Electrical Impedance Measurements in the Gastrointestinal Tract
Cancerous or Non-Cancerous?

Master of Science Thesis
For the degree of Master of Science in Biomedical Engineering at The Delft University of Technology

J.W.G. Remmerswaal
4387872

First Supervisor: Prof. Dr. J. Dankelman
Second Supervisor: Prof. Dr. B.H.W. Hendriks
Third Committee member: Prof. Dr. R. Dekker

In collaboration with
The Erasmus Medical Center and Phillips Eindhoven

3mE
Delft University of Technology
The Netherlands
August 16, 2023

Contents

List of Figures

List of Tables

Abstract

Objective: This work focuses on the feasibility of using Electrical Impedance Spectroscopy (EIS) to differentiate between cancerous and healthy tissues in real-time alongside the GI tract. This is to see whether it is possible to support surgeons when taking biopsies, or to see if it is possible to make biopsies obsolete with EIS. **Methods**: The study is limited to three tissue types representing the GI tract. These tissues are the esophagus, ileum, and colon tissue. Impedance measurements are taken with a 4-electrode probe on a Hewlett Packard 4192A LF HP Impedance Analyzer. All measurements are taken over a frequency range of 1 kHz to 7 MHz in 300 steps. Seven patients are included: three esophagus, two ileum, and two colon. Forty-eight ex vivo measurements are taken; 32 are on healthy tissue and 16 on cancerous tissue. Almost all measurements are verified by histological assessment (golden standard). In this work, a combination of parameterization with a classification method is made to create a classification strategy. These can be listed as the Cole impedance model, the two-pole Cole impedance model, the two-pole Cole impedance model in combination with a Constant Phase Element (CPE), and PCA. The Cole impedance models are combined with thresholding, and the PCA is combined with a SVM. **Results**: The thresholding algorithm is created in combination with the α_2 parameter ($P < 0.05$) from the two-pole Cole impedance model in combination with a CPE. The thresholding value is determined in a LOOCV approach via ROC curves created to search for the threshold that gives the most significant summation of sensitivity and specificity. In the combination of PCA and SVM, the PCA uses three principal components (containing 96.37% of the total variance), and the SVM applies an Radial Basis Function (RBF) kernel. For the 2-pole Cole impedance model with CPE in combination with thresholding, we find an accuracy 0.5208, sensitivity 0.4375, specificity 0.5625, Positive Predictive Value (PPV) 0.3333, Negative Predictive Value (NPV) 0.6666, and Mathews Correlation Coefficient (MCC) 0.0000. For the PCA in combination with the SVM, we find an accuracy of 0.4167, a sensitivity of 0.0000, specificity of 0.6250, PPV of 0.0000, NPV of 0.5556, and MCC -0.4082. **Conclusion**: It is concluded that we cannot create an algorithm that either supports surgeons or replaces a biopsy in the GI tract with the current setup. This is highly likely due to the extremely low amount of data that is included to train the algorithm.

Acknowledgements

I would like to thank my supervisors Jenny & Benno for their support and feedback during the thesis period. Additionally, I would like to thanks Tessa, Mahesh, and Anh for their support at the Erasmus MC, and at last I would like to express my gratitude to Ronald for being the third committee member.

Delft, University of Technology
August 16, 2023

J.W.G. Remmerswaal

“Het begint met willen.”

— *Gerard Remmerswaal*

Glossary

List of Acronyms

EIS	Electrical Impedance Spectroscopy
EC	equivalent circuits
AUC	Area under the ROC Curve
ROC	Receiver Operator Characteristic
CI	Confidence Interval
SVM	support vector machine
PCA	principal component analysis
SVD	Singular Value Decomposition
RBF	Radial Basis Function
GI	gastrointestinal
CPE	Constant Phase Element
LOOCV	Leave-One-Out Cross Validation
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
PPV	Positive Predictive Value
NPV	Negative Predictive Value
MCC	Mathews Correlation Coefficient
MSE	Mean Square Error

Chapter 1

Introduction

This introductory chapter defines the problem-statement & motivation for this thesis work, which transformed into several research questions that are posed in the objectives for this work. The introduction is concluded with the outline of this report document.

1-1 Problem-Statement and Motivation

Combining a biopsy in the gastrointestinal (GI) tract with cytology renders the highest detection rates for cancer [45, 20]. During this procedure, a surgeon uses an endoscope to remove a small tissue sample from the affected area. After this, a pathologist looks at the specimen and tries to find unique cancer cell patterns under a microscope.

However, it can be challenging to determine where to take the biopsy from (as sometimes cancer cells are spread and not visible), and therefore, using an endoscope can be seen as a limited technique where the surgeon needs more guidance toward achieving the best biopsy. Additionally, taking a biopsy is rather time-consuming, as multiple specialized individuals are needed to determine whether the tissue specimen is cancerous or non-cancerous. Thus, real-time feedback on whether the tissue is cancerous or non-cancerous during a biopsy in the GI tract could be a solution that tackles both issues.

Electrical Impedance Spectroscopy (EIS) is a minimally-invasive, quantitative method for determining cellular changes in tissue as a function of frequency, which can potentially facilitate real-time classification between cancerous and healthy tissue, resulting in better procedural outcomes [35].

Along the entire GI tract, the walls have the same four primary layers (mucosa, submucosa, muscularis externa, and serosa) [24], it can be hypothesized that the EIS measurement results yield similar outcomes for healthy tissue along the GI tract, making it possible for classification of healthy versus non-healthy tissue along the entire tract. Therefore, EIS may be a supportive tool when taking biopsies or can even replace biopsies.

Before this work, a literature survey was conducted on using EIS techniques in real-time cancer detection during surgery and on techniques that are most appropriate for the data analysis [35]. This work focuses on the feasibility of using EIS to differentiate between cancerous and healthy tissue identification in real-time alongside the GI tract. It should be noted that we limit ourselves to three tissue types that are represented in the GI tract, which are: esophagus, ileum, and colon tissue. Additionally, this work is a collaboration between the Department of Surgical Oncology and Gastrointestinal Surgery, the Erasmus MC Cancer Institute, Philips, and the TU Delft.

1-2 Objectives

To ensure the relevance of this thesis work, the author has created various research questions in combination with a summary of the question's motivation.

Research Questions

- ***Is it possible to create classification algorithms that can be used on the esophagus, colon, and ileum tissue to detect cancerous versus non-cancerous tissue?***
  By answering this research question, the possibilities within the realm of tissue classification with EIS for tissues of the GI tract is understood.
- ***Which classification strategy has better performance, the more classical Cole Impedance parameterization in combination with thresholding, or the newer principal component analysis (PCA) in combination with a trained support vector machine (SVM) algorithm.***
  The answer to this question shows different techniques to achieve the same end goal. This question is interesting because it tests two different approaches. A more classical approach that uses parameterization via data-fitting with a threshold to determine whether a sample is cancerous or non-cancerous, and a more new-school approach that uses PCA and a SVM algorithm.
- ***Is EIS a technique that can be used in combination with a parameterization and classification algorithm to replace or support biopsies in the GI tract?***
  This question serves as the most critical question that is answered in this thesis work. It is also the most crucial question as it examines the current golden standard technique: a biopsy of the GI tract.

1-3 Document Outline

This document is structured as follows. Chapter 2 discusses the theory that is required to understand this work. It ranges from the physiology of the GI tract to the theory behind EIS. Chapter 3 gives an overview of the methods that are used to perform this research. This section elaborates on the data acquisition, parameterization, and classification methods.

In Chapter 4, the results from the measurements are shared and the performance of the classification strategies are shown. In Chapter 5, a discussion is given on this work, where all the significant findings and the meaning behind those findings are discussed. Finally, Chapter 6 concludes all the work and findings presented in this document. Also, some suggestions regarding future research in this direction are made, where the author points out the main challenges and questions to be answered.

Chapter 2

Theory

This chapter encompasses the theory that is required to create this work. We start off by covering the physiological and clinical theory, followed by more technological theory about Electrical Impedance Spectroscopy (EIS) and how EIS can be used in biological tissue.

2-1 Physiology and Clinical Biopsy of the GI tract

Food is ingested into the digestive system, which breaks it down into nutritional molecules, absorbs them into the bloodstream, and excretes the indigestible waste from the body. In Figure 2-1, one can observe the gastrointestinal (GI) tract, commonly known as the gut. It is a continuous muscular tube that winds through the body, which starts at the mouth and ends at the anus. The alimentary canal (GI tract) consists of the mouth, pharynx, esophagus, stomach, and small and large intestines that lead to the anus [24].

The walls of the alimentary canal have the same four primary layers: mucosa, submucosa, muscularis externa, and serosa. These four can be observed in Figure 2-2. Each layer has a dominating tissue type that contributes in a particular way to the breakdown of food [24].

The inner layer is the mucose that lines the GI tract lumen from the mouth to the anus. Its functions are to secrete mucus, digestive enzymes, hormones, absorb the end nutrients of digestion into the bloodstream, and act as a protection mechanism against infectious disease. The mucus that is produced protects specific digestive organs from being digested themselves. Additionally, it aids in the passage along the entire tract. In the stomach and small intestine, the mucose contains enzyme-synthesizing and hormone-secreting cells. The submucosa lines the mucosa and is a connective tissue that contains many blood and lymphatic vessels, lymphoid follicles, and nerve fibers to supply the surrounding tissue of the GI tract wall. Surrounding the submucosa is the muscularis external. The responsibility of this layer is to ensure segmentation and peristalsis. The muscularis externa typically consists of an inner circular layer and an outer longitudinal layer of smooth muscle cells. This layer acts as valves in specific segments of the GI tract to control the food's propulsion and prevent backflow. In the GI tract's canal organs, the serosa's outermost layer is formed of rigid and flexible tissue

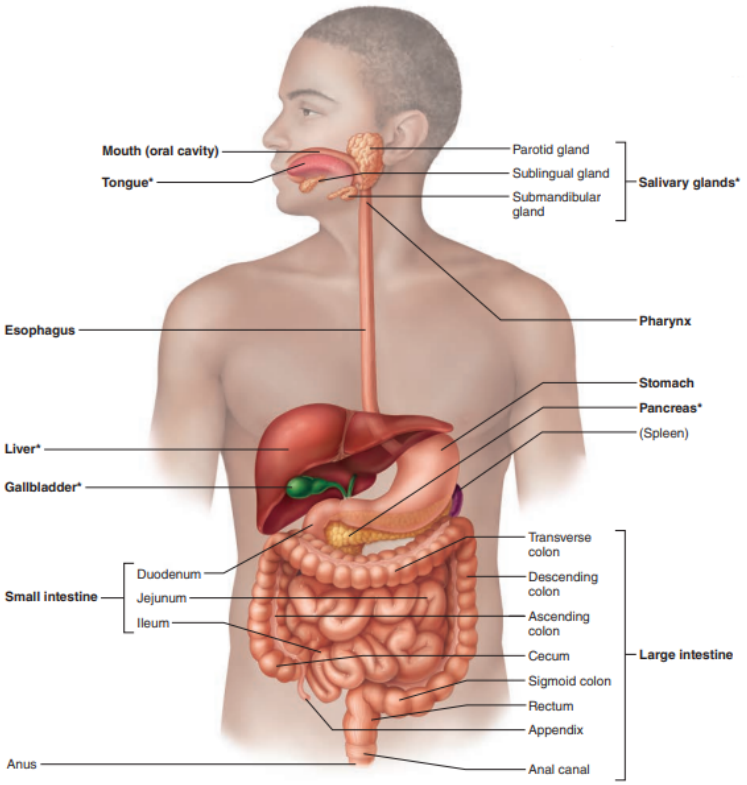


Figure 2-1: The GI tract or gut with all its organs [24].

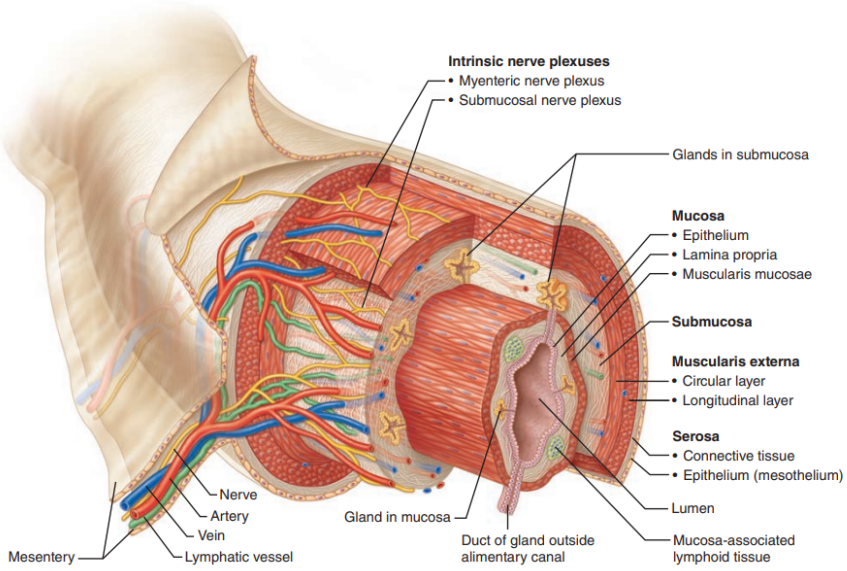


Figure 2-2: The four layers that can be found within the entire GI tract [24].

(areolar connective tissue) covered with a single layer of squamous epithelial cells [24]. For this work we focus on thee parts of the GI tract: the esophagus, the ileum, and the colon, therefore, we will give a more detailed description of of the esophagus, the small intestine and the large intestine.

2-1-1 The Esophagus

The esophagus begins at the pharynx and runs through the diaphragm down to the stomach. The canal is approximately 25 centimeters long and functions as a passageway to the stomach for ingested food. Additionally, the esophagus aids in food transport by antegrade peristaltic contractions that push the food bolus toward the stomach. The esophagus contains two sphincters at the proximal and distal ends of the esophagus. When no food is ingested, the esophagus relaxes, and its upper and lower sphincters are contracted to prevent acid reflux. Additionally, the upper sphincter prevents aerophagia (excessive and repetitive air swallowing) during inhalation. [8].

The esophagus can be divided into the cervical, thoracic, and abdominal parts, which can be found in Figure 2-3. The cervical segment starts at the cricoid cartilage's lower border and finishes at the thoracic inlet of the jugular notch [8]. The cervical segment is positioned between the anterior trachea and posteriorly prevertebral layer of the cervical fascia posteriorly [8]. The thoracic segment is subdivided into three parts: upper, middle, and lower. The upper thoracic esophagus spans the length of the superior mediastinum between the thoracic inlet and the carina [8]. The middle and lower part of the thoracic esophagus lies in the posterior mediastinum and are divided by the middle between the tracheal bifurcation and the esophagogastric junction. The lower esophagus consists of, the lower thoracic esophagus and the abdominal part of the esophagus. The abdominal segment ranges from the diaphragm to the gastro-esophageal junction.

2-1-2 The Small Intestine

The small intestine is the longest part of the GI tract and can be seen as a long twisted tube extending from the pyloric sphincter to the ileocecal valve, where it joins the large intestine. The small intestine is considered the body's major digestive organ as this is where virtually all absorption occurs with the help of bile and pancreatic enzymes [24].

The small intestine is divided into the duodenum, the jejunum, and the ileum. The duodenum is the initial part of the small intestine; it curves around the head of the pancreas and is approximately 25 centimeters long (see Figure 2-1). The jejunum is approximately ten times the size of the duodenum and extends from the duodenum to the ileum. The ileum is approximately 3.6 meters long and joins the large intestine. This happens at the ileocecal valve. Both the jejunum and ileum hang in a coil-like shape in the central and lower part of the abdominal cavity; here, the distal parts of the small intestine are encircled by the large intestine [24].

The primary function of the small intestine is to absorb nutrients. The intestine is highly adapted to this task due to its long length. Additionally, the small intestine has three structural modifications that amplify its absorptive function: circular folds, villi, and microvilli

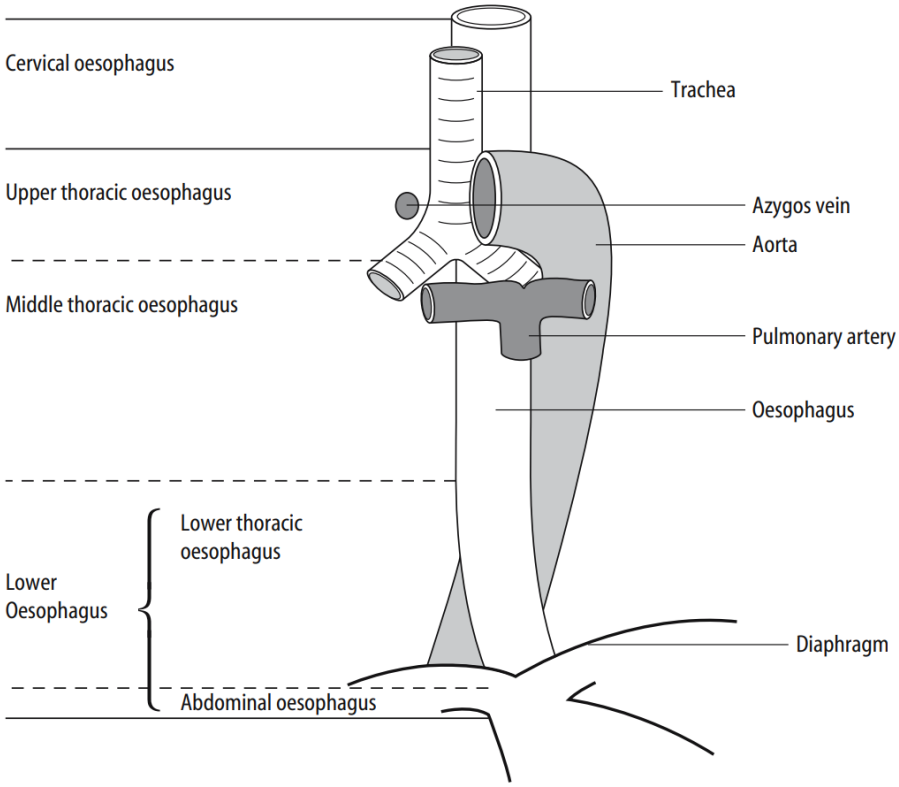


Figure 2-3: The three segments that make up the esophagus [8].

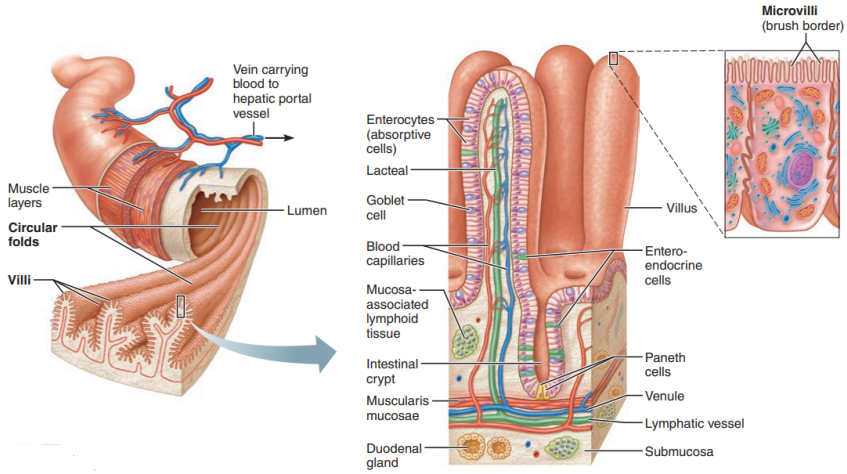


Figure 2-4: The circular fold, villi, and microvilli in the small intestine [24].

(see Figure 2-4) [24]. The folds in the wall of the small intestine are nearly 1-centimeter-deep permanent folds of the mucosa and submucosa, which force chyme (a semi-fluid mass of partly digested food expelled by a person's stomach) to propagate through the lumen spirally. It also slows down the chyme's movement, allowing for more digestion time and better absorption of nutrients in the small intestine. The villi are small extensions of the mucose that are over 1 millimeter in height. The villi are the largest in the duodenum since most of the nutrient absorption occurs here. Their size gradually decreases along the length of the small intestine. The core of a villus contains a dense capillary bed and a wide lymphatic capillary. This allows for optimal absorption of nutrients through the epithelial cells. On the villi, there are long and densely packed cytoplasmic extensions of micro absorptive cells of mucosa that give the mucous surface a furry-like appearance. These extensions are referred to as microvilli. The microvilli's plasma membrane contains brush border enzymes used to complete the digestion of carbohydrates and proteins [24].

2-1-3 The Large Intestine

The large intestine surrounds the small intestine on three sides (see Figure 2-5) [24]. The large intestine starts at the ileocecal valve and ends down at the anus. Its diameter is approximately seven centimeters, meaning it's much wider than the small intestine. However, the large intestine is about 1.5 meters long and much shorter than the 6-meter-long small intestine. The large intestine's main function is to absorb the remaining water of the chyme and to store the semi-solid chyme momentarily. After this, the feces are excreted via the anus and leave the body. Additionally, the large intestine absorbs metabolites (intermediate or end product of metabolism) produced by bacteria in the large intestine. These bacteria ferment carbohydrates that are not absorbed by the small intestine [24].

The large intestine contains three distinct features not seen elsewhere in the body: teniae coli, haustra, and epiploic appendages. The teniae coli is a muscle layer that runs in the middle over the entire span of the large intestine and consists of three bands of smooth

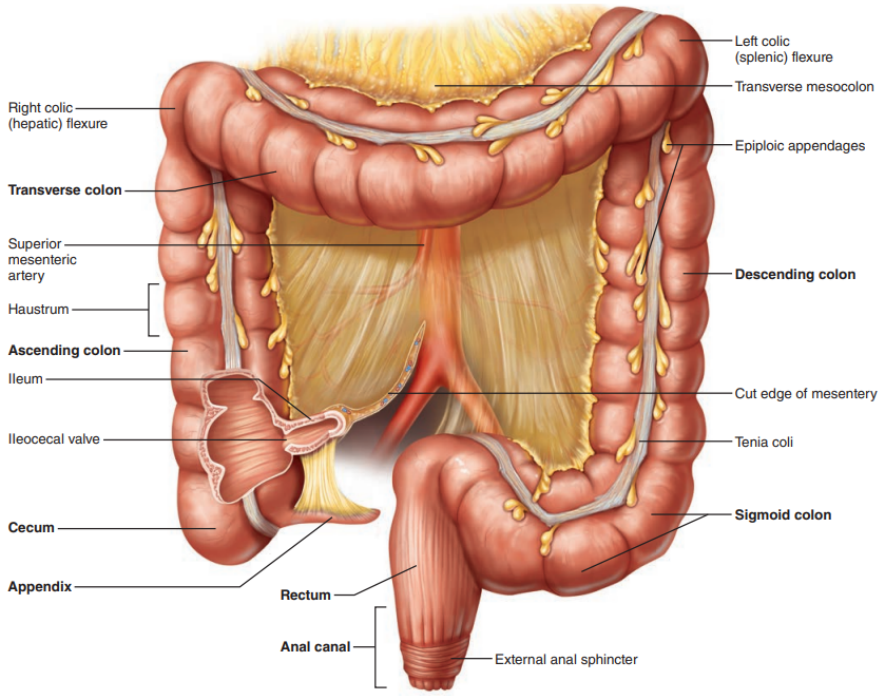
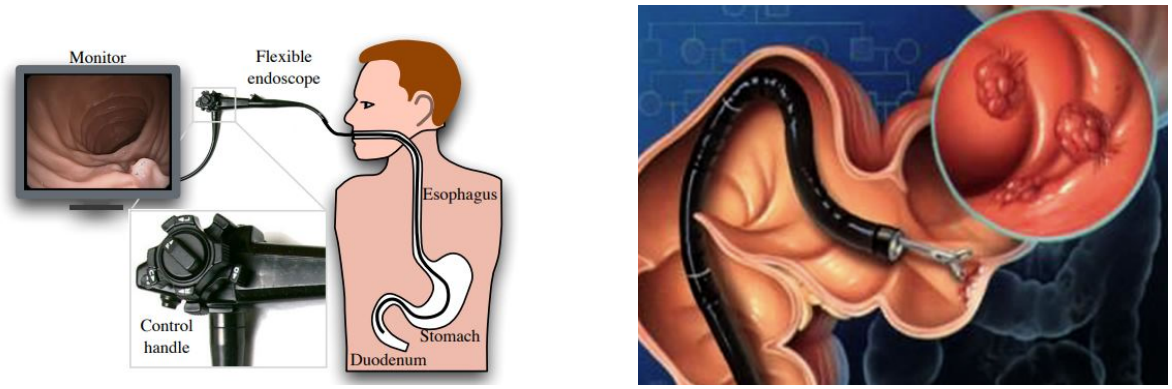


Figure 2-5: The gross anatomy of the large intestine [24].

muscle. The shape of the intestine wall is exhibited as pocketlike sacks sewed together called haustra. Another distinct feature of the large intestine is its epiploic appendages, which are little pouches filled with fat hanging from the intestinal wall. Their function is still unknown to humans to this date [24].

The large intestine can be subdivided into the following parts: the cecum, appendix, colon, rectum, and anal canal. the cecum has a saclike shape that lies right below the ileocecal valve. Attached to the cecum is the appendix. The appendix plays an important role in the body's immunity as it contains much lymphoid tissue. Furthermore, it stores microorganisms and recolonizes the stomach as necessary. However, the appendix has a significant structural flaw, its twisted form makes it prone to obstruction [24].

The colon has a subdivision of its particular regions. The ascending colon ascends to the right side of the abdominal cavity to the level of the right kidney. At the right colic (hepatic) flexure, the colon turns right and becomes the transverse colon, which proceeds to cross the abdominal cavity. The right colic (hepatic) flexure turns right-angled at this point, and the transverse colon proceeds across the abdominal cavity. Next, it ascends the right side of the abdominal cavity at the same angle as the ascending colon to the level of the right kidney. After traversing the abdominal cavity, the colon bends sharply at the left colic (splenic) flexure in front of the spleen. Then it begins as the descending colon, along the left side of the posterior abnormal wall. Then it enters the pelvis inferiorly, transforming into the S-shaped sigmoid colon. The sigmoid colon joins the rectum, which runs posteroinferiorly immediately in front of the sacrum at the level of the third sacral vertebra in the pelvis. The rectum has three bends, each representing a transverse fold called a rectal vale. These valves prevent the passage of excrement combined with gas. The anal canal is the last segment of



(a) The workings of an endoscope through the mouth [34].

(b) A biopsy being taken in the intestine [41].

Figure 2-6: The endoscope and the biopsy.

the large intestine and lies external to the abdominopelvic cavity. The anal canal is about three centimeters long and begins where the rectum penetrates the pelvic floor muscles and opens up to the body's exterior via the anus. The anal canal contains two sphincters, an involuntary internal anal sphincter and a voluntary external anal sphincter. The first is composed of smooth muscle, and the latter is composed of skeletal muscle, both sphincter act to open and close the anus [24].

2-1-4 Gastrointestinal Endoscopy and Biopsy

Gastrointestinal endoscopy, combined with the acquisition of a tissue sample, is an essential technique used to later diagnose various diseases in and of the digestive system [30].

During a biopsy, a small piece of tissue is taken from the lining of the GI tract for further examination under a microscope [32, 38]. The biopsy is typically performed using an endoscope, a long and flexible fiberoptic, or a video tube [34]. The endoscope contains a camera and a light at the end of it and is inserted through the mouth or nose [34]. After this, it is guided through the GI tract towards the biopsy site [34, 32]. Then, by watching the endoscopic images on the monitor, the doctor steers the endoscope through the patient's gastrointestinal tract using the control handle [34] (see Figure 2-6a).

Once in position, a small biopsy catheter is passed through the endoscope to remove a tiny piece of tissue from the GI tract's lining (see Figure 2-6b), which is later assessed by a pathologist [38, 31]. The procedure is generally safe and well-tolerated, but there is a small risk of bleeding or infection at the biopsy site [31, 38]. Patients may be asked to avoid eating and drinking for a while before the procedure and may receive sedation or anesthesia to help them relax and minimize discomfort [38].

Performing an endoscopy can be divided into two classes: upper endoscopy and lower endoscopy [30]. Recommended biopsies for upper endoscopies are Barrett's esophagus, gastroesophageal reflux disease, and infectious esophagitis. [30] Recommended biopsies for the lower endoscopy are primary intestinal malabsorption and other intestinal diseases, including celiac disease, Crohn's disease, ulcerative colitis, and cancer [31]

2-2 Electric Impedance Spectroscopy

EIS is an effective and powerful method for studying the electrical characteristics of a wide range of tissue and biomaterials. It involves applying a voltage or current to the sample to measure its impedance over an appropriate chosen frequency range [35]. In actuality, a fitting is made to an equivalent electrical model, which represents the sample's electrical fingerprint, providing an understanding of its characteristics and actions [3, 35]. EIS is a fast and straightforward solution in various applications. The essential fundamental theoretical concepts behind impedance spectroscopy are examined in this section. EIS was conceptualized and worked out by Oliver Heaviside in 1886 and has turned into an effective tool that is currently being applied in multiple domains [12, 23, 35]. Electrical impedance is the opposition to alternating current that consists of a combination of resistance and reactance [23]. The opposition to the flow of electric currents is resistance, and the opposition to the flow of electric currents by inductance or capacitance is reactance [23, 35]. Although reactance and resistance both create resistance to the current flow, resistance results in the dissipation of electrical energy in the form of heat, while this is not the case for reactance [12]. A resistor continuously loses energy as time passes, whereas reactance stores energy and returns it to the circuit every one-fourth-cycle [12]. Reactance is measured in ohms (Ω) like resistance, with negative values representing capacitive reactance and positive values representing inductive reactance [12].

2-2-1 Impedance and Permittivity

Impedance is a complex-valued electrical resistance, which varies with the voltage frequency [40, 35]. Impedance is usually written as Z or $Z(\omega)$, emphasizing that is dependent on the frequency [40, 3]. Admittance, or complex-valued conductivity, is closely related to impedance and is a quantity that is defined as $\frac{1}{Z}$ and typically denoted as Y [40, 3].

$$Z(\omega) = \frac{1}{Y(\omega)} = |Z|e^{j\phi} = R + jX. \tag{2-1}$$

Here, R represents the resistance, and X represents the reactance [40]. Additionally, $|Z|$ is the impedance magnitude, ϕ is the impedance phase, $R = |Z|\cos\phi$ is the real part of the impedance, and $X = |Z|\sin\phi$ is the imaginary part of the impedance [40, 35]. Both forms are shown in Figure 2-7.

As mentioned earlier, impedance can also be denoted as admittance as [40],

$$\begin{aligned} Y(\omega) = \frac{1}{Z(\omega)} &= G + jB(\omega), \\ &= G + jC\omega. \end{aligned} \tag{2-2}$$

Here G represents the conductance, and $B(\omega)$ represents the susceptance measured in Siemens, and is expressed in capacitance and frequency [40, 35]. It should also be noted that the capacitance and the conductance are related to the relative permittivity (ε_r) and conductivity (σ) [12, 35]. The dielectric constant, also known as relative permittivity, is the permittivity

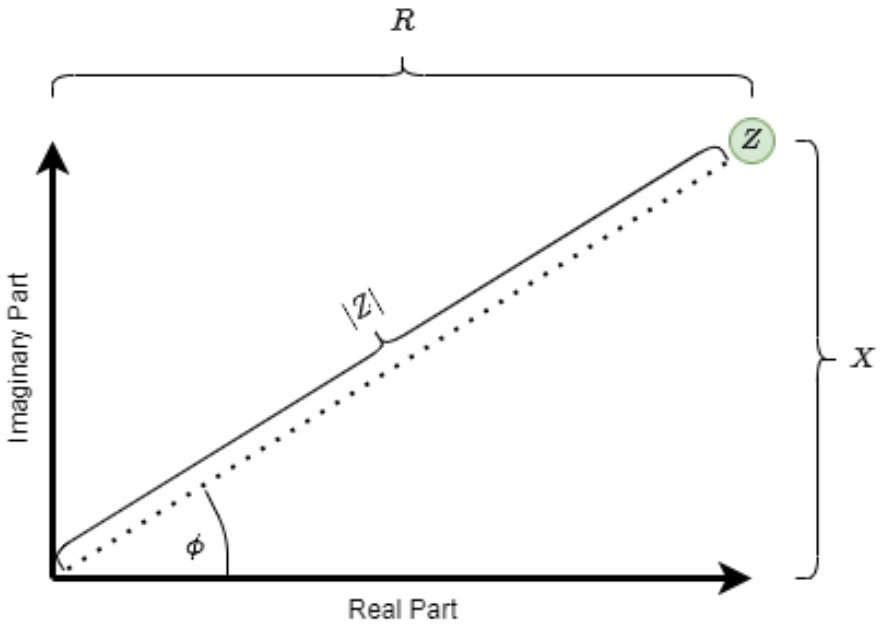


Figure 2-7: The complex-valued electric impedance [3].

of a material when it is expressed as a ratio of the electric permittivity that is being held in a vacuum (vacuum permittivity, ε_0) [12].

The absolute permittivity, or permittivity, measures the dielectric's electric polarizability in electromagnetism [12, 35]. When a material has a high permittivity it will polarizes more in response to an applied electric field than a material with a low permittivity, allowing it to store more energy [12, 35]. The permittivity plays an important role, In electrostatics, when determining the capacitance of a capacitor [12, 35].

Relative permittivity represents the factor by which the electric field between the charges is reduced relative to a vacuum [12, 35]. In addition, relative permittivity is the ratio of the capacitance of a capacitor using that material as a dielectric, compared with a similar capacitor when using a vacuum as its dielectric [12, 35]. For a simple electrode configuration of two plates that are placed in parallel, the relation is given as [12, 35].

$$\varepsilon_r = \frac{Cd}{\varepsilon_0 S} = \frac{\varepsilon}{\varepsilon_0}, \tag{2-3}$$

$$\sigma = \frac{Gd}{S}. \tag{2-4}$$

Here the distance between the two parallel plates is represented by d, the cross-sectional area of the electrodes is S, and the vacuum permittivity is ε_0 [12, 35]. For time-varying electromagnetic fields, this relative permittivity becomes frequency-dependent [12, 35]. As a result, it is is expressed as [12, 35]

$$\varepsilon_r(\omega) = \frac{\varepsilon(\omega)}{\varepsilon_0}. \tag{2-5}$$

In contrast to the response of a vacuum, the response of normal materials to external fields is mainly determined by the frequency of the field [12, 35]. This frequency dependence reflects that when an electric field is applied, the polarization of a material does not change immediately [12, 35]. A phase difference can represent the response, which is always causal, and therefore arises after the applied field [12, 35]. This results in the permittivity being frequently modeled as a complex function of the applied field's frequency ω [12, 35].

Since the response of materials to alternating electric fields are characterized by a complex permittivity, it is sensible to split it up into real and imaginary parts [9, 35]. This is done by convention in the following way [9, 35].

$$\hat{\varepsilon}(\omega) = \varepsilon'(\omega) - i\varepsilon''(\omega) \tag{2-6}$$

with $\varepsilon'(\omega)$ as the relative permittivity [9, 35]

$$\varepsilon'(\omega) = \varepsilon_r(\omega), \tag{2-7}$$

and $\varepsilon''(\omega)$ as the out-of-phase loss factor such that [9, 35]

$$\varepsilon''(\omega) = \frac{\sigma}{\varepsilon_0 \omega}. \tag{2-8}$$

In most cases, the complex permittivity is a complicated function of frequency ω as it is a superimposed portrayal of dispersion phenomena happening at multiple frequencies [9].

