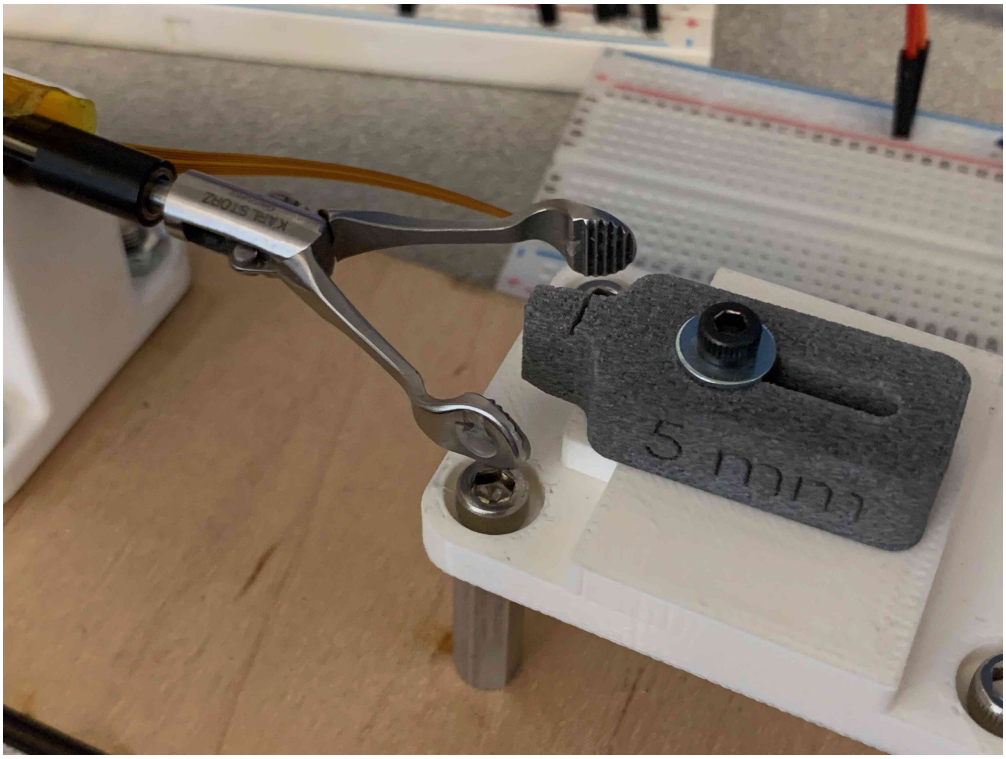



Figure 4.3: Position calibration subassembly with 5 mm block.

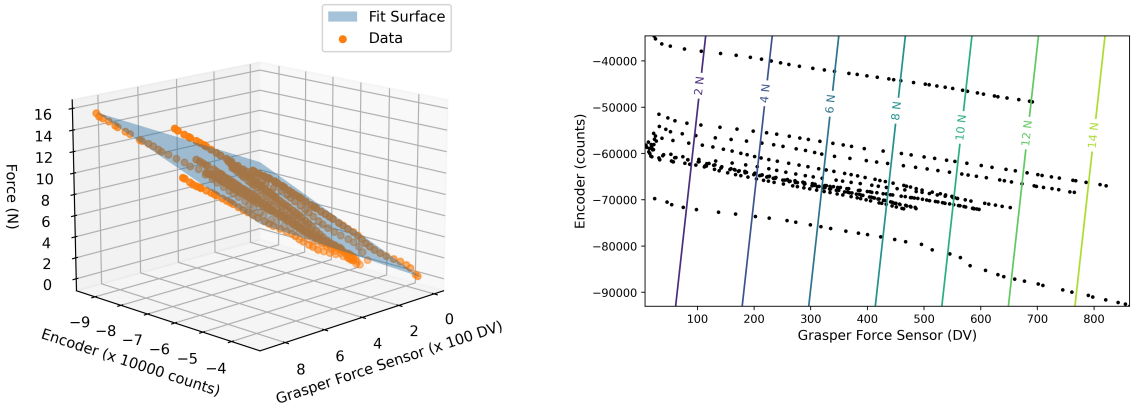
reach a mechanical stop. The point at which the jaws cannot open any wider is made the zero point. This was done at the start of each run and every time a new block was put into the position fixture. Zeroing the encoder every time a new block was put into the position fixture prevented error accumulating when manually aligning the jaws to the measurement line. However, during the animal experiments, described in Chapter 6, we found out that the best way to zeroing the encoder is on closed position and not open position, because it's more reliable.

### *4.3 Force and Position Fusion*

The Python 3.8 function `scipy.opimize.curve_fit` was used for all surface fitting. Both quadratic and linear surface fits were tried for the position and the force fit.

(a) Measured data in 3D

(b) 2D contour plot




(c) Distribution of the error between the measured and calculated force for a given force-encoder pair

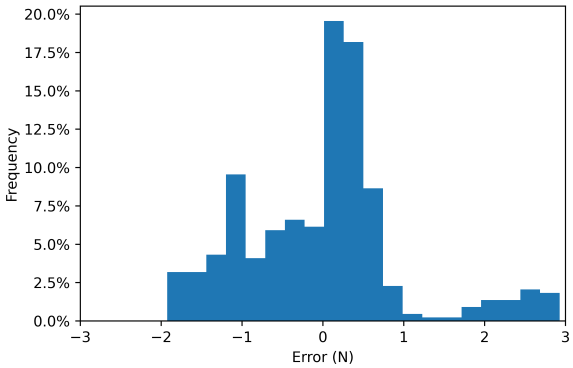



Figure 4.4: Planar surface fit for force calibration.

The purpose of this regression analysis is to find a calibration function that takes the sensor readings as inputs and outputs a value close enough to the true value to be usable. Both overfitting and underfitting optimization curves can reduce their predictive power. Given the relatively small size of the data sets, the risk of overfitting is greater than the risk of underfitting. Therefore, even though the quadratic better fit the current data, the linear fit was chosen. The improvement was not enough to justify the increased risk of overfitting. The best fit planar surface for the force calibration is defined by:

$$F(x, y) = \alpha + \beta x + \gamma y \tag{4.2}$$

where $x$ is the internal force sensor reading in DV, $y$ is the encoder reading in counts. The values for $\alpha$, $\beta$, and $\gamma$ can be found in Table 4.1 and the resulting surface can be seen in Fig. 4.4. This fit has an RMSE of 0.980 N and a mean absolute error (MAE) of 0.721 N. Given the quality of the fit found in the previous section, Eq. (4.1) was treated as true for the purpose of determining the error of the final planar fit.

Table 4.1: Grasper Force Reading to Jaw Tip Force Coefficient Values

| Coefficient | Value | Units |
|---|---|---|
| $\alpha$ | $-0.484$ | N |
| $\beta$ | $1.70 \times 10^{-2}$ | N per Grasper DV |
| $\gamma$ | $-1.54 \times 10^{-5}$ | N per encoder count |

The error between the measured and calculated force values exerted by the Grasper's jaws appears to be approximately normally distributed around zero (the actual calculated mean error is $-5.11 \times 10^{-9}$ N with a standard deviation of 0.980 N. The small value for $\gamma$ in equation (4.2) indicates that the encoder value plays little role in determining the force applied at the Grasper tip.

Using 160 kPa as the upper limit on pressure at the jaw tips as suggested in [20] limits the maximum applied force to 9 N. Heijnsdijk *et al.* found the minimum force to securely

grasp tissue to be approximately 3 N for surgical graspers with similar profiles [36]. These bounds leave a 6 Neffective operating range, and large enough window even with a standard deviation of approximately 1 N. If future work requires more precise force measurements then either the calibration process needs to be refined or a better method of force sensing needs to be implemented.

For position surface fit, we can use the following equation:

$$d(x, y) = k_0 + k_1 x + k_2 y \tag{4.3}$$

where $x$ is the internal force sensor reading in DV, $y$ is the encoder reading in counts, and $d$ is jaw distance in millimeters. Values for the $k_0$, $k_1$, and $k_2$ coefficients can be found in Table 4.2. The fit has an RSME of 0.742 mm and a MAE of 0.575 mm. The error between

Table 4.2: Position Coefficient Values

| Coefficient | Value | Unit |
|---|---|---|
| $k_0$ | 80.3 | mm |
| $k_1$ | $1.88 \times 10^{-2}$ | mm per force sensor DV |
| $k_2$ | $1.07 \times 10^{-3}$ | mm per encoder count |

the measured and calculated position values for a given set of force and encoder readings (Fig. 4.5c) appears to be approximately normally distributed around zero (the calculated mean error is $-1.03$ mm. This implies that equation (4.3) can be rewritten as:

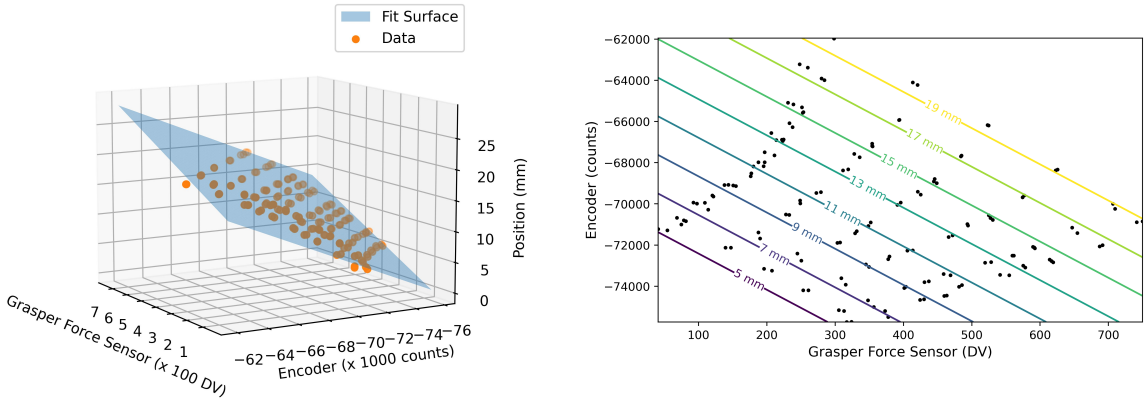
$$d(x, y) = k_0 + k_1 x + k_2 y + \epsilon(\mu, \sigma) \tag{4.4}$$

where $\epsilon$ is a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.743$ mm. The formulation in equation (4.4) can then be propagated forward when evaluating sensor readings with position-dependent modalities such as ultrasound.

Roan [4] use the linkage geometry to get a set of equations relating the force measured by the internal load cell to the force at the jaw tip. These equations do indeed depend on the

(a) Measured data in 3D

(b) 2D contour plot




(c) Distribution of the error between the measured and calculated jaw distance for a given force-encoder pair

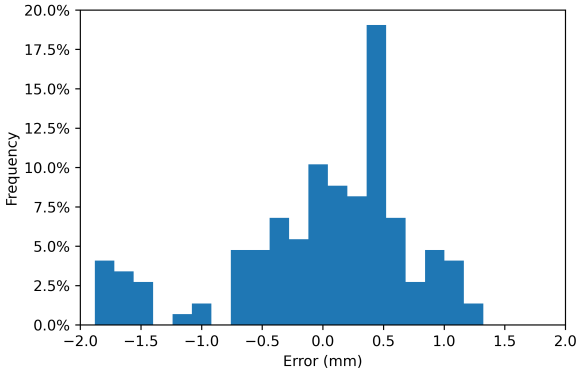



Figure 4.5: Planar surface fit for position calibration.

jaw angle $\theta$ which is directly related to the jaw distance $d_{\text{jaw}}$. However, these equations do not account for disturbances in the system such as backlash, friction, and deformation. Roan *et al.* add an "adjusted jaw displacement factor" to better fit their equations to the data, but they attribute this fudge factor only to "measurement errors in link length" [4]. Rather than try to model all possible deviations from the perfectly rigid, frictionless environment assumed in [4], we used the data generated in previous section to determine the relationship between the applied force and the measured force. Our standard deviation of 0.99 N is worse than the 0.30 N standard deviation claimed in [4]. Nonetheless, we believe our results are a more accurate reflection of the precision possible in the system.

Like [4], we use regression to determine an equation relating the measured force and encoder values to the jaw distance. Our measured standard deviation of 0.743 mm is larger than Roan's reported standard deviation of 0.2834 mm. Nonetheless, we believe our results are more robust for current state of the Grasper. Whereas we used a machined fixture to collect position data, while Roan *et al.* used cardboard "structures" to set the jaw distance for calibration [4]. Accurately measuring the jaw distance without a proper calibration fixture is difficult, so there are likely meaningful measurement errors that are unaccounted for in the stated standard deviation.

### *4.4 Temperature*

The thermistor was calibrated with a hot bath experiment, where the Grasper's sensor set and a Fluke multimeter's thermocouple were placed inside a container of hot water and then the resistance was measured by the Wheatstone bridge configuration. The Steinhart–Hart equation was used to convert resistance to temperature:

$$1/T = A + B \times ln(R) + C \times (ln(R))^3 \tag{4.5}$$

where $T$ - temperature in Kelvins, $A$, $B$, $C$ - the Steinhart–Hart coefficients and defined in Table 4.3 derived by fitting equation (4.5) to the calibration resistance and measured temperature values. Then, the temperature obtained was converted from Kelvins to Celsius.

Table 4.3: The Steinhart–Hart coefficients

| | |
|---|---|
| A | 0.00187 |
| B | 0.00011 |
| C | $6.021 \times 10^{-7}$ |

A uniform mean filter with size 25 samples was applied to the thermistor data to smooth the signal and to reduce the high frequency noise due to oversampling.

### *4.5 Electrical Impedance*

Analog Devices AD5933 SoC was used to measure bioimpedance. It must first be calibrated to a known impedance to define the gain factor. The impedance should be approximated to the measured impedance range, otherwise the SoC will have a big error. The resistance chosen for calibration was 512 Ω. The gain factor is 0.0000045, adjusted to reduce the error and compensate for parasitic capacitance. The frequency swap was used from 30 kHz to 70 kHz with a step of 1 kHz. Electrical impedance was verified with two resistor circuits: 328.8 Ω and 1 kΩ, shown in Figure 4.6. The distance between pads was designed so the Grasper could grasp them and the impedance could be verified. On the other hand, Roan, in his work, used resistor networks on an Analog Front End board and not on Grasper's tip, which does not adjust to the parasitic capacitance and can cause significant errors in the impedance measurements because there is a long wire (approximately 1 m) between Grasper and the processing electronics box.

Measured values from the Grasper's bioimpedance system for the resistor 328.8 Ω are in the range of $367.13 \pm 3.27\,\Omega$, and give average error of 11.95 % across all of the measured frequencies. Measured values from the Grasper's bioimpedance for the resistor 1 kΩ are within range of $1094.31 \pm 28.83\,\Omega$ with average error of 9.67 % across of all the measured

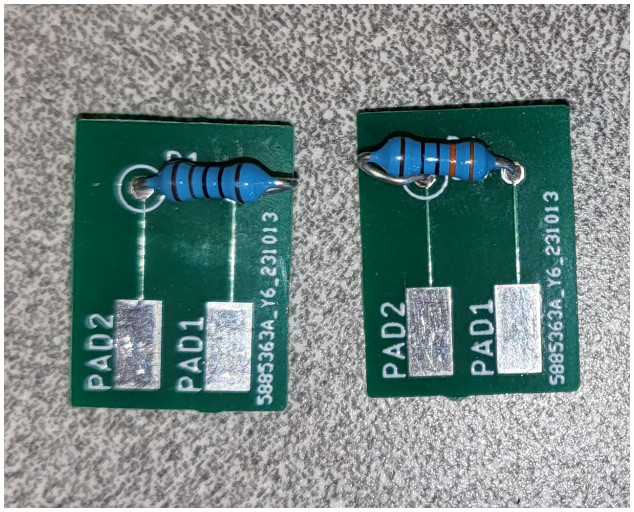



Figure 4.6: Two resistor circuits for impedance verification.

frequencies.

### *4.6 Optical Spectroscopy*

We verified the optical system based on the pulse measurement with the commercial pulse oximeter Innove Deluxe IP900AP. Pulse oximetry is a non-invasive technique that measures changes in blood volume to produce a photoplethysmogram (PPG) signal. This signal is then processed to approximate heart rate. Pulse-oximeter produces this PPG signal by leveraging a photodiode with different colors of LEDs. The LEDs emit a light with specific wavelengths relating to the color of the LED. These waves then transmit through the skin of the subject. However, not all the waves can transmit entirely through the subject's skin. Part of the waves reflect towards the photodiode, where a current signal is generated. Commonly, this current signal is put through two amplification stages to produce the PPG signal, like in our case of utilizing MSP430 SoC. First, the current signal is put through a trans-impedance amplifier (TIA). The TIA converts the current signal into a voltage signal. Then, the voltage signal is inputted to an inverting gain amplifier. It is necessary to observe small changes in the signal that are needed to approximate heart rate and oxygen level. In current work, we

could only verify heart rate. However, we needed special expensive equipment to verify the oxygen level.

The algorithm for approximating heart rate relies on counting the number of peaks of the PPG signal over a specified time frame. Since PPG measures the change in blood volume, the signal has a periodic component with a frequency that matches the subject's heart rate. The frequency is then used to approximate the subject's heart rate.

The high frequency noise of the PPG signal makes it impossible to determine the heart rate without filtering. The low-pass filter was applied to the PPG signal with a cut-off frequency of 2 Hz. It helps to clear the signal from high frequency noise, and actual peaks of the signal is more visible to count them in specific time frame, 60 s. It then compared with a commercial pulse oximeter Innove Deluxe IP900AP for 6 trials resulting the heart rate percentage error of 7.48 %.

# Chapter 5

# ULTRASOUND SENSOR: RINGDOWN ARTIFACT REMOVAL

## *5.1 Ringdown Artifact Overview*

Miniaturized ultrasound transducers have not been widely used in surgical graspers. In one application, a transmitter-receiver transducer pair on Martin forceps was used to detect soft tissue cancers, by evaluating the echo attenuation [16]. Our use differs by using a 1D ultrasound transducer (Steminc SMD063T07R111) in bidirectional A-mode, which switches between transmitter and receiver functionalities. Operation in bidirectional mode saves space and gives an opportunity to add other sensors on the surgical grasper's jaws by using only one transducer instead of two. However, it is also a drawback: the system records exponentially decaying "ringing" (aka, the "ringdown artifact") of the transducer after excitation. When a surgical grasper's jaws operate on a tissue, the distance between them can be under 1 cm. As seen in Fig. 5.1, at these short distances relative to the speed of sound, the ringdown artifact blends with the received echo, making it impossible to identify the TOF without a method to remove the artifact. Short-distance ultrasound, in which the echo can blend with the ringdown, is applicable in many fields. These include the estimation of the degenerative loss of skeletal muscles [40], obstacle avoidance and mapping ultrasound sensors for underwater robots [41], thickness and defect detection applications in non-destructive testing [42], and other applications. All of these applications fundamentally suffer from the ringdown artifact when an A-mode bidirectional ultrasound transducer is used. In spite of all the above-mentioned applications, limited published literature exists on ringdown artifact removal from ultrasound signals, and there does not appear to be applicable work on AI-based noise removal algorithms for 1D ultrasound waveforms.

One application of a ringdown artifact removal from an ultrasound signal is a gastroin-

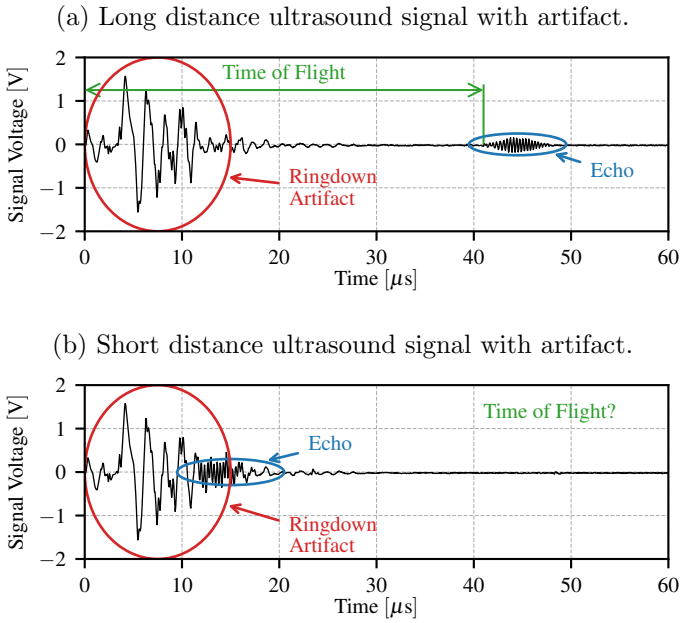



Figure 5.1: Raw ultrasound signals with ringdown artifacts.

testinal capsule with a 1D ultrasound for microanatomical diagnostics [43]. The capsule uses a ringdown-compensating filter by subtracting a moving average of the ringdown portion of the signal from the combined signal. This approach can help at longer TOFs, but not at short ones: at short TOFs, the ringdown blends with the echo, and the moving average starts including the echo as well.

Besides 1D applications like the surgical grasper or the diagnostic capsule above, ringdown artifact removal is also needed for short-range ultrasound imaging, in which the ringdown artifact adds a blind spot or dead space. A patent by Barlow, et al. claims a method for such removal by obtaining multiple "reference scans" (scans taken in an echo-free space), averaging them in a specific way, and subtracting them from the combined scan [44]. This method could potentially be extended to a single bidirectional transducer, but its main limitation (specified by the inventors) comes from "ringdown drift": the changing of the ringdown of a sensor over time. Applying such a method to a transducer of a surgical grasper would therefore require frequent (and potentially lengthy, depending on the number of reference

scans required) recalibration, which is not feasible during a surgery.

The need to remove ringdown from short-range ultrasound signals also arises in ultrasonic thickness measurement systems with bidirectional transducers. Most such systems use multiple transducers, but some, such as the Elcometer Ultrasonic Precision Thickness Gauges, use single bidirectional transducers. These systems are able to measure short-range echoes by placing a "delay line" material between the transducer and the material being inspected, which artificially elongates TOF, thus giving the ringdown time to decay prior to receiving the echo. The signal is also filtered through frequency selective filters, which further helps remove the ringdown, and also Gaussian noise. However, this delay line is also at least 9 mm thick, which makes it prohibitively bulky for a surgical grasper.

### *5.2 Denoising Algorithms Description*

Fig. 5.2 illustrates the two-stage process we designed with which the ringdown artifact is removed from the ultrasound signal, and TOF is then estimated from the filtered signal. In this section, six noise removal algorithms for the first stage are compared: three traditional, and three deep learning ones for comparison. For traditional algorithms, the Bandpass filter, the Adaptive LMS filter, and the Spectrum Suppression method (SPS) were chosen. These traditional algorithms were chosen because these algorithms are used for denoising conventional audio signals, speech, and Natural Language Processing.

The Bandpass filter was designed with passband frequencies of 2.5 MHz to 3.5 MHz to bracket it over the 3 MHz resonant frequency. The Adaptive LMS filter uses the target signal to optimize the coefficients for the filter. From preliminary attempts, the LMS filter order of 13 with a step size of 0.004 was selected for use on all signals of interest. Initial coefficients were based on the coefficients of the above Bandpass filter with the Hamming window.

The SPS used here was adapted from a paper by Boll [45]. The method relies on a separate noise signal (i.e., the ringdown artifact measured from a free transducer) and the blended signal. The signals are transformed to the frequency domain via Fast Fourier Transform (FFT), and in the frequency domain, the noise is subtracted from the blended signal. An

inverse FFT is then used to obtain the denoised signal in the time domain. Boll followed this sequence with a lowpass filter; in this paper, the above Bandpass filter was used on the denoised signal instead.

For deep learning algorithms, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) were chosen for comparison, because these algorithms work with time series data, and have previously been used for denoising speech signals. Of the three, RNN is the simplest and slowest to train, LSTM has the most representational power and is generally cheaper to train than standard RNNs, and GRU is a streamlined, cheaper to train version of LSTM. Chollet presents the theory of the three methods [46]. In this work, the tanh activation function was used for all three deep learning networks, with the Adam optimizer with the learning rate of $5 \times 10^{-4}$ and 330 training epochs. The Euclidean norm of the error between the input and target signals was used as the loss function. All trained networks had two layers with 512 and 256 units. All three networks used a bidirectional architecture.

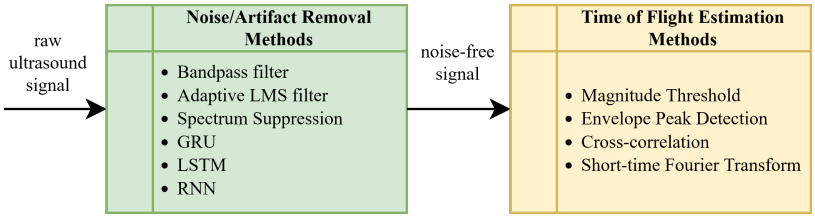



Figure 5.2: Architecture for method comparison in two-stage TOF estimation.

### *5.3 Time of Flight Estimation (TOF) Methods*

In the second stage of the two-stage TOF estimation process in Fig. 5.2 the actual TOF estimation algorithm is applied, producing the TOF from the cleared signal. TOF is the time required for a sound wave to travel through a medium. TOF depends on the distance traveled, and the medium's properties, including density, compressibility, and rigidity, which are temperature-dependent.

Four ToF estimation methods are utilized in this work.

*Magnitude Threshold Method*, used widely in hardware setups for automatic detection of TOF. When the amplitude of the received signal reaches a preset threshold, it records the time. The threshold should bypass the maximum noise level expected. This method is sensitive to noise fluctuations and signal decay with distance traveled, but is the most straightforward to implement [47].

*Envelope Peak Detection* consists of approximating the echo signal as an analytical signal with complex components, from which a Hilbert envelope is computed. Then, peaks are found in the signal's envelope. The TOF is then detected either by locating the peak with the largest amplitude. Such envelope signal is more robust to fluctuations of the echo's peaks, as well as a variable white noise level [48].

*Cross-correlation Method (CC)*, in its simplest form, consists of cross-correlating the transmitted and received signals, and estimating the TOF as the maximum in the correlation function [49].

*Short-time Fourier Transform Method* (STFT) for TOF estimation uses the time-frequency representation of ultrasound signals. STFT is obtained with a sliding window over the time domain signal, and taking the Fourier transform of the window. Then, only windows with the frequency of interest are considered for the estimation of TOF [50].

### *5.3.1 Data Collection and Preprocessing*

Apart from the sensorized surgical grasper (Fig. 3.2), another experimental setup (Fig. 5.3) was developed for ultrasound data collection, consisting of an acrylic container with an attached ruler to measure the level of liquid in the container – the one-way distance traveled. The bottom acrylic thickness of the container is 2 mm. The same SMD063T07R111 transducer as the one used on the surgical grasper was glued to the bottom of the acrylic container. Similarly, the same data acquisition setup was used: TI TDC1000 to drive the transducer at 3 MHz with a pulsed square waveform with 6 impulses, Siglent SDS 1104X-E oscilloscope to provide the external clock and acquire the signal. The data was recorded by the oscilloscope

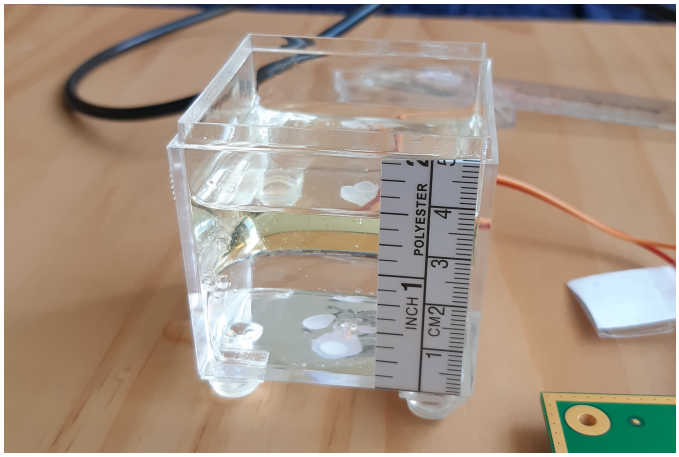



Figure 5.3: Acrylic container with an attached ultrasound transducer at the bottom.

in *csv* format, combined and aligned in MATLAB. Then, the dataset was downsampled by a factor of 26 to a new sampling rate of 19.23 MHz, leading to 1279 samples per waveform.

In order to use the dataset for training in deep learning models, it is required to have target signals to compare with and to adjust weights accordingly. When generating target signals, at distances of 2 cm and longer, the ringdown and the echo were sufficiently far apart, which made it possible to isolate the echo by zeroing the ringdown time segment. The waveforms obtained at shorter distances were not used as target signals for the training dataset, due to the difficulty of extracting the echo from such signals. The training dataset contains 993 waveform pairs (consisting of a raw signal, and a target signal with only the echo), with an approximately equal number of waveforms per distance. A separate test dataset was collected at 9 distances from 0.5 cm to 4.0 cm, using liquid water as the medium. This testing dataset has 270 waveforms, 30 waveforms per distance. Both datasets were saved in *mat* format for further processing in Google Colaboratory. Both the downsampled and original resolution datasets have been published on IEEE DataPort [7].

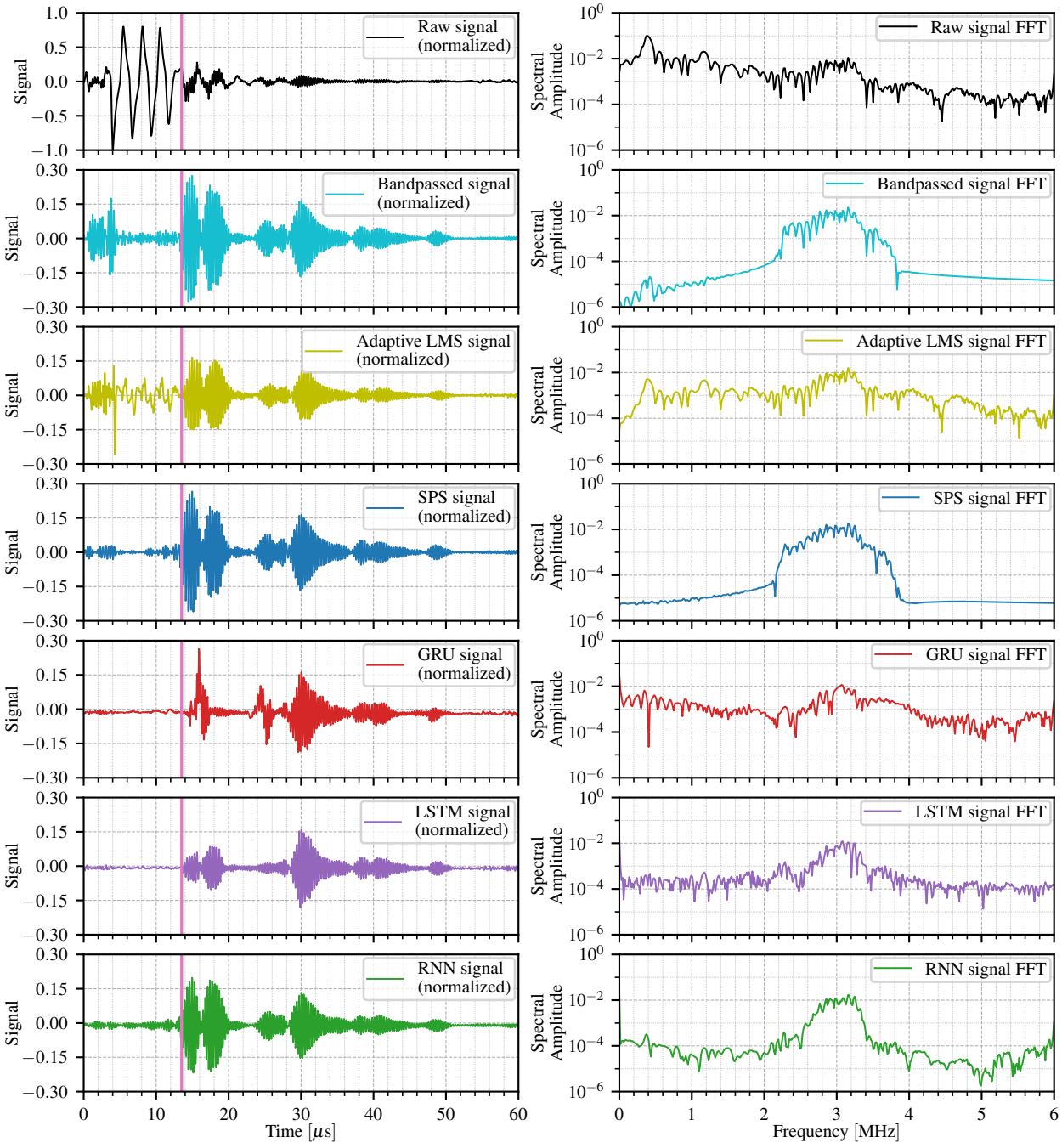



Figure 5.4: Denoising results for the Bandpass filter, the Adaptive LMS filter, Spectrum Suppression, GRU, LSTM and RNN. Left: the results of the denoising methods in time domain. Vertical pink line shows the true TOF with the travel distance of 1 cm. Right: the results of the denoising methods in the frequency domain.

### *5.4 Noise Removal Qualitative Comparison*

Each artifact removal method's results for a single raw ultrasound signal at 1 cm flight distance (here and below, this refers to the distance traveled until reflection and return to the transducer) are presented on Fig. 5.4. At this distance, the echo is capable of traveling back and forth multiple times, without decaying enough to become immeasurable; this multi-echo phenomenon can be observed in all filtered signal plots, at approximately 13.5 µs, 27.0 µs and 40.5 µs. The training dataset did not have any data with multi-echo.

The upper left plot of Fig. 5.4 shows the raw signal, corrupted by the ringdown artifact; the echo's amplitude is smaller than the ringdown artifact's by about a factor of 3. At a shorter flight distance, this corruption would become more severe, thus making it even harder to extract the time of flight from the blended signal. Because both the ringdown and the echo are physically generated by the oscillating transducer, part of the ringdown's frequency spectrum is similar to the echo's, and due to the short TOF, they blend with each other in time domain as well. For this reason, the Bandpass filter still retains a lot of noise, particularly in the first 5 µs of the waveform. The adaptive LMS filter exhibits the same behavior. The SPS method, however, despite its reliance on the frequency domain, cleaned the signal much better, with only negligible noise remaining in the first 5 µs. The performance of deep learning methods was more varied. As Fig. 5.4 shows, at this short distance, GRU significantly distorted the signal, although the ringdown was removed completely. LSTM showed some distortion as well, although echoes remained visually recognizable. RNN performed the best among the three deep learning methods.

### *5.5 Time of Flight Comparison*

Following the denoising via the six methods discussed in Section 5.2, TOF was estimated via the four methods discussed in Section 5.3. We evaluated the true TOFs based on the velocity of sound in water and the known distance. Mean relative errors in TOFs, as estimated by each method pair for each distance measured, are presented on Fig. 5.5.