The low-frequency limit of permittivity, also known as the static permittivity ε_s, expresses a medium's reaction to static electric fields [13, 35]. At the high-frequency limit, the complex permittivity is generally referred to as ε_∞ [13, 35]

$$\varepsilon_\mathrm{s} = \lim_{\omega \to 0} \hat{\varepsilon}(\omega), \tag{2-9}$$

$$\varepsilon_\infty = \lim_{\omega \to \infty} \hat{\varepsilon}(\omega). \tag{2-10}$$

2-2-2 Electrical Impedance Spectroscopy in Biological Tissue

The capability to investigate the acquired impedance spectra using the equivalent circuits (EC) approach is one of the main benefits of EIS as an experimental technique [40]. The main idea behind these methods is to construct an equivalent electrical circuit with a similar impedance as the sample that is being analyzed [40]. The membrane found around living cells is semi-permeable and allows specific ions to pass while obstructing others [24]. This behavior mimics a capacitor [3, 40]. A capacitor is a charge-storing device made of two conducting plates with an insulating layer in between [44]. Due to the lipid bilayer in the cell membrane, the conductive salt solutions of the intracellular and extracellular fluids are separated. This makes the cell membrane act as a capacitor [3]. The parallel RC circuit is an excellent equivalent circuit for a living cell membrane since it has very high conductivity at a high-frequency current and a low conductivity at a constant current [3, 40, 26]. The

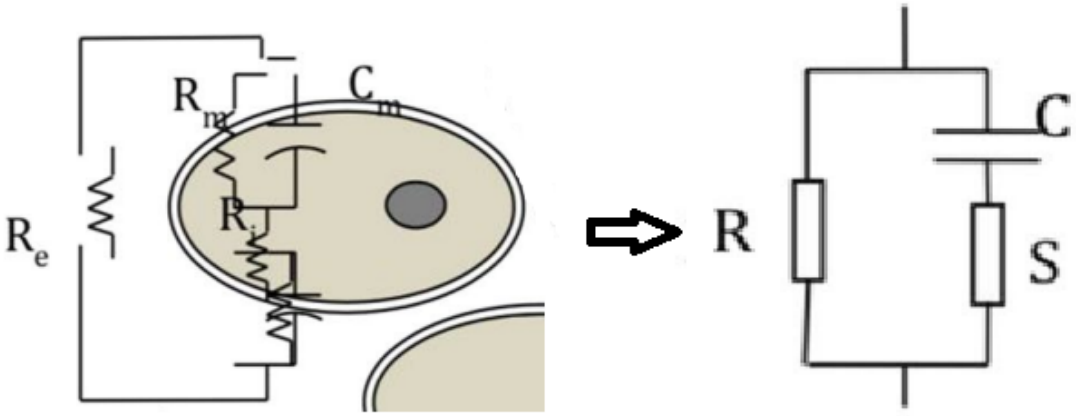


Figure 2-8: Transfer from live cell to electrical circuit [29].

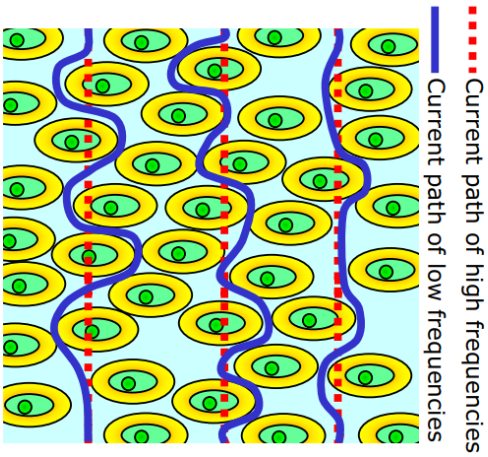


Figure 2-9: Pathway for low and high-frequency currents in living cells [3]

resistor represents the membrane's resistance, while the capacitor describes the geometrical capacitance of the cell membrane [3, 40, 26]. As a result, the equivalent circuit method lets the user create a direct association of the impedance spectrum with physical, chemical, and biological phenomena in living cells and extract its properties [40].

Biomaterials will vary in electrical properties, tissue structure, and chemical composition [3, 26]. This results in different frequency characteristics of each biological tissue [3, 26]. Overall, the extracellular space in the cell influences the low-frequency region, whereas the intracellular space influences the high-frequency region [3, 26]. Low-frequency currents cannot penetrate cell membranes because of their high capacitance. Therefore, they must pass around the cells, i.e., through the extracellular area [3, 26]. In contrast, high frequencies can pass through cell membranes and other electronic barriers. Therefore, these can pass through the cell and intracellular area [3, 26]. As a result, different rudimentary properties of the cells affect different frequency regions [3, 26]. The difference in the pathway due to low and high-frequency currents is illustrated in Figure 2-9 [3, 26].

The electrical impedance spectra of biological tissue include various frequency regions where

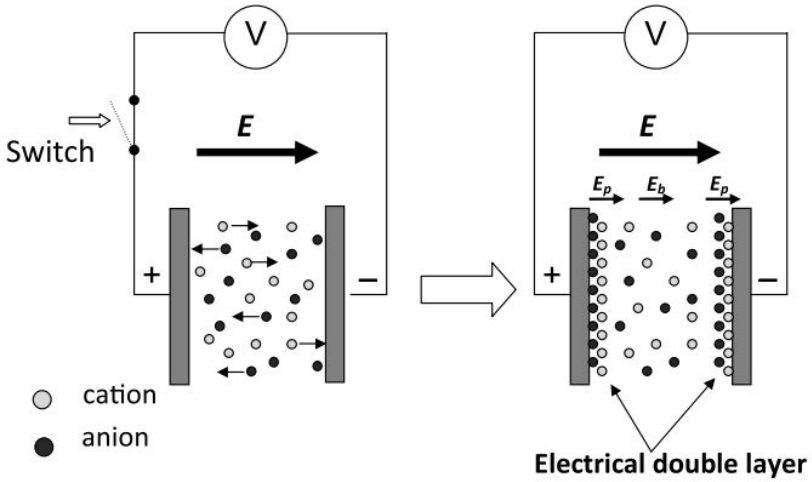


Figure 2-10: Electrode polarization [14].

the impedance declines with a rising frequency [3]. The areas where a drop in impedance is observed correspond to specific electrochemical processes named dispersions [3]. In 1957 Schwan discovered three main dispersions of electro-bioimpedance spectra [37]. These three dispersion regions are the α-, β-, and γ-dispersion, and can be seen in Figure 2-11 [3]. The α-dispersion reflects the polarization of ionic clouds around the cells and is observed in the low-frequency regions between the Hz to tens of kHz region [3, 9]. It should also be noted that when performing impedance measurements in lower frequency ranges, electrode polarization can occur [16]. Electrode polarization occurs because of the free ions from the tissue sample around the electrode [14]. These free ions advance towards the electrode/tissue sample interface when exposed to an electric field [14]. This results in an ionic double layer on the surface of the electrode (see Figure 2-10) [14]. This layer will then acts as a capacitor, which results in an electrode impedance [14]. The amount of electrode polarization depend on several aspects, e.g., the type of tissue sample under study, shape of the electrode, structure, the radius, the roughness, the distance between the electrode, and the temperature [16, 14].

The β-dispersion represents cell membrane polarization and structural shifts to the membrane[3, 9]. Therefore, this region is usually interesting because it includes the most clinically relevant data [3]. The β-dispersion is active between the KHz and hundred MHz regions [3, 9]. The γ-dispersion represents water molecules' relaxation and is seen over the hundred MHz regions [3, 9]. later a fourth dispersion was discovered, the δ-dispersion, which is in in the low GHz region [3].

The geometry of the electrodes has a considerable impact on the probe's performance [7]. The length between the electrodes, the tissues' physical properties, and the excitation signal frequency influence the depth of current penetration in biological tissue [7]. This phenomenon is shown in Figure 2-12, where the measurement's deepness increases as the length between the electrodes grows [7].

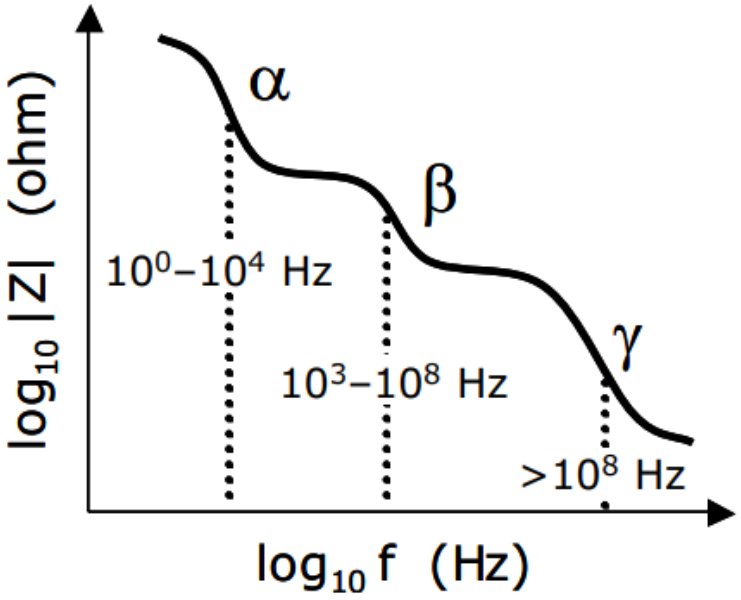


Figure 2-11: Different dispersion regions [3].

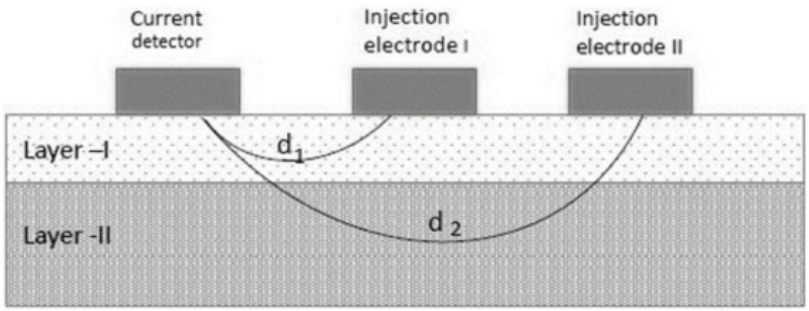


Figure 2-12: The depth penetration in layered tissue depends on the depth selectivity principle, which notes that the length between the injection electrode and the current detector determines the depth of penetration. [7].

Chapter 3

Methods

This Chapter covers the methods that are used during this work. Therefore it shows the measurement methods used to obtain the impedance, permittivity, and conductivity data. As well as the data analysis techniques and their performance metrics.

3-1 Measurement Methods

This section discusses the measurements of tissue in the Erasmus MC. Taking measurements is divided into two trials: trial 1 and trial 2. In the first trial, measurements of different tissue types are taken, and in the second trial, a focus on the esophagus, colon, and ileum tissue is conducted. These specimens also represent the data of the gastrointestinal (GI) tract. In this work, a combination of the tissue data of trials 1 and 2 is made to create a more extensive final data set. It should be noted that although the period of trial 2 focused on the esophagus, colon, and ileum tissue, no ileum tissue samples were available during this period.

3-1-1 Data Collection Setup

All tissue measurements are conducted on the Hewlett Packard 4192A LF HP impedance analyzer and an HP Power Pavilion 15-cb093nd laptop; both are depicted in Figure 3-1 [48].

For all measurements, a 4-electrode probe is used. Different combinations of measurements can be created, e.g., between the first and second, between the first and third, between the second and fourth, etc. However, in this work, all measurements are taken between the second and third electrodes, shown in Figure 3-2. The second electrode is connected to the low potential output, and the third electrode is connected to the high potential output [42]. The length and distance between each electrode are 5 mm, and the radius of each electrode is 0.5 mm [22].

Before a measurement can be taken, the entire system needs to warm up for at least 30 minutes; this is done to prevent noise in the measurements [48]. An increase in noise is

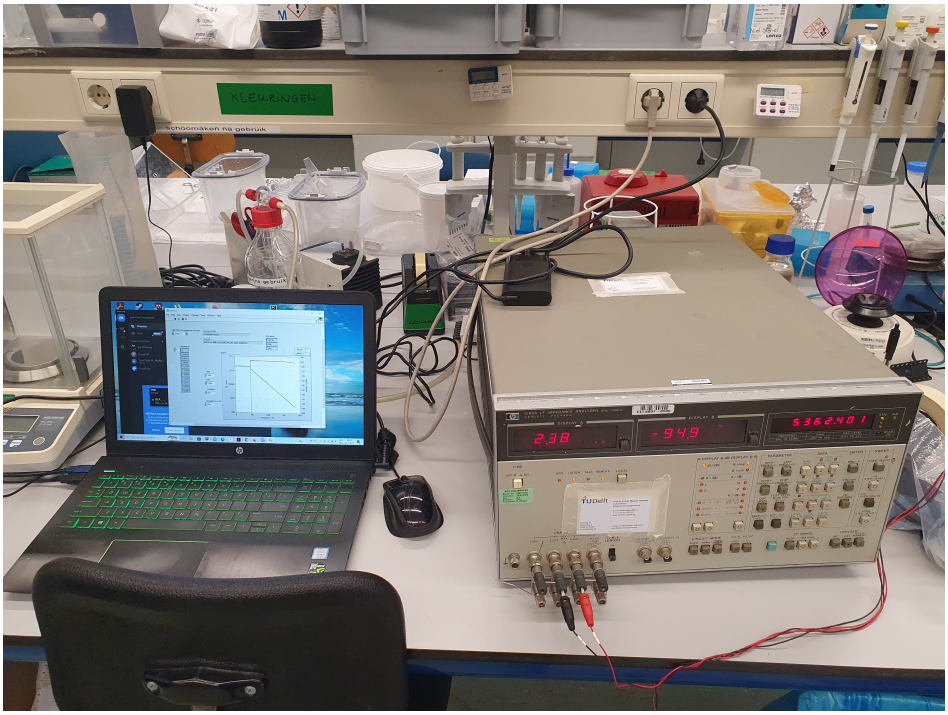
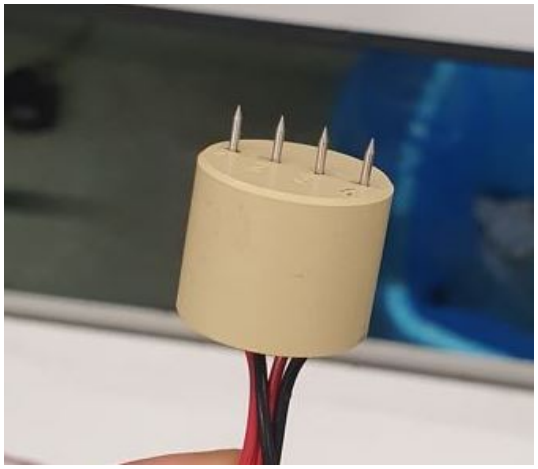
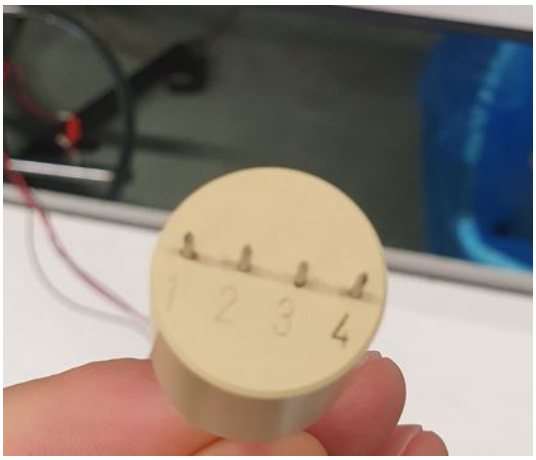


Figure 3-1: The 4192A LF HP impedance analyzer and an HP Power Pavilion 15-cb093nd laptop, with the 4-electrode probe. The second electrode is connected to the low potential output, and the third electrode is connected to the high potential output.



(a) The 4-electrode probe that is used in this work.



(b) The 4-electrode probe, with the electrode's number on the probe itself.

Figure 3-2: The 4-electrode probe with which all measurements are taken.

observed whenever the system is not warmed up properly. For all measurements, the probe is fixated on the tissue. In all cases, this is done by pressing the electrodes deep into the tissue and ensuring no movement of the 4-electrode probe is detected. The reason for this fixation is to mitigate the noise obtained by the hand's trembling during a measurement. Additionally, the fixation allows all measurements to be of similar depth with minimal variance.

The Hewlett Packard 4192A LF HP impedance analyzer can either measure the absolute impedance and the phase angle, or the capacitance and conductance. Running two dedicated software packages in NI LabVIEW 2018 (LabVIEW) allows one to decide what information to extract during the measurement time. In LabVIEW, one can also determine the frequency range over which the measurements are made and the number of steps measured. In this work, the measurements are taken in the α and β dispersion over a frequency range between 1 kHz and 7 MHz, with 300 steps. After the measurement is completed, the data is analyzed via a custom MATLAB script in MATLAB R2023A. The absolute impedance, |Z|, and phase angle, θ, are transformed into the real and imaginary impedance as [3]

$$\mathrm{Z_{re}} = |\mathrm{Z}| \cdot \cos(\theta), \tag{3-1}$$

and

$$\mathrm{Z_{im}} = |\mathrm{Z}| \cdot \sin(\theta), \tag{3-2}$$

with the real and imaginary impedance in Ohms (Ω). Measuring with the 4-electrode probe between two electrodes, the configuration can be seen as two parallel cylinders of length L, with radius r, and distance b [33, 22]. If $b \gg r$, the capacitance C, in Farads, can be calculated as

$$\mathrm{C} = \frac{\pi \varepsilon_0 \varepsilon_r}{\ln\left(\frac{b}{r}\right)}\mathrm{L}, \tag{3-3}$$

and the conductance, G in Siemens S can be calculated as

$$\mathrm{G} = \frac{\pi \sigma}{\ln\left(\frac{b}{r}\right)}\mathrm{L}, \tag{3-4}$$

with the vacuum permittivity, $\varepsilon_0 = 8.854 \times 10^{-12}$, and σ as the electrical conductivity in S/m [33, 47, 22]. To analyze the relative permittivity, ε_r, and the electrical conductivity, σ. The analysis uses the capacitance and conductance from Equation (3-3) and (3-4) is used to obtain the relative permittivity, and the electrical conductivity as [33, 47, 22]

$$\varepsilon_\mathrm{r} = \frac{\mathrm{C}\ln\left(\frac{b}{r}\right)}{\pi \varepsilon_0 L}, \tag{3-5}$$

and

$$\sigma = \frac{\mathrm{G}\ln\left(\frac{b}{r}\right)}{\pi L}. \tag{3-6}$$

It should be noted that in this work, only the real and imaginary impedance are used.

3-1-2 Measurement Protocol

The work conducted regarding the measurements was split up into two trials. Trial one was used to find suitable tissues to conduct research for this work, and trial two was used to find more data on the esophagus, ileum, and colon tissue.

The first trial started on 25-08-2022 and ended on 14-10-2022. During this period, different tissues of different patients were measured. All tissues were measured within 24 hours of resection. During this trial, the resemblance in esophagus, ileum, and colon tissue was observed, which can be because these three tissues are all in the GI tract and therefore have similar cell characteristics [24]. During the first trial period, impedance, phase, conductance, and capacitance measurements were made over 1 kHz to 7 MHz in 300 steps. Three measurements were made on the healthy part of the tissue, and three measurements were made on the cancerous part of the tissue. After measurements, the pathologist cut out the tissue on which the measurements were conducted for further research by histological analysis (the golden standard). The measurement protocol for the first trial can be found in Figure 3-3. At all times, the measurement process was started by turning on the impedance analyzer so it could warm up for at least 30 minutes. After this, the 4-electrode probe was cleaned with 96 % alcohol and rinsed it with distilled water. To check if the system is operable, a test measurement was ran that consisted of holding the probe still and measuring the impedance of the surrounding air. If no anomalies were detected, official measurements were taken. The pathologist would bring in and assess the tissue sample, showing areas that could be measured. In trial one, the measurements were conducted at two locations: healthy and unhealthy, i.e., non-cancerous and cancerous. After all the measurements were conducted, the pathologist would cut away both locations and verify them by histological assessment.

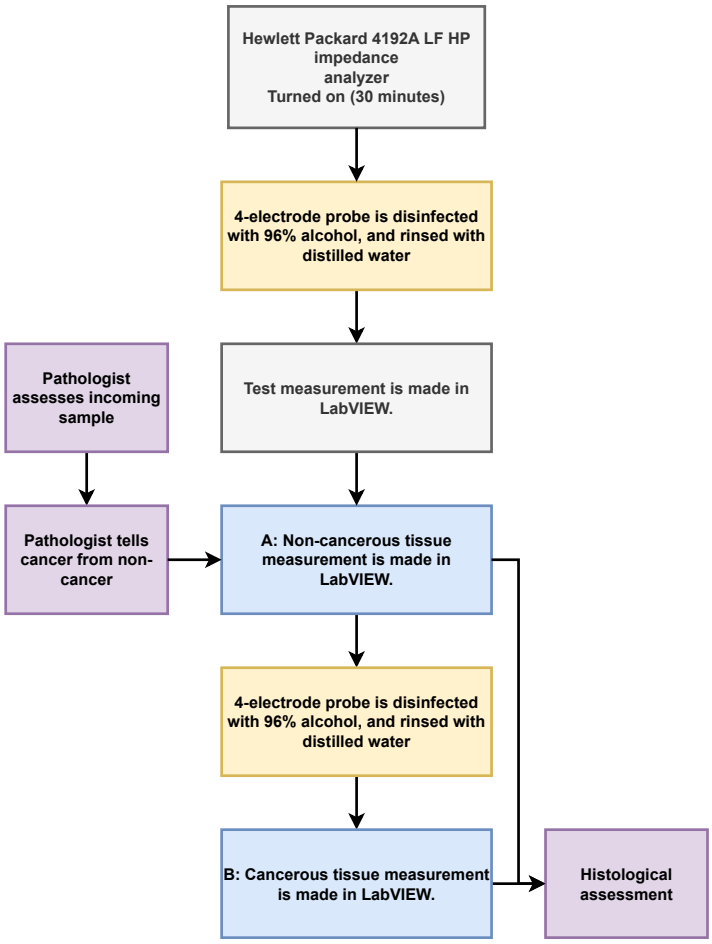


Figure 3-3: The measurement protocol for the first trial that was conducted at the Erasmus MC.

The second trial, trial 2, lasted from 31-01-2023 until 09-06-2023. Here, all measurements were conducted following a slightly different measurement protocol than trial one (see Figure B-1). During trial 2, the system was warmed up, the probe was cleaned, and the system was checked. However, in trial 2, at least three places were measured: A, B, and C. If time permitted, a fourth place, D, was included. Locations A and B would always be non-cancerous tissue, whereas location C would always be cancerous tissue.

After identifying all locations, the 4-electrode probe was carefully inserted into the first location (A), ensuring that electrodes 2 and 3 were in the tissue. After this, we first measured the absolute impedance and phase angle twice, and then we measured the capacitance and conductance. This process was then repeated for the remaining locations. During the trial period, impedance, phase, conductance, and capacitance measurements were made over a frequency range of 1 kHz to 7 MHz in 300 steps.

The histological assessment would also verify the tissue at locations A, B, and C. Due to time and resource constraints, this was not done for location D. For a more visual understanding of the different locations, please see Figure 3-5.

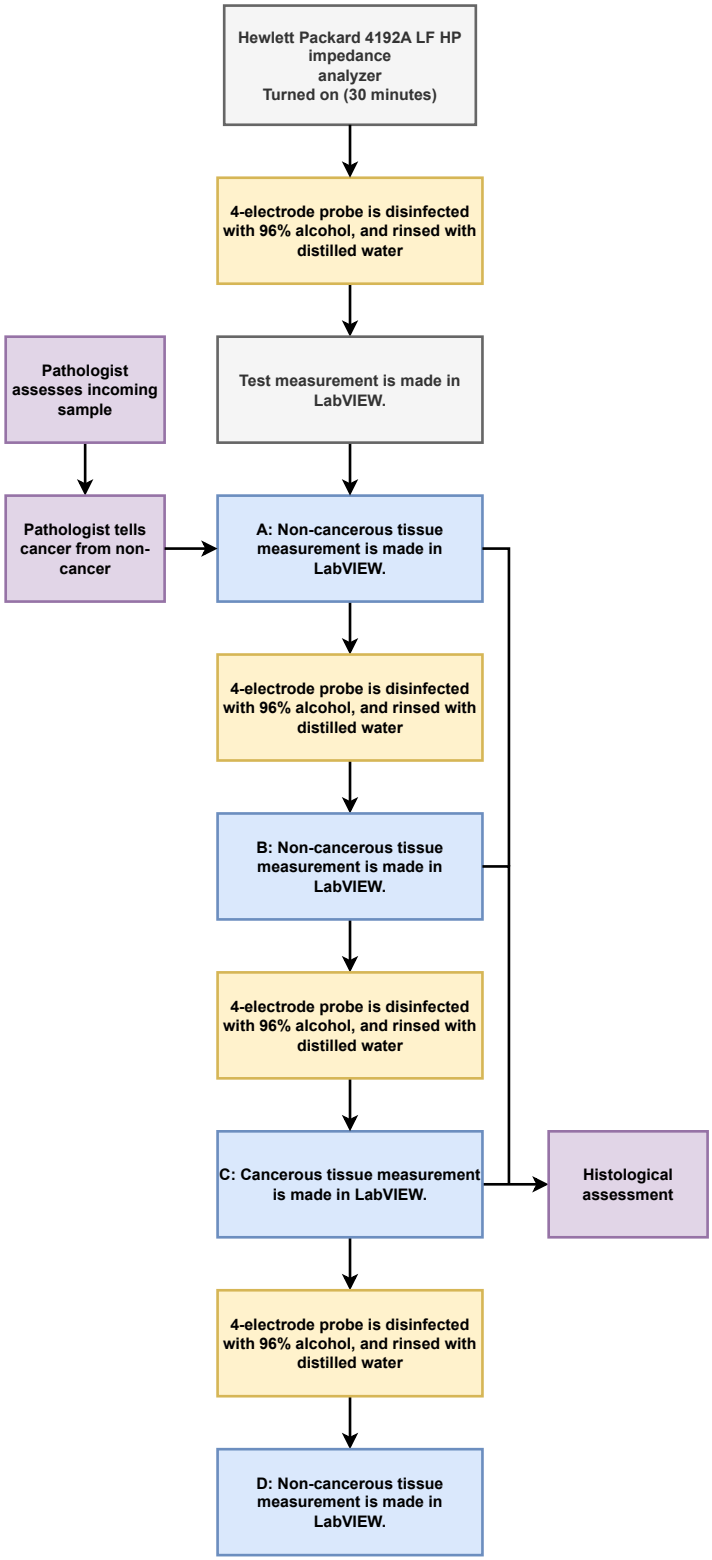


Figure 3-4: The measurement protocol for the second trial that was conducted at the Erasmus MC.

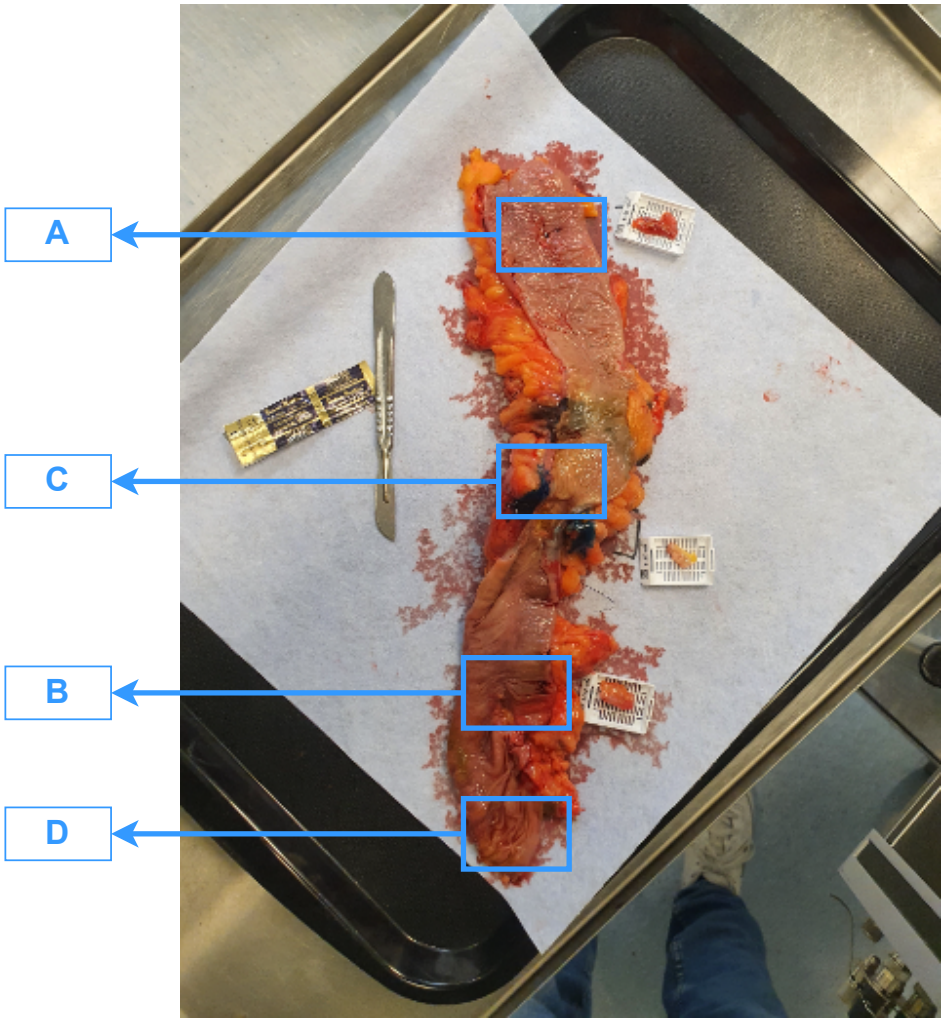
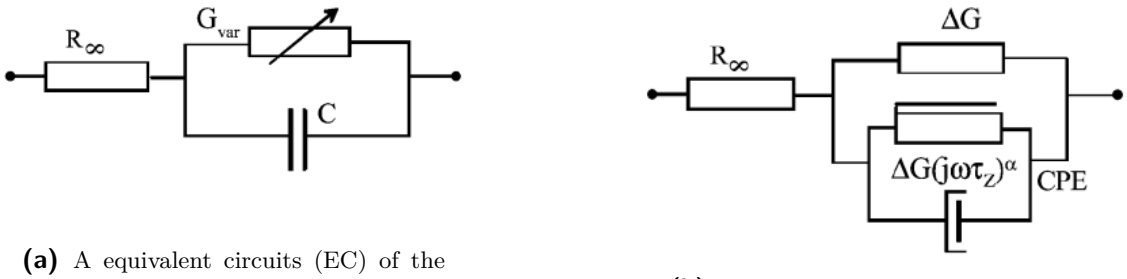


Figure 3-5: Different measurement locations the pathologist cuts out at the Erasmus MC. Location A, B, and D are healthy tissue, and location C is cancerous tissue. Only locations A, B, and C will be verified by histological analysis.

3-2 Parameterization Methods

The analysis of biological impedance spectra presents several challenges because the data is multivariate, the impedance is complex, and two numbers represent each data point, i.e., the magnitude and phase, or the real and imaginary parts [3]. Furthermore, complex numbers are difficult to numerically analyze [3]. Therefore, to interpret bio-impedance data, the raw data must be fitted into a model or, in some other way, the data must be simplified to a small number of clinically relevant parameters [3]. When the data is simplified, post-processing is frequently necessary to interpret the data [3]. This section describes two essential parameterization techniques used in this work: the Cole Impedance Modeling, and derivations of it, and principal component analysis (PCA).



(a) A equivalent circuits (EC) of the Debye model .

(b) A EC of the Cole impedance model.

Figure 3-6: The Debye and Cole impedance model ECs [10].

3-2-1 Cole Impedance Modeling

The Cole impedance model is a theoretical model that describes the tissue's impedance being studied. It is similar to the Cole-Cole model; however, that model describes the permittivity [10]. The aim is to identify equivalent electronic circuits with characteristics similar to the material that is being studied [40]. In this work we fit our impedance data to the Cole Impedance model with the fminsearchbnd algorithm from Matlab [27]. The models reduce the impedance spectra to a limited number of electronic components with a small number of parameters, producing a set of specific electronic elements [40].

In 1940 Kenneth Cole created a model named the Cole impedance model, which was based on the Debye model. However, the Cole impedance model replaces the ideal capacitor in the Debye model by a Constant Phase Element (CPE), as is depicted in Figure 3-6.

The impedance of the Debye circuit can be expressed as

$$Z = R_\infty + \frac{1}{G_{\text{var}} + G_{\text{var}} j\omega\tau}, \quad \tau = \frac{C}{G_{\text{var}}}. \tag{3-7}$$

In equation 3-7 Z represents the complex impedance, R_∞ the resistance at an infinite frequency, j is the imaginary number, ω the angular frequency, τ is the characteristic relaxation time constant that corresponds to the characteristic angular frequency, which is $\omega = \frac{1}{\tau}$, C represents the parallel capacitance, and G_{var} is the conductance [10]. The idea behind changing the ideal capacitor with a CPE came after findings in electrochemistry and tissue and cell suspension research. It was found that the impedance loci in the complex plane of living tissue had depressed circle centers in their shape, which was best mimicked by introducing a CPE [6, 10]. The Cole Impedance equation with the CPE is

$$Z = R_\infty + \frac{R_0 - R_\infty}{1 + (j\omega\tau)^\alpha}. \tag{3-8}$$

Here R_0 is the resistance at zero frequency, τ again represents the characteristic time constant of the system corresponding to a characteristic angular frequency $\omega = \frac{1}{\tau}$, α is a dimensionless exponent that lies between zero and one. The exact meaning behind α has still not reached a consensus. Although consensus on the meaning behind α is not yet achieved, many believe

that it is caused by the heterogeneity of the cell size shape of living tissue [4, 19]. In [15], it is stated that when a magnitude phase plot is made, the value of α influences the minimal spectral width as $\alpha = 1$ corresponds to the minimal spectral width, and it tends to broaden when α becomes lower (see Figure 3-7). Additionally, it is stated that using α makes it possible to detect tissue structural alterations as it was observed that a decrease of alpha was observed due to modification of the extracellular medium. Subsequently, the value for α can also be regarded as the derivation from the Fricke-Morse model, which is obtained when $\alpha = 1$ [15].

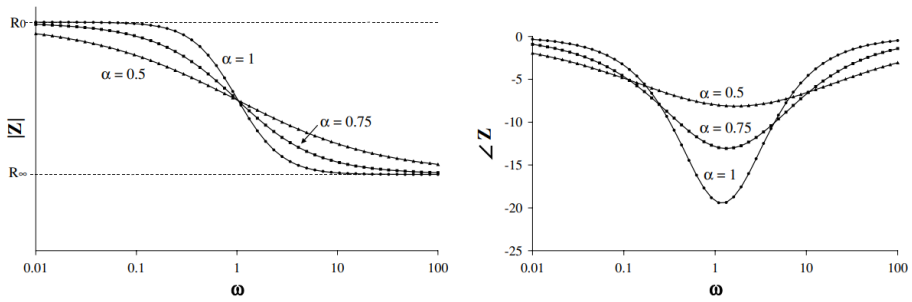


Figure 3-7: The absolute impedance and phase angle from the Cole equation for different values of α [15].

The product $\omega\tau$ is also dimensionless, and the combination of $(j\omega\tau)^{\alpha}$ represent a CPE when α is a constant value because [10]

$$j^{\alpha} = \cos(\alpha\pi/2) + j\sin(\alpha\pi/2). \tag{3-9}$$

The Cole Impedance model from Equation 3-8 is a single-dispersion model [10], meaning that it accounts for one dispersion. However, living tissue usually exhibits the dispersions of α and β. When one would like to use the model but over a frequency range, e.g., the α and β dispersions, the Two Pole Cole Impedance model can be a good option by creating a series of Equation 3-8 [27, 15].

$$Z = R_{\infty} + \frac{R_1}{1 + (j\omega\tau_1)^{\alpha_1}} + \frac{R_2}{1 + (j\omega\tau_2)^{\alpha_2}}. \tag{3-10}$$

In [27], the two pole cole model is also brought in series with an additional CPE model. A CPE model, as mentioned before, is used to mitigate the negative effects (increase in impedance) due to electrode polarization as explained in Section 2-2-2. In this work,a similar model to [27, 36] is used.

$$Z_{CPE} = K(j\omega)^{-m} \tag{3-11}$$

According to [36], K represents the ion concentration that is found at the electrode-tissue interface, and m is related to the surface roughness of the electrode. If one combines the two-pole Cole impedance model with the CPE model from Equation 3-11, they obtain [27]

$$Z = R_{\infty} + \frac{R_1}{1 + (j\omega\tau_1)^{\alpha_1}} + \frac{R_2}{1 + (j\omega\tau_2)^{\alpha_2}} + K(j\omega)^{-m}. \tag{3-12}$$

3-2-2 Principal Component Analysis

This section's main objective is to explain the principal component analysis method, which is used to reduce the dimensionality of multivariate data. The calculations show how to use singular value decomposition and orthogonal diagonalization.

Eigenvalues and Eigenvectors

Eigen is the German word for "latent" or "characteristic". In both pure and applied mathematics, the fundamental ideas of eigenvectors and eigenvalues come into play in contexts that are highly varied. For example, in addition to being used in engineering design, Eigenvalues are used to study various equations and continuous dynamical systems.

The eigenvector of an $n \times n$ matrix A is a nonzero vector x such that

$$Ax = \lambda x, \tag{3-13}$$

for some scalar λ. A scalar λ is named an eigenvalue of A if there is a nontrivial solution (any solution in which at least one variable has a nonzero value) x of $Ax = \lambda x$, such that x is named the eigenvector corresponding to λ [39].

(3-13) can be rewritten into so that λ is an eigenvalue of the $x \times n$ matrix A if and only if the equation

$$(A - \lambda I)x = 0, \tag{3-14}$$

has a nontrivial solution. The set of all solutions of (3-14) is the null space of the matrix $A - \lambda I$. So this set is a subspace of $\mathbb{R}^n$ and is called the eigenspace of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors correspondingly to λ [21]

The Singular Value Decomposition

The Singular Value Decomposition (SVD) is one of the most useful matrix factorization techniques in linear algebra. SVD can be used for non-square real and complex matrices, where it generalizes the eigendecomposition of a normal square matrix with an orthonormal eigenbasis to any $m \times n$ matrix. The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretch or shrink the eigenvectors. As previously stated $Ax = \lambda x$, if it is assumed that the x is of unit length, i.e., $\|x\| = 1$, then

$$\|Ax\| = \|\lambda x\| = |\lambda|\|x\| = |\lambda|. \tag{3-15}$$

If the eigenvalue with the largest magnitude is found, λ_1, then a corresponding unit eigenvector v_1 shows the directions for which the stretch of A is the greatest. This entails that, the magnitude of Ax is the greatest when $x = v_1$, and $\|Av_1\| = |\lambda_1|$. Describing v_1 and $|\lambda_1|$ in this way is comparable to using the SVD for rectangular matrices [21].

If again matrix A is taken as an $m \times n$ matrix, then $A^\top A$ is symmetric and can be orthogonally diagonalized. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of the eigenvectors of $A^\top A$, and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A^\top A$ [39]. Then, for $1 \leq i \leq n$,

$$\begin{aligned} \|Av_i\|^2 &= (Av_i)^\top Av_i = v_i^\top A^\top Av_i, \\ &= v_i^\top (\lambda_i v_i), \\ &= \lambda_i. \end{aligned} \tag{3-16}$$

Since the eigenvalues of $A^\top A$ are all positive, they can be reordered in such a way that it is assumed

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0. \tag{3-17}$$

The singular values, $\sigma_1, \ldots, \sigma_n$, of matrix A are the square roots of the eigenvalues of $A^\top A$, arranged in deceasing value ($\sigma_i = \sqrt{\lambda_i}$ for $1 \leq i \leq n$). By observing (3-16), it can be seen that the singular values are equal to the length of matrix $Av_1, \ldots, Av_n$, i.e., $\|Av_i\|$ [21].

The decomposition of A involves an $m \times n$ matrix Σ as

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}. \tag{3-18}$$

here D is an $r \times r$ diagonal matrix for some r that does not exceed m and n.

For the SVD A is used, which is an $m \times n$ matrix of rank r. Then there is a matrix Σ of the same dimensionality as A, for which the diagonal entries of D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and there exists an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$A = U\Sigma V^T \tag{3-19}$$

Any matrix that is factorized as $A = U\Sigma V^T$, with U and V orthogonal, and Σ as in (3-18), is called the Singular Value Decomposition of matrix A. Both matrices U and V are not uniquely determined by A. However, the diagonal entries always are the singular values of A. Furthermore, the columns of U are referred as the left singular vectors of A and the columns of V as the right singular vectors of A [21].

The workings of the SVD can be proven by first supposing that the $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $A^\top A$, arranged in such an order that the first eigenvector corresponds to the largest eigenvector of $A^\top A$. Additionally, suppose that matrix A has r nonzero singular values. Then the Span $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis for the column space Col A, and the rank of A equals r. If each $A\mathbf{v}_i$ is normalized to obtain an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, where

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{v}_i \tag{3-20}$$

and

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad (1 \leq i \leq r) \tag{3-21}$$

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is extended to $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ in $\mathbb{R}^m$, and let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix}$ and $V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. By construction of these matrices, they are both orthogonal matrices. From (3-21), the matrices can be filled in and extended after r entries by zero [21].

$$AV = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{3-22}$$

Let D be the diagonal matrix with the singular values on the diagonal, and let Σ be as in (3-18). Then

$$\begin{aligned} U\Sigma &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \left[\begin{array}{cccc|c} \sigma_1 & & & 0 & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right] \\ &= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \\ &= AV \end{aligned} \tag{3-23}$$

Since V is an orthogonal matrix $U\Sigma V^T = AVV^T = A$.

When the Σ contains columns and rows that contain only zeros and the rank of A is r, a so-called Reduced Singular Value Decomposition can be used to decompose matrix A [39]. If this is the case, the matrices U and V can be partitioned into sub-matrices as

$$\begin{aligned} U &= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \\ V &= \begin{bmatrix} V_r & V_{n-r} \end{bmatrix} \end{aligned} \tag{3-24}$$

Then U_r is $m \times r$ and V_r is $n \times r$. s.) By filling in these matrices in (3-19), it is obtained:

$$A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T. \tag{3-25}$$

Because the diagonal elements of matrix D are all nonzero, they can be inverted. The following matrix is called the Moore–Penrose inverse or pseudoinverse of A [21]

$$A^+ = V_r D^{-1} U_r^T \tag{3-26}$$

Prinipal Component Analysis

Any data that consists of a list of measurements taken on a group of items or people can be subjected to PCA. PCA, can be used to examine challenging to interpret multivariate data. The equations below demonstrate how to use the SVD and orthogonal diagonalization to create the principal components of a data set [21].

An example of an observation vector is X_n, with $n = [1, 2, \ldots, N]$. Each observation vector consists of two variables with variable α_n and β_n, then the data matrix X contains a total of N measurements and is given in (3-27).

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_N \\ \beta_1 & \beta_2 & \cdots & \beta_N \end{bmatrix} \tag{3-27}$$

To prepare a PCA, the data matrix $\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_N \end{bmatrix}$ is a $p \times N$ matrix of observation (in the case of (3-27) $p = 2$). The sample mean, M, is then calculated as

$$\mathbf{M} = \frac{1}{N}\left(\mathbf{X}_1 + \cdots + \mathbf{X}_N\right) \tag{3-28}$$

for, $k = 1, \ldots, N$, let

$$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M} \tag{3-29}$$

The columns of the $p \times N$ matrix

$$B = \begin{bmatrix} \hat{\mathbf{X}}_1 & \hat{\mathbf{X}}_2 & \cdots & \hat{\mathbf{X}}_N \end{bmatrix} \tag{3-30}$$

have a sample mean equal to zero, and B is said to be in mean-deviation form. The (sample) covariance matrix is the $p \times p$ matrix S defined by

$$S = \frac{1}{N-1} BB^T. \tag{3-31}$$

The covariance is a measure of the joint variability of two random variables. The covariance of the diagonal is the covariance of the variable with itself and is called the variance. Variance serves as a measure for dispersion, or the degree to which a set of numbers deviates from their mean [21].

The sum of the variances along the diagonal of S is the total variance of the data set. In general, the sum of the diagonal entries of a square matrix is called the trace of that matrix [21]. Thus,

$$\{ \text{ total variance } \} = \text{tr}(S) \tag{3-32}$$

Whenever an entry in the covariance matrix S equals 0, these variables are said to be uncorrelated. Multivariate data analysis is greatly simplified when most or all variables are uncorrelated, that is when the covariance matrix S is diagonal or nearly diagonal [39].

For simplicity, it is assumed that our data matrix is already in mean-deviation form. PCA attempts to find an orthogonal $p \times p$ matrix (a square matrix with the size of the number of variables) $P = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{bmatrix}$ that determines a change of variable $\mathbf{X} = P\mathbf{Y}$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} \tag{3-33}$$

with the property that the variables $y_1, \ldots, y_p$ are uncorrelated (meaning that its covariance matrix is diagonal) and arranged in order of decreasing variance.

The orthogonal change of $\mathbf{X} = P\mathbf{Y}$ means that each observation vector $\mathbf{X}_k = P\mathbf{Y}_k$. The new variable Y_k can then be expressed as

$$\mathbf{Y}_k = P^{-1}\mathbf{X}_k = P^T\mathbf{X}_k \tag{3-34}$$

for $k = 1, \ldots, N$. It is not hard to see that for any orthogonal P, the covariance matrix of $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ is P^TSP. The wanted orthogonal matrix P makes P^TSP diagonal [21].

If D is let to be a diagonal matrix with the eigenvalues of S on its diagonal, arranged in such a manner that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and matrix P is let to be orthogonal, with its columns being the corresponding unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$. Then $S = PDP^T$ and $P^TSP = D$. The unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$, of the covariance matrix S are called the principal components of the data. Therefore, the first principal component is the eigenvector that corresponds with the largest eigenvalue of S, the second principal component is the eigenvector that corresponds with the second largest eigenvalue of S, and so on [39].

With the first principal component, u_1, the new variable can be created in the following way. Here $c_1, \ldots, c_p$ are the entries in u_1. Because $u_1^\top$ is the first row of $P^\top$, equation (3-34) shows that

$$y_1 = \mathbf{u}_1^T\mathbf{X} = c_1x_1 + c_2x_2 + \cdots + c_px_p. \tag{3-35}$$

This means that y_1 is a linear combination of the weights from the eigenvector u_1 and the original variables $x_1, \ldots, x_p$. Similarly, u_2 contains the weights that can be linearly combined with the original variables to create y_2, and so on.

It can be shown that an orthogonal change of variables $\mathbf{X} = P\mathbf{Y}$ does not change the total variance of the data . .) This means that $S = PDP^T$ then,

$$\left\{ \begin{array}{c} \text{total variance} \\ \text{of } x_1, \ldots, x_p \end{array} \right\} = \left\{ \begin{array}{c} \text{total variance} \\ \text{of } y_1, \ldots, y_p \end{array} \right\} = \operatorname{tr}(D) = \lambda_1 + \cdots + \lambda_p \tag{3-36}$$

The variance y_j is λ_j, and the quotient $\lambda_j / \operatorname{tr}(S)$ measures the fraction of the total variance that is "explained" or "captured" by y_j [21].

The SVD can also be used to compute the principal components. If B is a $P \times N$ matrix that is in the mean-deviation form, and $A = \left(1/\sqrt{N-1}\right) B^\top$, then $A^\top A$ will equal the covariance

matrix S as seen in (3-31). The singular values of A are the p eigenvalues of S, and the right singular vectors of A are the principal components of the data [21].

3-3 Classification Methods

In this our two classification methods are treated, thresholding and support vector machine (SVM). The thresholding algorithm is used for the parameters that are obtained via the Cole impedance model, the SVM is used as a classification method in conjunction with the PCA.

3-3-1 Statistical Significance and Thresholding

To create a classification algorithm that takes a parameter or multiple parameters as its input, one must first examine whether it is even possible to make a classification algorithm with these parameters.

This work starts with the one-sample Kolmogorov-Smirnov test, which has the null hypothesis that the data in vector x comes from a standard normal distribution. This means that if the test is rejected, it indicates that the data in x does not come from a standard normal distribution, and when the test is accepted, it does come from a standard normal distribution. If it is found, via the one-sample Kolmogorov-Smirnov test, that the data in x belongs to a standard normal distribution, the t-test proceeds. The t-test has the null hypothesis that the data in vector x comes from a normal distribution with a zero mean (with an unknown variance). When this hypothesis is rejected, it would indicate that the data in x does not have a mean equal to zero. To determine which parameter can be used for the classification via thresholding, a parameter that does not have equal means when created with cancerous and non-cancerous data is required. To check this, the two-sample t-test is used, with the null hypothesis that the samples in x and y have equal means. If the null hypothesis is rejected for a parameter at a significant level ($P < 0.05$), it indicates it can be used for classification via thresholding.

If the null hypothesis of the one-sample Kolmogorov-Smirnov test is rejected for the data in x, i.e., the samples in x do not come from a normal distribution, it proceeds with a two-sided Wilcoxon signed rank test where the null hypothesis is that the data in x comes from a distribution whose median is equal to zero. To execute this test, it is assumed that the data in x comes from a continuous distribution symmetric and is symmetric about its median. We use the two-sided Wilcoxon rank sum test to determine which parameters can be used as a classification parameter. This tests the null hypothesis that vector x and y data are from a continuous distribution with equal medians. It is assumed that the two samples are independent. If the null hypothesis is rejected in the two-sided Wilcoxon rank sum test with statistical significance ($P < 0.05$), that parameter can be used to classify cancerous and non-cancerous tissue. An overview of the flow chart and methods can be seen in Figure 3-8 [27].

To determine the threshold value, the Receiver Operator Characteristic (ROC) curve is optimized by brute force. This entails testing a range of threshold values and creating the ROC curve. The moment the ROC curve is created, the point that is the furthest away from the

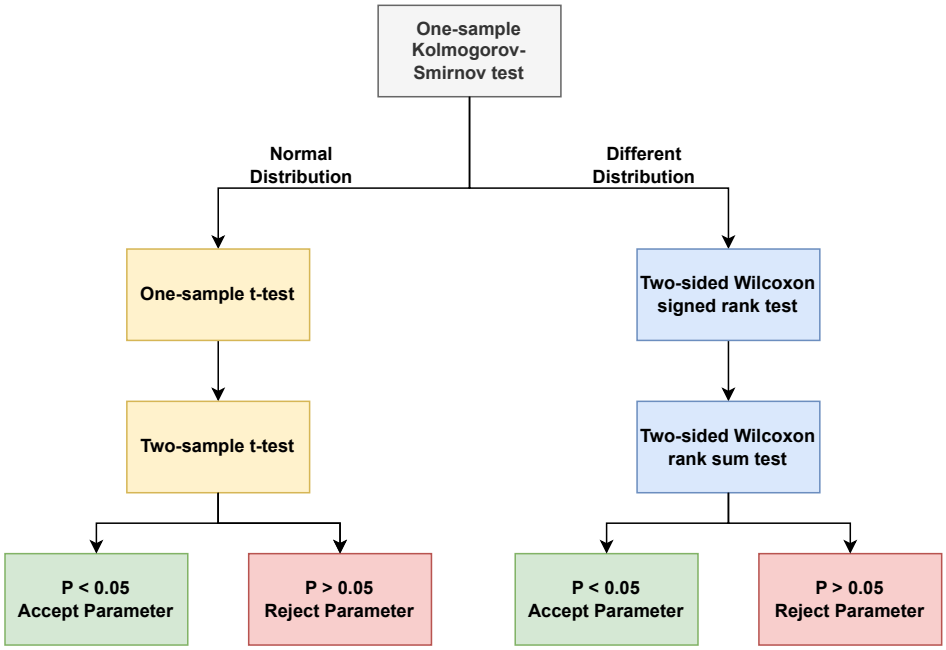


Figure 3-8: The flow chart to determine whether or not a parameter can be used for classification via thresholding to determine whether tissue is cancerous.

mid-line, which corresponds to the largest summation of specificity and sensitivity is found. The index that corresponds to that point is used to find the Threshold value.

Thresholding is frequently used in academic Electrical Impedance Spectroscopy (EIS) research. For instance, after colorectal cancer surgery, EIS is conducted on 22 freshly resected specimens of excised human colorectal tumor tissue and the matched normal colonic mucosa in [28]. The recorded measurements were not normally distributed (as determined by the Shapiro-Wilk test), so the Wilcoxon rank sum test for paired non-parametric data of small sample sets was used to test for differences between matched normal and malignant tissue transfer impedance magnitudes at each frequency. Fourteen different frequencies were calculated to plot the ROC curve and to measure the corresponding Area under the ROC Curve (AUC). These were the transfer-impedance magnitude thresholds that, at all frequencies, would classify a sample as cancerous. Eventually, the authors found an AUC of 0.7105. The goal in [2] was to research how well EIS worked for finding cervical intraepithelial neoplasia using cervical impedance spectroscopy. The initial data set, consisting of 80 measurements, was averaged to produce mean impedance values for each of the 30 frequencies chosen. Then, R_0, R_∞, and F_c were calculated by fitting these data, which formed an impedance spectrum, to a Cole equation using the least square deviation method. There were 680 measurements from normal squamous epithelium, 178 from cervical intraepithelial neoplasia 2/3 (high grade), 39 from cervical intraepithelial neoplasia 1 (low grade); 28 points were designated as columnar tissue, 135 as mature metaplastic, and 79 as immature metaplastic. The parameters clearly distinguished the cervical intraepithelial neoplasia lesion from the squamous, mature metaplastic, and columnar epithelium. The authors obtained a sensitivity of 74% and specificity of 53%. In [43], the goal is to determine if EIS can aid in the diagnostic accuracy of colposcopy. EIS was assessed against colposcopic impression and

histopathology from biopsies taken from patients. Real-time EIS data collection and analysis was conducted. In total, 474 women were recruited, and 214 were eligible for analysis. Using a cut-off value, the accuracy of colposcopic impression to detect HG–CIN when using EIS as an adjunct at the time of examination improved the Positive Predictive Value (PPV) from 78.1% (95% Confidence Interval (CI) 67.5–86.4) to 91.5%. The Negative Predictive Value (NPV) was unchanged, but specificity increased from 83.5% (95% CI 75.2- 89.9) to 95.4%. However, sensitivity significantly decreased from 73.6% (95% CI 63.0-82.5) to 62.1%. In [25] EIS was used on 36 ex vivo human prostates. The magnitudes of the mean reactance and resistivity were significantly higher in cancerous tissues ($P < 0.05$). According to the ROC curves, the mean resistivity at 63.09 kHz had a AUC of 0.779, 75.4% specificity, and 76.1% sensitivity, making it the best frequency for separating cancerous tissues from healthy ones. The best frequency for using the mean reactance to distinguish between cancer and benign tissues was 251.1 kHz; this parameter had a AUC of 0.79, a 77.9% specificity, and a 71.4% sensitivity. The authors of [46] used EIS for differentiating cancerous from benign breast masses. The least squares algorithm was applied to the Cole-Cole model fitting. To evaluate the sensitivity and specificity, the final EIS results were compared to the frozen section and postoperative pathological examination results. In total 622 breast masses were identified, and 66 were determined to be malignant and 556 to be benign. The authors achieved a sensitivity of 92.4%, and a specificity of 96.0%. In [18], 53 patients were analyzed with EIS, 44 had lung cancer, 5 of whom had metastatic lung tumors, and 4 of whom had pneumonia. EC was used consisting of two resistors and a capacitor representing the extracellular and intracellular resistance and the cell membranes' capacitance. These three parameters were calculated using EIS on the lung tissue. The differences between mean values were assessed by a one-way analysis of variance and an unpaired Student's t-test.It was shown that malignant tumors had significantly higher extracellular resistance and intracellular resistance values compared to pneumonia, and the membrane capacitance was significantly less in malignant tumors. Thresholding was used to diagnose a malignant tumor and pneumonia. This resulted in a sensitivity of 100% and a specificity of 66%. A thresholding algorithm, with the use of a two-pole Cole impedance model with a CPE, was used in [27]. Here a four-electrode probe was used over a frequency range from 100 Hz to 1 MHz to measure ablated and non-ablated on three porcine hearts. A total of 340 measurements were conducted on the three porcine hearts. With the thresholding algorithm an accuracy of 81.4%, sensitivity of 84.4%, specificity of 78.6%, PPV of 77.8%, NPV of 85.0%, and a Mathews Correlation Coefficient (MCC) of 0.629 were created. It should be noted that these results indicate the proper performance of the algorithm; however, no Leave-One-Out Cross Validation (LOOCV) strategy was employed. This entails that the training or determination of the threshold was conducted with a priori knowledge of the testing data.

3-3-2 Support Vector Machine

This section discusses the SVM as a classification method. When given a data set consisting of two classes, the idea is to create an algorithm that can automatically classify to which of the two classes a new point belongs [11].

In its most basic form, a classification algorithm uses a linear function. If there is a set of training data, m, with different data points $x_i \in \mathbb{R}^n$, of the class y_i, which is either $y_i = 1$ or $y_i = -1$. Ideally, one would like to find a separation line between the positive and negative points. Suppose one can find a hyperplane $w^T x + b = 0$ that can separate the positive and

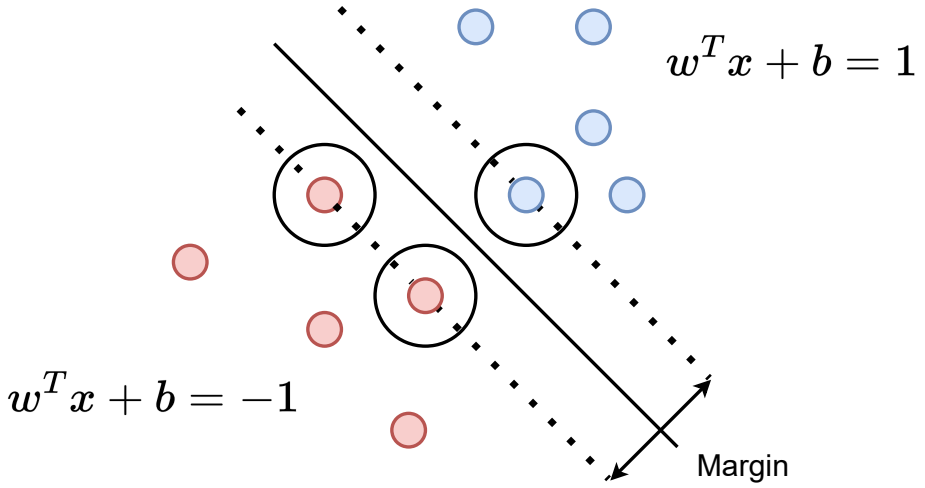


Figure 3-9: A linear hyperplane separating data.

negative points; this would require

$$\begin{aligned} w^T x_i + b &\geq +1 \quad \text{for } y_i = +1, \\ w^T x_i + b &\leq -1 \quad \text{for } y_i = -1. \end{aligned} \tag{3-37}$$

The hyperplanes aim to separate the positive from the negative points. To obtain the best results, one should create a hyperplane that separates the points as far apart as possible. The distance between the two hyperplanes, i.e., the separation margin, is $\frac{2}{\|w\|}$ [11]. Finding the hyperplanes that maximize the separation margin equals minimizing $w^T w$ [11].

$$\begin{aligned} &\text{minimize} \quad f(w,b) = \tfrac{1}{2} w^T w \\ &\text{subject to} \quad y_i \left(w^T x_i + b \right) \geq 1, \quad i = 1, \ldots, m. \end{aligned} \tag{3-38}$$

In Figure 3-9, a few data training points are found of either class $y_i = -1$ or $y_i = 1$; the training points are separated by a hyperplane that is created by a margin, which is determined by the training points that lie on the boundary of either of the hyperplanes. These are called support vectors and are indicated with a black circle. If a training point that is a support vector is removed, it will alter the coefficients of the hyperplane. Removing a training point that does not act as a support vector would keep the coefficients unchanged. This method is called SVM because of the support vectors used for classifying new data points as part of the machine learning process [11].

The example above assumes that the data is separable, meaning a hyperplane separates all points of a different class. However, this is often not the case. When the dataset is not separable, the approach must be refined. If one lets the equation of separation be violated but imposes a penalty for this violation, it is possible to create a SVM that can work with separable data. The non-negative variable is introduced,ξ_i, which represents the amount by which the point x_i does not adhere to the constraint at the margin [11].

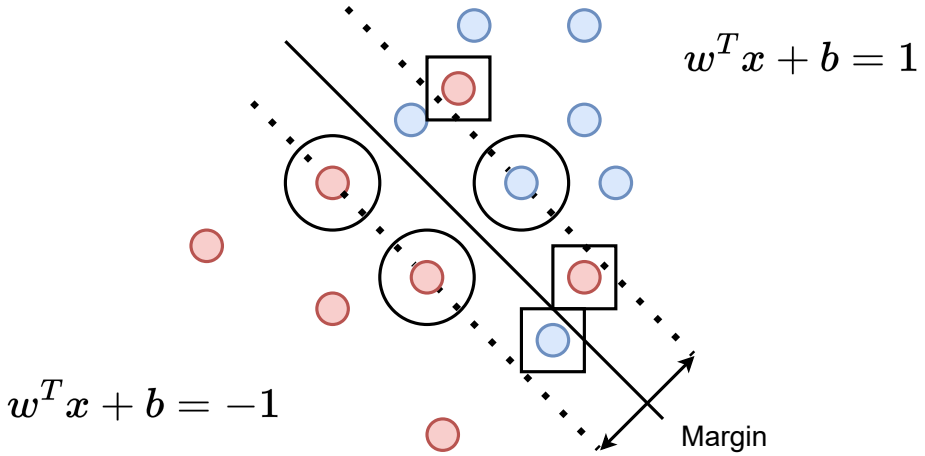


Figure 3-10: A linear hyperplane separating data for a nonseparable data set.

$$\begin{aligned} w^T x_i + b &\geq +1 - \xi_i \quad \text{for } y_i = +1, \\ w^T x_i + b &\leq -1 + \xi_i \quad \text{for } y_i = -1. \end{aligned} \tag{3-39}$$

To impose this penalty, one can add a term proportional to the sum of the violation to the objective function. This added penalty term is denoted as $C \sum_{i=1}^{m} \xi_i$, where C determines how large the penalty is for violating the initial equation of separation. The objective function becomes

$$\begin{aligned} \text{minimize} \quad & f(w, b, \xi) = \tfrac{1}{2} w^T w + C \textstyle\sum_{i=1}^{m} \xi_i \\ \text{subject to} \quad & y_i \left(w^T x_i + b \right) \geq 1 - \xi_i, \quad i = 1, \ldots, m \\ & \xi_i \geq 0. \end{aligned} \tag{3-40}$$

The image in Figure 3-10 shows a linear hyperplane in the case of a nonseparable case. In this case, three points are classified wrong since they are on the wrong side of the hyperplane $w^T x + b = 0$.

SVM can handle both linearly separable and non-linearly separable data using the kernel trick. The kernel trick transforms the input data into a higher-dimensional feature space, where it becomes possible to find a hyperplane that separates the data [11]. This allows SVM to work effectively with complex, non-linearly separable data. In this thesis work, we work with a Radial Basis Function (RBF) kernel. RBF kernel is a non-linear function that maps the input data into a higher-dimensional feature space, allowing SVM to handle non-linearly separable data effectively. The RBF kernel function calculates the similarity between two data points by computing the distance between them in the feature space. The RBF kernel function is defined as

$$K\left(X_1, X_2\right) = \exp\left(-\frac{\left\|X_1 - X_2\right\|^2}{2\sigma^2}\right). \tag{3-41}$$

Here X_1 and X_2 are two data points, σ represents the hyperparameter that controls the width of the kernel, and $||X_1 - X_2||^2$ is the squared Euclidean distance between the two data points

in the feature space [11]. The RBF kernel allows SVM to capture complex patterns in the data by creating decision boundaries that are curved or irregular. The RBF kernel also allows SVM to handle data that is not linearly separable by projecting it into a higher-dimensional space, where it may become linearly separable. The choice of the hyperparameter gamma in the RBF kernel can affect the performance of SVM. A high gamma value creates a narrow kernel, which can lead to overfitting. A low gamma value creates a wider kernel, which can lead to underfitting. Therefore, the gamma value must be carefully tuned to achieve the best performance of SVM with RBF kernel.

SVM has several advantages over other classification algorithms, including its ability to handle high-dimensional data, its effectiveness in handling non-linearly separable data, and its ability to avoid overfitting by maximizing the margin between the classes. SVM is currenly already employed within EIS cancer research. For example in [5], a study is developed that introduces a method with the potential to intraoperatively localize pulmonary nodules in-depth. In vitro investigation of the bioimpedance data from 286 lung tissue samples was performed using a bioimpedance probe. Based on the bioimpedance phase and magnitude, an analysis tool was developed to distinguish between tumoral and healthy lung tissue. In [5] SVM was combined with PCA, and the classifier accuracy was over 97%. In [17], an ex vivo investigation of 14 human prostates is conducted with a total of 23 cancerous and 53 benign EIS measurements. After the measurements, the Composite Impedance Metrics were calculated over various frequencies. Using a leave-one-patient-out cross-validation strategy, the predictive value of complex impedances measured data for differentiating between benign and malignant tissue was assessed. A SVM classifier was trained using a RBF kernel. The predictive accuracy was 90.8%, with a specificity of 94.3% and a sensitivity of 82.6%.

3-4 Performance Analysis Methods

For the performance analysis of our data fitting algorithm, which makes use of Cole impedance modeling the Mean Square Error (MSE) for every data point of the fit is calculated and summed up. The MSE which can be calculated as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \tag{3-42}$$

For our classification technique analysis LOOCV is used. With this technique, our data set is split into two sets, a training set, and a validation set. In Figure 3-11, an example is created where we treat five patients. For every run of the classification algorithm, we separate one patient (leaving that one out) and train our thresholding or SVM algorithm on the remaining patients. After the threshold is determined and the SVM is trained, the algorithms with the new data from the patient that was left out are tested.

The same metrics are used for both classification methods to analyze our results from the LOOCV. The performance of the classification is described in a confusion matrix, which can be found in Figure 3-12

In the confusion matrix, the following options are observed: a classification predicted as cancerous that actually is cancerous is considered a True Positive (TP). When a prediction

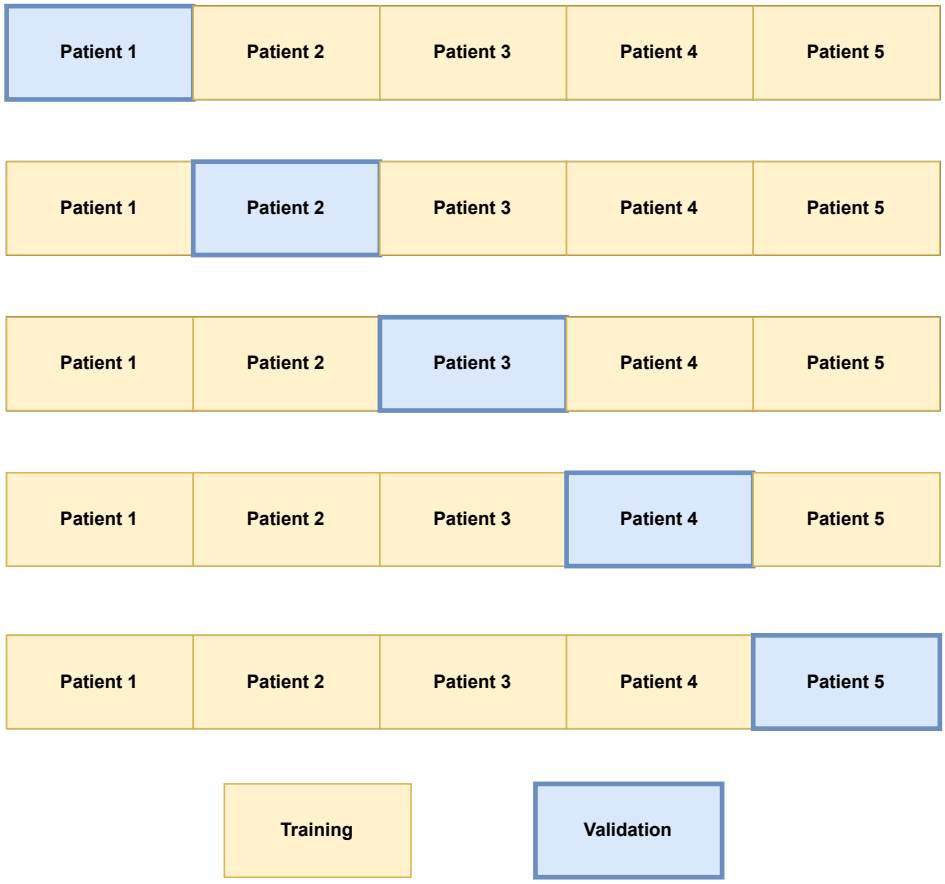


Figure 3-11: A example of a LOOCV with a 5 patient population.

		Actual outcome	
		Cancerous (1)	Non-cancerous(0)
Predicted outcome	Cancerous (1)	True Positive (TP)	False Positive (FP)
	Non-cancerous (0)	False Negative (FN)	True Negative (TN)

Figure 3-12: The confusion matrix.

is non-cancerous, and it is, this is considered a True Negative (TN). Both TP and TN are correct classifications. A False Positive (FP) is when a prediction states that the tissue sample under test is cancerous but it is not, and a False Negative (FN) occurs when the classification algorithm predicts that the tissue sample is non-cancerous but it is cancerous. In cancer research, one can easily argue that a FN is the worst possible outcome as it may lead to no treatment when treatment is required.

After obtaining the amount of TP, TN, FP, and FN a further analysis can be performed that showcases the capabilities of the classification algorithms under study. In this work, the Accuracy, Sensitivity Specificity, PPV, NPV, and MCC are provided which are calculated as follows

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3-43}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3-44}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3-45}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3-46}$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{3-47}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{3-48}$$

The accuracy depicts the ratio between correctly classified tissue samples and the number of predictions made. The sensitivity gives insight into the ability to detect cancerous tissue correctly, while the specificity gives insight in detecting non-cancerous tissue. The PPV and NPV give the proportion of positive and negative predictions that are truly positive or negative, respectively. The MCC is used as a general measure for the classification quality.

Chapter 4

Results

This chapter, discusses the results and all the relevant findings. It starts off with the measurement results of trials one and two. It is followed by the creation of the final data set. After this, the two classification strategies are discussed: the Cole impedance analysis in combination with thresholding and principal component analysis (PCA) in combination with support vector machine (SVM)

4-1 Measurement Results

In this section, an overview of the measurement results is given. As mentioned before, two trials were conducted, trial one and trial two. Trial one focused on collecting different cancerous and non-cancerous tissue types, and trial two focused on collecting tissue that could represent the gastrointestinal (GI) tract. The esophagus, colon, and ileum tissue were measured. In Table 4-1, one can find an overview of the measurements conducted during the first trial period. For an even more extensive description of trial 1, please refer to Appendix A. During the second trial, three patients were included, with their information in Table 4-2. During trial two, images of every cut-out at the pathologist department were made, which also indicate the different measurement areas (A, B, C, and D). These figures can be seen in Figure 4-5, Figure 4-7, and Figure 4-9. For an even more extensive description of trial 2, please refer to Appendix B.

4-2 Final Data Set

After conducting the two trials, the esophagus, ileum, and colon data were merged into a final data set that is used within this thesis work. The complete data set can be found in Table 4-3. As can be seen from the table, measurements on a total of seven different specimens were taken, three esophagus, two colon, and two ileum.

#	Date of Measurement	Gender (M/F)	Year of Birth	Tissue
1	25-08-2022	F	1971	Adrenal Gland
2	22-09-2022	F	1963	Retroperitoneal Soft Tissue
3	27-09-2022	F	11971	Breast
4	29-09-2022	M	1952	Liver, Ileum and Colon
5	29-09-2022	F	1971	Ileum
6	30-09-2022	M	1942	Glute
7	30-09-2022	M	1951	Esophagus
8	11-10-2022	M	1951	Pancreas
9	11-10-2022	F	1947	Lung
10	14-10-2022	M	1947	Esophagus

Table 4-1: The year of birth, gender, and date of the measurement per patient and the corresponding tissue the measurements were conducted during trial 1.

#	Date of Measurement	Gender (M/F)	Year of Birth	Tissue
1	31-01-2023	F	1979	Colon
2	28-03-2023	M	1960	Colon
3	09-06-2023	M	1943	Esophagus

Table 4-2: The year of birth, gender, and date of the measurement per patient and the corresponding tissue the measurements were conducted during the second trial.

After the histological assessment (golden standard) it was observed that there were multiple deviations from what was intended to be measured to what actually was measured. It was found by the pathology department that there were deviations in the fifth, the sixth and the seventh patient. In the fifth patient measured on 31-01-2023 the measurements conducted on section B also contained cancerous tissue. For the sixth and seventh patient the measurements conducted on location C (the location where the tumor should be located) no cancerous tissue was found. This means that the final data set with histological assessment is as found in Table 4-4. These faulty measurements could be the fault of the probe, as there was no complete control of the probe. The measurement protocol described that the probe is fixated deep into the tissue, allowing minimal variance in the measurements. However, what could have occurred is that the positioned at the surface (superficial), and the measurement was conducted deep into the tissue.