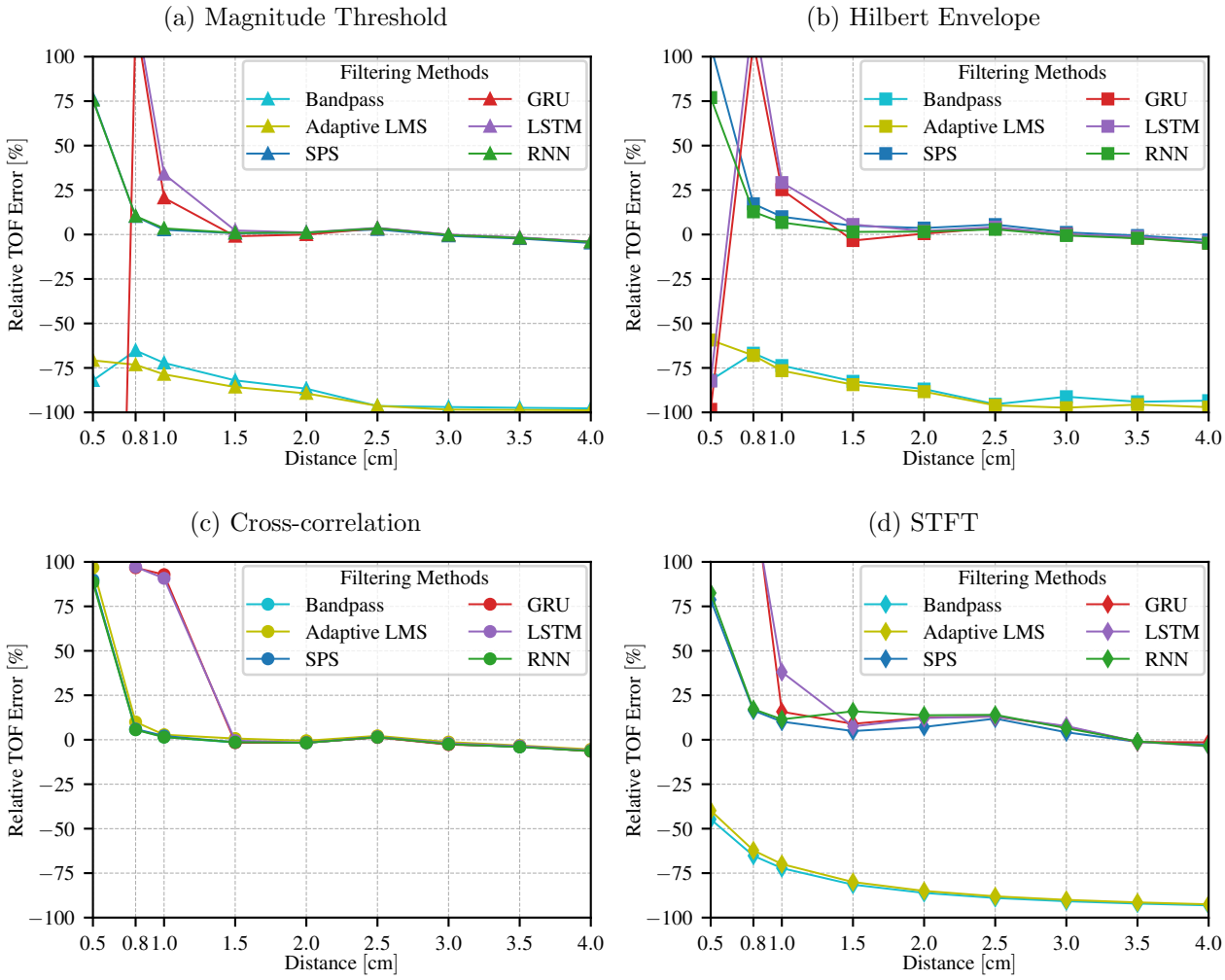



Figure 5.5: TOF mean percentage errors by different filtering and TOF estimation method. Errors over 100 % indicate method failure, and are cropped out.

As we can see from Fig. 5.5, all pairs of methods fail at the distance of 0.5 cm, which indicates the system's limitation. It's observable that deep learning denoising methods LSTM and GRU with all TOF estimation methods started to fail at distances lower than 1.5 cm. Distances lower than 2 cm were never shown during the training of deep learning denoising methods, and LSTM and GRU fail to generalize the solution for short-distance echo waveforms. It may be possible to use them with additional training data on short distances. On the other hand, standard RNN showed remarkably good results in spite of not having short-distance echo waveforms shown during the training.

The Bandpass and the Adaptive LMS filters had relative errors of more than 75 % paired with three of the four methods for TOF estimation at all distances. This is likely because as Fig. 5.4 illustrates, after these filters are applied, the remaining artifact is larger in amplitude than the echo, which causes threshold-based methods to fail. However, the Bandpass and the Adaptive LMS filters performed well with the CC TOF estimation method, because it does not depend on thresholding, but on signal correlation. CC method is also immune to white noise more than other methods.

Root Mean Square Error (RMSE) over all valid distances (i.e., 0.8 cm to 4.0 cm) was calculated for all 24 pairs of methods. 6 pairs demonstrated an RMSE of less than 5 %:

1. SPS with Magnitude Threshold: RMSE of 4.31 %.

2. RNN with Magnitude Threshold: RMSE of 4.39 %.

3. Bandpass with CC: RMSE of 3.62 %.

4. Adaptive LMS with CC: RMSE of 4.40 %.

5. Spectrum Suppression with CC: RMSE of 3.70 %.

6. RNN with CC: RMSE of 3.60 %.

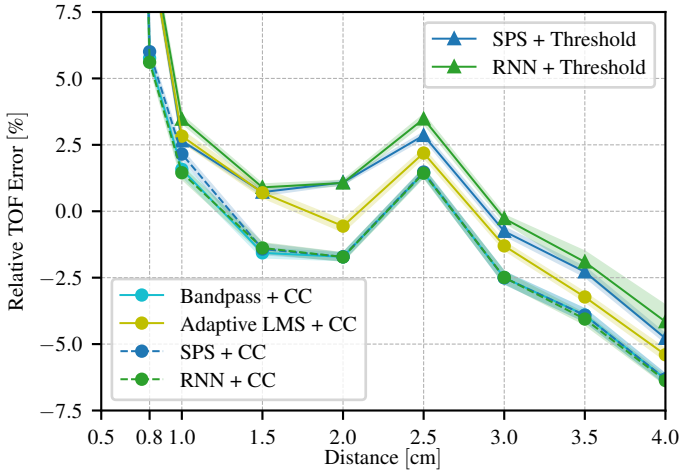



Figure 5.6: Relative TOF errors for best method pairs across all distances.

These methods' RMSEs for each distance, and their standard deviations over all samples taken at each distance are illustrated on Fig. 5.6. Based on their RMSEs, three pairs performed particularly well: Bandpass with CC, Spectrum Suppression with CC and RNN with CC. RNN with CC has the lowest RMSE error. The error variances are also small, which confirms that because the noise removal algorithms used in this work do not rely on subtracting a particular ringdown signal to remove it, they are not sensitive to ringdown drift.

All 24 method pairs compared are usable in real-time, each providing a TOF result in under 15 ms. The bulk of this time is taken by the denoising stage.

## *5.6 Discussion*

The RNN denoising method, combined with the CC TOF Estimation method, appear to be best suited for ringdown artifact removal, followed by TOF extraction. In literature, LSTM and GRU are typically considered superior methods to RNN due to the vanishing and exploding gradient problems [46]. At longer TOFs, with the echo and the ringdown artifact completely separated, GRU did indeed outperform all other methods. The poor

performance of LSTM and GRU at shorter distances may be explained by the limitations of the training dataset, generated by manually removing the ringdown, did not have any examples of the multi-echo, and may have lacked other features that the networks have relied on.

RNN, on the other hand, showed optimal performance at both long and short distances, with results comparable to the traditional Spectrum Suppression and Bandpass methods. It is conventional to compare neural network performance according to loss function values achieved on training sets, and their training time. Training time-wise, the results were indeed as expected, with RNN taking significantly longer than the other two, and GRU being the fastest. For this problem, the significant difference between the training and testing sets makes the loss function evaluated on the training set irrelevant; instead, average errors across the testing set, presented in the previous section, are more illustrative about the methods' relative performance.

Although RNN is one of the best denoising methods, it has a disadvantage—it requires adding data from each new medium to prevent performance deterioration. And RNN training takes a whole day. We decided to use the Bandpass filter for further research on animals.

# Chapter 6
# ANIMAL EXPERIMENTS DESCRIPTION

After developing the Smart Grasper instrument, we run several several different experiments for data collection for tissue characterization. We gradually increased complexity of the experiments, where we started from structured lab environment and butcher shop chicken breast and thigh, then continued to surgical room euthanized pigs, and finished with *in-vivo* under general anesthesia pigs. All *in-situ* and *in-vivo* pig's experiments are done under protocol IACUC: 4329-02: EMRL Pig with a collaboration to Dr. White's Emergency Medicine research group at UW Harborview Medical Center.

## *6.1 Abnormality Detection Experiment with ex-vivo Chicken Tissue*

### *6.1.1 Experiment Description*

The goal of the experiment is to show that the Smart Grasper can simultaneously detect abnormalities (e.g. shrapnel), and identify the tissues. During the experiment, chicken thighs and chicken breasts were bought at the local store. Each piece of chicken was cut and labeled into about one-inch pieces. It is important to keep each piece larger than 1.5 cm to ensure the Smart Grasper tip can be covered by the piece. Also, to avoid creating a jaw displacement bias, the thickness of the pieces within each tissue type needs to vary. For example, chicken breasts are usually thicker than chicken thighs, so some pieces of chicken breast might be thinly cut to introduce variable thickness. Ten pieces of each type of tissue were used for the experiment.

We ran ten grasps on the same spot for each piece. If there was slippage, where the piece of chicken was squeezed out of the grasper's tip, we repeated the trial. The numerous trials cause the tissue to compress by 10 % after 10 grasps. After all 20 pieces were done, we then

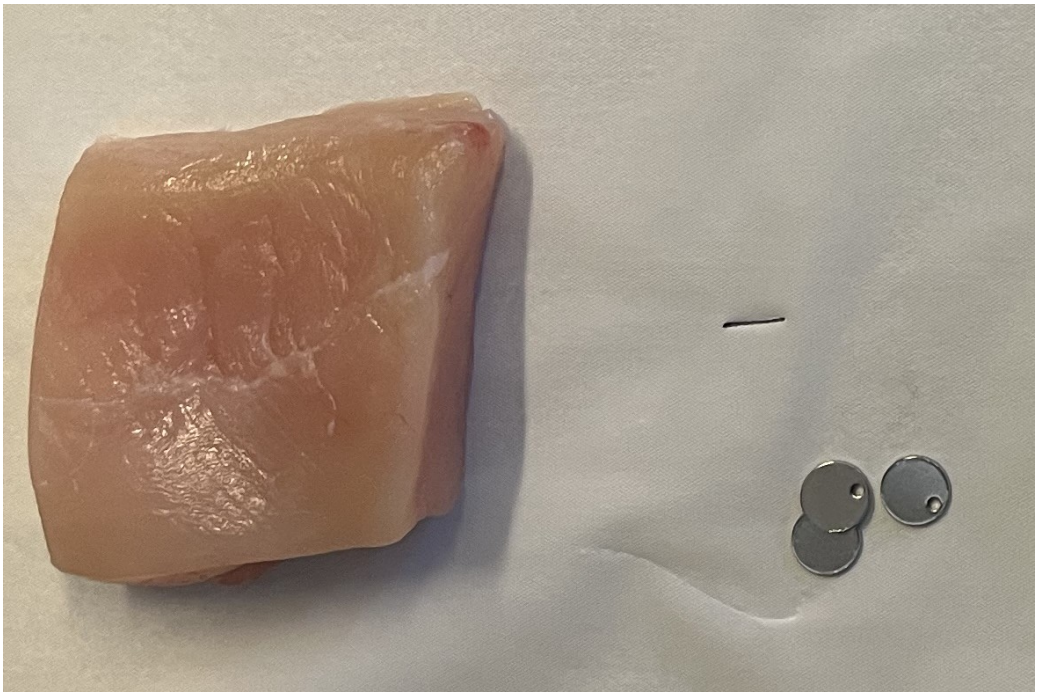


Figure 6.1: Single piece of chicken breast for a trial. Metal disk (5 mm diameter) abnormality is shown on the bottom left when not inserted into the meat.

introduced the abnormality and repeat the experiments.

The abnormality we used were three small metal discs, stacked together inside of tissue. In order to introduce the abnormality, we cut a small side-slit along the area the piece was grasped and inserted the three stacked metal discs. We made sure there were no air bubbles inside the piece by squeezing the slit shut. Then we ran the experiment for each piece of chicken. When running the trials, we made sure the grasper tip was directly over the abnormality.

### *6.1.2 Dataset Description*

The dataset resulting from our experiments on chicken tissue has time series data from indirect force sensor $f$, displacement of jaws $d$ (encoder readings), temperature data $t$ (ther-

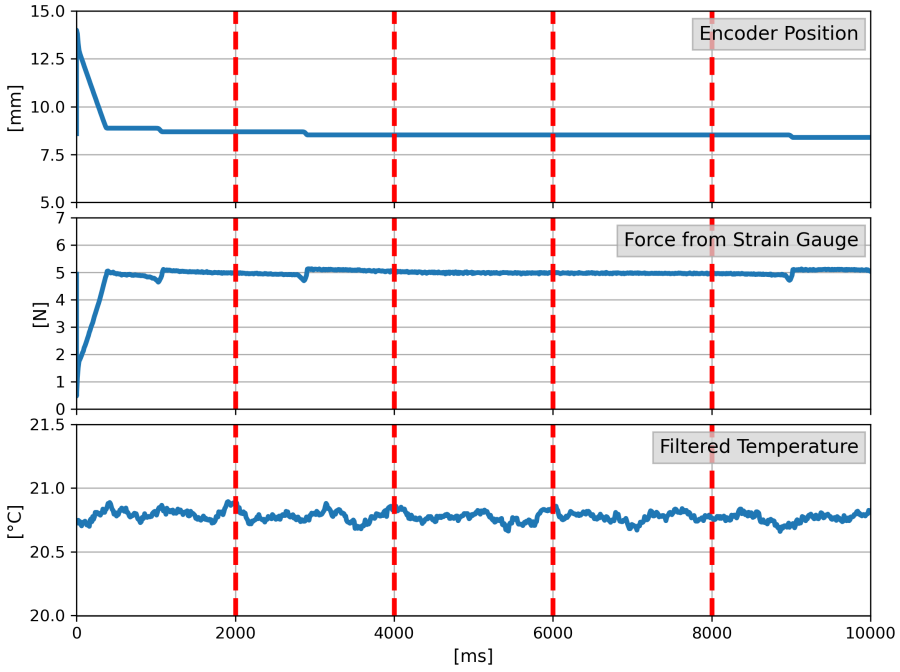



Figure 6.2: Encoder, strain gauge, thermistor signals obtained from Smart Grasper's sensor set. Red lines in encoder, strain gauge and thermistor plots indicate when ultrasound signals were collected.

mistor readings) (Fig. 6.2), and both raw ($u$), and filtered with the Bandpass filter from the ringdown artifact ($\overline{u}$) ultrasound waveforms (Fig. 6.3). The data was divided into 10 s grasp periods by detecting peaks in the displacement readings. The sampling rate, length of the signal per grasp, and number of time series signals per grasp are presented in Table 6.1.

Ultrasound signals have a much higher sampling rate compared with other modalities. Because of that, ultrasound waveforms were collected every 2 sec during the grasp, which led to 5 ultrasound waveforms per grasp with total of 1970 ultrasound waveforms. The red dashed lines in Fig. 6.2 show when the ultrasound was collected during the grasp. Fig. 6.3 presents one example of raw and filtered ultrasound waveforms.

The dataset was divided by grasps: 44 grasps made up the testing dataset, and a separate

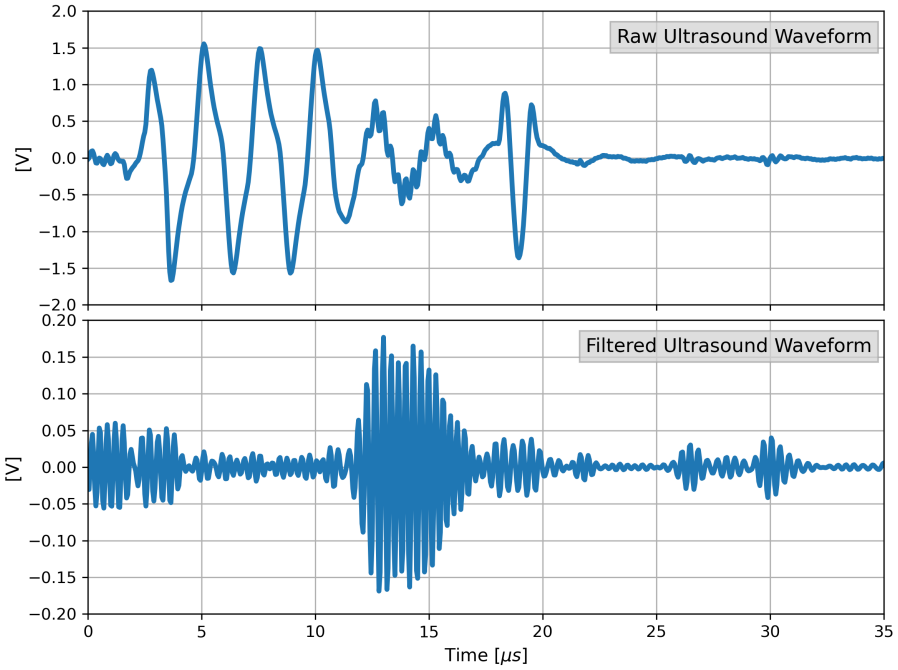



Figure 6.3: Raw and filtered ultrasound signals obtained from Smart Grasper's sensor set.

350 grasps were used for the training dataset. Four different classes were defined to classify time series signals: chicken breast/normal tissue (no metal inside), chicken thighs / normal tissue (no metal inside), chicken breast / abnormality (metal inside), and chicken thighs / abnormality (metal inside).

The dataset has been published in IEEE DataPort database [6].