4-2-1 Patient 1

The measurement on patient one was conducted on 29-09-2022, during the first trial, and is a male who is 71 years old. The cancerous tissue was adenocarcinoma, the most common type of cancer. It arises from epithelial cells. These epithelial cells are the top layer of tissue and can be found all over the body, both inside and outside. It can be observed from Figure 4-1 that the impedance of the non-cancerous (A) and cancerous (B) tissue are different. The real impedance of the cancerous tissue started about 300 Ω higher than that of the non-

#	Date of Measurement	Gender (M/F)	Year of Birth	Tissue
1	29-09-2022	M	1952	Ileum
2	29-09-2022	F	1971	Ileum
3	30-09-2022	M	1951	Esophagus
4	14-10-2022	M	1947	Esophagus
5	31-01-2023	F	1979	Colon
6	28-03-2023	M	1960	Colon
7	09-06-2023	M	1943	Esophagus

Table 4-3: The year of birth, gender, and date of the measurement per patient and the corresponding tissue the measurements were conducted during the trial measurement period in the Erasmus MC.

#	Tissue	Measurement Location	Histological Analysis
1	Ileum	A: Normal B: Tumor	A: Normal B: Tumor
2	Ileum	A: Normal B: Tumor	A: Normal B: Tumor
3	Esophagus	A: Normal B: Tumor	A: Normal B: Tumor
4	Esophagus	A: Normal B: Tumor	A: Normal B: Tumor
5	Colon	A, B: Normal C: Tumor	A: Normal **B, C: Tumor**
6	Colon	A, B: Normal C: Tumor	A, B: Normal **C: Normal**
7	Esophagus	A, B: Normal C: Tumor	A, B: Normal **C: Normal**

Table 4-4: The results from the histological assessment per patient.

cancerous tissue. However, as the frequency increased toward 7 MHz, the real impedance of the cancerous and non-cancerous tissue seemed to cross. In the imaginary plane, it seemed that the cancerous and non-cancerous tissue have a similar trajectory with little deviation from each other.

4-2-2 Patient 2

The measurement on patient two was conducted on 29-09-2022 and was an ileum measurement on a 53-year-old female. It was part of the first trial that was conducted for this work. The tumor of patient 1 was a neuroendocrine tumor grade 2, which can arise from the excessive dividing of cells in the neuroendocrine system. The amount of differentiation determines the grade. Tumor grade 2 was intermediate, with one being the lowest and three being the highest. In Figure 4-2, it is observed that, like patient one, the cancerous part (B) starts

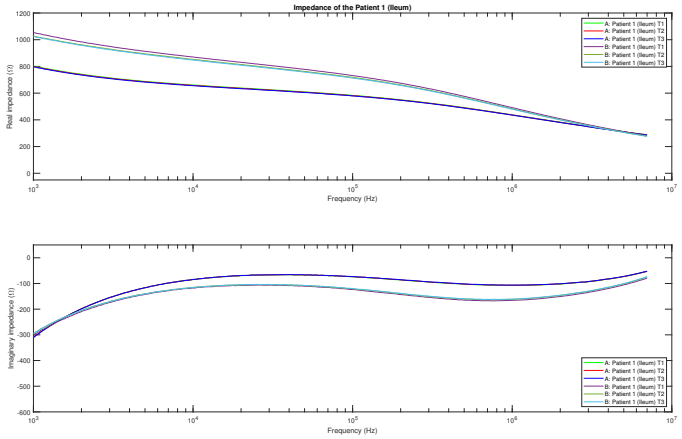


Figure 4-1: Impedance Patient 1 (Ileum)

with a higher real impedance and a similar imaginary impedance. However, as the frequency increases, the real impedance of the non-cancerous tissue (A) becomes larger.

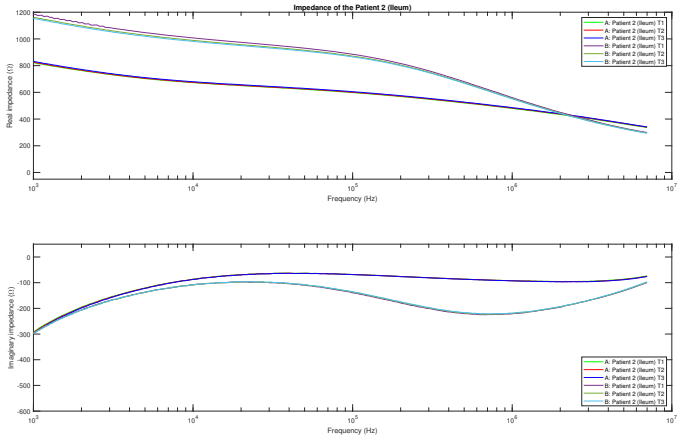


Figure 4-2: Impedance Patient 2 (Ileum)

4-2-3 Patient 3

The third patient was analyzed on 30-09-2022 and belonged to the first trial. Patient three is a male of 72 years old and was an esophagus specimen. The tumor was both adenocarcinoma combined with signet ring cell carcinoma. Signet ring cell carcinoma is a rare and highly malignant adenocarcinoma that produces mucin [49]. Mucins are proteins that help the functioning of healthy cells and tissues. In mucinous carcinoma, mucin becomes part of the tumor. In Figure 4-3, it is observed that the real part of the impedance is higher and the imaginary part of the impedance is lower, which is a clear difference from patient one and patient two. Additionally, the trajectory of the real part of the impedance for the cancerous

tissue (B) is very gradual, while the non-cancerous tissue is more steep.

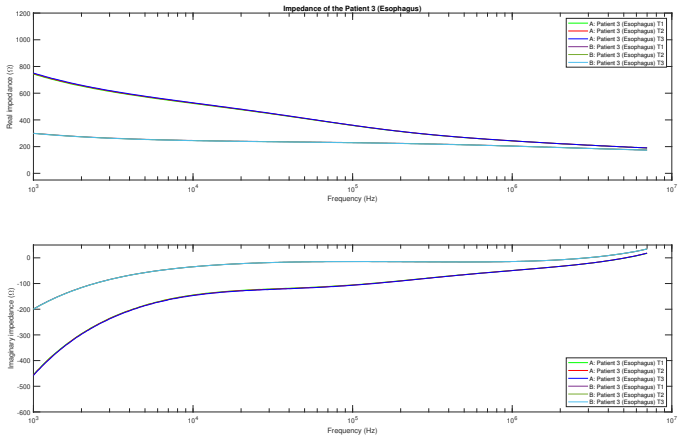


Figure 4-3: Impedance Patient 3 (Esophagus)

4-2-4 Patient 4

The measurements on the fourth patient were conducted on 14-10-2022, and belong to the measurements of trial one. The measurements were esophagus measurements of a male specimen that was 76 years of age. The tumor of patient four is dysplasia, passing into squamous cell carcinoma. Dysplasia is an abnormal development of cells, which separates them from normal tissue regarding size, shape, and structural organization of the tissue. Squamous cell carcinoma develops in the squamous cells that make up the middle and outer layers of the tissue. Most variation in the data is observed in between 1 kHZ and 5 kHz in the imaginary axis. For the real part of the impedance both the non-cancerous (A), and cancerous part (B) follows a similar trajectory.

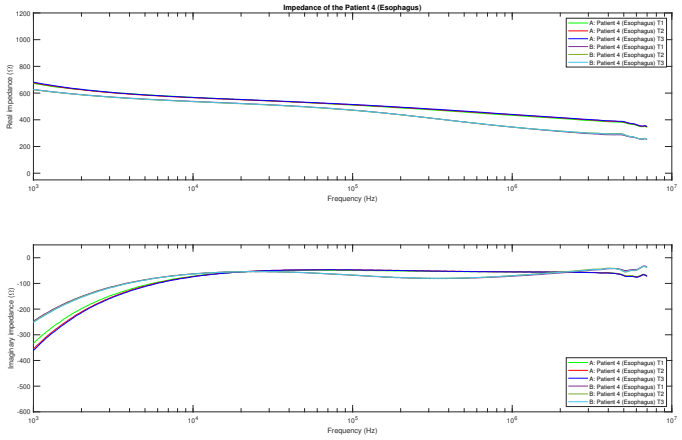


Figure 4-4: Impedance Patient 4 (Esophagus).

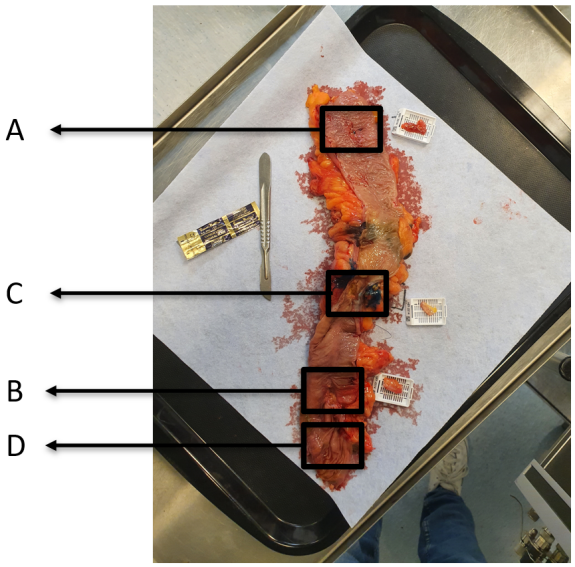


Figure 4-5: Different measurement locations of the fifth patient that is a colon specimen.

4-2-5 Patient 5

Patient five is the first patient that was included in this work that came from the second trial. The measurements on patient five were conducted on 31-01-2023, and the tissue specimen was colon tissue belonging to a 44 year old female patient. As mentioned before during the second trail images were made by the author of the tissue specimen. For patient five the colon specimen can be observed in Figure 4-5.

After histological assessment the pathologist concluded that the measurements on location B (intended to be a healthy location) were cancerous. The tumor in location B and C were both adenocarcinoma. It is observed in Figure 4-6, that the location C, which is cancerous has a higher real impedance and lower imaginary impedance than that of location B, which is also cancerous.

4-2-6 Patient 6

The sixth patient included in this work is another colon specimen from a 63-year-old male. It was the second patient from the second trial. The measurement was conducted on 28-03-2023. The measured locations can be observed in Figure 4-7. The black dye observed on the colon specimen is used for the pathologist's work and does not interfere with any of the measurements taken.

The histological assessment showed that all the measurements, including location C, were healthy non-cancerous tissue measurements. This means that for patient six there are no cancerous tissue measurements. It is also observed in Figure 4-8 that all measurement are highly similar. It can be seen that the measurements of location C (intended to be cancerous) fall in between the measurements of location A, B, and D. In this measurement a similar trajectory to the two ileum measurements from patient one and patient two is observed. This makes sense as these tissues are very close to each other in the body.

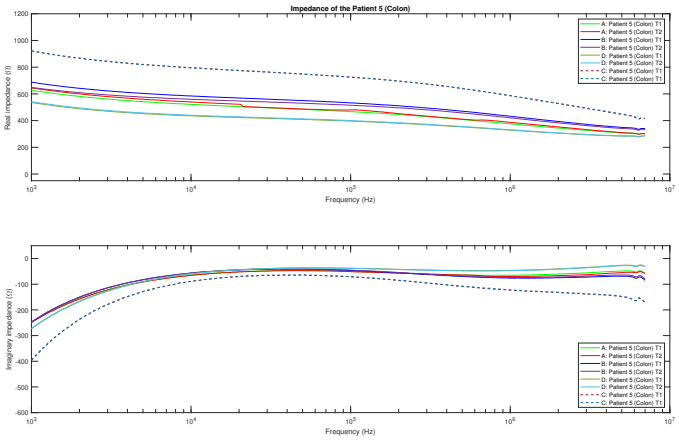


Figure 4-6: Impedance Patient 5 (Colon).

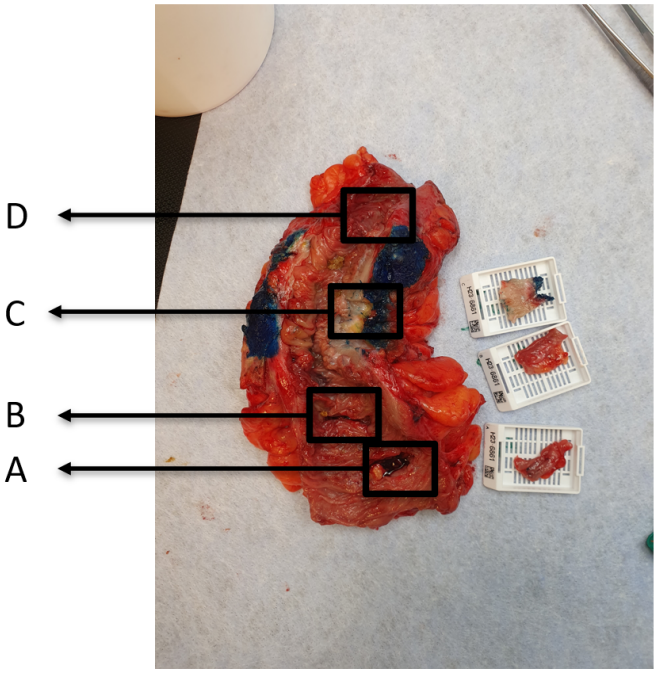


Figure 4-7: Different measurement locations of the sixth patient that is a colon specimen.

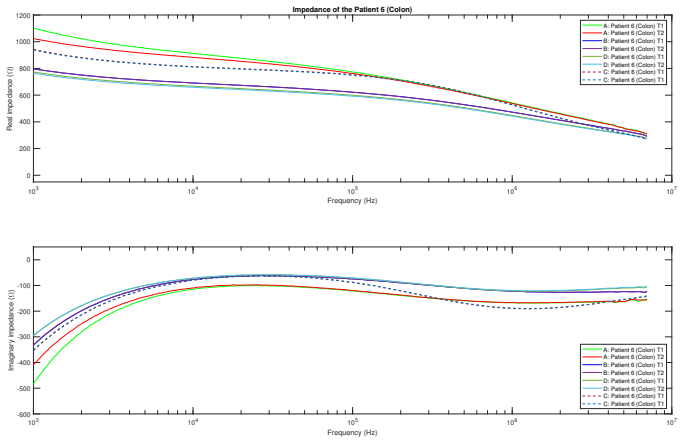


Figure 4-8: Impedance Patient 6 (Colon).

4-2-7 Patient 7

The seventh patient was the last patient of this work and the second trial. It was conducted on an esophagus specimen obtained from a 60-year-old male on 09-06-2023. The specimen is depicted in Figure 4-9, and it should be noted that this esophagus specimen was pinned to a board and cut up in between for another research project that another researcher did parallel to this work. Additionally, it should be noted that their work does not influence our measurements.

After the pathological assessment conducted at the Erasmus MC, it was found that similar to patient 6, the measurements of patient seven were also all on non-cancerous tissue. In Figure 4-10, the measurements can be observed. It is seen that although pathological assessment states that the measurements are all non-cancerous, the measurements at location C (intended to be cancerous) are slightly deviating from the measurements at the other locations. We also observe similar trajectories in the measurements compared to the other esophagus measurements from patients three and four.

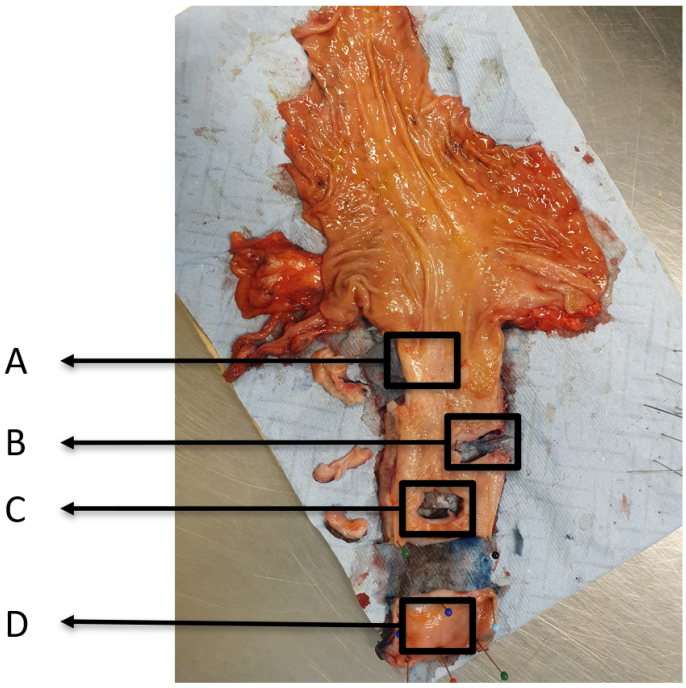


Figure 4-9: Different measurement locations of the seventh patient that is an esophagus specimen.

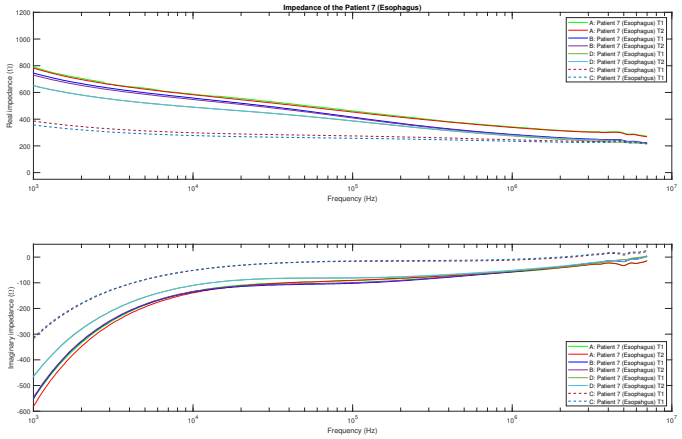


Figure 4-10: Impedance Patient 7 (Esophagus).

4-3 Cole Impedance Analysis in Combination with Thresholding

This section covers the analysis that is conducted with the Cole Impedance parameterization in combination with thresholding. Section 4-3-1 and 3-3-1 cover the way that the parameters are obtained with the Cole impedance parameterization method and how those parameters are subsequently analyzed via the Receiver Operator Characteristic (ROC) to find the threshold. It should be noted that the ROC analysis is only conducted if the data fit of the parameters is satisfactory and if a parameter that is statistically significant is found, i.e., $P < 0.05$.

4-3-1 Cole Impedance Model

The first analyzed model is the Cole impedance model that is shown in Equation 3-8. It has four parameters that were fitted via the fminsearchbnd algorithm in Matlab [27]. The algorithm was run for a maximum of 100000 iterations, and it stops when the change in the objective function has reached 1e-9. The initial values were determined in a few rounds of running the algorithm, to empirically determine suitable starting values. It was determined that $R_0 = 4000\ \Omega$, $\tau = 3e-4$, $\alpha = 0.6$, and $R_\infty = 100\ \Omega$ were good values. Convergence of the data fit for every measurement was observed. In total there were 32 non-cancerous measurements and 16 cancerous measurements. For the fitting of the non-cancerous tissue an average Mean Square Error (MSE) of 9.8208e+05 was found, and for the cancerous tissue an average MSE of 9.7343e+05 was found. This is a rather high value for the MSE. The fits of the data for the non-cancerous data can be found in Figure 4-11, 4-12, 4-13, and 4-14. The fits for the cancerous data can be found back in Figure 4-15 and 4-16. It can be observed that all Cole impedance model fits, fit the data to a certain degree, but are not able to capture the detailed curvature of the data.

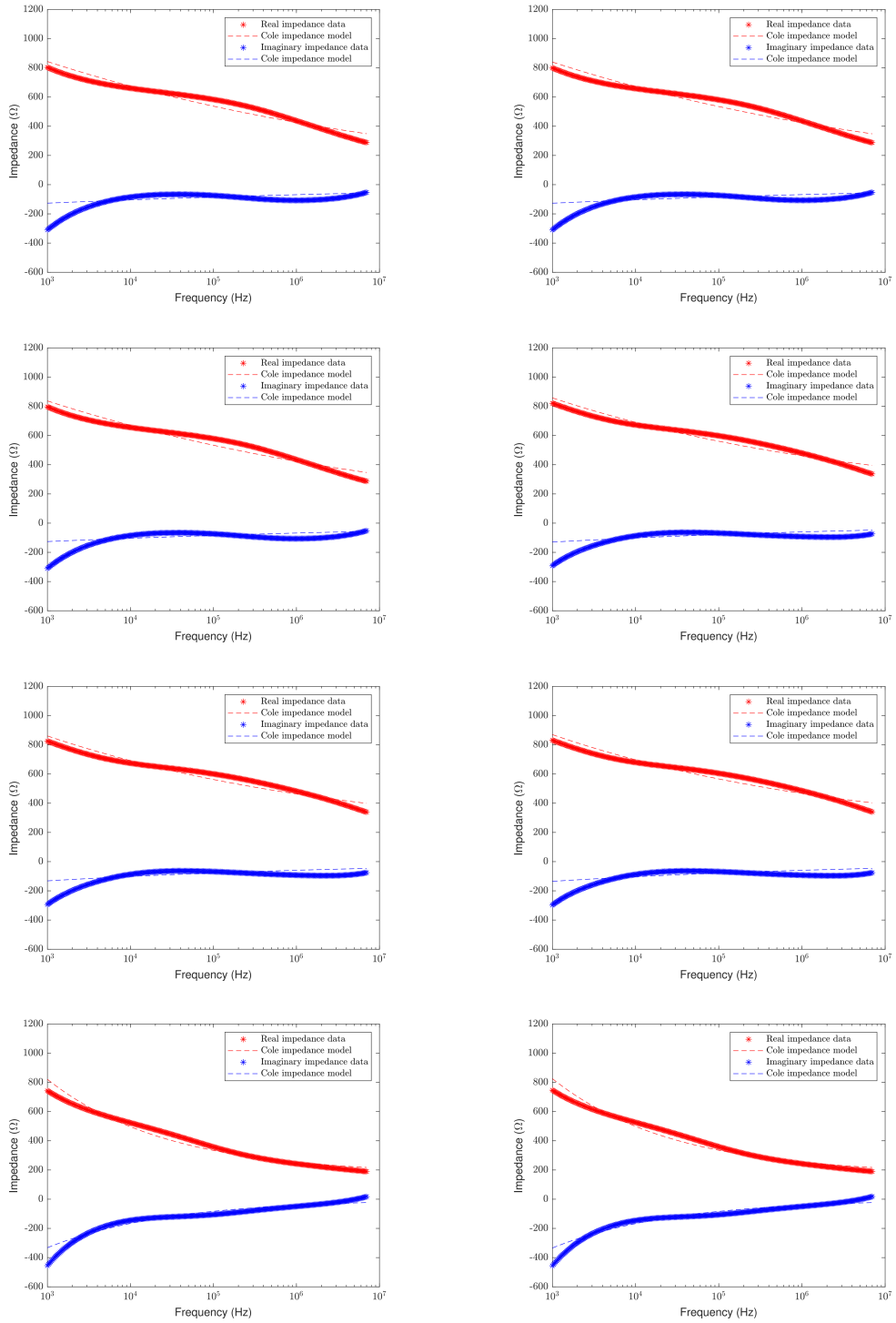


Figure 4-11: Cole impedance model fit for non-cancerous measurement 1A1, 1A2, 1A3, 2A1, 2A2, 2A3, 3A1, 3A2 (top left to right bottom).

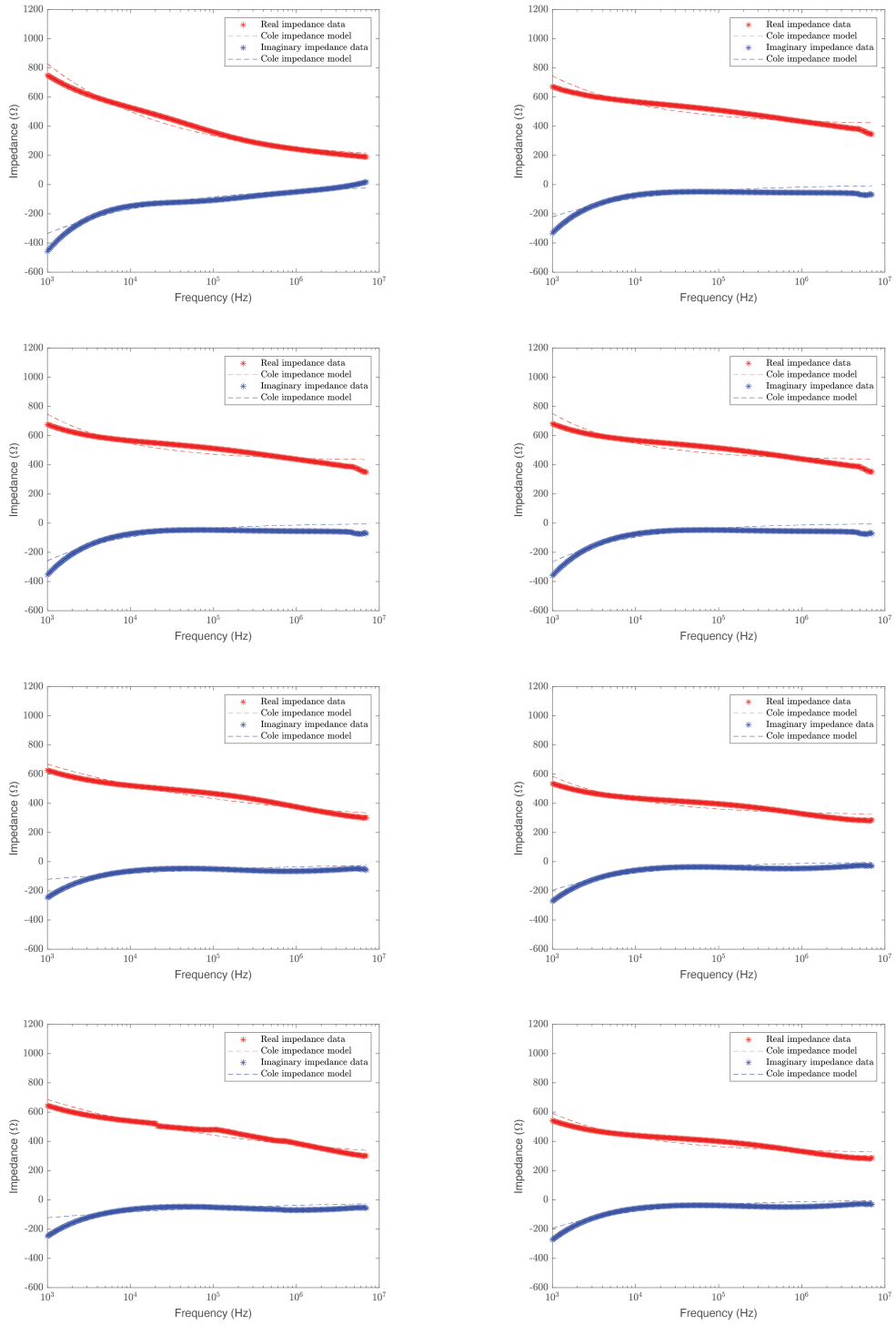


Figure 4-12: Cole impedance model fit for non-cancerous measurement 3A3, 4A1, 4A2, 4A3, 5A1, 5D1, 5A2, 5D2 (top left to right bottom).

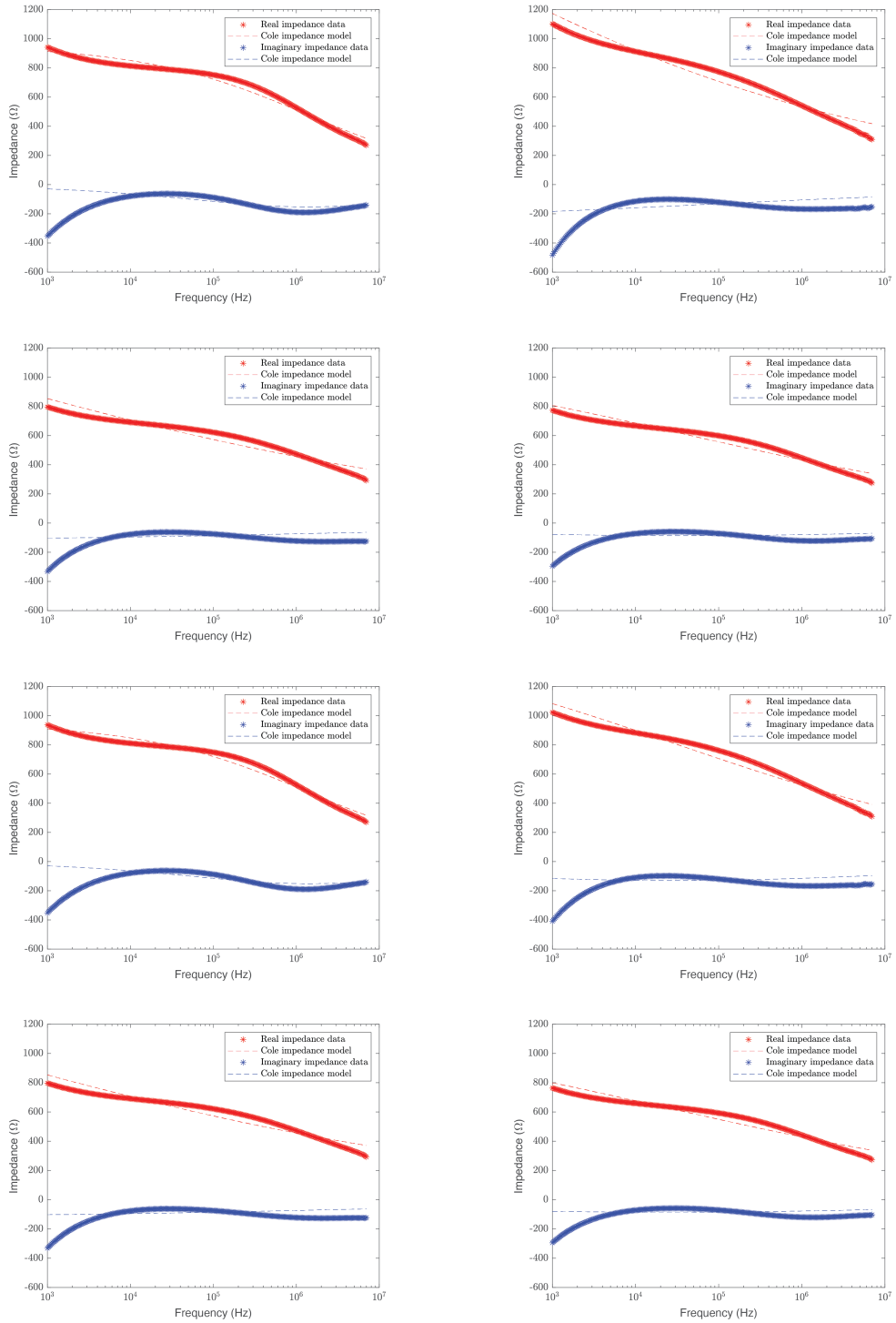


Figure 4-13: Cole impedance model fit for non-cancerous measurement 6C1, 6A1, 6B1, 6D1, 6C2, 6A2, 6B2, 6D2 (top left to right bottom).

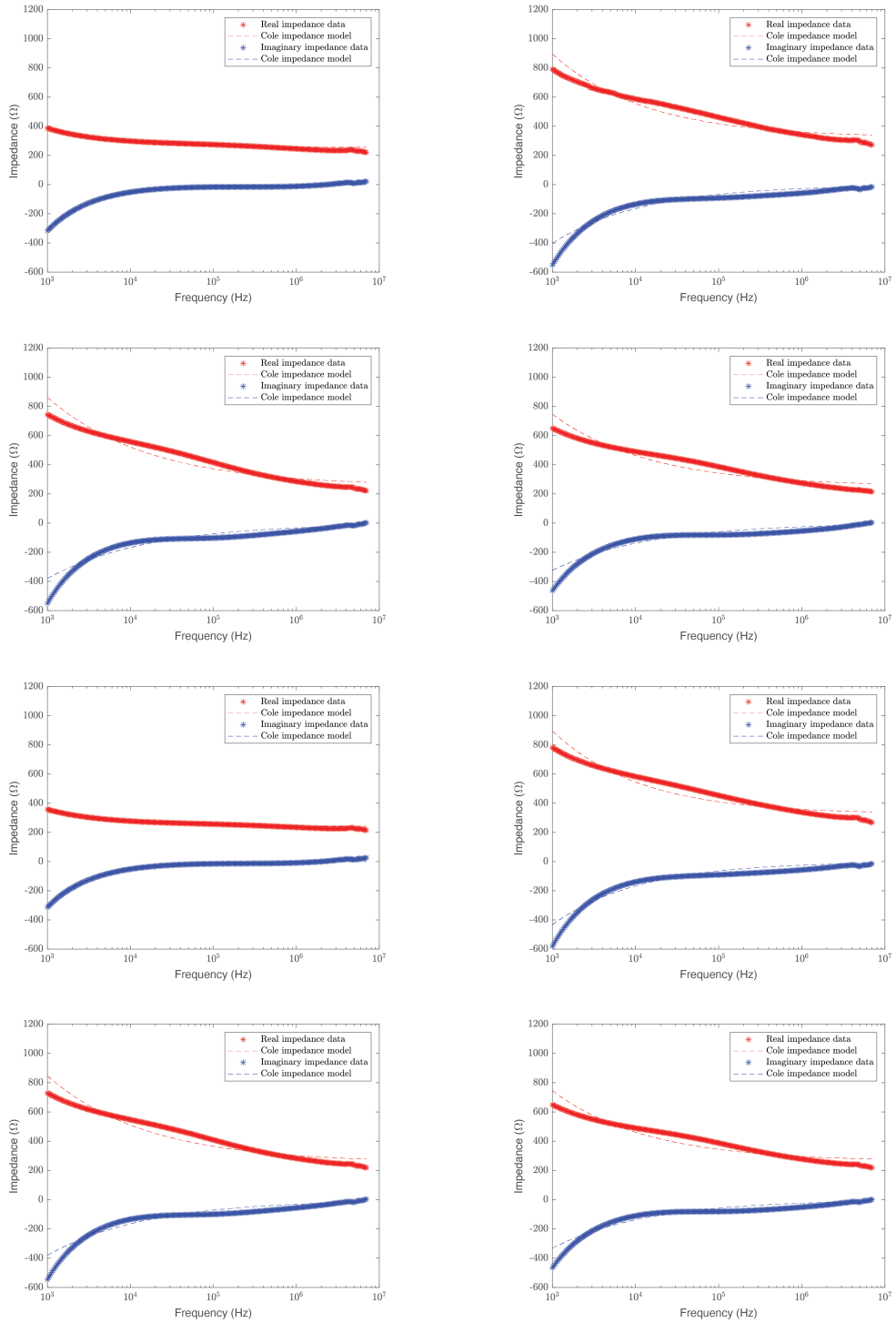


Figure 4-14: Cole impedance model fit for non-cancerous measurement 7C1, 7A1, 7B1, 7D1, 7C2, 7A2, 7B2, 7D2 (top left to right bottom).

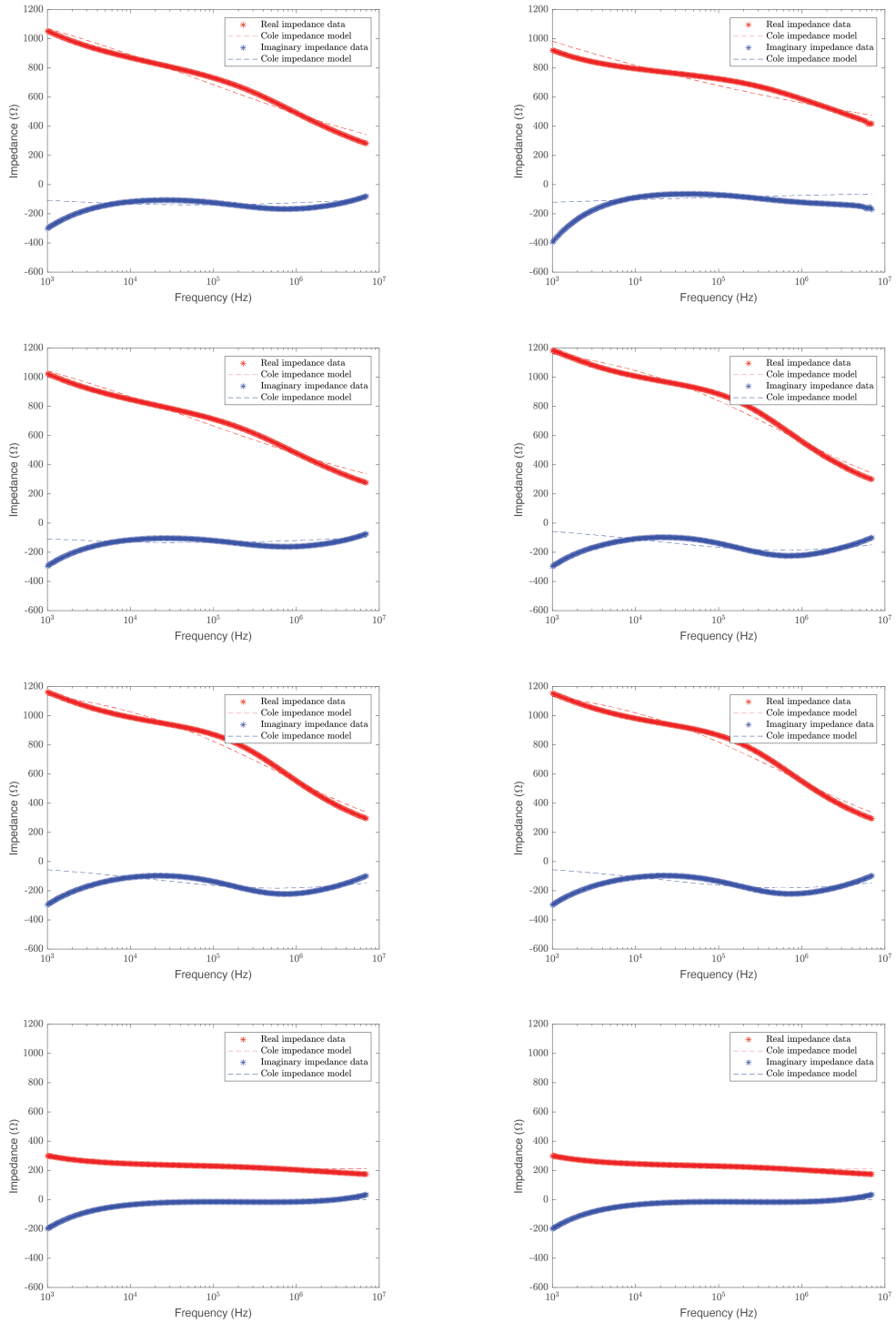


Figure 4-15: Cole impedance model fit for cancerous measurement 1B1, 1B2, 1B3, 2B1, 2B2, 2B3, 3B1, 3B2 (top left to right bottom).

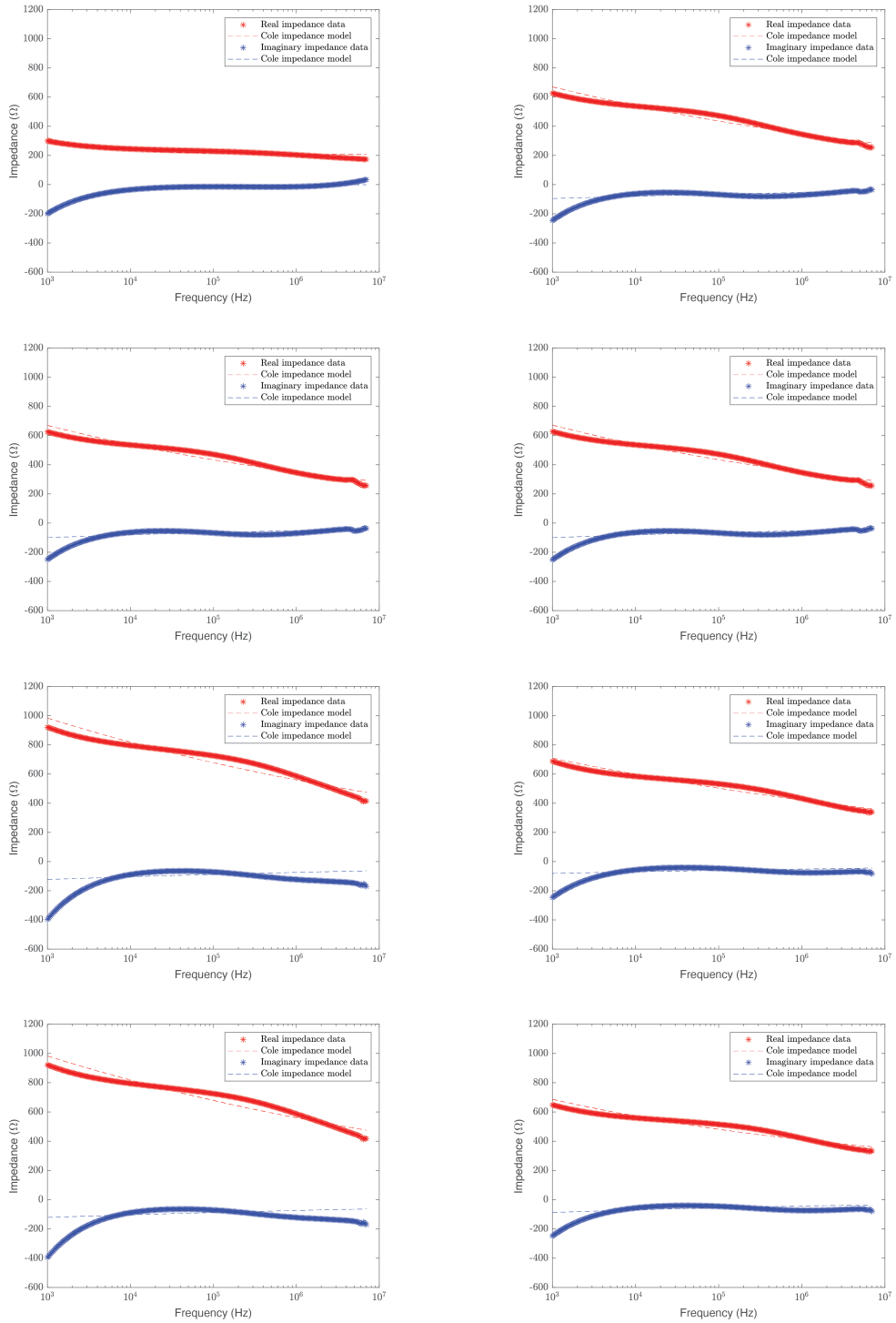


Figure 4-16: Cole impedance model fit for cancerous measurement 3B3, 4B1, 4B2, 4B3, 5C1, 5B1, 5C2, 5B2 (top left to right bottom).

ID	R_0	τ	α	R_∞
1B1	1452	2.51	0.2405	2.86
1B2	1424	2.52	0.2392	2.71
1B3	1438	2.90	0.2357	1.78
2B1	1261	2.80	0.3638	1.03
2B2	1237	2.74	0.3640	6.25
2B3	1229	2.74	0.3628	7.34
3B1	4309565	697	0.6484	210
3B2	6096296	1129	0.6504	210
3B3	4616395	721	0.6512	209
4B1	4653	3006	0.1058	0.44
4B2	5941	23886	0.1126	45.7
4B3	6214	46047	0.1107	38.1
5C1	7517	170907	0.0907	0.09
5B1	4142	8795	0.0896	18.8
5C2	6995	64323	0.0911	0.73
5B2	6100	81915	0.1159	164

Table 4-5: Four Cole impedance model parameters for the cancerous measurements.

After fitting the data, the parameters were extracted and stored in a separate table. For the parameters made, with the non-cancerous data, one can look at Table 4-6. For the data that was created with the cancerous data, one can observe Table 4-5. Please note that the ID stands for the measurement and should be read as Patient-Location-Measurement, so 5D1 corresponds to patient five, location D, first measurement on that location. It is observed from both tables that there is quite some variation regarding the parameters, for example, τ, either the parameter is in the 10000 range or close to 0, the same hold for the R_∞ parameter, this parameter either goes towards zero, or around the 200 - 300 Ω.

The analysis to determine the statistical significance of the parameters can now be executed as was described in Section 3-3-1 and Figure 3-8. The parameters are not normally distributed in the one-sample Kolmogorov-Smirnov test. This leads to the use of the Two-sided Wilcoxon signed rank test and the Two-sided Wilcoxon rank sum test.

It was found via the Two-sided Wilcoxon signed rank test that the median is not zero for all parameters. Additionally, one statistically significant parameter was found. The test indicated that R_∞ comes from a continuous distribution where the medians of the parameter R_∞ when it comes from non-cancerous data are statistically different than when it comes from cancerous data. In Table 4-7, all parameters and their P-value of the Two-sided Wilcoxon rank sum test are shown.

Although a statistical significance was found, the fourth parameter R_∞, from the literature it is known that this model is a single dispersion model; and since the tissue measurements in this work were taken over a range of 1 kHz to 7 MHz, It is also known that the measurements are in both the α and β dispersion, this also reflects in the large MSE values for the data fit. Because of this, this model and these parameters are deemed unsuitable for further data analysis.

ID	R_0	τ	α	R_∞
1A1	6571	6391	0.1088	0.06
1A2	7925	67286	0.1068	0.00
1A3	7293	25592	0.1075	0.28
2A1	8017	11772	0.1269	151
2A2	9653	50505	0.1287	166
2A3	9837	37454	0.1318	176
3A1	419094	218258	0.3019	173
3A2	260500	44220	0.3019	172
3A3	367066	155361	0.2998	171
4A1	573991	45573	0.3732	412
4A2	2349968	73547	0.4338	430
4A3	1257678	11660	0.4426	433
5A1	30504	1996558	0.1818	250
5D1	871134	47225	0.4038	320
5A2	16708	84617	0.1770	244
5D2	482731	15009	0.3962	321
6C1	956	1.15	0.3935	3.67
6A1	3733	0.03	0.1448	1.92
6B1	2037	0.00	0.1314	8.51
6D1	1217	3.99	0.1800	3.60
6C2	953	1.14	0.3927	4.45
6A2	1617	5.36	0.2036	6.14
6B2	2019	0.00	0.1320	2.50
6D2	1254	6.45	0.1734	1.34
7C1	9347733	346	0.7018	256
7A1	1463664	53280	0.3894	322
7B1	845987	70410	0.3552	253
7D1	1109590	119033	0.3688	252
7C2	8591727	161	0.7356	243
7A2	2212794	43909	0.4134	326
7B2	1115037	101718	0.3633	254
7D2	1385973	104790	0.3825	261

Table 4-6: Four Cole impedance model parameters for the non-cancerous measurements.

	R_0	τ	α	R_∞
P-value	0.064	0.071	0.200	**0.0130**

Table 4-7: The P-values of the Cole impedance parameters, with only R_∞ being statistically significant (P$<$ 0.05).

4-3-2 Two-Pole Cole Impedance Model

The two-pole Cole impedance model is a model that makes use of two Cole impedance models in series. It was introduced in Equation 3-10 in Section 4-3-1. The two-pole Cole impedance model should be able to create a data fit for when two dispersion regions (in this case α and β) are being considered [15]. For the initial parameter values $R_1 = 400000\ \Omega$, $\tau_1 = 2e7$, $\alpha_1 = 0.6$, $R_2 = 600\ \Omega$, $\tau_2 = 1e-8$, $\alpha_2 = 0.6$, and $R_\infty = 200\ \Omega$ were picked. After running the data fitting algorithm, a much better fit was found in comparison to the Cole impedance model. This was to be expected since there are more parameters the data can fit. Additionally, [15] stated that the Cole impedance model could be placed in series with another Cole impedance model if multiple dispersion regions were being researched. A MSE of 4.1228e+04 was obtained for the non-cancerous data fits, and a MSE of 2.6642e+04 for the data fits conducted on the cancerous data. This makes the two-pole Cole impedance equation over ten times more accurate. It can also be observed in Figure 4-17, 4-18, 4-19, 4-20, 4-21, and 4-22, that the fit of the two-pole Cole model comes much closer to the actual data. In some cases, it is impossible to see the data fit because it is plotted under the real data (indicated with stars). A drawback of this method is however, that when the imaginary part becomes positive, meaning that the impedance changes from capacitive to inductive impedance, the algorithm is not able to track this change.

Seven different parameters were obtained after fitting the data to the two-pole Cole impedance model. These are showcased in Table 4-9, and 4-8. It can be seen that the numbers variate a large amount for parameters R_1 and τ_2. A possible reason could be that the solver gets stuck in a local minima that allows the data to be fitted correctly but creates big differences in parameter values.

ID	R_1	τ_1	α_1	R_2	τ_2	α_2	R_∞
1B1	24619	0.31	0.5480	622	2.00	0.6118	198
1B2	178717879	4434607	0.5437	604	2.03	0.6144	199
1B3	2977078	2765	0.5383	595	2.02	0.6176	200
2B1	1852129892	2074996758	0.5093	692	1.97	0.7257	243
2B2	226354171	2702995	0.5649	710	1.97	0.7065	229
2B3	2322085	746	0.5681	704	1.96	0.7079	229
3B1	2382416	149987579	0.3444	9.32	5.63	0.2265	172
3B2	1.e37	1.e36	0.8757	3690	3.e22	0.0448	34.9
3B3	215066021	7845	0.7814	64.4	1.18	0.5474	169
4B1	32768433	705	0.7677	300	3.55	0.6185	240
4B2	103679	0.37	0.7705	292	3.66	0.6164	246
4B3	127506	0.47	0.7708	295	3.58	0.6147	244
5C1	5881856	25.3	0.7993	810	2.23	0.4441	2.41
5B1	3460	0.00	0.7564	338	1.02	0.5437	241
5C2	2975066	15.0	0.7750	640	4.44	0.5127	152
5B2	1106665	4.70	0.8143	478	3.34	0.4001	105

Table 4-8: Seven two-pole Cole impedance model parameters for the cancerous measurements.

With the statistical analysis described in section 3-3-1, a similar result to the parameters of the Cole impedance model was found. The parameters are not from a normal distribution.

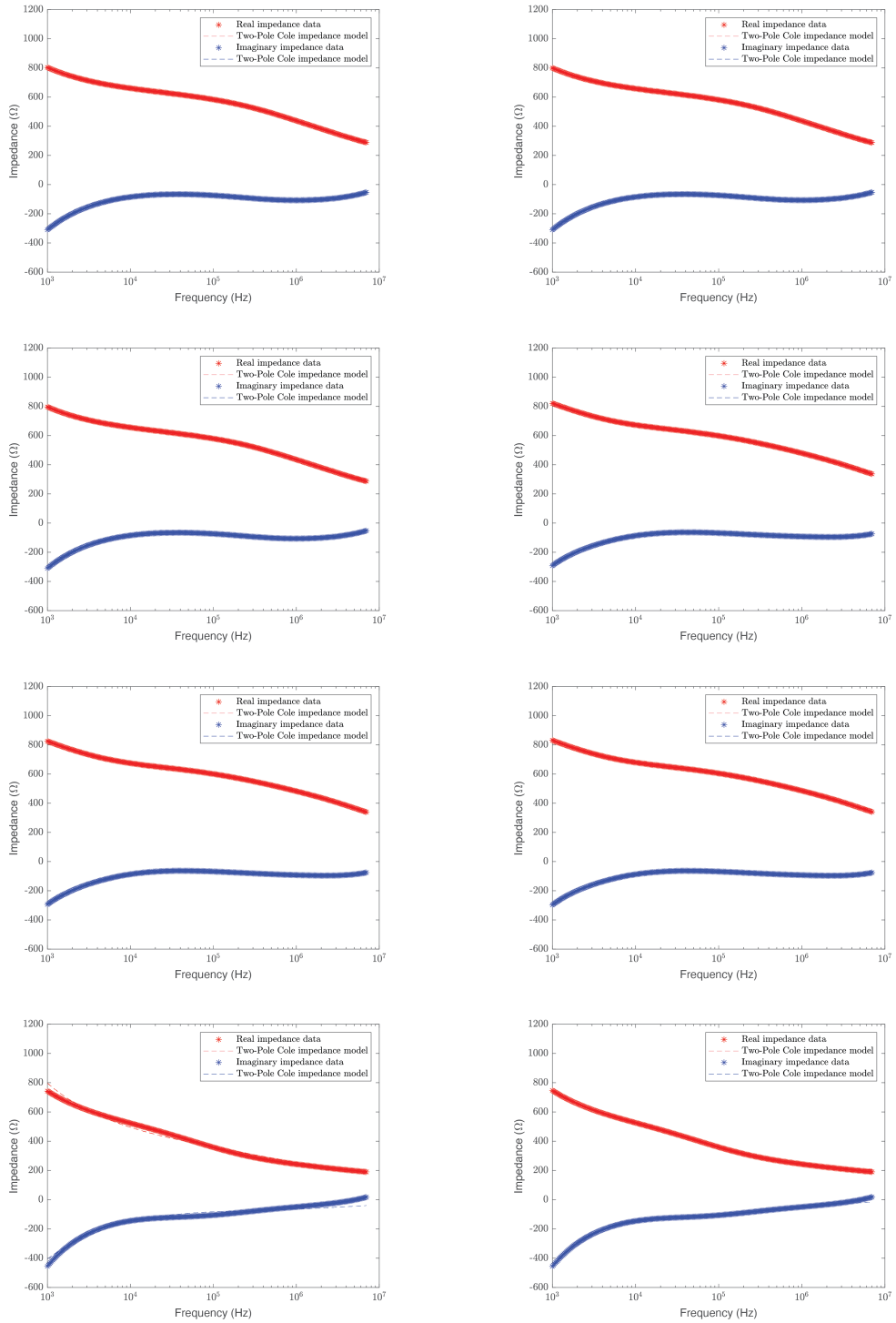


Figure 4-17: Two-pole Cole impedance model fit for non-cancerous measurement 1A1, 1A2, 1A3, 2A1, 2A2, 2A3, 3A1, 3A2 (top left to right bottom).

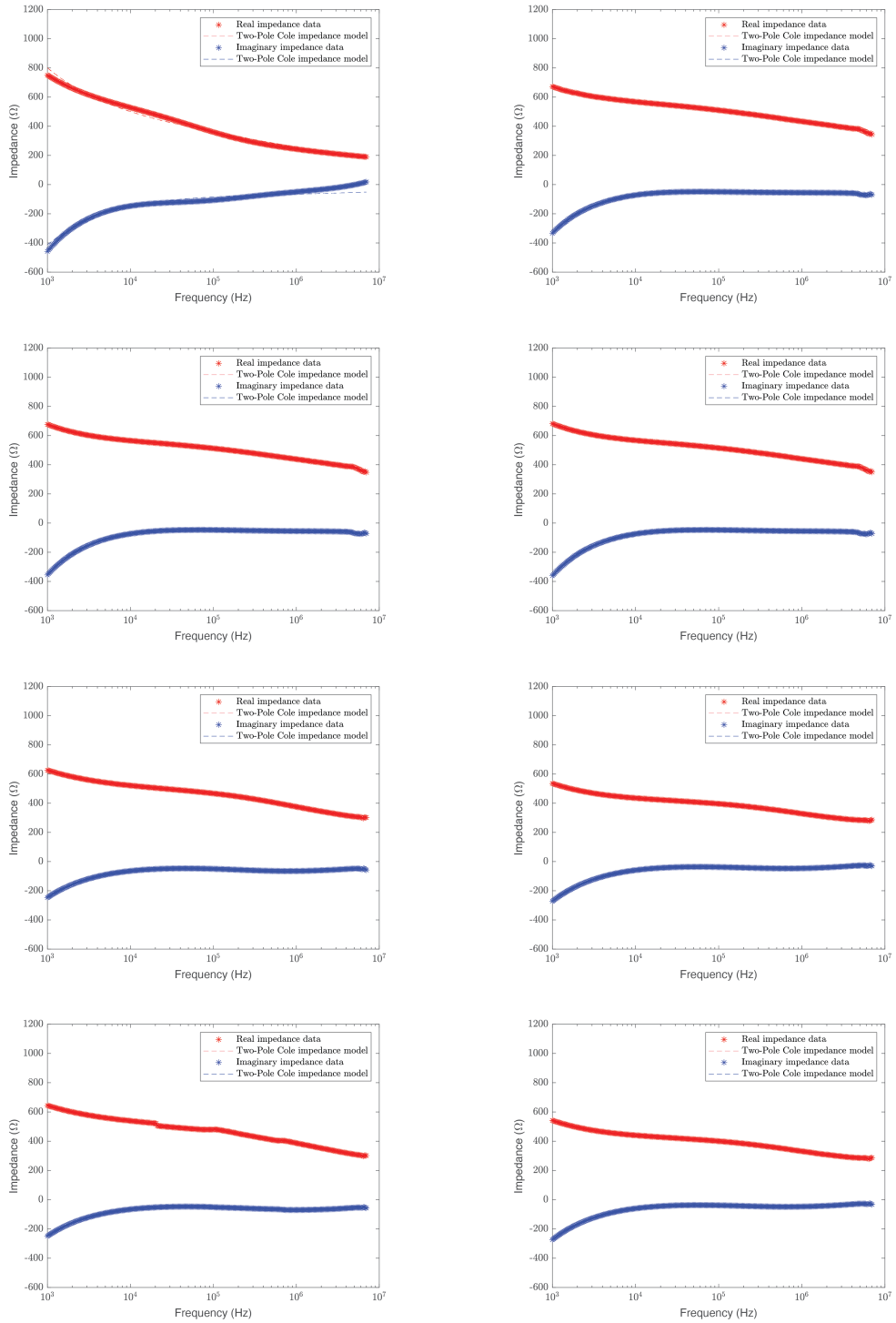


Figure 4-18: Two-pole Cole impedance model fit for non-cancerous measurement 3A3, 4A1, 4A2, 4A3, 5A1, 5D1, 5A2, 5D2 (top left to right bottom).

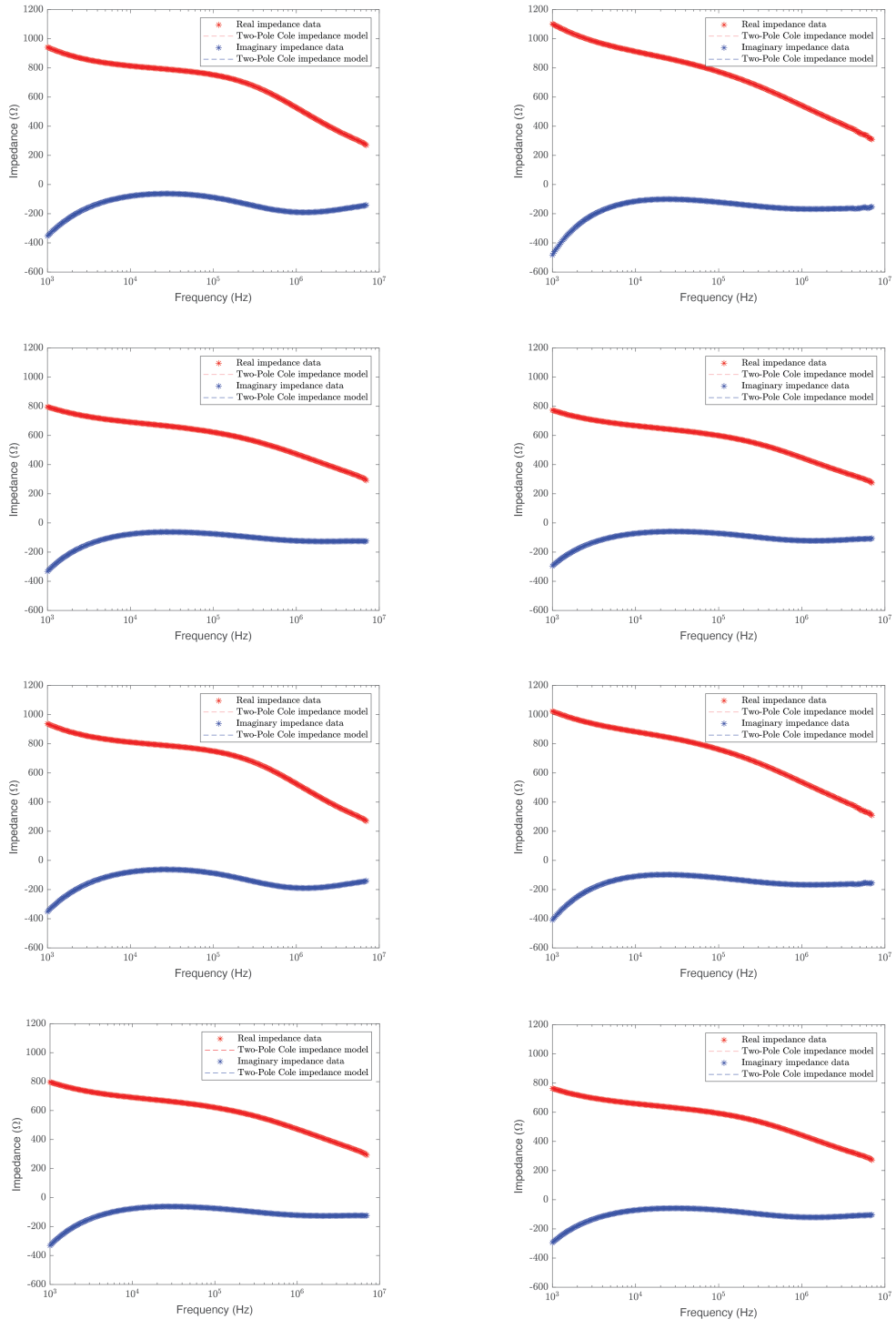


Figure 4-19: Two-pole Cole impedance model fit for non-cancerous measurement 6C1, 6A1, 6B1, 6D1, 6C2, 6A2, 6B2, 6D2 (top left to right bottom).

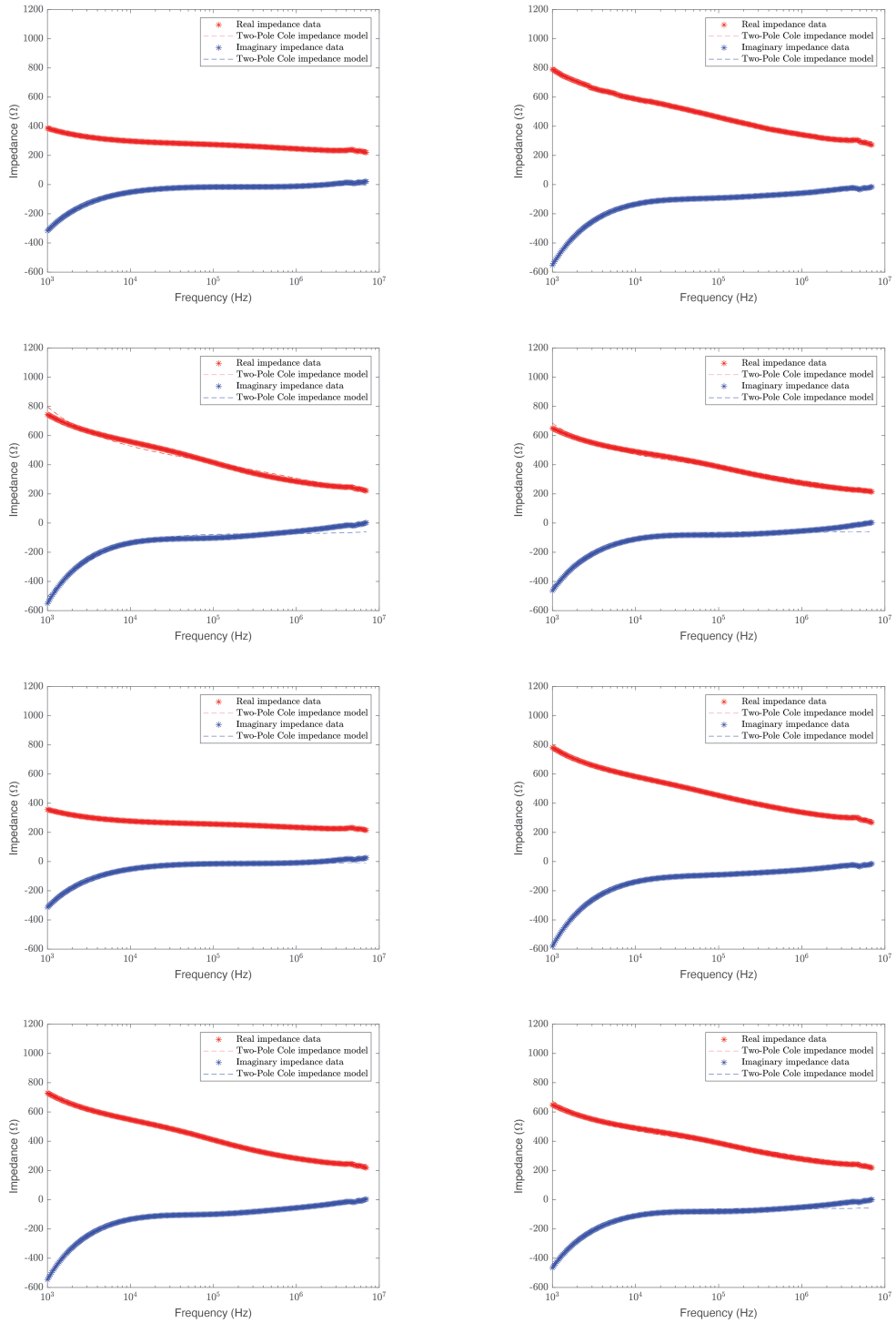


Figure 4-20: Two-pole Cole impedance model fit for non-cancerous measurement 7C1, 7A1, 7B1, 7D1, 7C2, 7A2, 7B2, 7D2 (top left to right bottom).

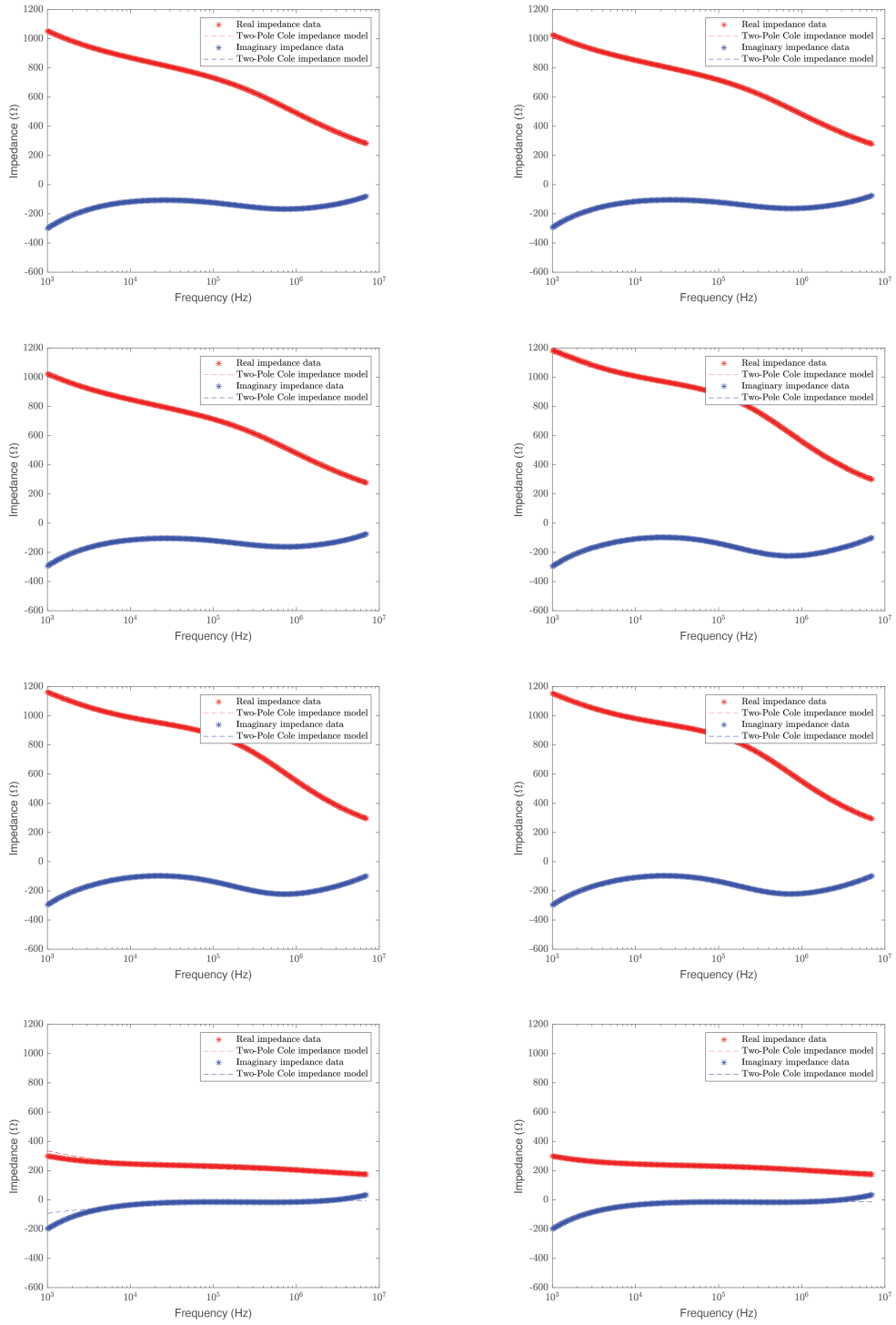


Figure 4-21: Two-pole Cole impedance model fit for cancerous measurement 1B1, 1B2, 1B3, 2B1, 2B2, 2B3, 3B1, 3B2 (top left to right bottom).

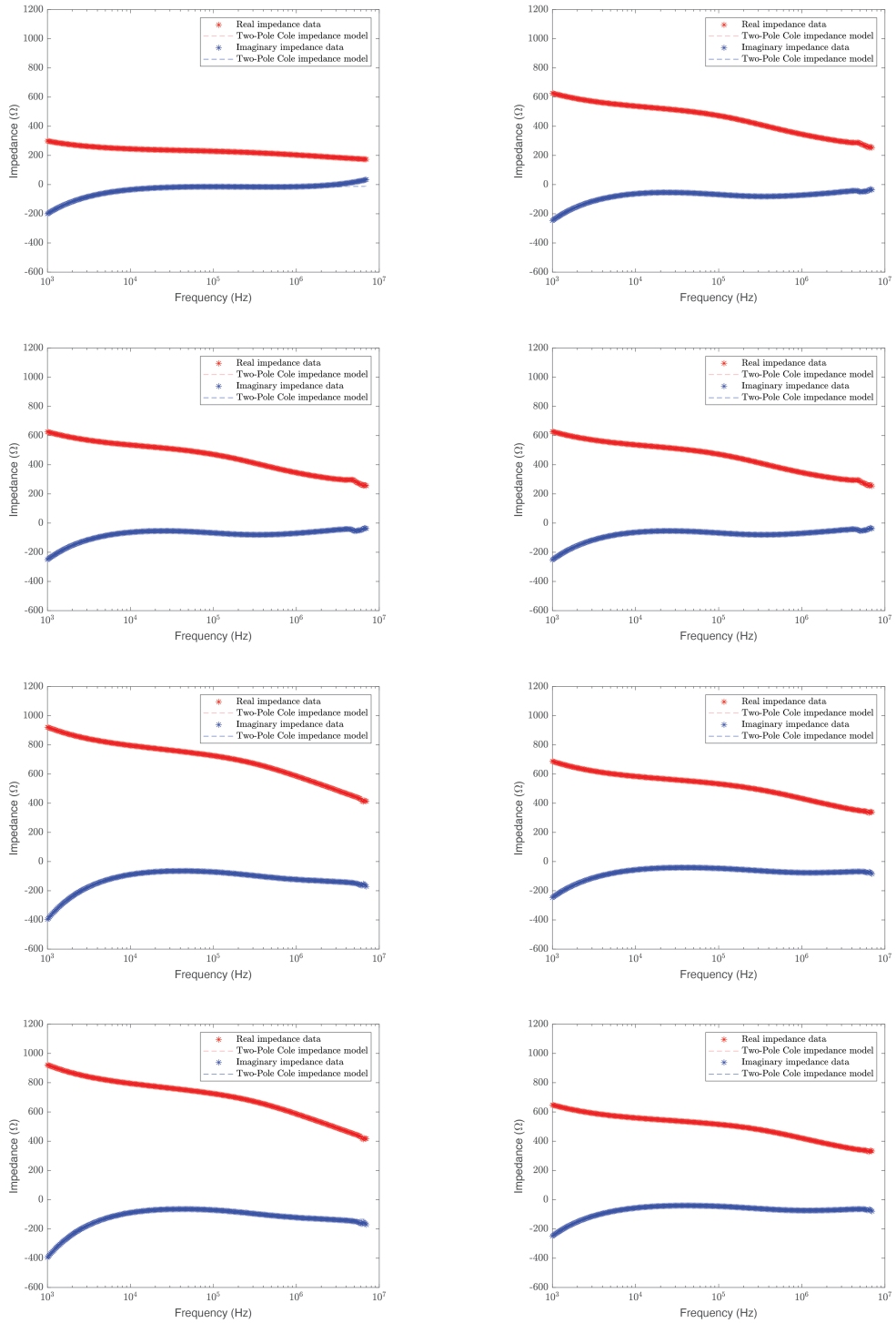


Figure 4-22: Two-pole Cole impedance model fit for cancerous measurement 3B3, 4B1, 4B2, 4B3, 5C1, 5B1, 5C2, 5B2 (top left to right bottom).