## *6.2 Tissue Classification Experiment with in-situ and in-vivo Porcine Tissue*

### *6.2.1 Problem Statement*

The goal is to use multimodal Smart Grasper instrument, and classify the different types of tissues based on developed deep learning model, described in Section 7.1. Classification of tissues task were done in many research with different approach: vision based, haptic based, and sensors based. However, limited multimodal research is presented currently for *in-situ*

Table 6.1: The time series datasets.

| Modality | Symbol | (Size, Length) | Sampling Rate |
|---|---|---|---|
| Force | $f$ | (394, 993) | 100 Hz |
| Displacement | $d$ | (394, 993) | 100 Hz |
| Temperature | $t$ | (394, 993) | 100 Hz |
| Ultrasound | $u$ | $(5 \times 394, 794)$ | 20 MHz |
| Filtered Ultrasound | $\overline{u}$ | $(5 \times 394, 794)$ | 20 MHz |

and *in-vivo* animal experiments, and even less datasets are available for open use.

### *6.2.2 Experiment Description*

The experiments were conducted at UW Harborview Medical Center, Emergency Medicine, with Dr. Nathan White's research group under protocol IACUC: 4329-02: EMRL Pig. Several different experiments from different research labs were run under this protocol. We have to share our time with an animal accordingly to ensure that everybody can run their experiments with animals without interacting with each other or different devices, instruments and/or sensors. All collected datasets are presented in Table 6.2.

Datasets 1 and 2 from Table 6.2 were collected with the help of a subset of sensors ($f$, $d$, $t$, $u$, $\overline{u}$), and were conducted on euthanized three months old female pigs, and one *in-vivo* female pig. Most of the data was collected from the following tissues: liver, fat, leg muscle, small intestine and stomach. The data for grasp was disregarded if the Grasper slips off the tissue.

After adding more sensors (bioimpedance and optical sensing) and substituting the force sensor to OMEGA LCL-020, 20 lb sensor, we run additional experiments on both *in-vivo* and *in-situ* porcine tissue. The dataset with bioimpedance and optical sensor's data is marked as

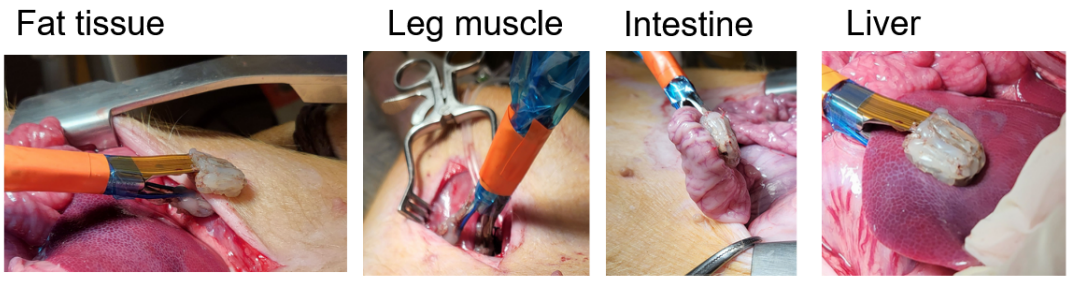



Figure 6.4: Fat, leg muscle, small intestine and liver tissues during the experiments at UW Harborview Medical Center.

number 3 in Table 6.2. The grasped tissues are the liver, fat, big intestine, and leg muscle.

Availability for *in-vivo* experiments on the pigs was related to the pig's health and if the experiments run at UW Harborview Medical Center can accommodate the Smart Grasper grasping abdominal organs. For example, when the powder was experimented with the treatment of acute liver trauma, the Smart Grasper couldn't be used due to powder substances that can damage the Grasper. Sometimes, Dr. White's research group conducted more serious trauma research, and pigs could be euthanized early. We could collect only post-mortem data. That explains that the majority of the data is post-mortem data, and *in-vivo* data is limited.

The experiment involved the grasping by Smart Grasper 10–15 times per pig's tissue or organ (depending on the availability of a pig) for 10 s. For every grasp, we tried to move the Grasper slightly over the tissue, not to grasp the same part of the tissue over and over again, and to avoid tissue excessive compression. Fig. 6.4 shows how we collected data with Smart Grasper at UW Harborview Medical Center. Because optical sensing, bioimpedance, and thermistor are located on the opposite sides of the ultrasound, we have to rotate Grasper when we collect data for fat tissue to make sure we collected all of the data from inside fat tissues and surface skin because hairy skin has low penetrability for ultrasound, optical sensing, and bioimpedance.

### *6.2.3 Datasets Description*

From the data collected over 15 pig's surgeries, several datasets were formed, presented in Table 6.2. Sensor's modalities, presented in Table 6.2 are force ($f$), displacement of jaws ($d$), temperature ($t$), ultrasound ($u$), ultrasound filtered ($\overline{u}$), optical sensing ($o$), bioimpedance ($b$).

Table 6.2: The datasets collected from the porcine tissue.

| N | Dates | Sensor's Modalities | Tissue | Number of grasps per tissue |
|---|---|---|---|---|
| 1 | 4/26/23 - 9/1/23 | $f$, $d$, $t$, $u$, $\overline{u}$ | *In-situ* fat, liver, small intestine, stomach, muscle | (63, 100, 91, 102, 105) |
| 2 | 8/31/23 | $f$, $d$, $t$, $u$, $\overline{u}$ | *In-vivo* liver, small intestine, stomach, muscle, spleen | (14, 14, 9, 15, 19) |
| 3 | 4/10/24 - 4/19/24 | $f$, $d$, $t$, $u$, $\overline{u}$, $b$, $o$ | *In-vivo* and *in-situ* liver, fat, muscle, big intestine | (58, 60, 61, 60) |

# Chapter 7

# DATA FUSION AND CNN–GRU RESULTS

## *7.1 CNN–GRU Model Description*

The most common neural networks for time series multimodal data learning are Recurrent Neural Networks (RNN), Fully Connected Neural Networks (FCNN), Convolutional Neural Networks (CNN), Graph Neural Networks (GNN), Autoencoders, Transformers, and their variations. Different challenges are raised when multimodal biomedical data is used, such as small dataset size, missing data from modalities, an imbalance of dimensionality between modalities, and what data fusion technique to use. [35]

In the dissertation, a hybrid multi-channel CNN-GRU model, Fig. 7.1, was used with an intermediate structure for the fusion of sensing modalities ($f$, $d$, $t$, $u$ or $\overline{u}$). We expanded the model structure previously used in [51] for unimodal electrocardiogram data to form a parallel multi-channel structure for multimodal time series data, similar to [52, 53]. The parallel structure of the CNN-GRU model was chosen not to mix data from different unrelated modalities, and learn features independently, and only then concatenate the features to classify them with the GRU layer.

The model (Fig. 7.1) consists of thirteen layers of 1D Convolution Layers (CONV1D), Batch Normalizations, Rectified Linear Units (ReLu), Dropouts, and one GRU and Dense layer for classification. The convolution layers used 1 (first convolution layer), 2 (from 2nd to $7^{\text{th}}$ layer), and 4 (from $7^{\text{th}}$ to $13^{\text{th}}$ layer) dilation rates to speed up the training process by inserting zero samples into input signals and cover a larger portion of the input signal without losing the resolution of the signal. Adam optimization algorithm with batch size 24, 350 epochs, and variable learning rate (0.00001 was set as a minimum) were used for training the model. The training was accomplished on the GPU of the Google Colaboratory.

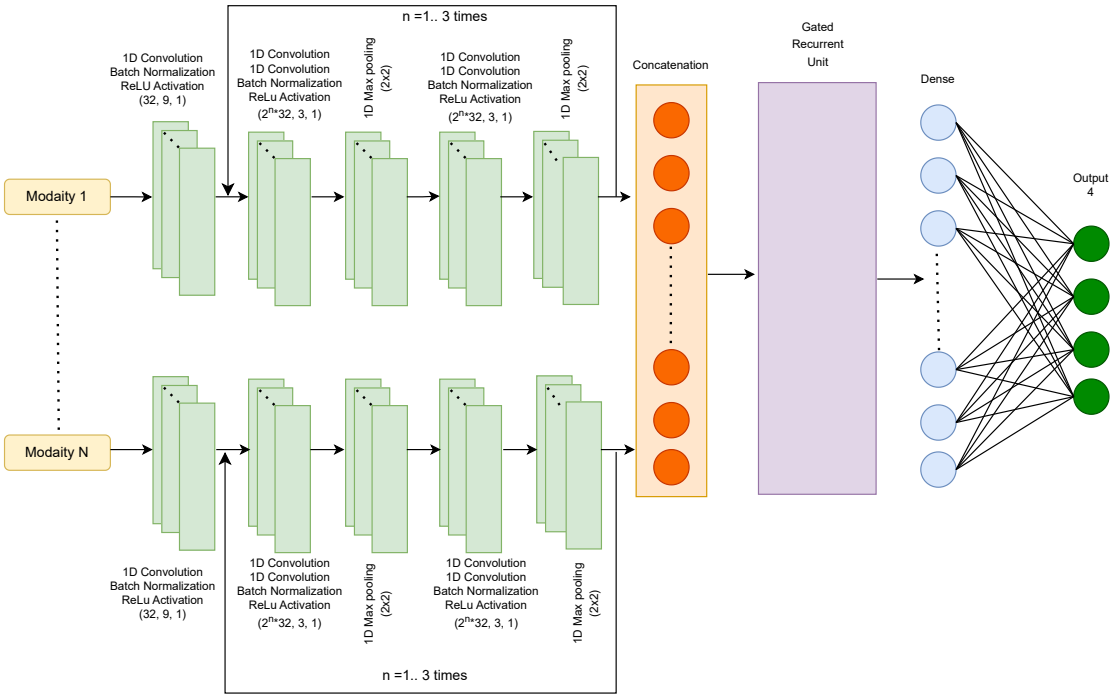



Figure 7.1: Multi-channel hybrid structure of CNN-GRU network

## *7.2 Results for Abnormality Detection Experiment with ex-vivo Chicken Tissue*

### *7.2.1 Different Data Representation*

Due to different sampling frequencies and received times of the ultrasound signals, we looked at two different data representations for the same deep learning model, Fig. 7.1:

- **Approach 1.** Full waveforms from ($f$, $d$, $t$) were used, and fused with all five ultrasound waveforms in the array for each grasp.

- **Approach 2.** Windowed points over ten neighbor samples from ($f$, $d$, $t$) were taken at the time when the ultrasound waveforms were collected, red dashed lines on Fig. 6.2. Each ultrasound was treated independently from the grasp period, and all ultrasound waveforms were fed independently to the deep learning model.

Each approach has advantages and disadvantages for training purposes. Approach 2 has the biggest dataset because all ultrasound waveforms, and windowed points from ($f$, $d$, $t$) values are considered independent. However, Approach 1 utilizes the full waveforms from ($f$, $d$, $t$) and allows the DL layers to learn features from the grasp transient state, when the jaws are closing but not yet steadily holding the tissue. In terms of training time, Approach 1 is three times faster compared with Approach 2 due to 5 times less amount of data.

### *7.2.2 Results for Abnormality Detection Experiment*

The results for any combination of *(f, d, t)* for both approaches are shown in Table 7.1, and for any combination of ($f$, $d$, $t$, $u$, $\overline{u}$) is in Table 7.2 over 5 training sessions. The symbol ✓ is used in Tables 7.1 and 7.2 to show which modality or a combination of the modalities was used for training and testing.

From the results of Table 7.1, we can conclude that Approach 1 has better performance (better mean accuracy, and mean F1 score) for unimodal data, and for the combinations of ($f$, $d$) and ($f$, $t$). However, the same performance is noticeable for a combination of ($f$, $d$, $t$) for both approaches.

Ultrasound raw ($u$) and filtered ($\overline{u}$) modalities were introduced in Table 7.2 for both approaches. Ultrasound data significantly improved the mean accuracy and mean F1 score for the classification task. In Table 7.1 Approach 2 significantly outperforms Approach 1 in both ultrasound data and fused with other modalities, which implies that the quantity of the data is more important than the richness of the data for training. Interesting to find that the raw ultrasound performs very well, and is comparable with filtered ultrasound waveforms, even outperforming the filtered ultrasound in Approach 2. This is related to the ringdown artifact, which has distinct features due to sound wave propagation through different tissues. The results also show that we don't have to provide displacement $d$ together with ultrasound data to the DL model to distinguish between different types of tissue and define the abnormality to compare with the conventional approach. In the conventional approach, the velocity of sound and attenuation were calculated to classify tissue, for example in Doyle work [16].

Table 7.1: Average accuracy and F1 score results after 5 training sessions for any combinations of ($f$, $d$, $t$) modalities for 2 data fusion approaches. (Best options bolded.)

| Modalities | | | Approach 1 | | Approach 2 | |
|---|---|---|---|---|---|---|
| $f$ | $d$ | $t$ | $\overline{Acc}$ **/SD** | $\overline{F1}$ **/SD** | $\overline{Acc}$ **/SD** | $\overline{F1}$ **/SD** |
| ✓ | – | – | 0.68(0.06) | 0.64(0.2) | 0.27(0.01) | 0.21(0.13) |
| – | ✓ | – | **0.72(0.02)** | **0.68(0.17)** | 0.43(0.02) | 0.38(0.08) |
| – | – | ✓ | 0.49(0.07) | 0.45(0.1) | 0.5(0.01) | 0.47(0.04) |
| ✓ | ✓ | – | 0.67(0.04) | 0.63(0.2) | 0.43(0.02) | 0.39(0.07) |
| ✓ | – | ✓ | 0.67(0.04) | 0.63(0.21) | 0.5(0.008) | 0.47(0.04) |
| ✓ | ✓ | ✓ | **0.72(0.05)** | **0.67(0.14)** | **0.7(0.02)** | **0.66(0.07)** |

Both parameters require to know tissue thickness, $d$. Filtered ultrasound also performs well in Approach 2, but worse in Approach 1, showing that echo itself has distinct features that could help to classify tissue and find abnormalities without tissue thickness provided. The highest mean accuracy (higher than 90 %) and mean F1 score (higher than 82 %) are for the Approach 2 data representation combinations: ($d$, $t$, $u$), ($f$, $t$, $u$), ($t$, $u$), ($f$, $u$), ($u$). All four modalities ($f$, $d$, $t$, $\overline{u}$), with filtered ultrasound, give the highest accuracy of 88 % and F1 score 82 % in Approach 2 data representation. We can see from the results that ultrasound is a dominant sensor for tissue identification and abnormality detection.

### *7.3 Results for Tissue Classification Experiment with in-situ Porcine Tissue*

The dataset 1 from Table 6.4 was used for processing with CNN–GRU model with ($f$, $d$, $t$, $u$ $\overline{u}$) modalities, described in Section 7.1. During the first dataset collection for porcine model, the most data was collected on liver, stomach, leg muscle and small intestine tissues. These tissues were used for classification procedure, combining to four different classes: muscle, intestine, liver, stomach. The dataset is quite different from what we observed previously in

Table 7.2: Average accuracy, F1 score and loss results after 5 training sessions for *(f, d, t, u, $\overline{u}$)* data fusion. (Best options bolded.)

| Modalities | | | | | Approach 1 | | Approach 2 | |
|---|---|---|---|---|---|---|---|---|
| $f$ | $d$ | $t$ | $u$ | $\overline{u}$ | $\overline{Acc}$ **/SD** | $\overline{F1}$ **/SD** | $\overline{Acc}$ **/SD** | $\overline{F1}$ **/SD** |
| ✓ | ✓ | ✓ | ✓ | – | 0.72(0.08) | 0.68(0.2) | 0.85(0.06) | 0.79(0.13) |
| – | ✓ | ✓ | ✓ | – | 0.65(0.06) | 0.64(0.2) | **0.92(0.02)** | **0.85(0.13)** |
| ✓ | – | ✓ | ✓ | – | 0.70(0.09) | 0.68(0.2) | **0.92(0.4)** | **0.85(0.15)** |
| ✓ | ✓ | – | ✓ | – | 0.71(0.09) | 0.67(0.2) | 0.88(0.07) | 0.82(0.15) |
| – | – | ✓ | ✓ | – | 0.6(0.04) | 0.58(0.2) | **0.92(0.04)** | **0.85(0.15)** |
| – | ✓ | – | ✓ | – | 0.52(0.03) | 0.52(0.2) | 0.91(0.15) | 0.85(0.15) |
| ✓ | – | – | ✓ | – | 0.66(0.06) | 0.63(0.2) | **0.93(0.04)** | **0.86(0.2)** |
| – | – | – | ✓ | – | 0.5(0.06) | 0.52(0.22) | **0.94(0.03)** | **0.87(0.14)** |
| ✓ | ✓ | ✓ | – | ✓ | **0.81(0.04)** | **0.75(0.13)** | 0.88(0.02) | 0.82(0.14) |
| – | ✓ | ✓ | – | ✓ | 0.75(0.04) | 0.7(0.12) | 0.88(0.01) | 0.82(0.13) |
| ✓ | – | ✓ | – | ✓ | **0.81(0.02)** | **0.76(0.13)** | 0.87(0.02) | 0.81(0.14) |
| ✓ | ✓ | – | – | ✓ | 0.73(0.03) | 0.69(0.15) | 0.85(0.02) | 0.79(0.14) |
| – | – | ✓ | – | ✓ | 0.75(0.05) | 0.7(0.13) | 0.89(0.02) | 0.82(0.13) |
| – | ✓ | – | – | ✓ | 0.7(0.05) | 0.67(0.14) | 0.86(0.01) | 0.81(0.15) |
| ✓ | – | – | – | ✓ | **0.81(0.07)** | **0.77(0.16)** | 0.86(0.01) | 0.8(0.15) |
| – | – | – | – | ✓ | 0.64(0.05) | 0.62(0.15) | 0.85(0.02) | 0.80(0.2) |

chicken *ex-vivo* data. We noticed that the pig's tissues were much slimmer compared with chicken tissues, it took longer for Grasper to close jaws (250 ms to 350 ms), and meaningful data collected during the contact grasp is much smaller to compare with chicken dataset data. It means that the only one or two last ultrasound waveforms are good for data processing during the grasped time. Last ultrasound from five ultrasound waveforms was used for

Table 7.3: Average accuracy, F1 score and loss results after 5 training sessions for any combinations of ($f$, $d$, $t$) modalities for porcine tissue, with 4 classes: 0 – muscle, 1 – intestine, 2 – liver, 3 – stomach. (Best options bolded.)

| Modalities | | | Approach 1 | | |
|---|---|---|---|---|---|
| $f$ | $d$ | $t$ | $\overline{Acc}$ **/SD** | $\overline{F1}$ **per class** | $\overline{Loss}$ **/SD** |
| ✓ | – | – | **1.0(0.0)** | **[1.0, 1.0, 1.0, 1.0]** | **0.0009(0.0001)** |
| – | ✓ | – | 0.87(0.04) | [0.87, 1.0, 0.74, 0.87] | 0.32(0.07) |
| – | – | ✓ | 0.38(0.02) | [0.45, nan, 0.41, 0.37] | 1.65(0.78) |
| ✓ | ✓ | – | **1.0(0.0)** | **[1.0, 1,0, 1.0, 1.0]** | **0.001(0.0004)** |
| ✓ | – | ✓ | **1.0(0.0)** | **[1.0, 1,0, 1.0, 1.0]** | **0.001(0.0002)** |
| ✓ | ✓ | ✓ | **0.99(0.01)** | **[1.0, 1.0, 0.98, 0.98]** | **0.011(0.01)** |

classification to make sure it's a valid signal. Because of these differences, we combined the data by Approach 1, but we used only one last ultrasound, instead of all five combined. The total number of waveforms for tissue porcine used for training is 358 waveforms/modality, and for testing is 40 waveforms/modality. We made sure that we have a good mix of different days, pigs experiments in both training and testing dataset. Validation dataset was used as a part of testing dataset in TensorFlow training routine – 10 % of training dataset.