ID	R_1	τ_1	α_1	R_2	τ_2	α_2	R_∞
1A1	2292817	91.8	0.6618	399	1.61	0.6266	234
1A2	10983373	954	0.6629	397	1.61e-07	0.6268	234
1A3	9006204	692	0.6639	395	1.61	0.6272	233
2A1	260752926	308988	0.6330	426	8.79	0.5303	222
2A2	2095144250	8297387	0.6327	430	8.37	0.5276	220
2A3	8572704	1361	0.6330	433	8.40	0.5287	222
3A1	39384	0.94	0.4829	238	2.32	0.5163	134
3A2	1.e17	5.65e14	0.6997	346	2.04	0.6166	194
3A3	6296291	18840	0.5016	363	1.70	0.3652	65.1
4A1	540351	1.10	0.8368	544	1.78	0.2845	76.4
4A2	1237079	3.17	0.8228	526	1.24	0.3063	73.4
4A3	1031909	2.73	0.8143	471	1.94	0.3335	122
5A1	25727	0.10	0.7083	272	1.63	0.5521	240
5D1	389350	1.24	0.8095	381	2.92	0.2928	93.4
5A2	6652131	391	0.6863	314	1.20	0.5019	213
5D2	17004	0.04	0.7453	170	2.20	0.6301	256
6C1	30352855	443	0.7610	644	1.14	0.6852	159
6A1	1447294	5.03	0.7758	950	9.38	0.4426	8.24
6B1	428877	1.12	0.8050	615	6.62	0.5036	88.4
6D1	160341143	5618	0.7571	522	9.74	0.5600	144
6C2	1584302	9.08	0.7620	642	1.13	0.6832	158
6A2	1111687	2.56	0.8161	899	9.66	0.4582	31.0
6B2	16901573	115	0.8016	613	6.68	0.5045	89.9
6D2	1104035	7.39	0.7610	512	9.92	0.5630	146
7C1	326200060	7737	0.7795	38.7	4.84	0.9348	238
7A1	323477	0.60	0.7690	331	1.13	0.5670	271
7B1	5339011	249	0.6393	422	1.23	0.3894	82.1
7D1	27836	0.06	0.6892	420	8.18	0.3551	53.7
7C2	1082815	3.05	0.8257	455	2.72	0.0466	1.81
7A2	1750581382	22501	0.7940	352	1.32	0.5310	262
7B2	7806850	28.3	0.7893	338	1.36	0.6131	225
7D2	2780632	32.1	0.7096	476	7.64	0.3073	28.1

Table 4-9: Seven two-pole Cole impedance model parameters for the non-cancerous measurements.

This allowed for the use of the Two-sided Wilcoxon signed rank test, which showcased that all parameters in both the cancerous and non-cancerous groups have a mean that is not equal to zero. The Two-sided Wilcoxon rank sum test tested whether or not a parameter from the two-pole Cole impedance model can be used for classification. However, a single parameter with any statistical significance ($P < 0.05$) was not found. Therefore, this model and these parameters were not continued in our classification strategy. The P-values for all parameters can be found in Table 4-10.

	R_1	τ_1	α_1	R_2	τ_2	α_2	R_∞
P-value	0.718	0.309	0.330	0.319	0.241	0.241	0.185

Table 4-10: The P-values of the two-pole Cole impedance parameters, with no statistically significant parameter (P$<$ 0.05).

4-3-3 Two-Pole Cole Impedance Model with CPE

The last equivalent circuits (EC) model we analyzed was the Two-Pole Cole impedance model in combination with the Constant Phase Element (CPE) model from Equation 3-11. This led to a combined model that was displayed in Equation 3-12 and was also used in [27]. This model contained nine parameters because of the additional two from the CPE model. The initial parameters were set at $R_1 = 101\ \Omega$, $\tau_1 = 3e-4$, $\alpha_1 = 0.7$, $R_2 = 253\ \Omega$, $\tau_2 = 1e-8$, $\alpha_2 = 0.8$, $K = 2.5e-6$, $m = 0.9$, and $R_\infty = 188\Omega$. A very satisfying fit was obtained with MSE of 1.68402e+04 and 1.1296e+04 for the non-cancerous and cancerous data, respectively. This means that the data fit with the combined model is roughly twice as accurate as the data fit from the two-pole Cole impedance model. The fitted data is displayed for the non-cancerous data in Figure 4-23, 4-24, 4-25, and 4-26 and the cancerous data in Figure 4-27, and 4-28. These figures show the contribution of the CPE model and the two-pole Cole impedance model separately. It can be observed that the main contribution of the CPE is in the lower frequency range from 1 kHz until 10 kHz. This was expected as a CPE model is usually introduced to mitigate the negative effect of electrode polarization, which occurs in the α dispersion [14]. However, in some cases, it can also be observed that the CPE component does not do anything; for example, in 2B3, the imaginary part of the CPE remains zero, while for the second measurement on the similar patient at the similar location, 2B2, both the imaginary and real part of the CPE component is non-zero. Also, here, it was observed that with this model it is impossible to fit the data to the imaginary impedance when it changes from capacitive to inductive impedance.

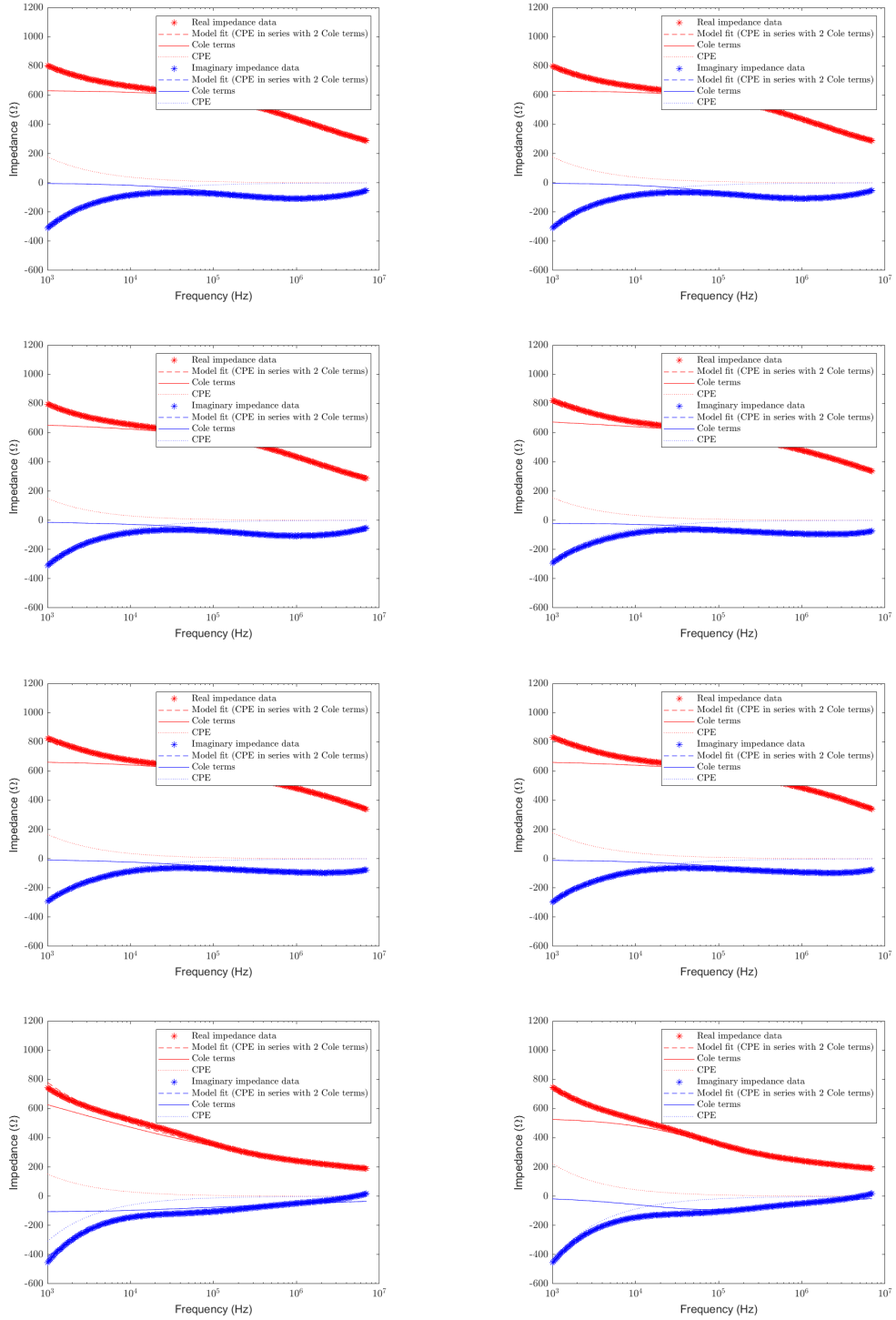


Figure 4-23: Combined CPE and 2-pole Cole impedance model fit for non-cancerous measurement 1A1, 1A2, 1A3, 2A1, 2A2, 2A3, 3A1, 3A2 (top left to right bottom).

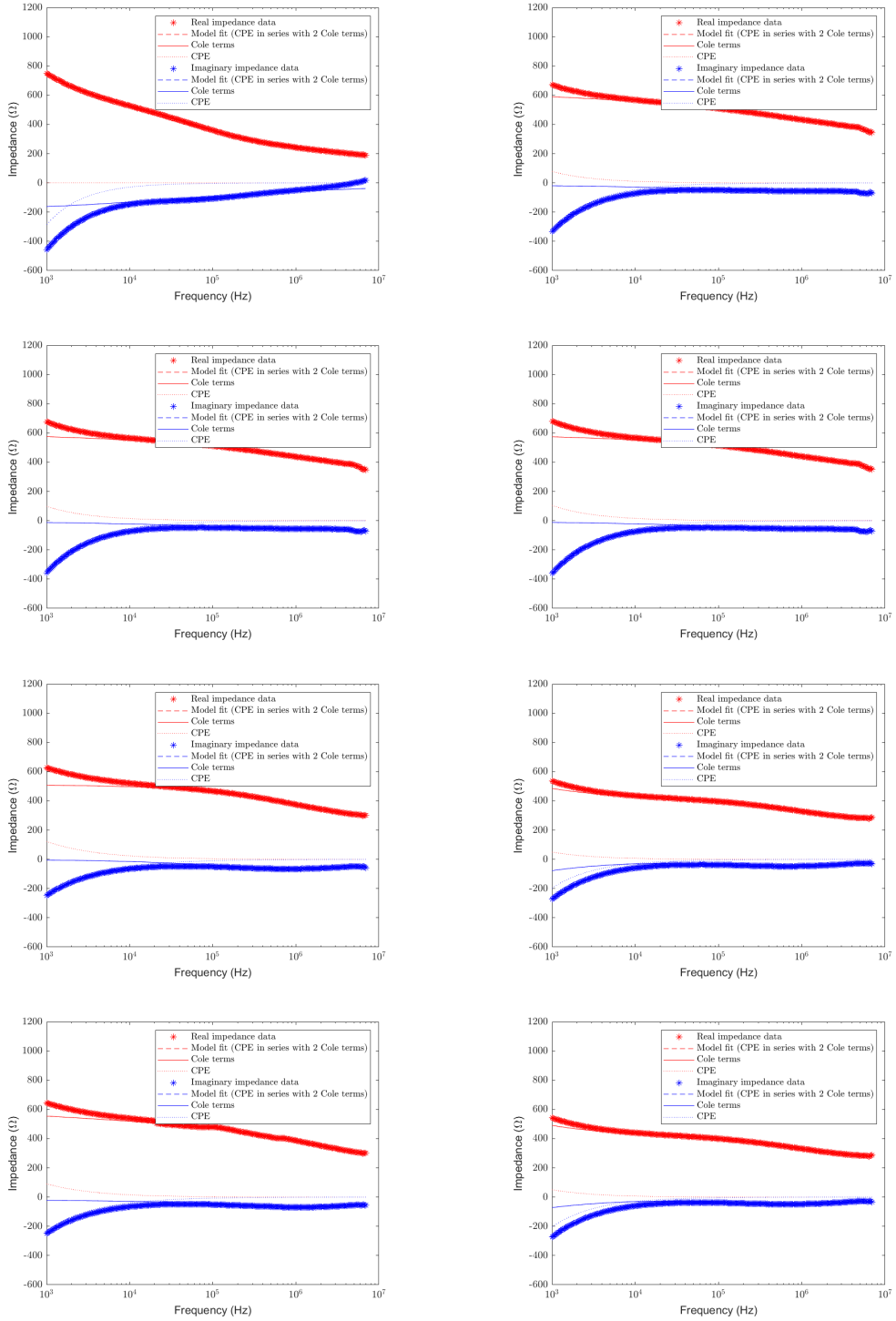


Figure 4-24: Combined CPE and 2-pole Cole impedance model fit for non-cancerous measurement 3A3, 4A1, 4A2, 4A3, 5A1, 5D1, 5A2, 5D2 (top left to right bottom).

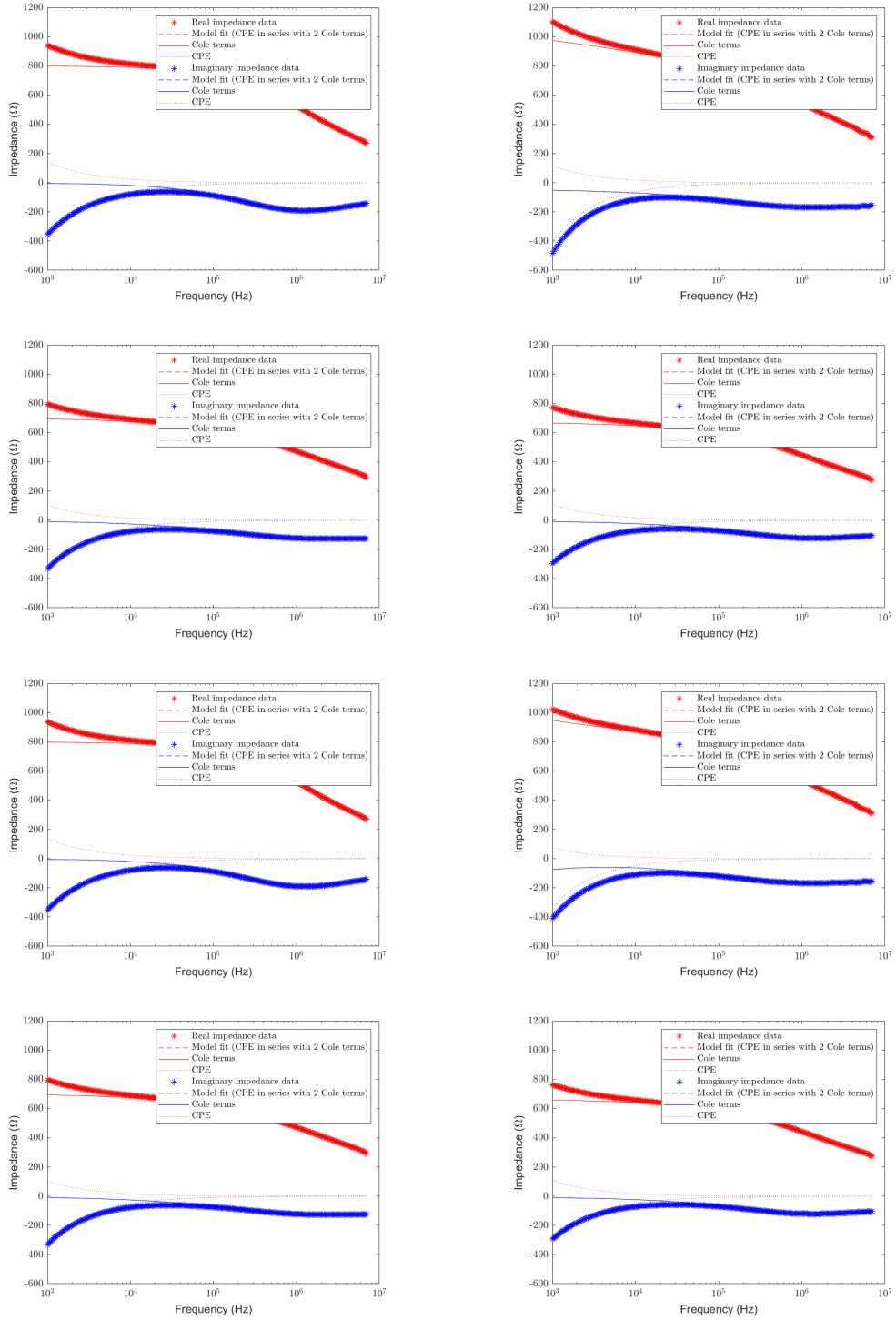


Figure 4-25: Combined CPE and 2-pole Cole impedance model fit for non-cancerous measurement 6C1, 6A1, 6B1, 6D1, 6C2, 6A2, 6B2, 6D2 (top left to right bottom).

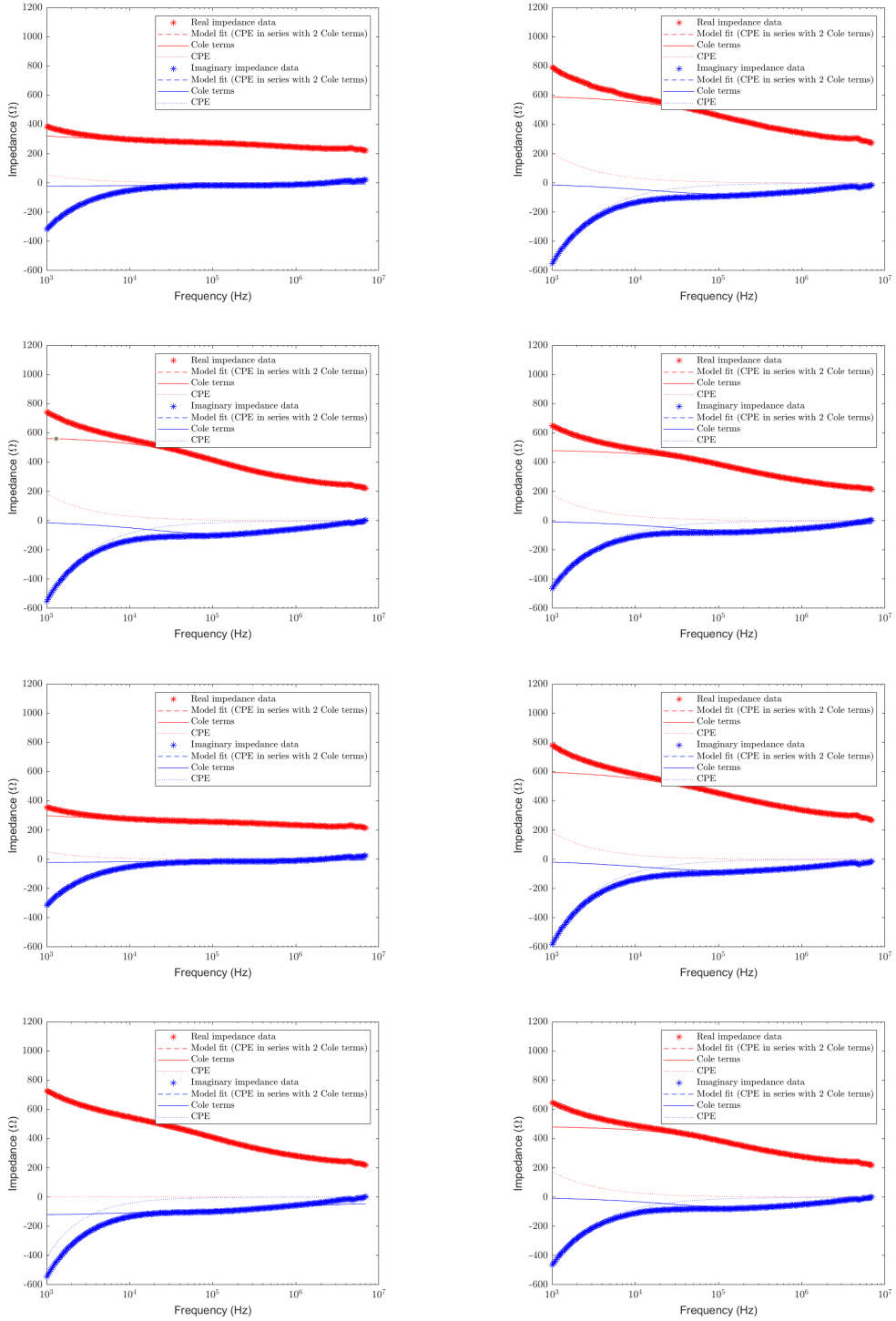


Figure 4-26: Combined CPE and 2-pole Cole impedance model fit for non-cancerous measurement 7C1, 7A1, 7B1, 7D1, 7C2, 7A2, 7B2, 7D2 (top left to right bottom).

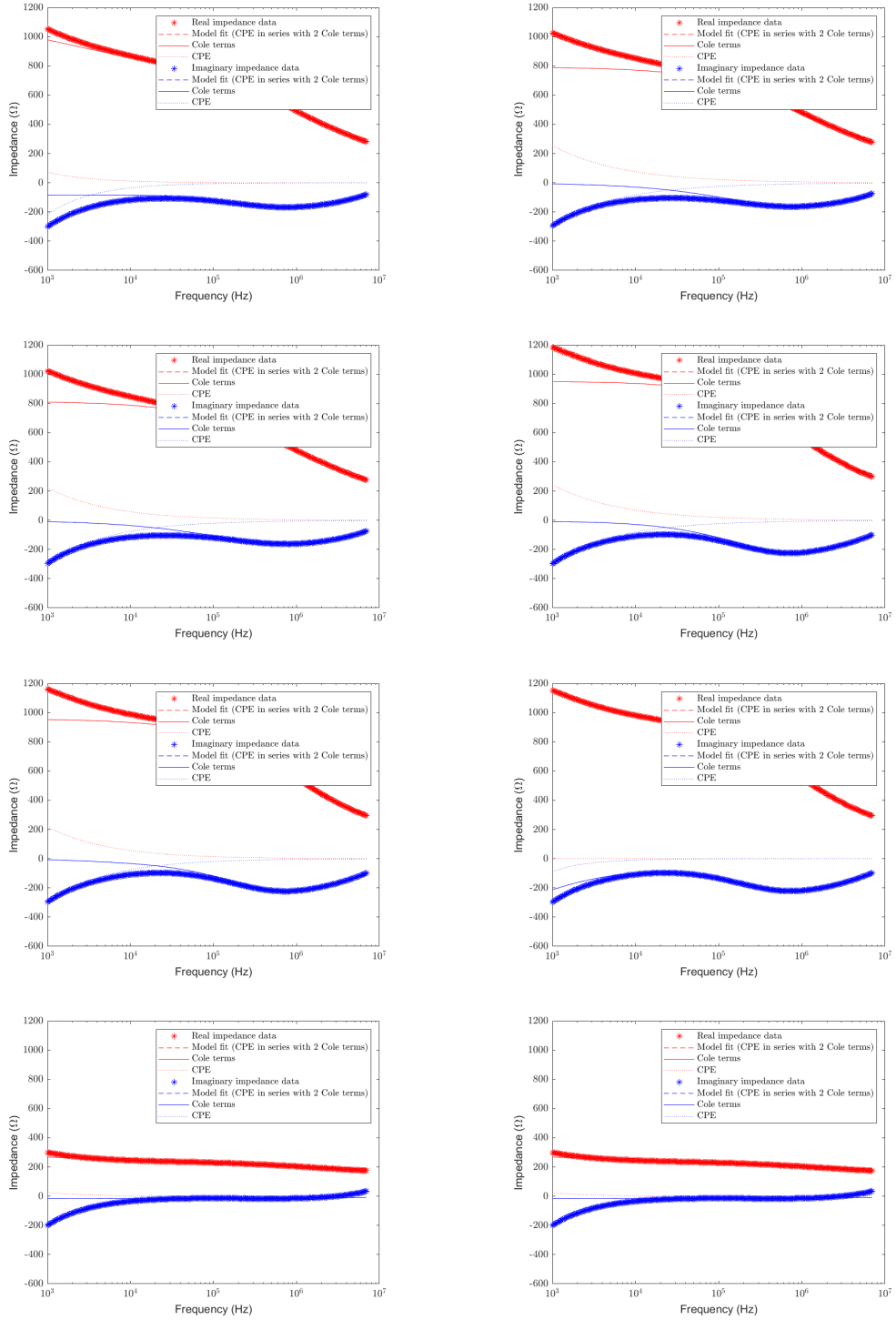


Figure 4-27: Combined CPE and 2-pole Cole impedance model fit for cancerous measurement 1B1, 1B2, 1B3, 2B1, 2B2, 2B3, 3B1, 3B2 (top left to right bottom).

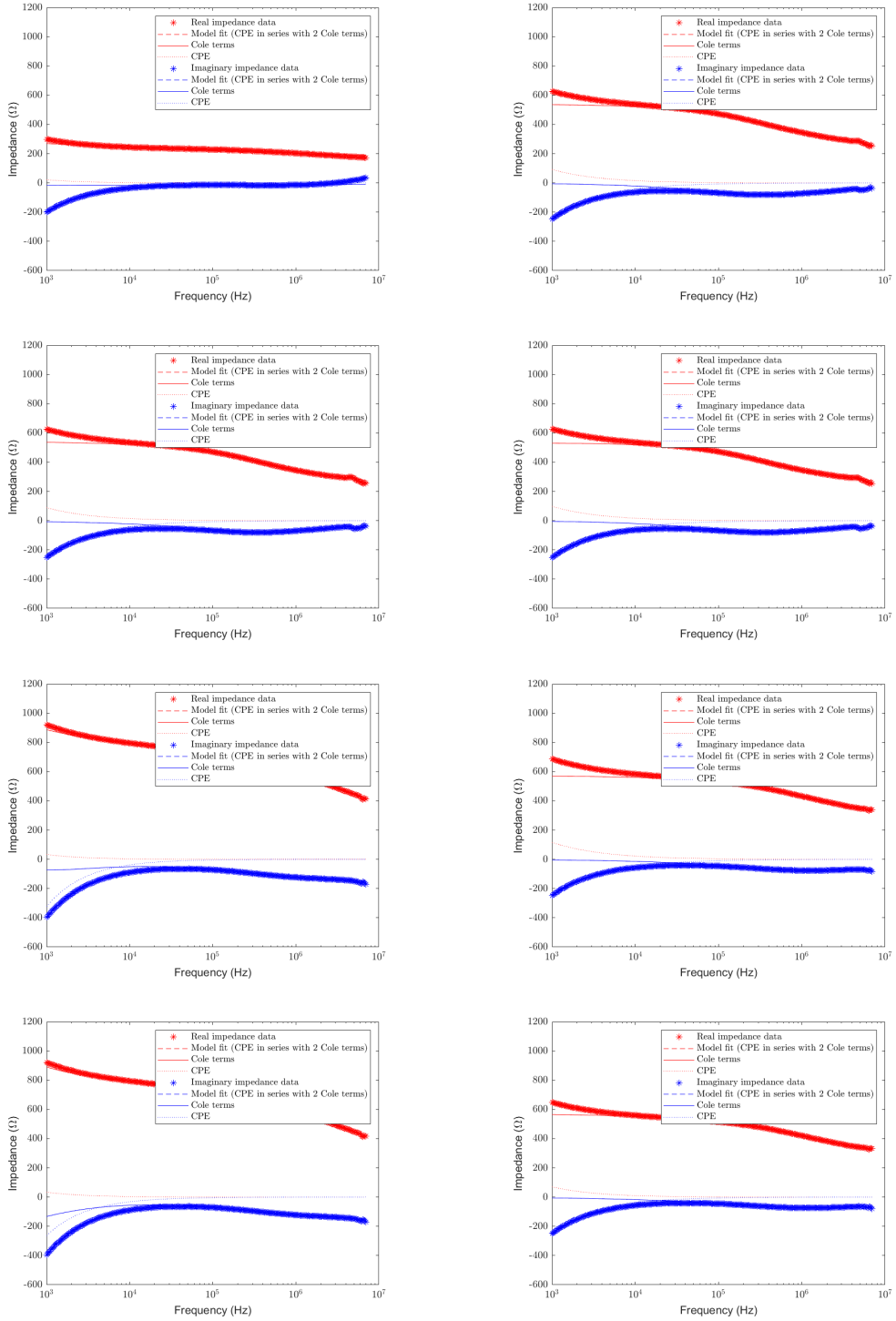


Figure 4-28: Combined CPE and 2-pole Cole impedance model fit for cancerous measurement 3B3, 4B1, 4B2, 4B3, 5C1, 5B1, 5C2, 5B2 (top left to right bottom).

Fitting the data with the combined model from Equation 3-12 yielded the following parameters for the non-cancerous data (Table 4-12) and for the cancerous data (Table 4-11). The fact that there was so much variation in the parameters could be because of the complexity and the many local minima the fitting algorithm converges to. Again by taking a glance at 2B2 and 2B3. It was observed that for 2B3, R_1 is relatively high, m is 1 (the maximum value), and for 2B2, R_1 was low, and m is approximately 0.5. By looking at equation 3-12, the behavior could be explained as it was seen in the data fit. The initial impedance in 2B3 stems from the Cole terms, and the initial impedance in 2B2 stems from the CPE. Again, these variations are likely due to the optimizer finding an optimum in local minima. Therefore, a true conclusion on the initial impedance in the α dispersion can not be made. It should be noted that similar behavior regarding the parameters was observed in [27].

ID	R_1	τ_1	α_1	R_2	τ_2	α_2	K	m	R_∞
1B1	575	0.00	0.3435	500	1.76	0.6798	4.17	0.7950	210
1B2	588	2.02	0.6272	307	1.11	0.0005	2.58	0.5287	46.1
1B3	458	4.15	0.6191	122	8.04	0.9717	1.70	0.5810	236
2B1	104	0.00	0.9975	718	2.00	0.7063	2.22	0.5470	235
2B2	11.0	2.26	0.9985	716	1.99	0.7042	1.56	0.5904	229
2B3	2911	0.01	0.5001	686	1.93	0.7179	1.83	1.0000	232
3B1	237	0.00	0.1657	282	7.90	0.0147	1.89	0.9102	14.0
3B2	220	0.00	0.1955	16.66	4.08	0.4148	1.73	0.9194	157
3B3	255	0.00	0.1739	29.82	8.58	0.9293	1.65	0.9248	121
4B1	115	0.00	0.0000	300	3.55	0.6185	4.68	0.7677	182
4B2	75	0.00	0.0826	278	3.67	0.6273	4.37	0.7752	225
4B3	274	4.05	0.6506	21.53	1.72	0.9994	4.84	0.7601	238
5C1	184	0.00	0.7611	807	2.33	0.4435	8.69	0.9341	6.25
5B1	84.1	3.01	0.0000	326	1.03	0.5619	7.59	0.7101	206
5C2	862	0.00	0.6756	800	2.18	0.4479	1.20	0.9192	4.35
5B2	216	2.25	0.2176	356	9.17	0.5044	2.73	0.8268	0.03

Table 4-11: Nine parameters from the two-pole Cole impedance model in combination with a CPE for the cancerous measurements.

After obtaining the parameters, the statistical analysis method described in Section 3-3-1, and Figure 3-8 was used. Again, via the one-sample Kolmogorov-Smirnov test, it was observed that parameters do not come from a normal distribution. It was found via the Two-sided Wilcoxon signed rank test that the median is not zero for all parameters. Furthermore, α_2 has a P-value lower than 0.05 in the Two-sided Wilcoxon rank sum test. This means a data-fitting equation has been found, Equation 3-12, that converges without large errors and gives a statistically significant parameter. In Table 4-13, all nine parameters with their P-value are displayed, and in Figure 4-29 a boxplot of the values of α_2 can be observed for the non-cancerous and cancerous data.

ID	R_1	τ_1	α_1	R_2	τ_2	α_2	K	m	R_∞
1A1	1.80	0.00	0.2809	398	1.61	0.6276	8.68	0.6617	234
1A2	84.1	1.41	0.7406	285	1.17	0.7602	8.39	0.6654	260
1A3	161	1.81	0.8304	415	4.17	0.2969	6.98	0.6955	106
2A1	146	0.00	0.3111	397	8.51	0.5519	9.00	0.6708	228
2A2	138	9.01	0.1026	508	6.22	0.4605	8.54	0.6689	25.3
2A3	385	8.57	0.5514	333	3.74	0.0537	1.02	0.6472	66.9
3A1	1023	0.00	0.2593	167	6.50	0.0007	5.89	0.7085	6.20
3A2	346	2.04	0.6166	9.94	5.71	0.0039	4.46	0.6997	194
3A3	1660	0.00	0.2576	5.10	3.57	0.1861	5.56	0.9999	93.7
4A1	295	5.07	0.0632	461	1.49	0.3021	1.89	0.8448	0.02
4A2	56.3	0.00	2.4424	526	1.24	0.3065	2.09	0.8226	45.5
4A3	210	0.00	0.0130	466	1.79	0.3332	2.17	0.8160	22.0
5A1	3.49	0.00	0.3503	269	1.63	0.5560	7.86	0.7028	241
5D1	1215	0.01	0.5437	152	2.11	0.6692	3.14	0.8411	260
5A2	123	1.35	0.7031	523	4.32	0.1430	5.52	0.7558	31.4
5D2	635	0.00	0.5584	159	2.09	0.6570	2.96	0.8440	259
6C1	209	1.69	0.7578	441	8.69	0.6551	3.53	0.7564	151
6A1	654	9.40	0.1838	566	8.94	0.5646	1.86	0.8175	2.78
6B1	112	1.65	1.1102	615	6.62	0.5038	2.57	0.8046	32.4
6D1	276	0.00	0.0330	497	9.80	0.5712	4.00	0.7669	34.9
6C2	167	1.86	0.0006	643	1.13	0.6827	3.35	0.7631	70.1
6A2	293	0.00	0.7320	905	9.60	0.4551	1.56	0.8604	27.0
6B2	41.9	2.079	0.0586	625	6.24	0.4977	2.71	0.7997	36.1
6D2	46.9	0.00	0.2079	492	9.85	0.5735	3.81	0.7726	147
7C1	262	0.00	0.2092	224	1.68	0.0000	1.58	0.8753	101
7A1	330	1.12	0.5676	534	2.33	0.0000	2.12	0.7676	3.83
7B1	348	1.38	0.6134	327	2.53	0.0000	1.83	0.7862	63.0
7D1	271	8.78	0.6257	266	1.27	0.0000	2.58	0.7633	80.5
7C2	202	0.00	0.2690	7.52	7.65	0.0002	1.46	0.8851	216
7A2	352	1.32	0.5310	71.1	2.78	0.0000	1.63	0.7940	227
7B2	1381	0.00	0.2207	39.6	1.66	0.0009	3.82	1.0000	70.9
7D2	60.6	1.21	0.5239	262	9.44	0.6322	2.48	0.7671	162

Table 4-12: Nine parameters from the two-pole Cole impedance model in combination with a CPE for the non-cancerous measurements.

	R_1	τ_1	α_1	R_2	τ_2	α_2	K	m	R_∞
P-value	0.9042	0.1786	0.5770	0.7678	0.5770	**0.0101**	0.4505	0.6382	0.3414

Table 4-13: The P-values of the two-pole Cole impedance in combination with a CPE, with α_2 being statistically significant parameter (P$<$ 0.05).

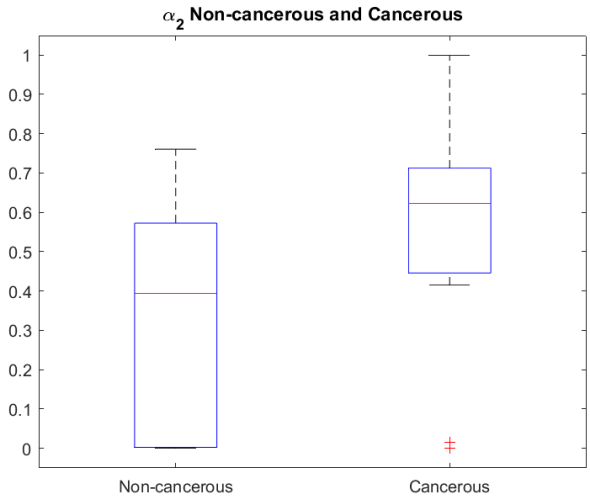


Figure 4-29: A boxplot of the α_2 parameter ($P < 0.05$) for the non-cancerous and cancerous data.

	Threshold Value
Patient 1	0.33320
Patient 2	0.33320
Patient 3	0.66930
Patient 4	0.66930
Patient 5	0.57360
Patient 6	0.33320
Patient 7	0.66930

Table 4-14: Threshold values.

Now that a statistically significant parameter was found, α_2, the ROC analysis with a Leave-One-Out Cross Validation (LOOCV) approach can be performed to find a suitable threshold. A threshold finding algorithm was run that iterated over a range of 0 to 1, in 10001 steps. It ran over seven data sets, as one patient is left out of every data set, for example, our first data set does not contain the data of patient one (as this is the one that is left out). The algorithm was ran over the remaining six patients, and the optimal threshold value was found by determining which value creates the largest summation in Specificity and Sensitivity. This index is depicted in the ROC curve in Figure 4-30.

After running our algorithm, the following threshold values were found, were to be tested on the left-out patient (Table 4-14). It can be observed that the threshold values are almost split into two groups: 0.33320 and 0.66930; the reason for this is the low amount of data. When a patient is left out, the algorithm's encounter of the first index, where the summation of the specificity and sensitivity is the largest, is often at the same index.

After obtaining the threshold values, testing them as intended with the LOOCV can be done. The data from patient one was tested with the threshold intended for patient one, and the data from patient two was tested with the threshold intended for patient two, and so on. In total, 7 True Positive (TP), 18 True Negative (TN), 14 False Positive (FP), and 9 False

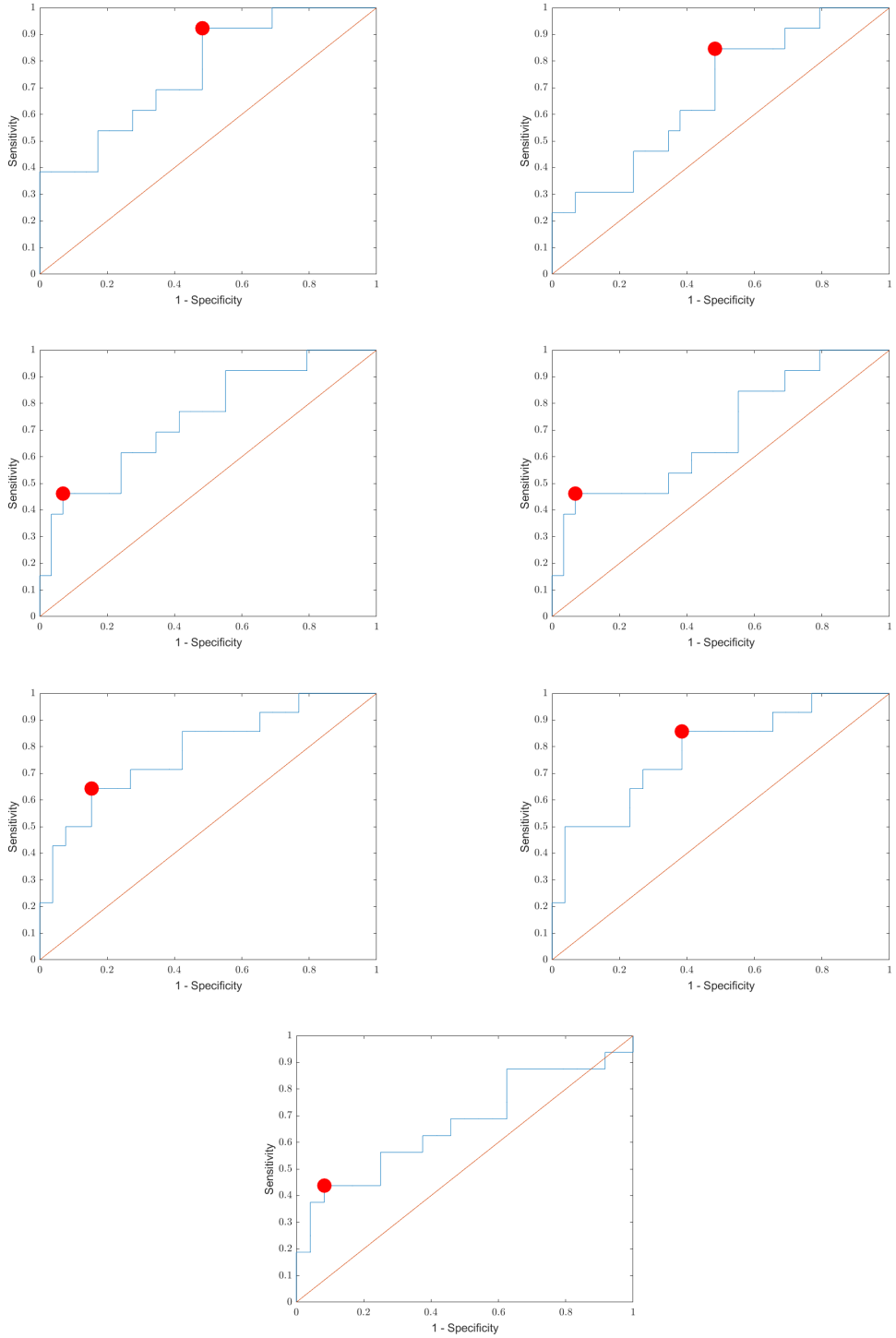


Figure 4-30: The ROC curve for our threshold determination algorithm, with the threshold index indicated as a red dot. Top left to bottom: TH1, TH2, TH3, TH4, TH5, TH6, and TH7

Accuracy	Sensitivity	Specificity	PPV	NPV	MCC
0.5208	0.4375	0.5625	0.3333	0.6666	0.0000

Table 4-15: The performance metric results for the LOOCV of the 2-pole Cole impedance parameterization model in combination with thresholding.

Negative (FN) were obtained. This gave the following performance analysis metrics in Table 4-15.

The results were unsatisfactory, and the current thresholding method with the current parameters could not be used to separately classify cancerous and non-cancerous data in the GI tract. The main issue is the low amount of data available, leading to low performance metrics scores. The value of the Mathews Correlation Coefficient (MCC) equals zero because, in the numerator, the TP and TN are summed and subtracted by the summation of the FP and FN since both equal 25 the MCC equals zero. The values for the sensitivity and Negative Predictive Value (NPV) were relatively high. However, it should be noted that this is due to the large amount of non-cancerous data measurements in the data set compared to the amount of cancerous data measurements.

4-4 Principal Component Analysis in Combination with Support Vector Machine

This section covers all the results that were obtained from the classification strategy where PCA in combination with SVM was used. The PCA is created via the Singular Value Decomposition (SVD), and the SVM makes use of a Radial Basis Function (RBF) kernel trick, which is preferred to be used when data in non-linear [11].

With the PCA one would like to maximize dimension reduction of the available data set while keeping the maximum amount of variance. It is possible to asses how much variance is captured by each principle component via Equation 3-36. To determine how many principal components were required for the parameterization algorithm, an initial PCA was ran on the entire data set. The results are displayed in Figure 4-31. It was found that for the first five principal components contained 0.6818, 0.2503, 0.0316, 0.0219, and 0.0080 accounted variance. This entails, that the first three principal components capture 0.9637 (96.37%) of the total variance in the data set. Therefore, the decision was made to run the principal component analysis during the LOOCV with three principal components

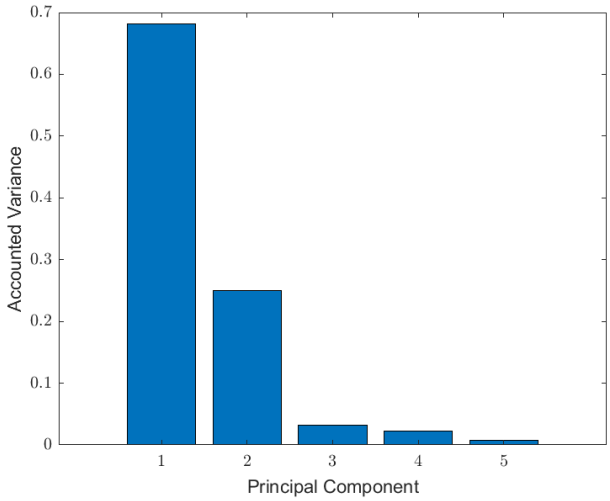


Figure 4-31: The accounted variance of the first five principal components.

We ran the PCA algorithm for seven data sets according to the LOOCV strategy. For example, in the first data set, patient one was omitted and the PCA was performed on the remaining six patients to reduce the dimensionality of the data matrix. When patient one is omitted, the data matrix's initial size is 42×600; after the PCA, this is reduced to a 42×3 matrix. This is a great benefit as smaller amounts of data are faster and easier to work with. After performing the PCA on the remaining six patients, the SVM was trained by giving the data to the fitcsvm function in Matlab. This process was repeated seven times for each patient. After running the trained SVM on the test subjects following the LOOCV strategy, a total of 0 TP, 20 TN, 12 FP, and 16 FN were found. The performance metrics are given in Table 4-16.

The results of the PCA combined with the SVM were also unsatisfactory. Creating a proper classification strategy in the current setting is impossible, mainly due to the low data avail-

Accuracy	Sensitivity	Specificity	PPV	NPV	MCC
0.4167	0.0000	0.6250	0.0000	0.5556	-0.4082

Table 4-16: The performance metric results for the LOOCV of the PCA in combination with the SVM.

able. The combination of the PCA with the SVM was not able to detect any cancerous tissue at all, meaning that targeting a biopsy in the GI tract is futile. It has a negative MCC score, indicating low performance. Compared to the two-pole Cole impedance model combined with a CPE model, it also performs less, except for the specificity score, which is higher for the SVM. The MCC score for the thresholding algorithm is 0.0000, while for the PCA in combination with an SVM, it is negative. Since the MCC is a general measure for classification, it is safe to say that the thresholding algorithm performed better.

Chapter 5

Discussion

This work explores different classification strategies for cancerous and non-cancerous tissue detection in the gastrointestinal (GI) tract to help surgeons who want to make taking a biopsy easier or even make is obsolete. This section discusses all significant findings and, consecutively, the meaning behind those findings.

5-1 Difference in Esophagus, Ileum, and Colon Tissue

In Chapter 4, it was observed from the figures in Figure 4-1 until Figure 4-10 that the ileum and colon tissue are similar, they have a similar trajectory and are roughly around the same impedance magnitude. However, the esophagus tissue deviates from these two tissues. In both colon and ileum measurements, it was observed that the cancerous tissue had a high real and imaginary impedance. At the same time, this was the other way around for the esophagus tissue. This leads to the belief that the hypothesis made in the introduction is false. It was stated that the esophagus, ileum, and colon tissue could be similar due to the same four tissue layers (mucosa, submucosa, muscularis externa, and serosa) in the GI tract, and therefore, it would be possible to make classification of cancerous and non-cancerous tissue easier.

During the second trial, it was discovered by the pathologist that the measurements on the fifth, sixth, and seventh patients contained faults. When measuring the tissue of patient five, and having the intention of measuring healthy tissue, a tumor was measured. A similar mistake occurred in patients six and seven, where it was intended to measure cancerous tissue. However, in both cases, normal tissue was measured. This could be the fault of the probe, as there was no complete control of the probe since it could not be seen in the tissue. The measurement protocol describes that the probe is fixated deep into the tissue, allowing minimal variance in the measurements. However, what could have occurred is that the tumor was highly superficial, and the measurement was conducted too deeply into the tissue. This is a clear disadvantage of a spiky electrode probe.

5-2 Findings Cole Impedance Model

In Section 4-3, the Cole impedance model from Equation 3-8, the two-pole Cole model from Equation 3-10, and the combined Constant Phase Element (CPE) model with the two-pole Cole model from Equation 3-12 were analyzed. It was found in Section 4-3-1 that the Cole impedance model is unsuitable for further data analysis because of the high error score in the data fitting process. An average Mean Square Error (MSE) value of 9.8208e+05 for the non-cancerous data fit was found, and for the cancerous tissue, an average MSE of 9.7343e+05 was found. These significant errors were also visually reflected in the figures created to showcase the model fit. In many cases, the model fit could not correctly follow the nuances of the impedance data, especially in the lower frequency regions. The reasoning behind this lack of performance is that the Cole impedance model is a single dispersion model [10]. This work's frequency range covers the α and β dispersion, as the measurements are taken from 1 kHZ to 7 MHz [3]. The parameters that were obtained via the Cole impedance model did yield a statistically significant ($P < 0.05$) parameter, R_∞. Nevertheless, due to the large average MSE of the data fit, it was concluded that this model is unsuitable for further analysis.

5-3 Findings Two-Pole Cole Impedance Model

By extending the existing Cole impedance model with another Cole impedance model, the two-pole Cole impedance model is created [15]. This model was analyzed throughout Section 4-3-2. The two-pole Cole impedance model created an accurate data fit with an average MSE of 4.1228e+04 and 2.6642e+04 for the non-cancerous and cancerous data, respectively. These numbers indicate that compared to the Cole impedance model, the two-pole Cole model fit is more accurate by a factor of 10. This is unsurprising as the second Cole element allows the model to cover two dispersion areas [15]. A drawback of the fitting methods is that the model does not account for changes in capacitive to inductive impedance (negative to positive). Although a proper fit with the two-pole Cole model was achieved. Via the two-sided Wilcoxon rank sum test not a single statistically significant ($P < 0.05$) parameter that could have been used as a classification parameter was obtained.

5-4 Findings Two-Pole Cole Impedance Model with CPE

The two-pole Cole impedance model in combination with a CPE was showcased in Section 4-3-3 and was used previously in [27]. The model showed a more accurate data fit than the 2-pole Cole impedance model, with an average MSE 1.68402e+04 and 1.1296e+04 for the non-cancerous and cancerous data, respectively. This indicates that this model fit was approximately twice as good as the model fit when the 2-pole Cole impedance model was used. Although a satisfying data fit was obtained, it was observed that in some measurements, the data is fitted more towards the Cole parameters and in others more towards the CPE model. These phenomena mainly occurred in the α dispersion region around 1 kHz to 10 kHz. This makes sense since the CPE is normally introduced to mitigate the adverse effects of electrode polarization in the α dispersion region [16, 36]. In some measurements, a significant value of R_1 and m were observed in Table 4-12, and 4-11. This indicates that the Cole parameters

are more dominant in the data fit, and therefore the contribution of the CPE is small. This can be explained by the complexity of the data and objective function, probably leading to the solver landing in a local minimum that gives good fitting results but does not consider real-world phenomena (like electrode polarization). With the combination of the two-pole Cole impedance model and a CPE, a statistically significant parameter ($P < 0.05$) was found via the two-sided Wilcoxon rank sum test, meaning that it was eligible for tissue classification via thresholding. The parameter α_2 had a P-value equal to 0.0101, and it was shown in Figure 4-29 that the values for α_2 for non-cancerous and cancerous data are different.

5-5 The Meaning and Use of α_2

In this work, α_2 is used to create a threshold algorithm. The interpretation of this parameter is somewhat difficult as current scientists have not yet reached a consensus on the meaning of the α parameter in the Cole impedance model. In [4, 19], it is stated that it depends on the heterogeneity of the cell size shape of living tissue. While in [15], it is stated that the value of α influences the minimal spectral width and that it can be regarded as the deviation from the Fricke-Morse model, which is obtained when $\alpha = 1$.

To use the thresholding parameter correctly, the data was separated according to a Leave-One-Out Cross Validation (LOOCV) strategy described in Section 3-4. In LOOCV, one leaves out one patient's data set and trains the algorithm on the remaining patient's data sets. This is iterated for the number of patients in the entire study. Receiver Operator Characteristic (ROC) curves were created for every data set according to the LOOCV strategy and iterated over threshold value between 0 and 1 in 10001 steps. The threshold was chosen for the value where the specificity and sensitivity summation were the largest. It was found that the threshold values (α_2) were almost split into two groups: 0.33320 and 0.66930. It is reasoned that this is due to the low amount of specimens. When a patient is left out due to the LOOCV strategy, the threshold finding algorithm crosses the first index where the sum of the specificity and sensitivity are the largest. It is often the case that these are the same.

5-6 The Performance of the Thresholding algorithm

The performance results of the thresholding algorithm were unsatisfactory. A total of 7 True Positive (TP), 18 True Negative (TN), 14 False Positive (FP), and 9 False Negative (FN) were obtained. Additionally, an accuracy of 0.5208, a sensitivity of 0.4375, a specificity of 0.5625, a Positive Predictive Value (PPV) of 0.3333, a Negative Predictive Value (NPV) of 0.6666, and Mathews Correlation Coefficient (MCC) of 0.0000 were obtained. These results are unsuitable for a classification algorithm that distinguishes cancerous from non-cancerous tissue in the GI tract. The low sensitivity and the low PPV stood out the most, as the sensitivity shows the ability to detect cancerous tissue correctly, and the PPV shows the proportion of true positive prediction. The literature has pointed out that Electrical Impedance Spectroscopy (EIS) was used to detect cancerous and non-cancerous tissue [1, 43, 25, 46, 18]. However, those performance results were more satisfactory. For example, in [2], the authors obtained a sensitivity of 74% and specificity of 53% by using a Cole equation and thresholding. However, it should be noted that over 800 measurements were conducted in this particular case. In

[43], the authors obtained a PPV of 91.5%, a specificity of 95.4%, and a sensitivity of 62.1%. However, it should be noted that, in total, 214 specimens entered this study. In [25], 36 specimens (prostate) were used. Thresholding was used to obtain a 75.4% specificity and a 76.1% sensitivity. [46] used EIS to differentiate cancerous from benign breast tissue. They employed the Cole-Cole model on 622 breast masses, and 66 were determined to be malignant and 556 benign. [46] achieved a sensitivity of 92.4%, and a specificity of 96.0%. In [18], 53 patients were classified with the help of an equivalent circuits (EC). Forty-four had lung cancer, 5 had metastatic lung tumors, and 4 had pneumonia. This resulted in a sensitivity of 100% and a specificity of 66%. In [27], three porcine hearts were measured with over 300 EIS measurements. A thresholding algorithm was created with the two-pole Cole impedance model with CPE. This led to an accuracy of 81.4%, a sensitivity of 84.4%, a specificity of 78.6%, an PPV of 77.8%, a NPV of 85.0%, and a MCC of 0.629. These results are excellent. However, it should be noted that the amount of measurements is relatively high compared to this work, and no LOOCV strategy was employed, meaning that the training threshold determination and testing were conducted on the same data (a priori data knowledge). The performance of our thresholding algorithm is not comparable with the current literature. However, it must be noted that there are almost no literature sources on EIS and cancer detection in the GI tract. Additionally, it should be noted that all literature sources reviewed in this work have access to significantly more data, and some create a classification algorithm with a priori data knowledge.

5-7 The Performance of the Support Vector Machine

In Section 4-4, the principal component analysis (PCA) in combination with a support vector machine (SVM) is displayed. The PCA is created to minimize the size of the data set while maximizing the captured variance. The PCA is executed on the entire data set to determine how many principal components are required for this work. This led to the discovery that the first three principal components accounted for 96.37% of the total variance of the data set.

To make the SVM compatible with the highly non-linear impedance data, the Radial Basis Function (RBF) kernel was used, which allows the SVM to capture complex patterns in the data by creating decision boundaries that are curved or irregular. After following the LOOCV strategy for the PCA with SVM, the following performance results were found: 0 TP, 20 TN, 12 FP, 16 FN. Additionally, the accuracy was 0.4167, the sensitivity 0.0000, the specificity, 0.6250. the PPV 0.0000, the NPV 0.5556, and the MCC -0.4082. The performance of the SVM is unsuitable for classifying cancerous and non-cancerous tissue during a biopsy in the GI tract. The zero-valued sensitivity and PPV resulted in an unusable algorithm. For comparison, SVM is currently used in EIS cancer research. For example, in [5], an SVM was combined with PCA, and the classifier accuracy was over 97%. In [17], 14 human prostates were researched with 23 cancerous and 53 benign EIS measurements. Using a LOOCV strategy, the predictive value of their parameterized data for differentiating between benign and malignant tissue was assessed. The predictive accuracy was 90.8%, with a specificity of 94.3% and a sensitivity of 82.6%. By comparing This work's PCA and SVM algorithms with current literature, it can be observed that the algorithms in this work are not performing as well as the ones found in current literature. However, it should be noted that the cause is likely to be the lack of data. If the number of data points measured were to be increased, it would probably be possible

to create a more successful classification strategy to showcase a proof-of-concept for EIS as a supportive tool for biopsies.

Chapter 6

Conclusion

This work presents the knowledge obtained during the master thesis project conducted at the TU Delft in combination with the Erasmus MC and Phillips. Two measurement trials were conducted, where the first trial was intended to explore different tissues, and the second trial was to converge to a more specific area. In this work, measurements were made on the esophagus, ileum, and colon tissue that represented the gastrointestinal (GI) tract. This work explored different parameterization methods, like the Cole impedance model, the two-pole Cole impedance model, and a combination of a Constant Phase Element (CPE) model with the two-pole Cole impedance model. This work also explored principal component analysis (PCA) as a parameterization technique. After the parameterization, the goal was to classify measurement data and predict whether the data was cancerous or non-cancerous. This was done either by setting a threshold determined via the Receiver Operator Characteristic (ROC) curve, where a threshold based on the maximal summation of specificity and sensitivity was chosen, or by using a support vector machine (SVM) with an Radial Basis Function (RBF) kernel. Via a Leave-One-Out Cross Validation (LOOCV) analysis, the findings of these methods were concluded. In this conclusion, the research questions created in the introduction are restated and answered accordingly.

6-1 Answer of the First Research Question

Is it possible to create classification algorithms that can be used on the esophagus, colon, and ileum tissue to detect cancerous versus non-cancerous tissue?

In Section 2-1, it was discussed that along the entire GI tract, the four primary layers (mucosa, submucosa, muscularis externa, and serosa) are present [24]. This led us to hypothesize that the esophagus, ileum, and colon tissue would respond similarly to Electrical Impedance Spectroscopy (EIS). However, this is not true. A clear difference was shown in Chapter 4, that esophagus, ileum, and colon tissue do not respond similarly to EIS. The trajectory of esophagus tissue differs from that of ileum and colon tissue. However, it should be noted

that the trajectory and magnitude of the impedance for the ileum and colon are similar. This makes sense as these two parts of the GI tract are connected.

In Section 4-3-3, the creation of a working algorithm that uses an equivalent circuits (EC) in the form of the two-pole Cole impedance model in combination with a CPE was showcased. The algorithm can classify cancerous and non-cancerous data. However, the performance of the classification algorithm is doubtful in this current setting, as the performance metrics displayed in Section 4-3-3 indicated that the algorithm performs rather poorly. In Section 4-4, a description was made of how to use PCA and SVM together to create a working classification algorithm for cancer detection. Similar to the algorithm that is based on the two-pole Cole impedance model and threshold, this algorithm's performance was poor.

The answer to this first research question is yes; it is possible to create classification algorithms that can be used on the esophagus, colon, and ileum tissue to detect cancerous versus non-cancerous tissue. However, in this work, the performance was relatively poor due to the current setup.

6-2 Answer of the Second Research Question

Which classification strategy has better performance, the more classical Cole Impedance parameterization in combination with thresholding, or the newer PCA in combination with a trained SVM algorithm.

In Section 4-3-3, the statistically significant parameter α_2 was found, which was used to create a thresholding algorithm. This algorithm was used via the LOOCV strategy to determine its performance. It was found that the algorithm performed unsatisfactorily since it obtained 7 True Positive (TP), 18 True Negative (TN), 14 False Positive (FP), and 9 False Negative (FN), which led to the following performance metrics.

Accuracy	Sensitivity	Specificity	PPV	NPV	MCC
0.5208	**0.4375**	0.5625	**0.3333**	**0.6666**	**0.0000**

Table 6-1: The performance metric of the thresholding strategy

For the PCA in combination with a SVM, we also employed the LOOCV strategy to evaluate the performance of this algorithm. We found that the combination of PCA and SVM also performed unsatisfactorily. For the combination between the PCA and SVM, we found a total of 0 TP, 20 TN, 12 FP, and 16 FN. This led to the following performance metrics

Accuracy	Sensitivity	Specificity	PPV	NPV	MCC
0.4167	0.0000	**0.6250**	0.0000	0.5556	-0.4082

Table 6-2: The performance metric results of the SVM strategy.

The accuracy, sensitivity Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Mathews Correlation Coefficient (MCC) of the thresholding algorithm were higher. This means that all but one performance metric is higher for the thresholding algorithm. This, with the additional fact that the MCC (which is arguably the most important metric as it

showcases the overall performance) is higher for the thresholding algorithm in comparison with the SVM algorithm it can be stated that the 2-pole Cole impedance modeling in combination with thresholding has better performance than the PCA in combination with SVM.

6-3 Answer of the Third Research Question

Is EIS a technique that can be used in combination with a parameterization and classification algorithm to replace or support biopsies in the GI tract?

During a biopsy, a small piece of tissue was taken from the lining of the GI tract for further examination under a microscope. It would be helpful if, with the aid of EIS, the biopsy could support in finding cancerous tissue, or even be replaced by a probe that can distinguish cancerous and non-cancerous tissue. In this work, EIS did not result as a technique that can be combined with a parameterization and classification algorithm to neither replace nor support surgeons in taking biopsies in the GI tract. The algorithms that are created in this work perform too poorly to consider them as reliable tools that can either support or replace biopsies.

6-4 Future Work

For future work, creating a more extensive data set is advised. This would undoubtedly help to increase the performance metrics as the algorithms will have more data available from which they can learn.

Sampling the ileum and colon tissue together could be possible as this work clearly showed the similarities between these two tissues. This work also showed the clear difference between intestine tissue (ileum and colon) and esophagus tissue. Therefore, it is suggested to individually perform the measurements on the ileum, colon, or esophagus tissue or to combine the ileum and colon tissue measurements.

Different probes should also be explored, as the current probe does not allow the user to know the measurements' depth precisely. This could lead to measuring the wrong tissue, e.g., measuring a layer of fat tissue while the researcher wants to measure muscle tissue or measuring deep non-cancerous tissue while the researcher intends to measure cancerous tissue. Therefore, the first suggestion for a different probe would be to use a surface probe. These probes do not contain spiky electrodes that stick into the tissue. Additionally, when EIS would be used in vivo, it would also be better to use a surface electrode as a spiky electrode could harm the patient.

Another future work reference could be to include the permittivity and conductivity data instead of only focusing on impedance data. Different EC models work with that data, which might be interesting. However, one must find a EC that again covers the α and β dispersion region if the measurements are made in the same range chosen in this work. It might be possible to place the Cole model in a series (similar to what is done in this work with the Cole impedance model). This work shows in Appendix A the permittivity and conductivity

data from the first ten tissue samples, and those figures indicate there is some differentiation possible. If, in any future work, the authors would like to use PCA with SVM, it is advised to concatenate the permittivity and conductivity data with the impedance data, as this creates a more extensive data matrix with more data. If used correctly, the PCA will be able to reduce the dimensionality while keeping the maximum amount of variance.

Appendix A

Trial One Measurement

This appendix will cover all measurements conducted during the measurement trial at the Erasmus MC in Rotterdam. This period started on 25-08-2022 and ended on 14-10-2022. During this period, different tissues of different patients were measured. All tissues were measured within 24 hours of resection. It can be observed that the ileum, colon, and esophagus tissue are most similar to one another. This can be because these three tissues are all in the gastrointestinal (GI) tract and therefore have similar cell characteristics [24]. Due to these observations, this work focused on the ileum and colon tissue.

During the trial period, impedance, conductivity, and permittivity measurements were made over 1 kHz to 7 MHz. Three measurements were made on the healthy part of the tissue, and three measurements were made on the cancerous part of the tissue. However, for the retroperitoneal soft tissue, this was not possible because the entire tissue was cancerous. After measurements, the pathologist cut out the tissue on which the measurements were conducted for further research by histological analysis (the golden standard). The measurement protocol for the first trial can be found in Figure A-1.

In Table A-1, one can find an overview of the measurements conducted during the trial period. Additionally, Table A-2 shows where the measurements were conducted and the results of the histological analysis performed in the Erasmus MC.

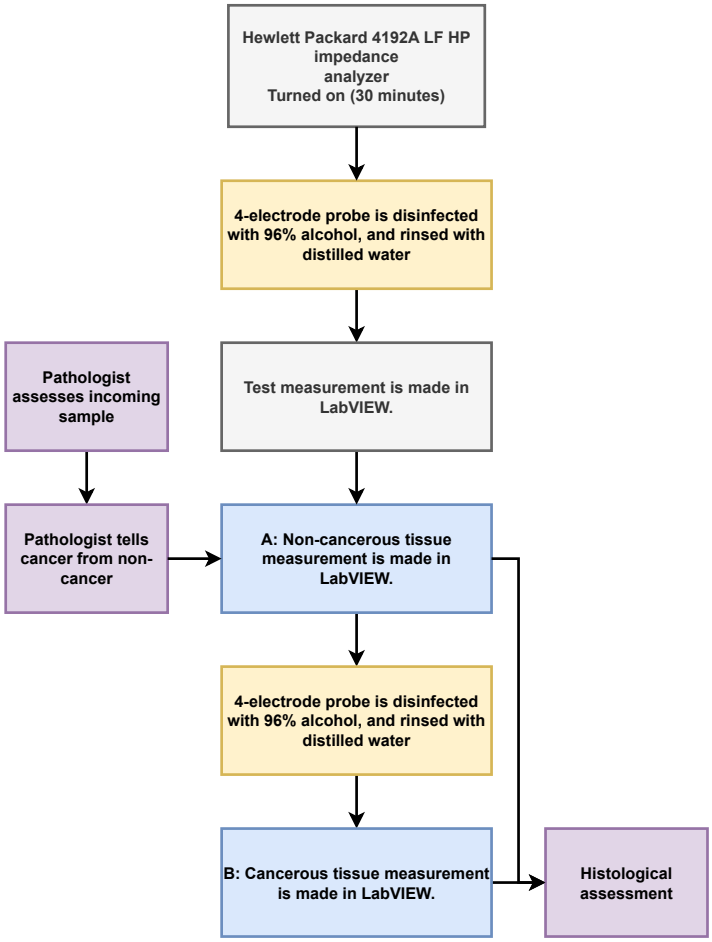


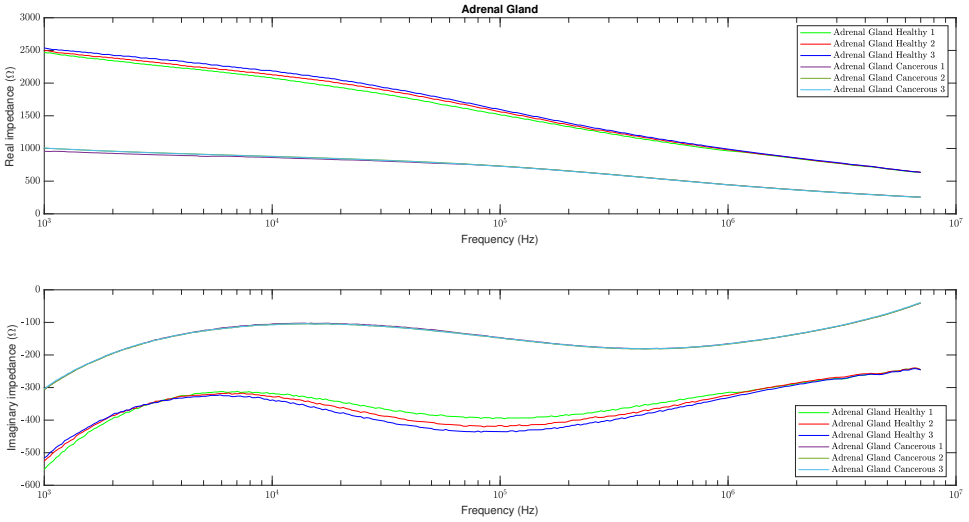
Figure A-1: The measurement protocol for the first trial that was conducted at the Erasmus MC.

#	Date of Measurement	Gender (M/F)	Year of Birth	Tissue
1	25-08-2022	F	1971	Adrenal Gland
2	22-09-2022	F	1963	Retroperitoneal Soft Tissue
3	27-09-2022	F	1971	Breast
4	29-09-2022	M	1952	Liver & Colon
5	29-09-2022	F	1971	Ileum
6	30-09-2022	M	1942	Glute
7	30-09-2022	M	1951	Esophagus
8	11-10-2022	M	1951	Pancreas
9	11-10-2022	F	1947	Lung
10	14-10-2022	M	1947	Esophagus

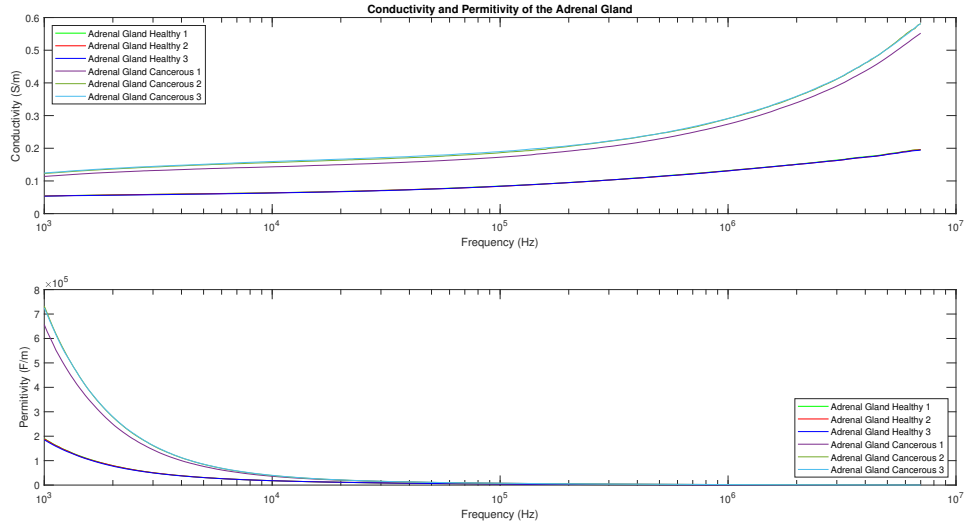
Table A-1: The year of birth, gender, and date of the measurement per patient and the corresponding tissue the measurements were conducted on during the trial measurement period in the Erasmus MC.

#	Tissue	Measurement Location	Histological Analysis
1	Adrenal Gland	A, B: Tumor C: Normal	A, B: Small cell lung carcinoma C: Normal and small cell lung carcinoma (macroscopically not visible)
2	Retroperitoneal Soft Tissue	A: Tumor	A: Differentiated Liposarcoma
3	Breast	A: Normal B: Tumor	A: Normal mammaparenchym B: Invasive carcinoma
4	I: Liver II: Liver III-B: Ileum, III-C: Colon	I-A: Normal liver I-B: Tumor liver II-A: Normal liver II-B: Tumor liver III-A: Tumor ileum III-B: Normal ileum III-C: Normal colon	I-A: Normal liver with small part adenocarcinoma I-B: Adenocarcinoom II-A: Normaal lever II-B: Adenocarcinoom III-A: Adenocarcinoom III-B: Normal ileum III-C: Normal colon
5	Ileum	A: Tumor B: Normal ileum	A: Neuroendocrine tumor grade 2 B: Normal ileum
6	Glute	A: Tumor B: Normal fat C: Normal muscle	A: Renal cell carcinoma B: Normal Fat C: Normal Muscle
7	Esophagus	A: Tumor B: Normal Esophagus	A: Adenocarcinoma and signet ring cell carcinoma B: Normal Esophagus
8	Pancreas	D: Normal pancreas E: Tumor	D: Normal pancreas and adenocarcinoma E: Normal pancreas
9	Lung	B: Tumor C: Normal lung	B: Adenocarcinoma C: Normaal lung
10	Esophagus	A: Normal Esophagus B, C: Tumor Esophagus	A: Normal Esophagus B: Dysplasia of the squamous epithelium C: Dysplasia into squamous cell carcinoma

Table A-2: The measurement tissue and the measured placed with the results from the histological results per patient.

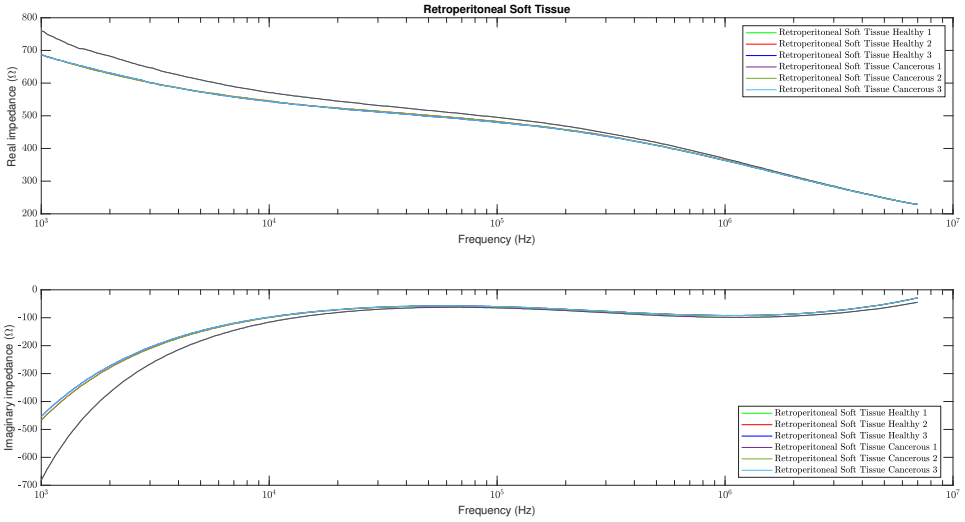


(a) The real part and the imaginary part of the resistance for the measured adrenal gland.