Table 7.3 shows that temperature data negatively influenced the results from the CNN–GRU model, when combined with other modalities (increasing loss and decreasing accuracy). Interestingly, compared with the chicken experiment, where force modality didn't play so much role, force has more impact in UW Harborview Medical experiments. It could be explained by the fact that the device was rigidly attached to the wooden platform during the chicken experiments (lab bench experiments). However, during the UW Harborview Medical Experiments, we hand-held the Smart Grasper over each grasp, and the device was rotated around the porcine body to accommodate the organ's access, which likely influenced the

Table 7.4: Average accuracy, F1 score and loss results after 5 training sessions for ($f$, $d$, $t$, $u$, $\overline{u}$) data fusion.

| Modalities | | | | | Approach 1 | | |
|---|---|---|---|---|---|---|---|
| $f$ | $d$ | $t$ | $u$ | $\overline{u}$ | $\overline{Acc}$ **/SD** | $\overline{F1}$ **per class** | $\overline{Loss}$ **/SD** |
| ✓ | ✓ | ✓ | ✓ | – | 0.94(0.02) | [0.82, 0.9, 0.97, 0.92] | 0.43(0.3) |
| – | ✓ | ✓ | ✓ | – | 0.87(0.02) | [0.8, 0.96, 0.77, 0.95] | 0.69(0.15) |
| ✓ | – | ✓ | ✓ | – | 0.97(0.016) | [0.99, 1.0, 0.96, 0.95] | 0.15(0.11) |
| ✓ | ✓ | – | ✓ | – | 0.95(0.01) | [0.95, 0.98, 0.96, 0.94] | 0.2(0.06) |
| – | – | ✓ | ✓ | – | 0.74(0.02) | [0.54, 0.74, 0.73, 0.95] | 1.58(0.26) |
| – | ✓ | – | ✓ | – | 0.85(0.015) | [0.79, 0.93, 0.76, 0.93] | 0.99(0.17) |
| ✓ | – | – | ✓ | – | 0.96(0.025) | [0.97, 0.99, 0.95, 0.93] | 0.27(0.16) |
| – | – | – | ✓ | – | 0.75(0.03) | [0.57, 0.8, 0.69, 0.95] | 1.4(0.24) |
| ✓ | ✓ | ✓ | – | ✓ | 0.93(0.0) | [0.89, 0.95, 0.93, 0.95] | 0.71(0.2) |
| – | ✓ | ✓ | – | ✓ | 0.93(0.0) | [0.87, 0.95, 0.95, 0.95] | 0.81(0.17) |
| ✓ | – | ✓ | – | ✓ | 0.93(0.0) | [0.99, 1.0, 0.96, 0.95] | 0.62(0.16) |
| ✓ | ✓ | – | – | ✓ | 0.93(0.0) | [0.89, 0.95, 0.95, 0.95] | 0.78(0.12) |
| – | – | ✓ | – | ✓ | 0.74(0.02) | [0.87, 0.95, 0.95, 0.95] | 1.58(0.26) |
| – | ✓ | – | – | ✓ | 0.93(0.0) | [0.87, 0.95, 0.95. 0.95] | 0.58(0.12) |
| ✓ | – | – | – | ✓ | 0.93(0.0) | [0.87, 0.95, 0.95, 0.95] | 0.63(0.08) |
| – | – | – | – | ✓ | 0.93(0.0) | [0.88, 0.93, 0.95, 0.95] | 0.43(0.08) |

force sensor's data readings. Also, when we cut the chicken into pieces in lab experiments, we tried to cut it into various sizes, but we couldn't change the pig's organ sizes; we could try to grasp different parts of the tissues. And we can see from Table 7.3 that the data is biased by size (displacement readings from encoder gives 87 % accuracy).

For raw and filtered ultrasound data, multimodal data fusion results from the CNN–GRU

model are presented in Table 7.4. We can see that raw ultrasound performs better in some combinations than filtered ultrasound. For example, for ($f$, $d$, $u$), ($f$, $t$, $u$), ($f$, $u$). Raw ultrasound alone performs much worse compared with filtered ultrasound. However, raw ultrasound performs better than filtered ultrasound with a combination of sensor modalities. It means other modalities, such as force and displacement, play a critical role in the model's accuracy in this case. We also noticed that the raw and filtered ultrasound combined with temperature is dropping significantly in accuracy because temperature does not play a critical role in tissue distinction and easily biased (heating lamp at surgical room plays a negative role). Filtered ultrasound is not changing in accuracy and F1 score with adding other modalities, meaning that filtered ultrasound is a dominant sensor in these models.

## *7.4 Testing on in-vivo data*

We decided to test the developed models from the previous Section 7.3, trained on *in-situ* data, on *in-vivo* data, collected on the same pig. Our results showed the failure of the generalization of the CNN–GRU deep learning model for most modalities and their combinations, except for filtered ultrasound. Filtered ultrasound showed a total accuracy of 64 %, and the F1 score showed 9 % for muscle, 94 % for the intestine, 59 % for the liver, and 97 % for the stomach. We can see that the model was relatively robust to detect some tissues still despite it having never trained on *in-vivo* data. Fig. 7.2 shows the confusion matrix for the best-performing model. We can see that the CNN–GRU model fails only in muscle data. And because it mismatched to the liver, it caused a fall in the F1 score for the liver. These results look promising for the deep learning model that can learn from *in-situ* filtered ultrasound data and be tested successfully on *in-vivo* data.

## *7.5 Combination of in-vivo and in-situ data*

Our previous results showed that the CNN–GRU model fails to generalize the solution except for filtered ultrasound modality. We combined the *in-vivo* and *in-situ* data and re-train it again to evaluate if the CNN–GRU model finds a standard solution for both *in-vivo* and

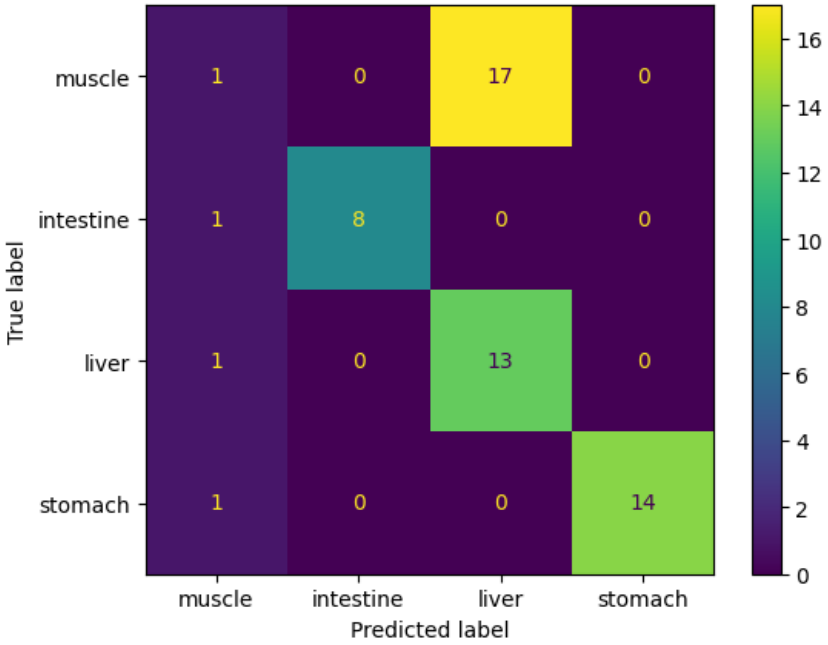



Figure 7.2: Confusion matrix for the best trained model CNN–GRU.

*in-situ* data. We also added three noise levels with mean 0 and standard deviations (0.01, 0.02, 0.03) to see when the CNN–GRU model fails the performance. Results for accuracy performance are presented in Fig. 7.3 for unimodal data and multiple modalities combination of data. We can see from Fig. 7.3 that the models trained on unimodal data fail faster with adding more noise; for noise level 0(0.03), unimodal trained models got lower than 70 % accuracy. If we compare with two modalities trained models, we could see that the two modalities models are much more robust to additive noise levels. Only three combinations started to fail lower than 70 % accuracy: ($t$, $u$), ($d$, $\overline{u}$), and ($t$, $\overline{u}$). The combination of three modalities lower than 70 % is ($d$, $t$, $\overline{u}$). Combining all four modalities is even more robust, and the lowest accuracy is 75 % for the highest noise data.

F1 score is presented in Fig. 7.4 for unimodal data. We can see from the plot that some modalities are more robust with additive noise than other modalities. For example, the F1 score is significantly reduced for ultrasound modality with additive noise because

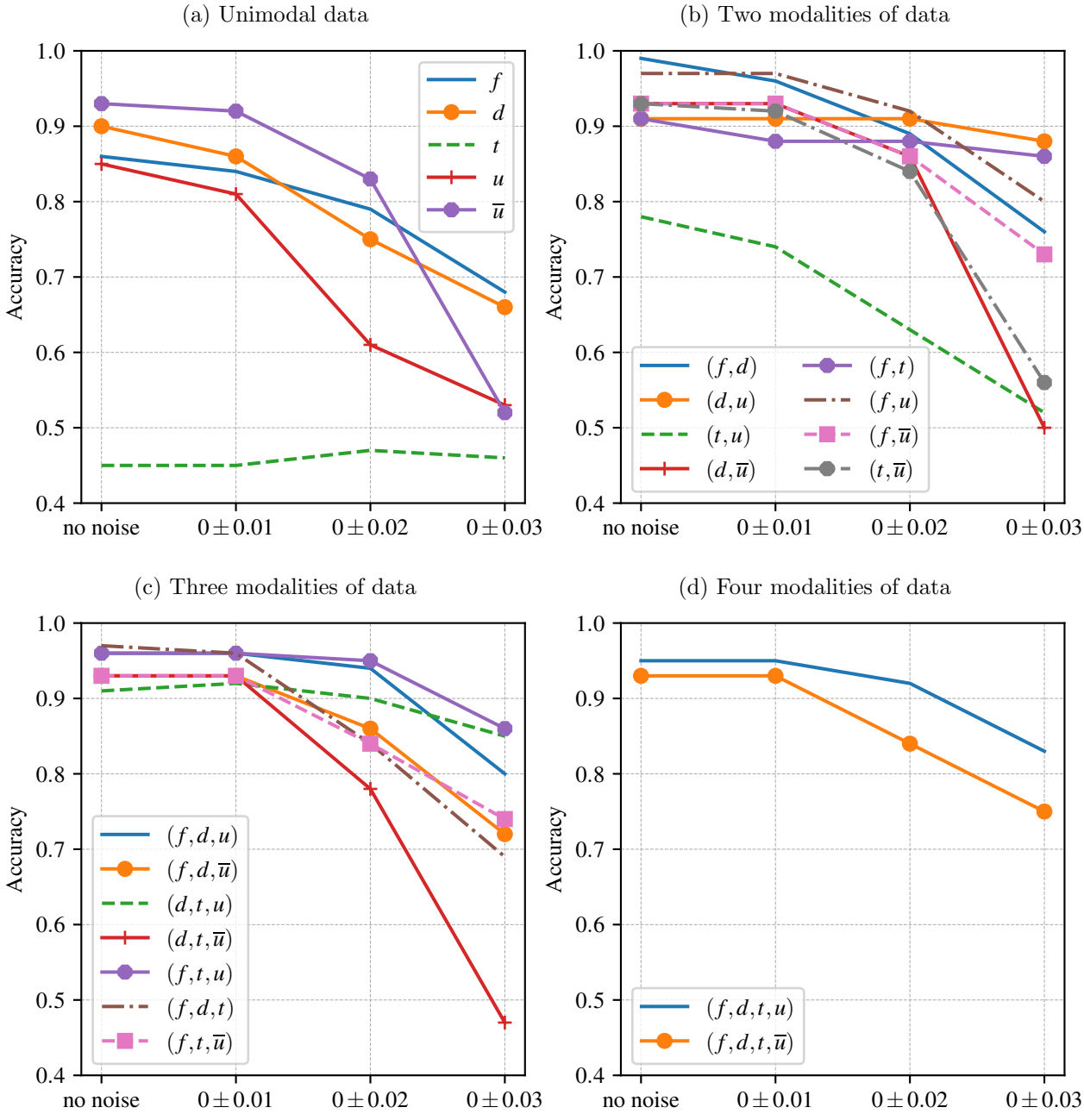



Figure 7.3: Average accuracy with additive noise for CNN–GRU model

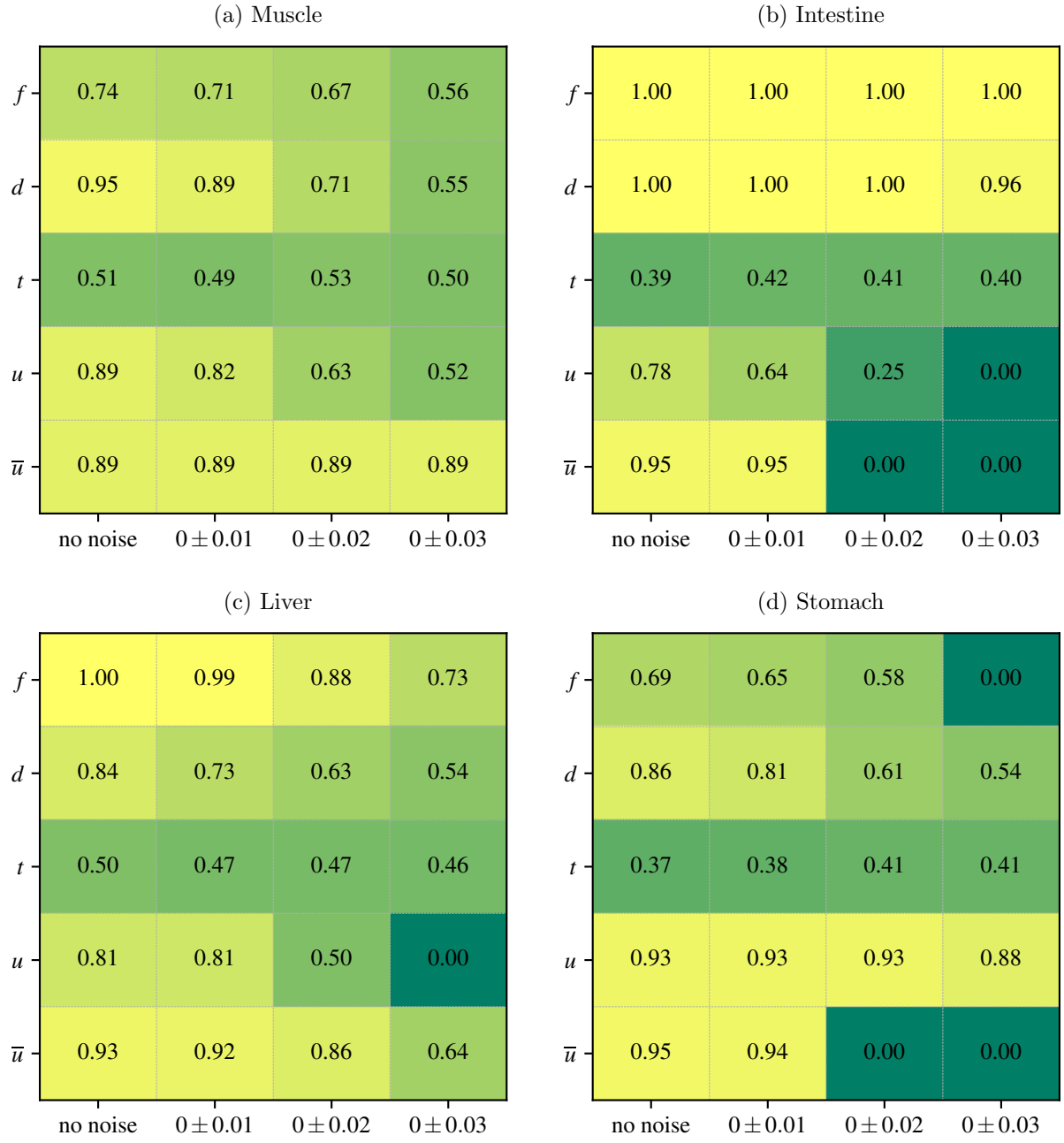



Figure 7.4: Average F1 score with additive noise for CNN–GRU model for separate sensors

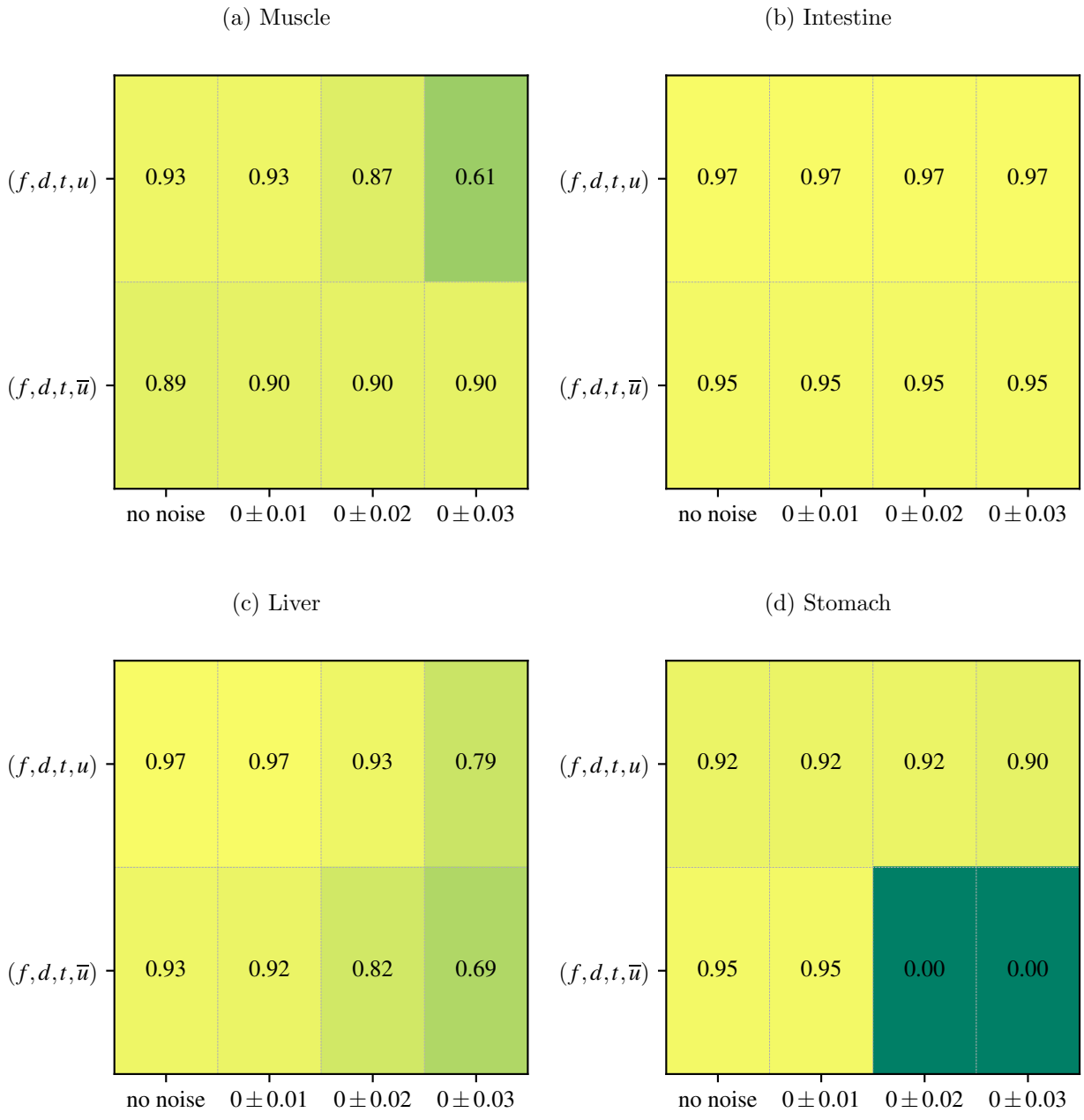



Figure 7.5: Average F1 score with additive noise for CNN–GRU model for a combination of four sensors

the ultrasound signal is small in amplitude, and adding too much noise makes the signal completely unrecognizable for the model.

Fig. 7.6 presents F1 scores for a combination of two different modalities for four different tissues: muscle, intestine, liver and stomach. We can observe that the best two modalities combinations are *(d, u)*, *(f, t)*, *(f, u)* these combinations of modalities are resistant to the additive noise across all of the tested porcine tissue. We can see when one modality (Fig. 7.4) fails, it leads a combination with this modality fail too (Fig. 7.6), e.g., ($f$, $d$), ($t$, $u$) and ($t$, $\overline{u}$) combinations.

From the combination of three sensor modalities, on Fig. 7.7, we can see that some combinations of sensors give more robust results due to combining with strong modalities. For example, if temperature and ultrasound combination started failing with additive noise, with an addition of displacement, the model was recovered. However, when force added to temperature and filtered ultrasound, nothing changed.

All four combinations of the sensors force, displacement, ultrasound and temperature with additive noise are presented on Fig. 7.5. We can see that the combination of sensors lead to robustness to the noise, except for stomach tissue. For some tissue, such as muscle, performance of the combination of four sensors ($f$, $d$, $t$, $\overline{u}$) never deteriorates.

### *7.6 Bioimpedance and Optical Sensing Modalities*

Additional data was collected, the third dataset from Table 6.2, on both *in-vivo* and *in-situ* female pigs at UW Harborview Medical Center with adding optical sensing and bioimpedance. To ensure that the model CNN–GRU still performs well, it was re-trained on a new dataset (the dataset 3 from Table 6.2) with fat, liver, big intestine, and muscle tissues. Unimodal performance of force, temperature, displacement, ultrasound, and ultrasound-filtered modalities with the model CNN–GRU showed very similar results as described in the previous section.

However, bioimpedance showed only 25 % accuracy. The failure of bioimpedance modality was also observed in Roan's previous work [26] and was likely caused by a high moisture level that is unavoidable inside of the body. Another reason is that a broader frequency swap

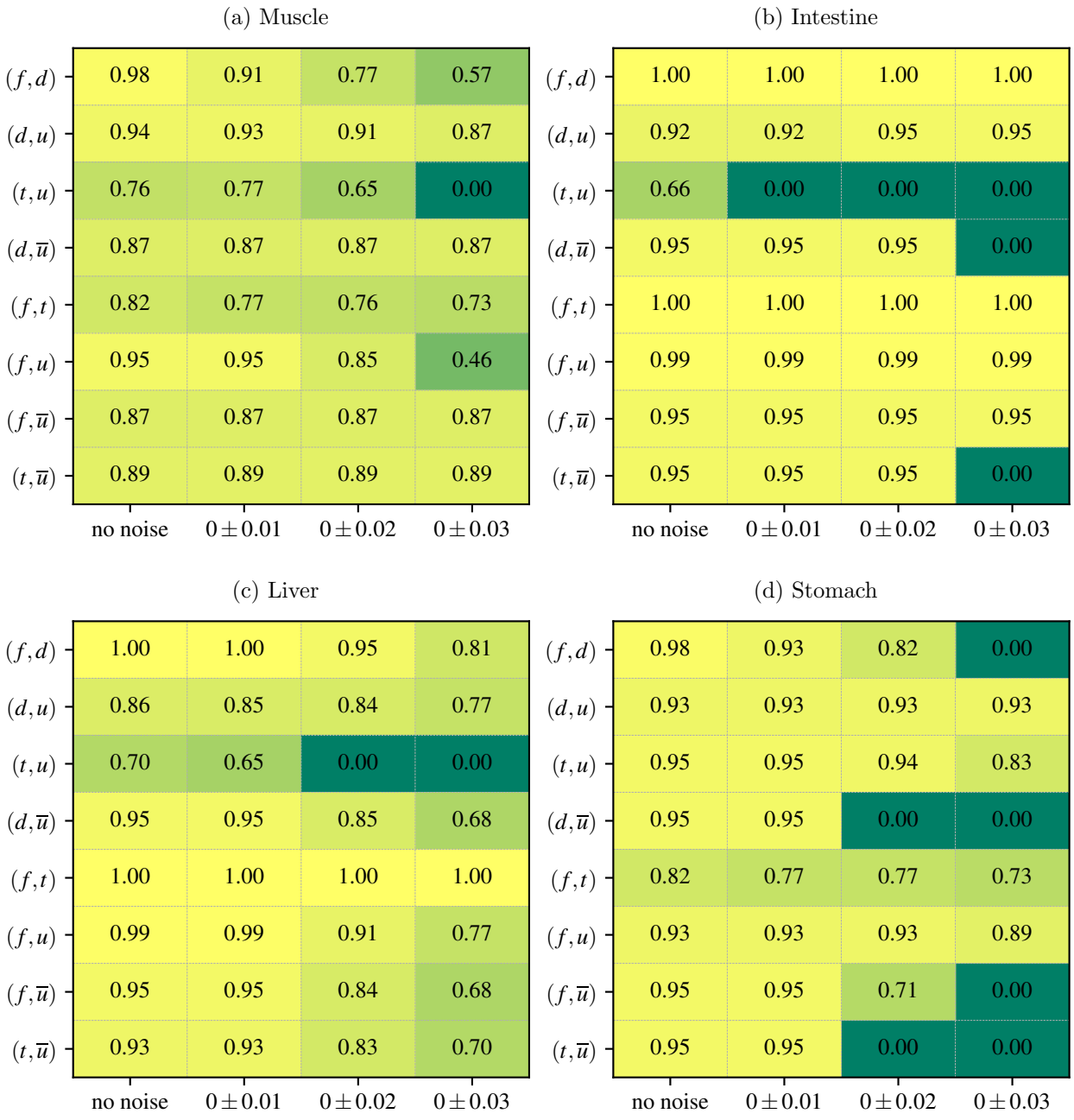



Figure 7.6: Average F1 score with additive noise for CNN–GRU model for a combination of two sensors

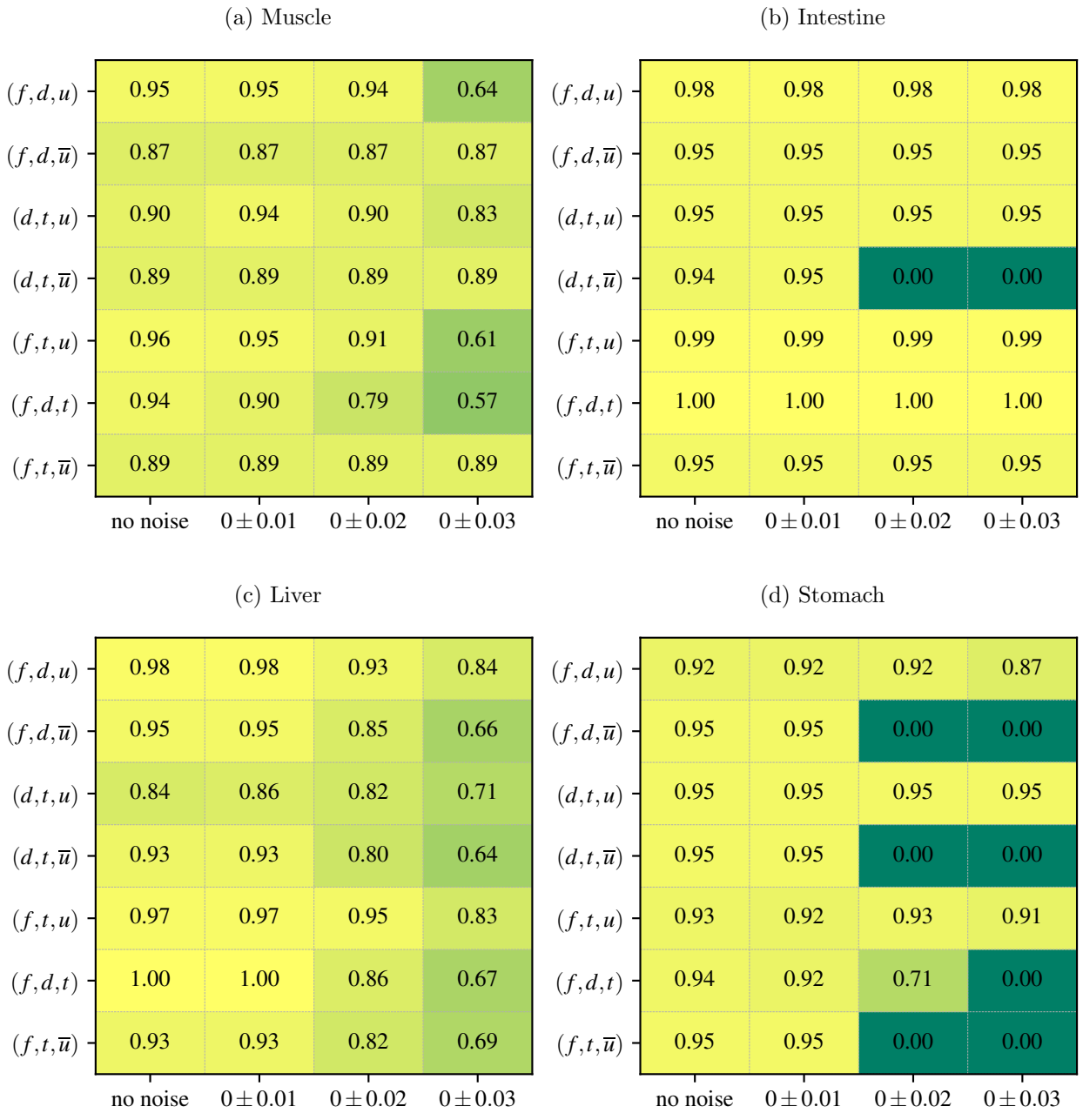



Figure 7.7: Average F1 score with additive noise for CNN–GRU model for a combination of three sensors

would be required to capture more variety of the signal. However, AD5933 SoC does not provide a broader swipe of frequencies and requires more power. To swipe more frequencies, the analog front end must be designed to work with AD5933 SoC.

Optical sensing was also experimented with a combination of responses from IR and RED LEDs. A combination of IR and RED LED responses gives 55 % accuracy. When training the deep learning model on separate IR and RED LEDs, IR LED provides 54 % accuracy, and RED LED gives 48 % accuracy. Such low response could be explained by high moisture inside of the body that can cause reflections, small available amount of optical data and the short length of the signals: only 150 samples length to compare with any other modalities (700+ samples/signal). Potentially adding more data and increasing the length of collected waveforms will increase accuracy and F1 score.

# Chapter 8

# DISCUSSION

The Smart Grasper multimodal instrument was upgraded with 1D ultrasound, two-electrode bioimpedance, 2-wavelength optical sensing, in-direct force sensor, jaw displacement, and temperature sensors. The motivation to have so many sensors was to test different modalities in the same environment and choose the best ones or combinations of modalities for specific surgical tasks: abnormality detection inside of tissue and *ex-vivo*, *in-situ* and *in-vivo* tissue identification.

The short distances that the Grasper is operating caused the ringdown artifact on the received 1D ultrasound signal. The ringdown artifact blended with the echo signal and made it impossible to define the time of flight, one of the significant characteristics of the medium. During the work, classical and deep learning methods were tried to clear the ultrasound signal from the artifact, such as the Bandpass filter, Adaptive LMS filter, Spectrum Suppression filter, RNN, LSTM, and GRU DL models. Over all of these applied methods, the Bandpass filter was chosen to continue to use in further tissue classification tasks due to the simplicity of utilization without re-training the filter with every new surgical task and medium.

Deep Learning CNN–GRU model was chosen for tissue identification tasks because it was used previously for other biomedical signals and has a fast training time to compare with Transformers models. CNN–GRU was trained with different combinations of the sensor modalities to observe if the multimodal system will help with accuracy and F1 score. However, we observed during the work that multimodality does not always help the model improve the accuracy and F1 score, and sometimes, some of the modalities mislead the model and deteriorates its performance. However, multimodality helps to enhance the robustness of the applied random noise.

Force and ultrasound sensors are the most successful modalities for abnormality detection and tissue identification in lab bench and UW Harborview Medical Center experiments. Moreover, raw and filtered ultrasound were evaluated, and we can conclude that the CNN–GRU model is more robust with filtered ultrasound than raw. When the CNN–GRU was trained on *in-situ* data and tested on *in-vivo* data, all modalities failed, except the filtered ultrasound modality. However, the CNN– GRU model, trained and tested on data from the same domain, can learn from raw ultrasound signals and sometimes performs better accuracy and F1 score than filtered. This result means that the ringdown artifact from the ultrasound has specific features unique to the medium.

Optical sensing requires further research, collecting more data, longer sequences, and using frequency spectrum analysis instead of the time series approach. The CNN–GRU model must be modified when frequency spectrum analysis is used. Similar to image processing, the first 1-D convolutional layers need to change to 2-D convolutional layers.

Bioimpedance is a surface measurement sensor that can't penetrate the tissue like optical or ultrasound sensors. Moisture significantly impacts bioimpedance readings, so we measured blood inside the body instead of the surface of the organs. Potentially wiping the tissue before bioimpedance usage can avoid the moisture influence on the data.

# Chapter 9

# CONCLUSION

The Smart Grasper laparoscopic instrument was upgraded during the dissertation to evaluate multiple miniature sensors for different surgical tasks. The surgical tasks involved simultaneous abnormality detection inside of tissue and tissue identification for *in-situ* and *in-vivo* animal experiments. The Smart Grasper has force, temperature, jaw displacement, bioimpedance, and optical sensors on a small area of flex PCB. Due to space limitations, a 1D ultrasound occupied another Grasper's jaw from another sensor's modalities. Peripheral onboard signal processing electronics were also developed to process the signals from the sensor set to the computer.

Deep learning and signal processing methods were applied to denoising ultrasound waveforms from the ringdown artifact that blends with the acoustic echo. Although RNN is one of the best methods, the Bandpass filter was chosen to denoise signals collected during animal experiments because it's easier to use and doesn't require re-training to different mediums.

The data for abnormality detection and tissue identification was collected in *in-situ*, *ex-vivo*, and *in-vivo* animal experiments from chicken and pig tissue. The CNN–GRU model was trained on the data, showing that it can successfully detect abnormality inside the tissue and identify up to four different tissues with 90 % accuracy.

The research showed that ultrasound and force sensors are the best modalities for abnormality detection and tissue classification because they sense density and tissue compressibility. The combination of force and ultrasound sensors gives 96 % accuracy for *in-situ* porcine tissue classification and 93 % accuracy for abnormality detection in *ex-vivo* chicken tissue. Optical sensing would require further research to develop signal processing methods and onboard electronics to collect longer sequences to increase accuracy and F1 score.

Bioimpedance was influenced by moisture inside the body and on the surface of the organs, and tissue was not classified correctly with deep learning training. Jaw displacement and temperature sensors are highly influenced by tissue size, shape, and grasp time. These sensor modalities should be carefully used because they can mislead deep learning models.

When paired with deep learning algorithms to denoise and fuse the data, such an instrument can detect tissue properties in real-time, help with faster diagnosis, and lower the probability of damaging healthy tissue. The purpose of a multimodal instrument could have a broader medical impact than described in the dissertation. For instance, the addition of a miniature ultrasound transducer to surgical grasper, described in the dissertation, opened possibilities for the development of new diagnostic equipment that wasn't possible due to the ringing artifact and the need to know the echo traveled distance. The dissertation reveals that to distinguish between different tissue types (and potentially, cancerous and healthy tissue), we don't need to know the echo traveled distance, and we can use classical or AI-based algorithms to clear the acoustic signals from the ringing artifact. The dissertation demonstrates that ultrasound could operate in very short distances. It also opens the room for an improvement in ultrasound imaging probes. Filtering algorithms developed throughout the dissertation could also be used for ultrasound imagining technologies.

Multimodal datasets collected during the experiments at UW Harborview Medical Center has an interest for AI community for the development of new multimodal deep learning models for processing and fusing biomedical time series signals.

# Chapter 10
# FUTURE WORK

The Smart Grasper was significantly reworked over the dissertation. However, there are many things that could be improved for future generations of the instrument in hardware, software, and data processing.

### *10.0.1 Hardware Improvement*

One aspect that could be improved in the ultrasound subsystem is the integration of functionality on one PCB board. We used Analog Discover 2.0 as a data acquisition tool and as external clock. Instead, one PCB could be developed responsible for the ultrasound subsystem. This PCB will include a 6 MHz external clock, TDC1000 SoC as a transducer driver, high-speed ADC with at least 10 MSp (e.g., LTC2386-16) to accommodate high carrier frequency (6 MHz) of the transducer. Integrating more functions onto a single PCB will also help with the integrity of the ultrasound signals. Piezo crystal transducer could be placed on flex PCB instead of glueing it directly to a Grasper's jaw to avoid long bulky wiring.

The optical subsystem's Analog Front End could be embedded in the Motor Controller board. The MSP430 Smart Combo, used for photodiode signal measurement, could be programmed via a different microcontroller and communicated via UART instead of the TI native microcontroller. This microcontroller could be used for multiple sensor modalities, not only optical sensing.

The bioimpedance subsystem could also be embedded in the Motor Controller board. Careful circuit design and layout are needed to decouple it from the power from another subsystem due to high surge current or add a current-limiter circuit to avoid current surges in the circuit. Utilizing surface electrodes printed on flex PCB could improve the signal's integrity.

However, it is most likely required to wipe the organ/tissue before using a bioimpedance sensor to avoid moisture influencing the results.

The motor could be substituted to more efficient one, and consuming less power.

### *10.0.2 Data Processing Improvement*

Saving data could be improved by saving in multidimensional *numpy* arrays instead of *.csv* format. Then *numpy* arrays can be saved in binary format to speed up and save the disk space. Preprocessing procedures could be improved, and parallel computing could be explored more. GPUs could also be included to run real-time deep learning algorithms and active learning algorithms. New multimodal deep learning models, such as hybrid CNN–Transformer neural networks, could be explored and compared with CNN–GRU model. The Transformer neural network model was tested on the data during this work and got promising results. However, the Transformer has almost tripled the training time on GPUs. Active learning could also be explored, where new data could be labeled and added for training during the data collection procedure.

### *10.0.3 Experiments*

The Smart Grasper could be improved with a better enclosure to avoid moisture trapped in the electronics box. A graphical user interface could be developed for the device to show the real-time data collection process. However, the data should be displayed at a much lower rate than the collected data. Usability research must be conducted among surgeons and medical workers to understand what data makes sense to display in real-time.

# Appendix A

# HANDLER HUB SCHEMATIC AND BOARD LAYOUT

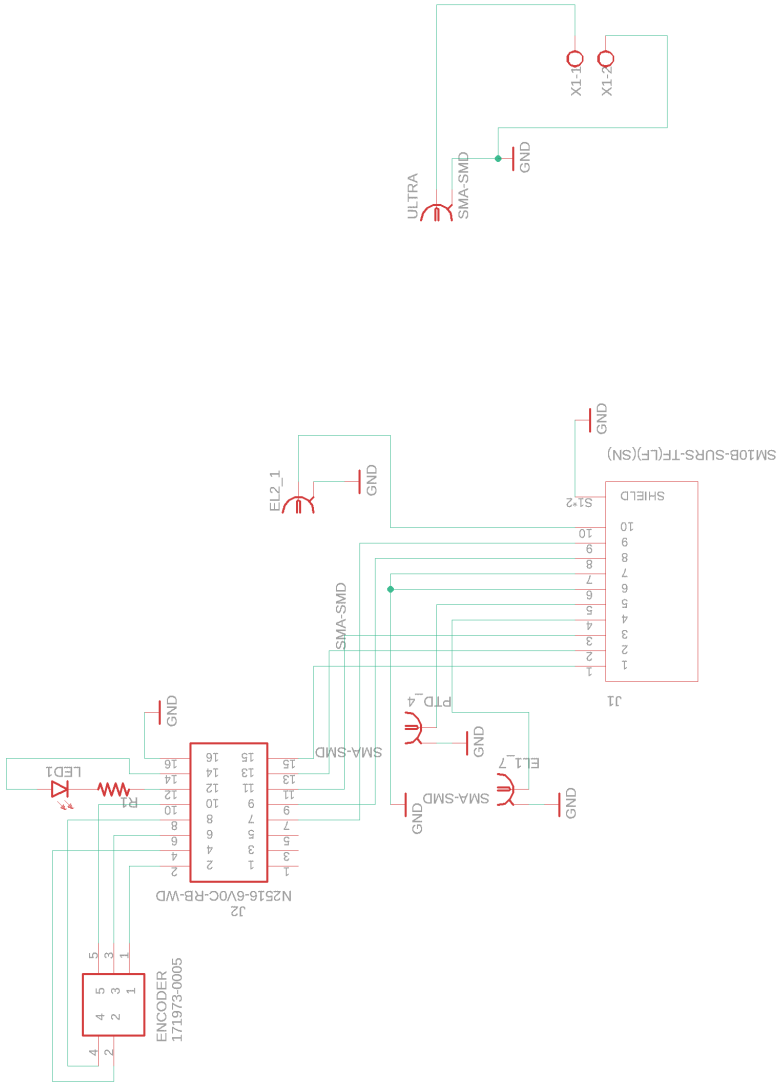



Figure A.1: Handler Hub Schematic

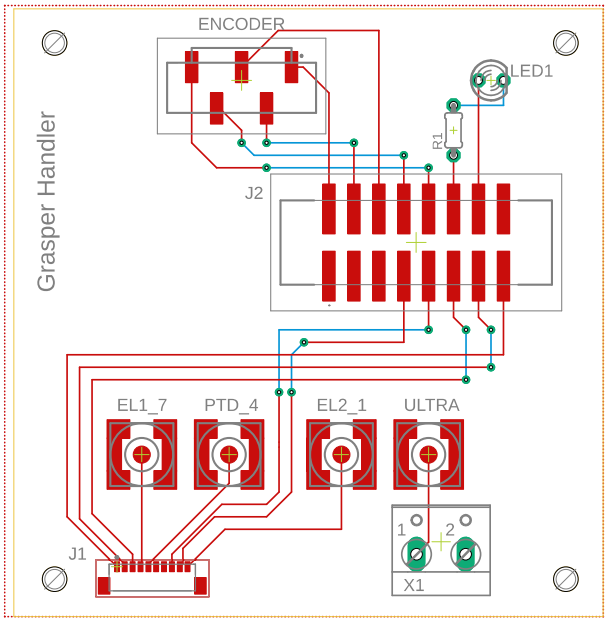



Figure A.2: Handler Hub Board Layout

# Appendix B

## SENSOR'S SET SCHEMATIC AND BOARD LAYOUT

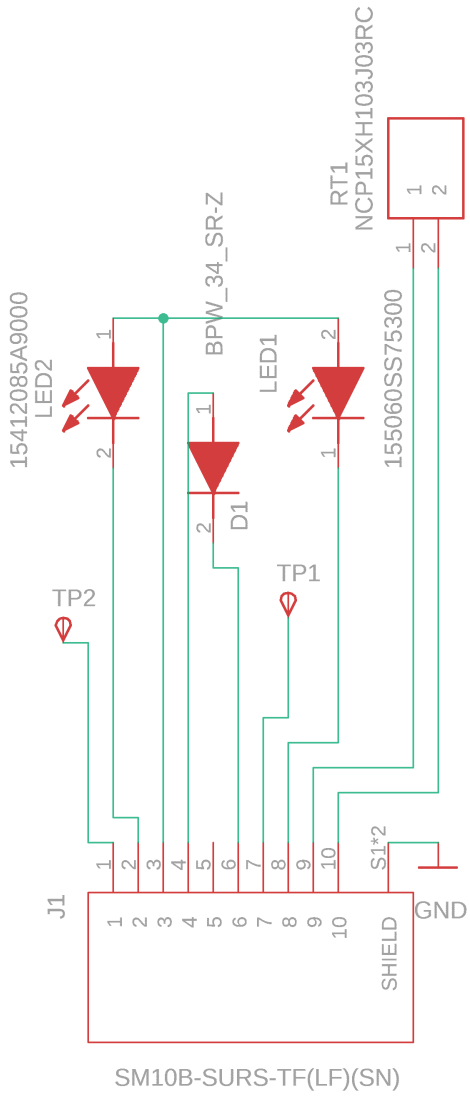



Figure B.1: Schematic for Sensor's set

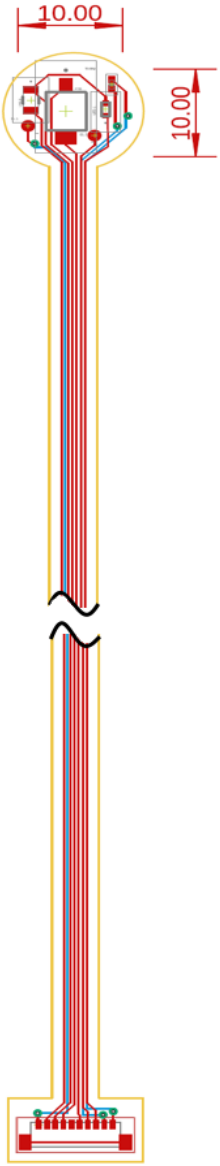



Figure B.2: Flex PCB Board Layout for Sensor's Set

# Appendix C

# MOTOR CONTROLLER AND MIXED SIGNAL FRONT END FOR SENSOR'S INTERFACE

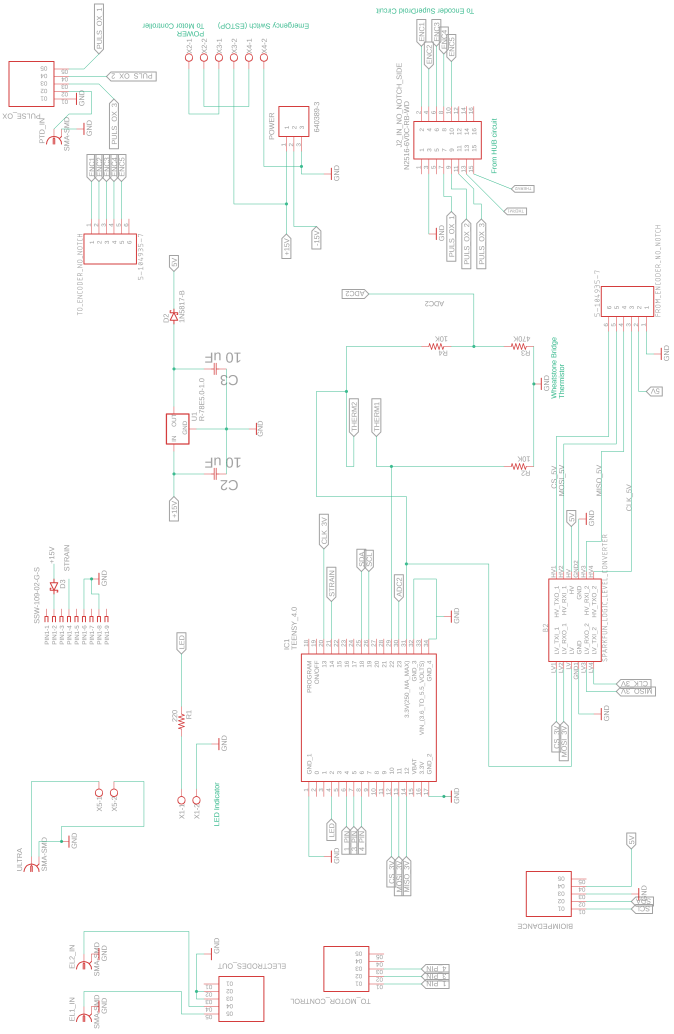


Figure C.1: Schematic for Motor Controller and Mixed Signal Front End for Sensor's Interface

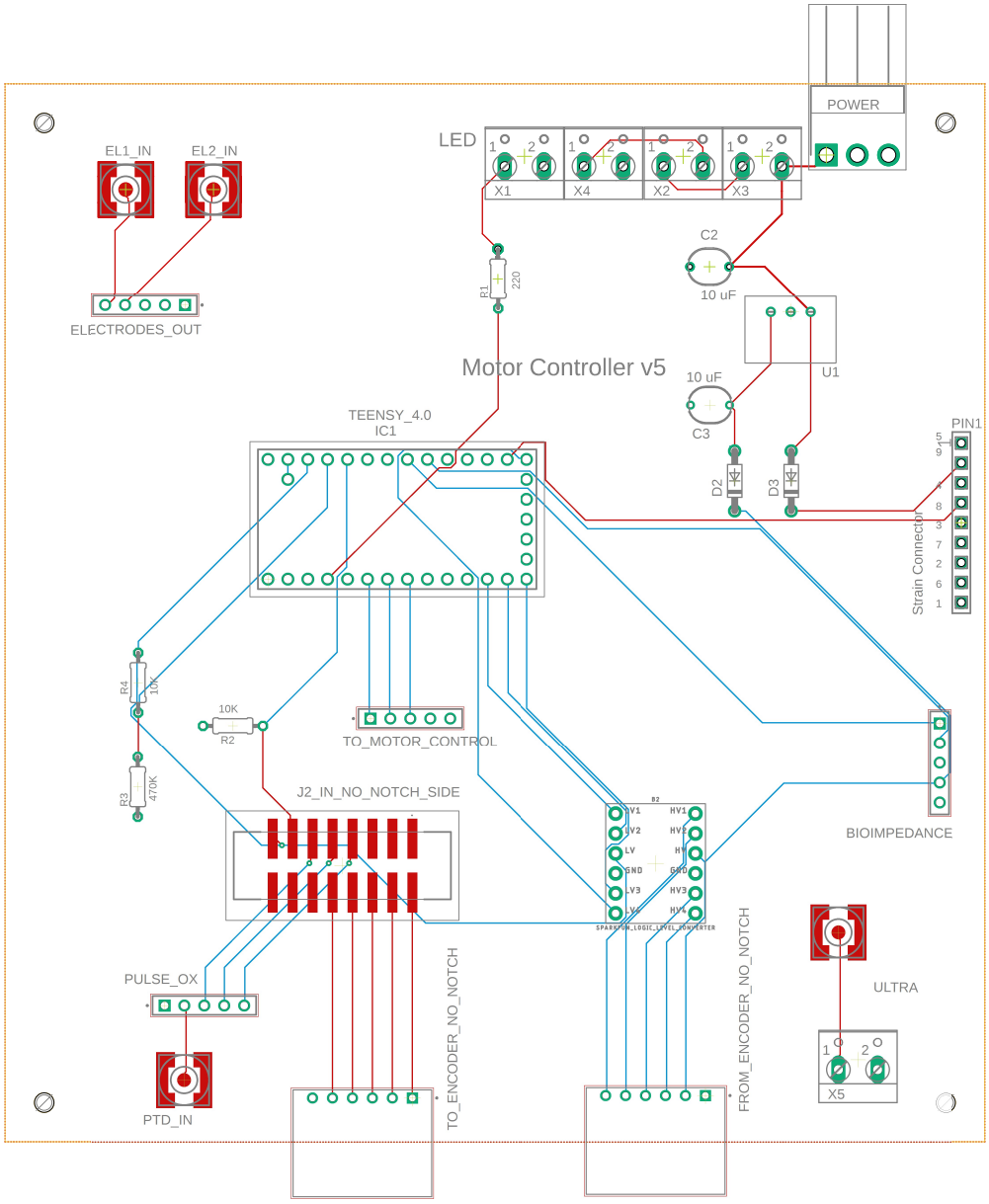



Figure C.2: Board Layout for Motor Controller and Mixed Signal Front End for Sensor's Interface

# Appendix D

# OPTICAL SENSING ANALOG FRONT END SCHEMATIC AND BOARD LAYOUT

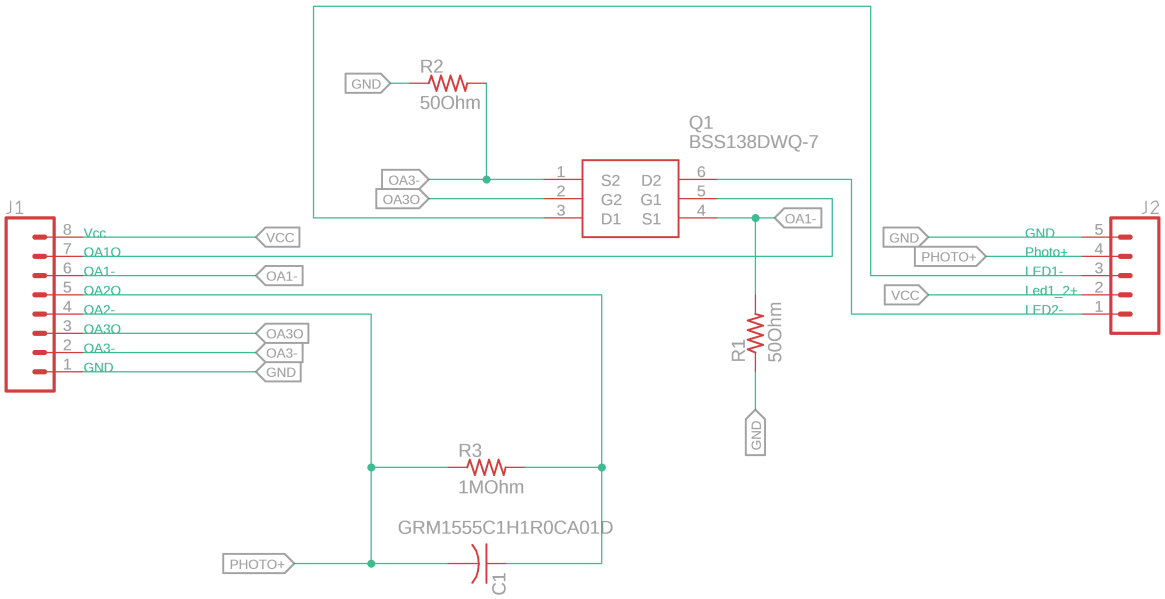



Figure D.1: Schematic for Analog Front End for Optical Sensing

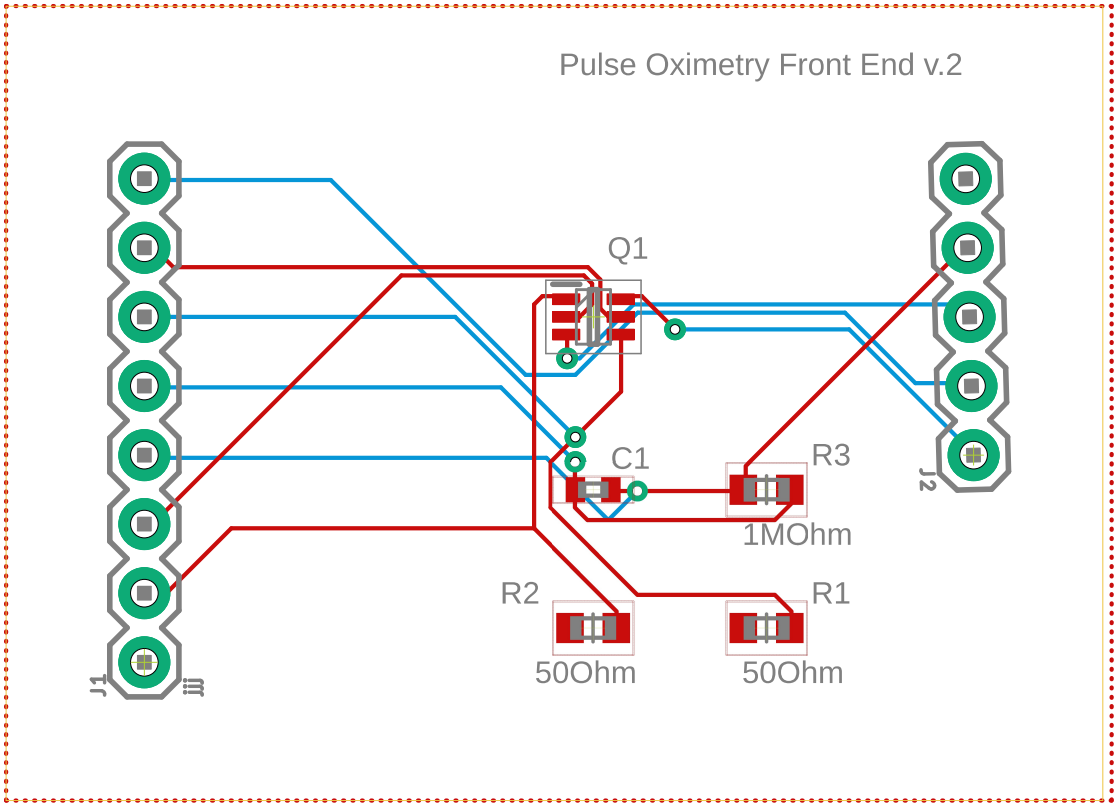



Figure D.2: Board Layout for Analog Front End for Optical Sensing

# BIBLIOGRAPHY

[1] T. D. Nagy and T. Haidegger, "Recent advances in robot-assisted surgery: Soft tissue contact identification," in *Proc. SACI*, Timisoara, Romania, May 29–31, 2019, pp. 99–106.

[2] S. Schostek, M. O. Schurr, and G. F. Buess, "Review on aspects of artificial tactile feedback in laparoscopic surgery," *Med. Eng. Phys.*, vol. 31, no. 8, pp. 887–898, Oct. 2009.

[3] J. Konstantinova, A. Jiang, K. Althoefer, P. Dasgupta, and T. Nanayakkara, "Implementation of tactile sensing for palpation in robot-assisted minimally invasive surgery: A review," *IEEE Sens. J.*, vol. 14, no. 8, pp. 2490–2501, Aug. 2014.

[4] P. R. Roan, A. S. Wright, T. S. Lendvay, M. N. Sinanan, and B. Hannaford, "An instrumented minimally invasive surgical tool: Design and calibration," *Appl. Bionics Biomech.*, vol. 8, pp. 173–190, 2011, Art. no. 940867.

[5] T. Mitchell, B. W. Newton, R. R. Terreberry, and M. X. VanCura, "Video dissections to determine the usefulness of synthetic cadavers vs. real cadavers," *J. Hum. Anat. Phys. Soc.*, vol. 20, no. 3, pp. 115–119, 2016.

[6] Y. Sosnovskaya, "Abnormality in chicken tissue," IEEE Dataport, 2024. [Online]. Available: https://doi.org/10.21227/87d5-kd86

[7] ——, "Ultrasound waveforms with and without ringdown artifacts," IEEE Dataport, 2022. [Online]. Available: https://doi.org/10.21227/z6v5-mf23

[8] F. A. Duck, *Physical properties of tissues.* San Diego, CA, USA: Academic Press, 1990.

[9] C. Rossmann and D. Haemmerich, "Review of temperature dependence of thermal properties, dielectric properties, and perfusion of biological tissues at hyperthermic and ablation temperatures," *Crit. Rev. Bioeng.*, vol. 42, no. 6, pp. 467–492, 2014.

[10] F. S. Knox III, T. L. Wachtel, G. R. McCahan, and S. C. Knapp, "Thermal properties calculated from measured water content as a function of depth in porcine skin," *Burns*, vol. 12, no. 8, pp. 556–562, 1986.

[11] K. Koskensalo, J. Raiko, T. Saari, V. Saunavaar, O. Eskola, P. Nuutila, J. Saunavaara, R. Parkkola, and V. K. A, "Human brown adipose tissue temperature and fat fraction are related to its metabolic activity," *Am. J. Clin. Nutr.*, vol. 4, no. 102, pp. 1200–1207, 2017.

[12] J. B. Bussemaker and J. Lindeman, "Comparison of methods to determine viability of small intestine," *Ann. Surg.*, vol. 176, no. 1, pp. 97–101, 1972.

[13] C. H. Seo, Y. Shi, S.-W. Huang, K. Kim, and M. O'Donnell, "Thermal strain imaging: a review," *Interface Focus*, vol. 1, 2011.

[14] F. R. Pereira, J. C. Machado, and F. S. Foster, "Ultrsound characterization of coronary artery wall in vitro using temperature-dependent wave speed," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 11, pp. 1474–1485, 2003.

[15] M. I. Daoud, P. Mousavi, F. Imani, R. Rohling, and P. Abolmaesumi, "Tissue classification using ultrasound-induced variations in acoustic backscaterring features," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, 2013.

[16] T. E. Doyle, A. P. Butler, M. J. Salisbury, M. J. Bennett, G. M. Wagner, H. A. Al-Ghaib, and C. B. Matsen, "High-frequency ultrasonic forceps for the in vivo detection of cancer during breast-conserving surgery," *ASME. J. Med. Devices.*, vol. 14, no. 3, Sep. 2020.

[17] E. J. M. Baltussen, H. J. C. M. Sterenborg, T. J. M. Ruers, and B. Dashtbozorg, "Optimizing algorithm development for tissue classification in colorectal cancer based on diffuse reflectance spectra," *Biomed. Opt. Express*, vol. 10, no. 12, 2019.

[18] F. J. Carter, T. G. Frank, P. J. Davies, D. McLean, and A. Cuschieri, "Measurements and modelling of the compliance of human and porcine organs," *Med. Image Anal.*, vol. 4, no. 5, pp. 231–236, 2001.

[19] M. P. Ottensmeyer and J. K. S. Jr., "In-vivo mechanical tissue property measurement for improved simulations," *Proc. SPIE*, vol. 4037, 2000.

[20] J. Rosen, J. D. Brown, S. De, M. Sinanan, and B. Hannaford, "Biomechanical properties of abdominal organs in vivo and postmortem under compression loads," *J. Biomech. Eng.*, vol. 130, no. 2, Apr. 2008, Art. no. 021020.

[21] A. Bosy-Westphal, S. Danielzik, R.-P. Dörhöfer, A. Piccoli, and M. J. Müller, "Pattern of biolectrical impedance vector distribution by body-mass index and age: implications for body-composition analysis," *Am. J. Clin. Nutr.*, vol. 82, no. 1, pp. 60–68, 2005.

[22] K. Ain, S. Soelistiono, R. A. Wibowo, and L. Muniroh, "Measurement of cholesterol concentration based on bioimpedance with AD5933-EVAL," in *Proc. ICICI-BME*, Bandung, Indonesia, November 6–7, 2017, pp. 251–254.

[23] P. S. H. Jose, K. Rajesekaran, P. Rajalakshmy, and B. Jebastine, "A non-invasive method for measurement of blood glucose using bio impedance technique," in *Proc. ICSPC*, Coimbatore, India, March 29–30, 2019, pp. 138–142.

[24] Y. Zou and Z. Guo, "A review of electrical impedance techniques for breast cancer detection," *Med. Eng. Phys.*, vol. 25, pp. 79–90, 2003.

[25] M. J. Buono, S. Burke, S. Endemann, H. Graham, C. Gressard, L. Griswold, and B. Michalewicz, "The effect of ambient air temperature on whole-body bioelectrical impedance," *Physiol. Meas.*, vol. 25, no. 1, pp. 119–123, 2004.

[26] P. R. Roan, "An instrumented surgical tool for local ischemia detection," Ph.D. Dissertation, Dept. Elect. Eng., Univ. Washington, Seattle, WA, USA, Apr. 2011.

[27] A. Sie, M. Winek, and T. M. Kowalewski, "Online identification of abnormal tissues in vivo for tissue-aware and injury-avoiding surgical robots," in *Proc. IROS*, Chicago, IL, USA, September 14–18 2014, pp. 2036–2042.

[28] A. Sie, "Online identification of abdominal tissues during grasping using an instrumented laparoscopic grasper," Master's Thesis, Dept. Mech. Eng., Minneapolis, Univ. Minnesota, MN, USA, 2013.

[29] S. Sokhanvar, M. Packirisamy, and J. Dargahi, "MEMS endoscopic tactile sensor: Toward in-situ and in-vivo tissue softness characterization," *IEEE Sens. J.*, vol. 9, no. 12, pp. 1679–1687, Dec. 2009.

[30] K. Sun, M. Li, S. Wang, G. Zhang, H. Liu, and C. Shi, "Development of a fiber Bragg grating-enabled clamping force sensor integrated on a grasper for laparoscopic surgery," *IEEE Sens. J.*, vol. 21, no. 15, pp. 16 681–16 690, Aug. 2021.

[31] G. S. Fischer, T. Akinbiyi, S. Saha, J. Zand, M. Talamini, M. Marohn, and R. Taylor, "Ischemia and force sensing surgical instruments for augmenting available surgeon information," in *Proc. BioRob*, Pisa, Italy, Feb. 20-22, 2006, pp. 1030–1035.

[32] G. Tholey and J. P. Desai, "A modular, automated laparoscopic grasper with three-dimensional force measurement capability," in *Proc. ICRA*, Rome, Italy, Apr. 10–14, 2007, pp. 250–255.

[33] D.-Y. Seok, Y. B. Kim, U. Kim, S. Y. Lee, and H. R. Choi, "Compensation of environmental influences on sensorized-forceps for practical surgical tasks," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 2031–2037, Apr. 2019.

[34] N. T. Burkhard, M. R. Cutkosky, and J. R. Steger, "Slip sensing for intelligent, improved grasping and retraction in robot-assisted surgery," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 4148–4155, 2018.

[35] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal deep learning for biomedical data fusion: a review," *Brief. Bioinform.*, vol. 23, no. 2, Mar. 2022.

[36] E. A. M. Heijnsdijk, H. de Visser, J. Dankelman, and D. J. Gouma, "Slip and damage properties of jaws of laparoscopic graspers," *Surg. Endosc.*, vol. 18, no. 6, pp. 974–979, Jun. 2004.

[37] J. D. Brown, J. Rosen, M. Moreyra, M. Sinanan, and B. Hannaford, "Computer-controlled motorized endoscopic grasper for in vivo measurement of soft tissue biomechanical characteristics," in *Proc. MMVR 02/10*, ser. Studies in Health Technology and Informatics, vol. 85, 2002, pp. 71–73.

[38] A. A. Pena, "A feasibility study of the suitability of an AD5933-based spectrometer for EBI applications," Master's Thesis, School of Engineering, University of Borås, Borås, Sweden, 2009.

[39] J. Kaplan, Y. Sosnovskaya, M. Arnold, and B. Hannaford, "Sensor fusion for force and position calibration of a motorized surgical Smart Grasper," in *Proc. ISMR*, 2021, pp. 1–7.

[40] X. Qu, T. Azuma, H. Lin, H. Takeuchi, K. Itani, S. Tamano, S. Takagi, and I. Sakuma, "Limb muscle sound speed estimation by ultrasound computed tomography excluding receivers in bone shadow," *Proc. SPIE*, vol. 10139, pp. 313–320, Mar. 2017.

[41] M. Wirtz and M. Hildebrandt, "IceShuttle Teredo: An ice-penetrating robotic system to transport an exploration AUV into the ocean of Jupiter's moon Europa," in *Proc. IAC*, Guadalajara, Mexico, Sep. 26–30, 2016.

[42] S. Wagle and H. Kato, "Real-time measurement of ultrasonic waves at bolted joints under fatigue testing," *Exp. Mech.*, vol. 51, no. 9, pp. 1559–1564, Nov. 2011.

[43] J. C. Norton, P. R. Slawinski, H. S. Lay, J. W. Martin, B. F. Cox, G. Cummins, M. P. Y. Desmulliez, R. E. Clutton, K. L. Obstein, S. Cochran, and P. Valdastri, "Intelligent magnetic manipulation for gastrointestinal ultrasound," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019.

[44] C. J. Barlow, R. J. Dickinson, and R. I. Kitney, "Medical ultrasound imaging," U.S. Patent 5 601 082, Feb. 11, 1997.

[45] S. F. Boll, "Suppression of acoustic noise in speech using spectral substraction," *IEEE T. Acoust. Speech.*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[46] F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2018.

[47] L. Svilainis, "Review of high resolution time of flight estimation techniques for ultrasonic signals," in *Proc. NDT*, Telford, England, UK, Sep. 10–12, 2013, pp. 1–12.

[48] J. F. Figueroa, "An ultrasonic ranging system for robot end-effector position measurement," Ph.D. dissertation, Dept. Mech. Eng., Penn. St. Univ., State College, PA, USA, 1988.

[49] L. Jia, B. Xue, S. Chen, H. Wu, X. Yang, J. Zhai, and Z. Zeng, "A high-resolution ultrasonic ranging system using laser sensing and a cross-correlation method," *Appl. Sci.*, vol. 9, no. 7, Apr. 2019, Art. no. 1483.

[50] Z. Lu, F. Ma, C. Yang, and M. Chang, "A novel method for estimating time of flight of ultrasound echoes through short-time Fourier transforms," *Ultrasonics*, vol. 103, Apr. 2020, Art. no. 106104.

[51] G. Yao, X. Mao, N. Li, H. Xu, X. Xu, Y. Jiao, and J. Ni, "Interpretation of electrocardiogram heartbeat by CNN and GRU," *Comput. Math. Methods in Med.*, Aug. 2021.

[52] J. S. Bang and S. W. Lee, "Motor imagery classification based on CNN-GRU network with spatio-temporal feature representation," in *Proc. ACPR*, Jeju Island, South Korea, November 9–12, 2022, pp. 104–115.

[53] L. Lu, C. Zhang, K. Cao, T. Deng, and Q. Yang, "A multichannel CNN-GRU model for human activity recognition," *IEEE Access*, vol. 10, pp. 66 797–66 810, Jun. 2022.