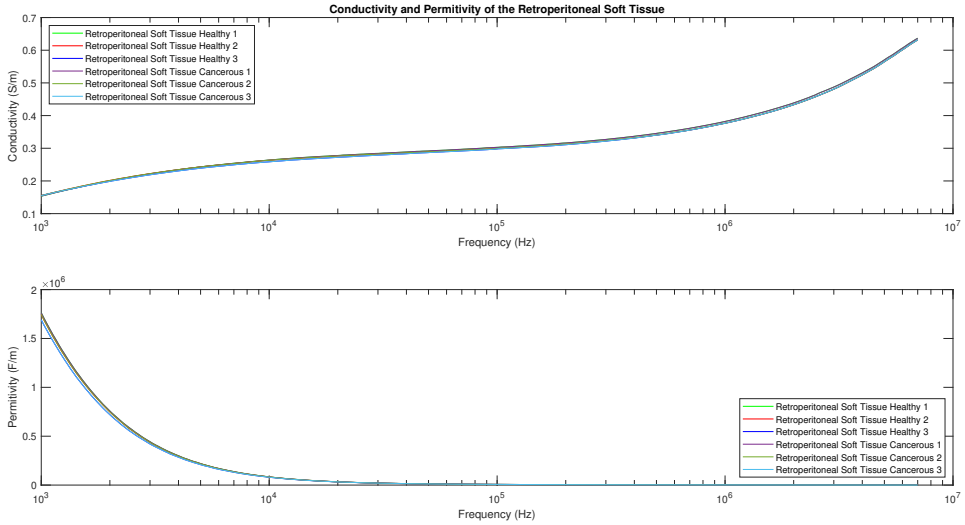


(b) The permittivity and conductivity for the measured adrenal gland.

Figure A-2: The measured impedance of the adrenal gland, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

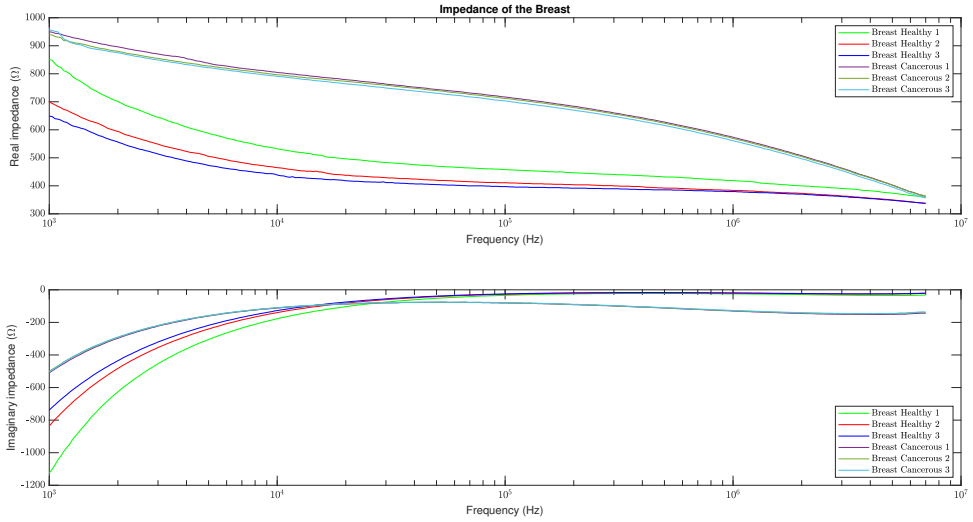


(a) The real part and the imaginary part of the resistance for the measured adrenal gland.

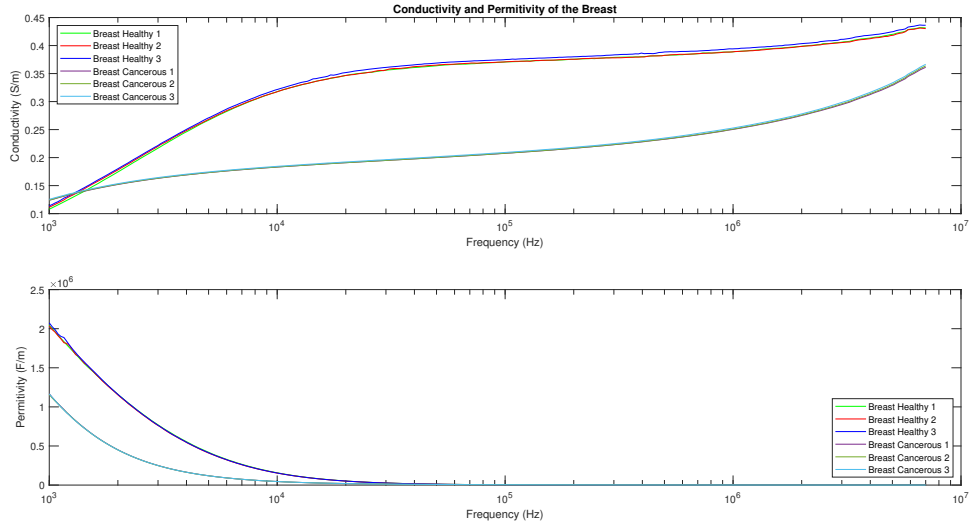


(b) The permittivity and conductivity for the measured adrenal gland.

Figure A-3: The measured impedance of the adrenal gland, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

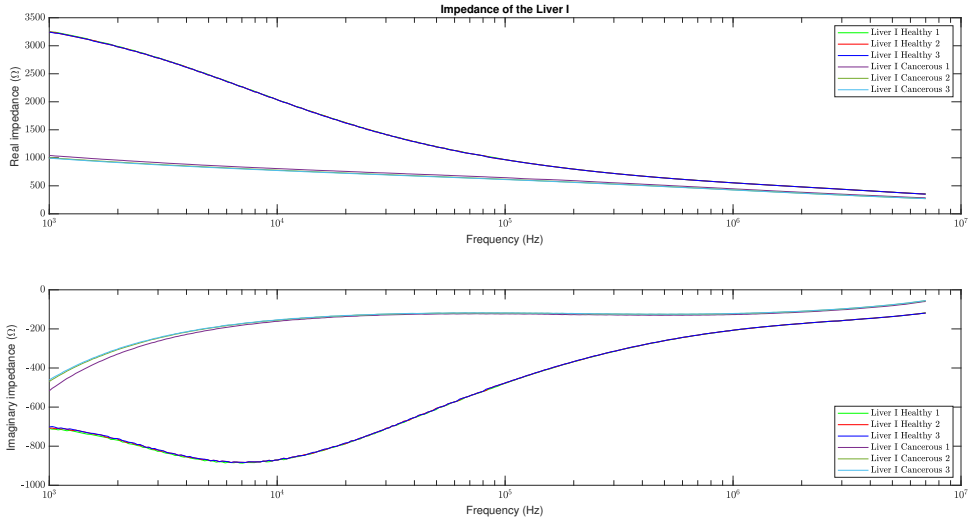


(a) The real part and the imaginary part of the resistance for the measured breast.

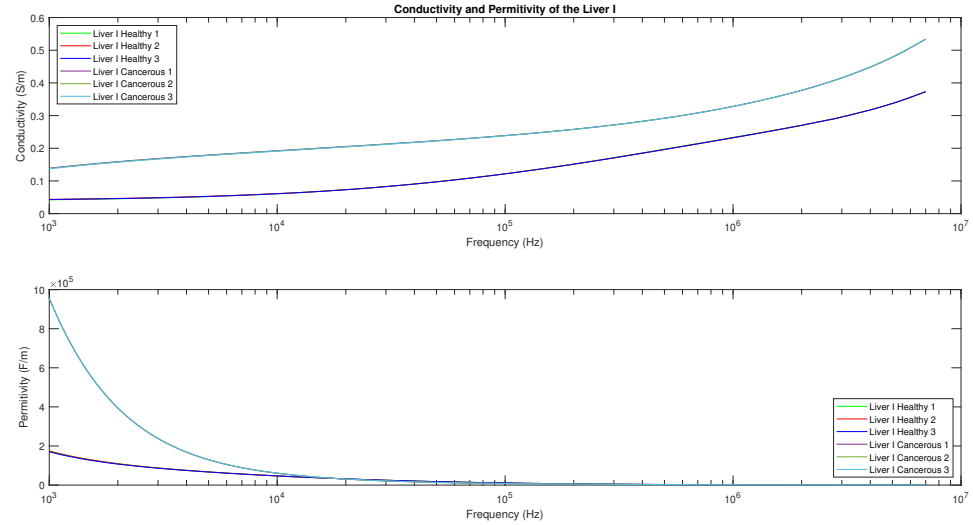


(b) The permittivity and conductivity for the measured breast.

Figure A-4: The measured impedance of the breast, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

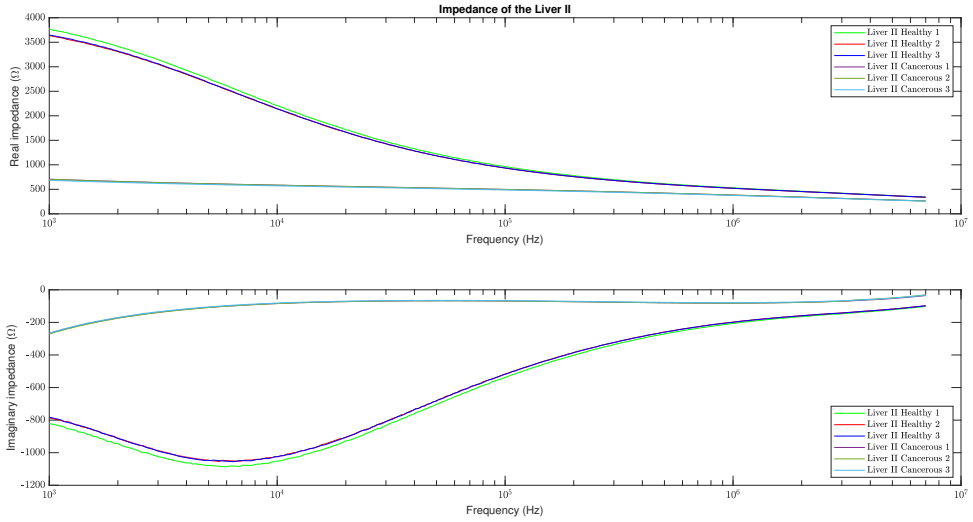


(a) The real part and the imaginary part of the resistance for the measured liver I.

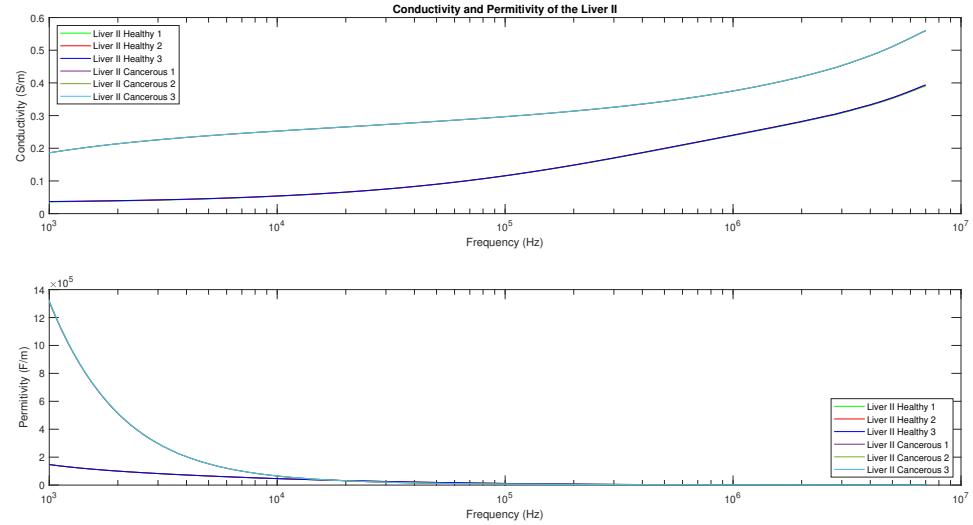


(b) The permittivity and conductivity for the measured liver I.

Figure A-5: The measured impedance of the liver I, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

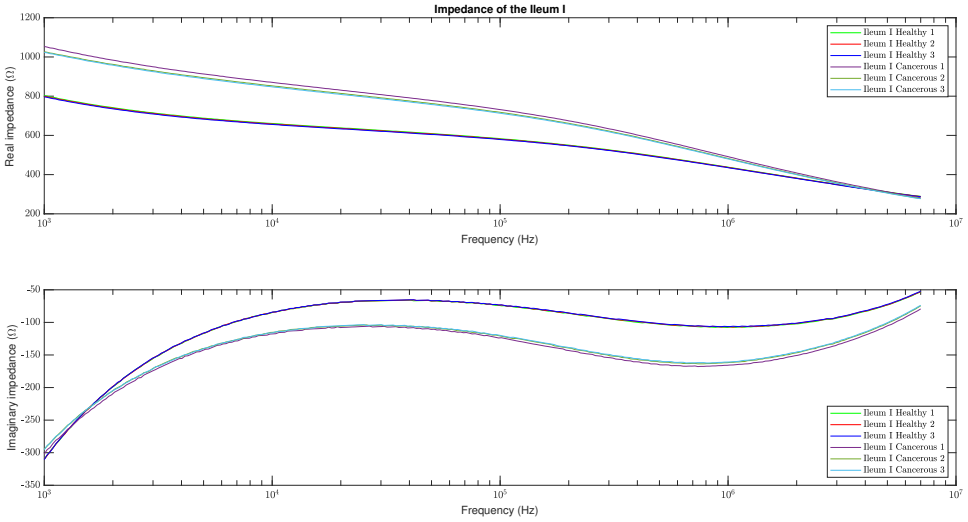


(a) The real part and the imaginary part of the resistance for the measured liver II.

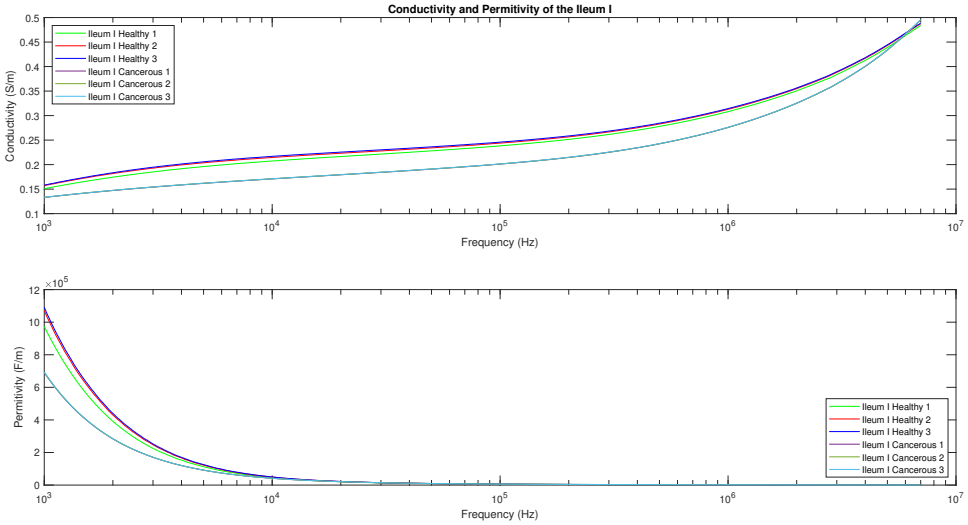


(b) The permittivity and conductivity for the measured liver II.

Figure A-6: The measured impedance of the liver II, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

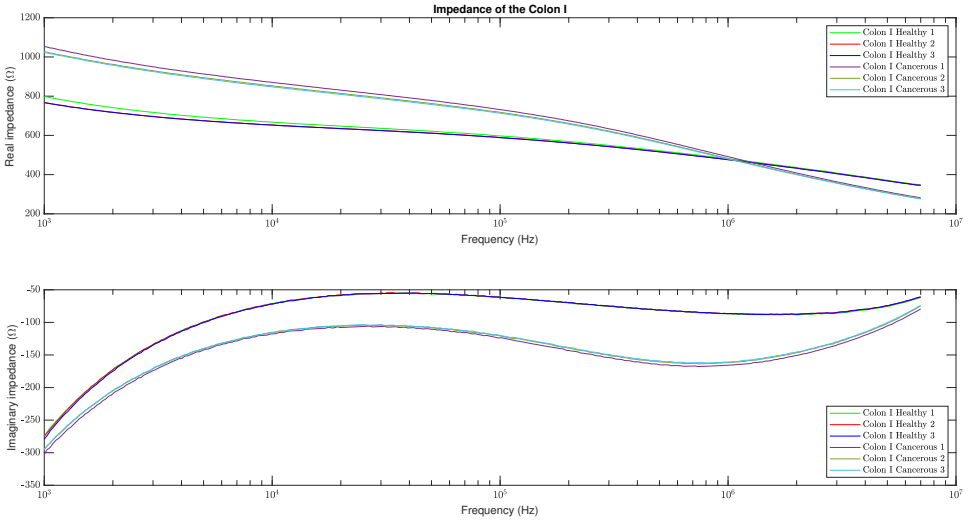


(a) The real part and the imaginary part of the resistance for the measured Ileum I.

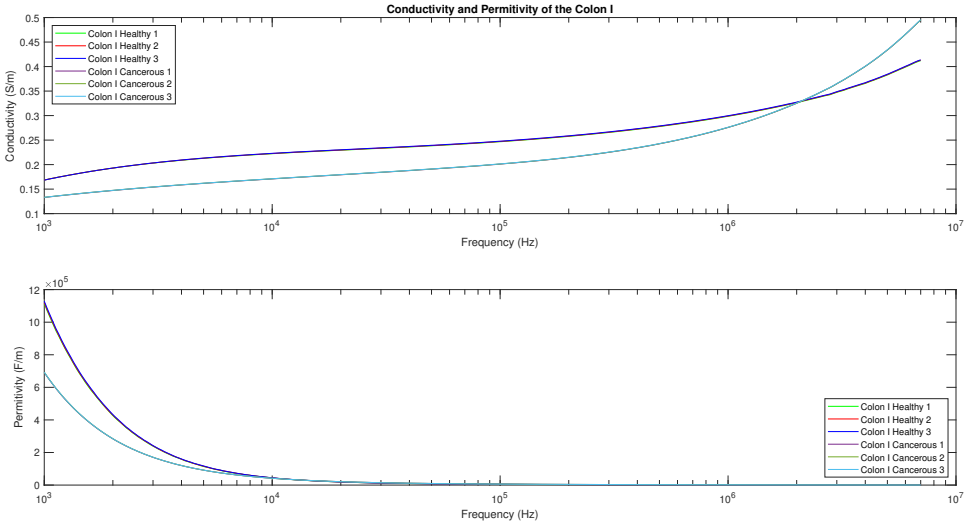


(b) The permittivity and conductivity for the measured Ileum I.

Figure A-7: The measured impedance of the Ileum I, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

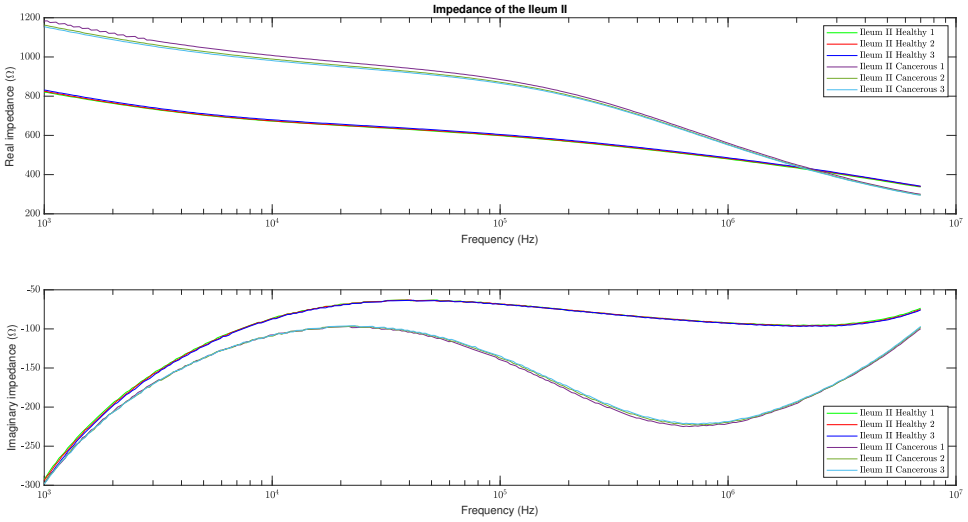


(a) The real part and the imaginary part of the resistance for the measured Colon.

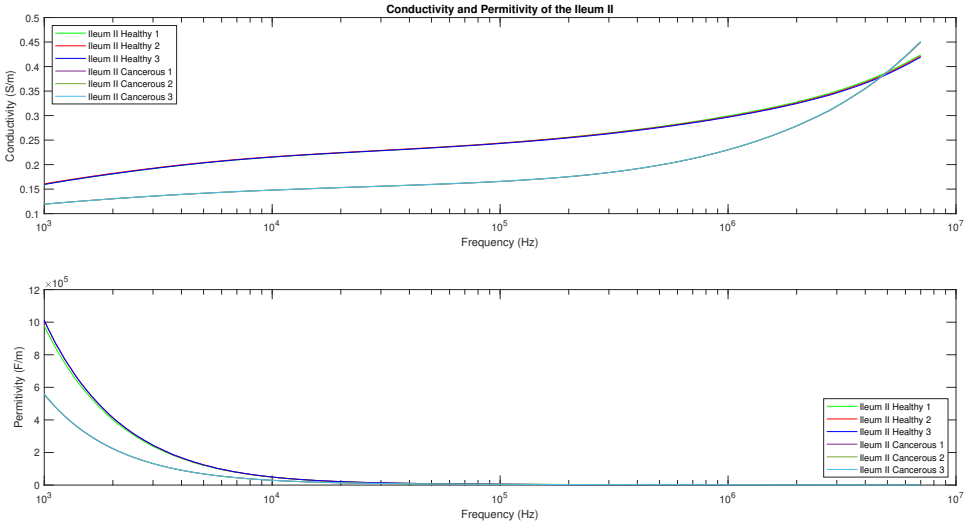


(b) The permittivity and conductivity for the measured Colon.

Figure A-8: The measured impedance of the Colon, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

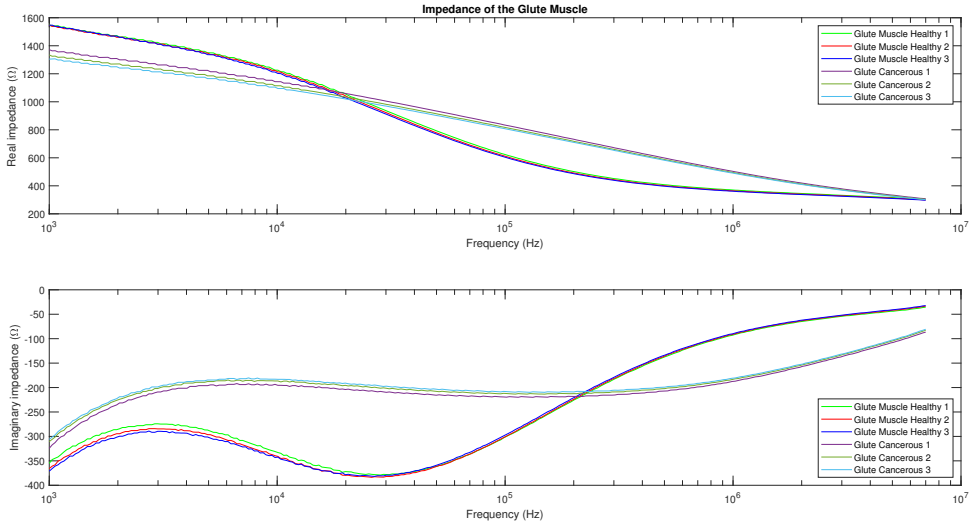


(a) The real part and the imaginary part of the resistance for the measured Ileum II.

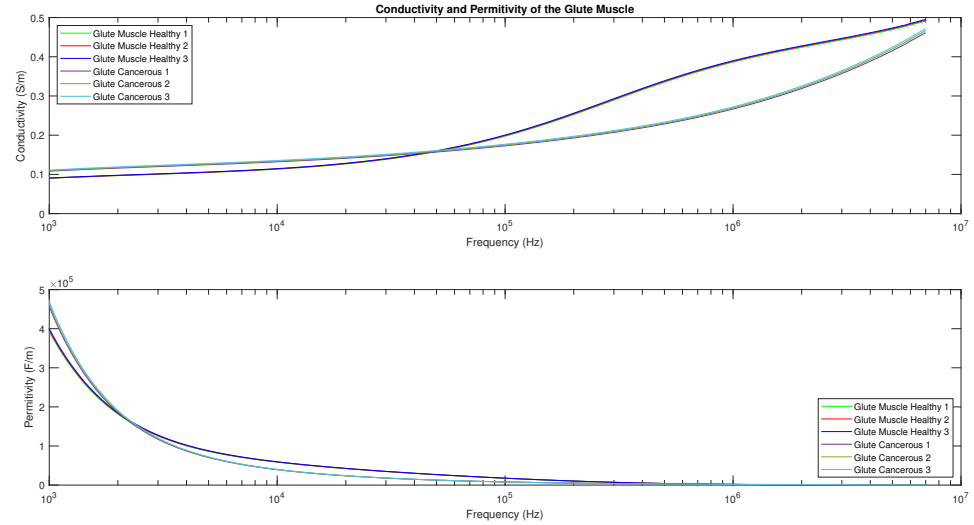


(b) The permittivity and conductivity for the measured Ileum II.

Figure A-9: The measured impedance of the Ileum II, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

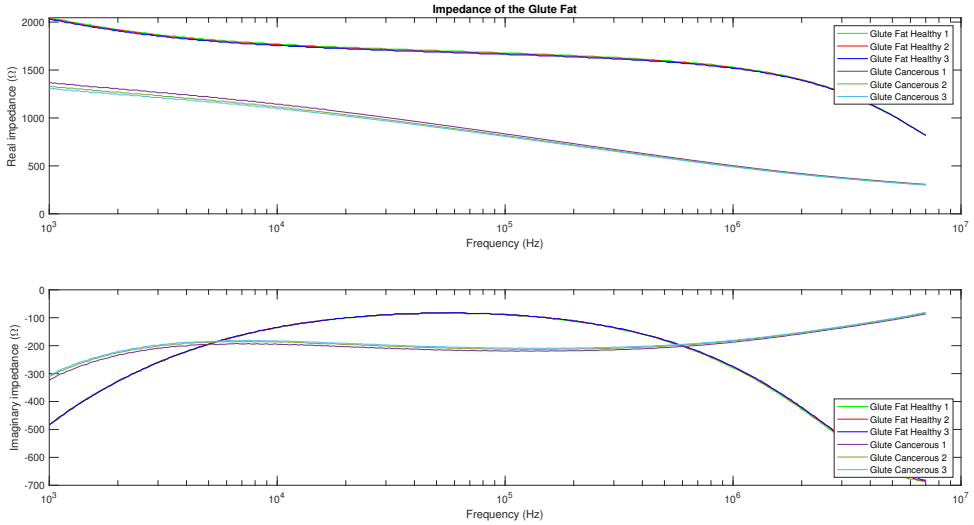


(a) The real part and the imaginary part of the resistance for the measured glute muscle.

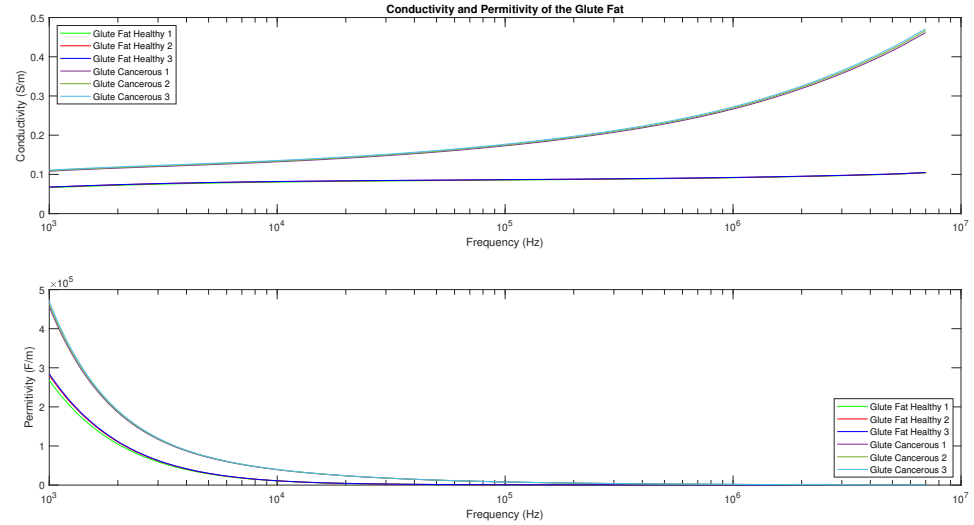


(b) The permittivity and conductivity for the measured glute muscle.

Figure A-10: The measured impedance of the glute musce, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

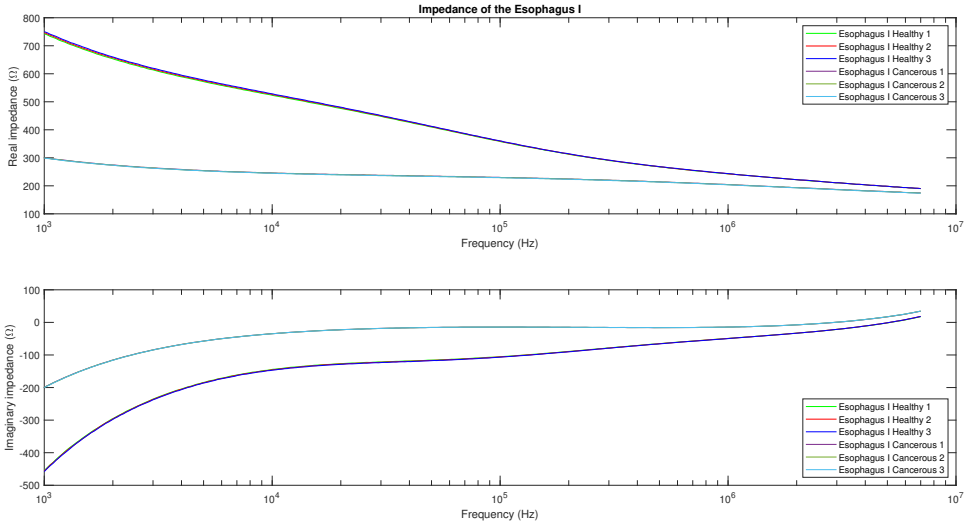


(a) The real part and the imaginary part of the resistance for the measured glute fat.

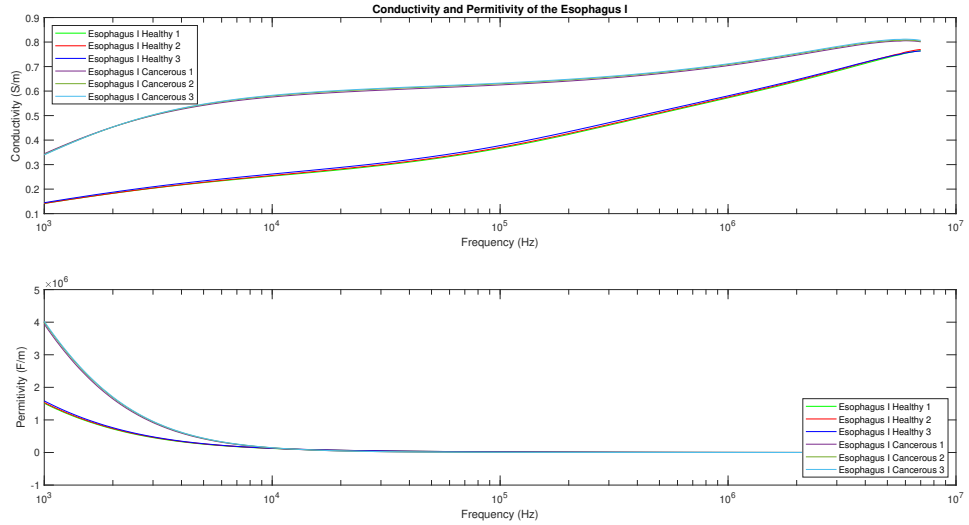


(b) The permittivity and conductivity for the measured glute fat.

Figure A-11: The measured impedance of the glute fat, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

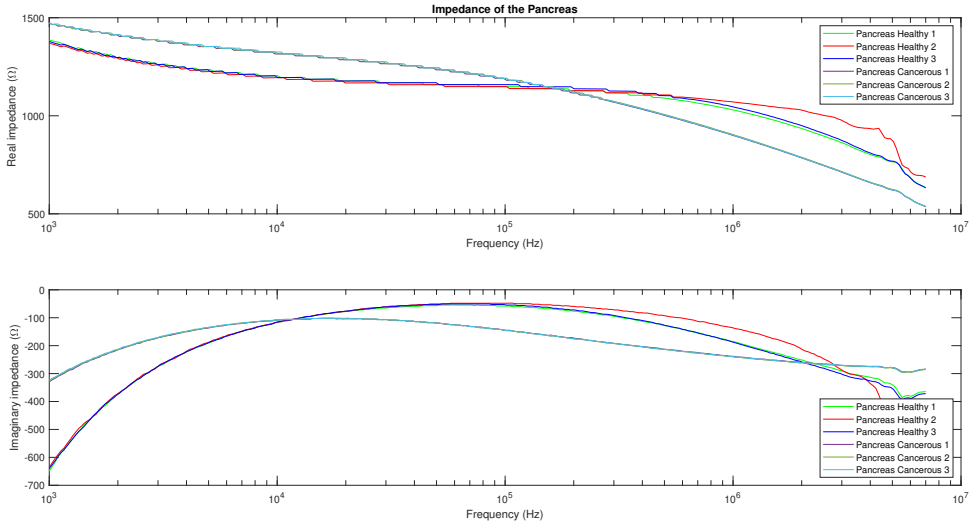


(a) The real part and the imaginary part of the resistance for the measured Esophagus I.

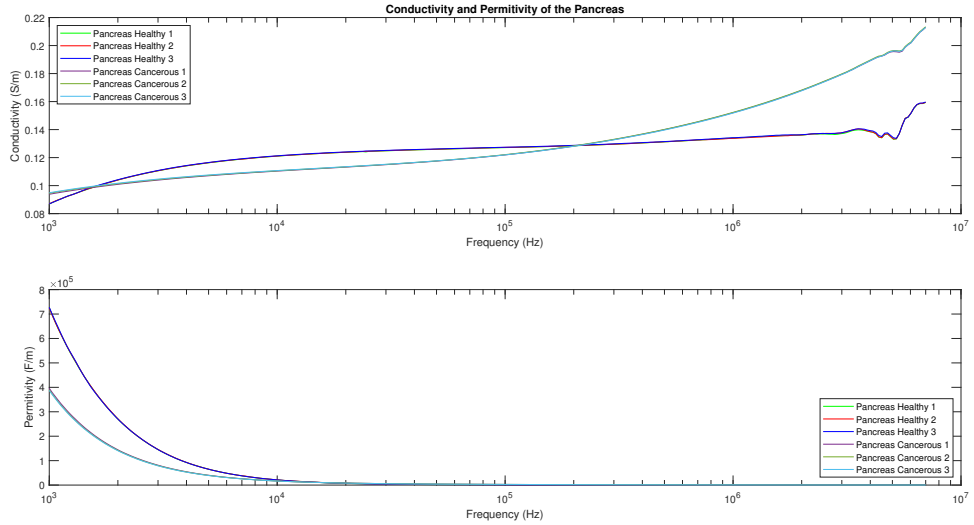


(b) The permittivity and conductivity for the measured Esophagus I.

Figure A-12: The measured impedance of the Esophagus I, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

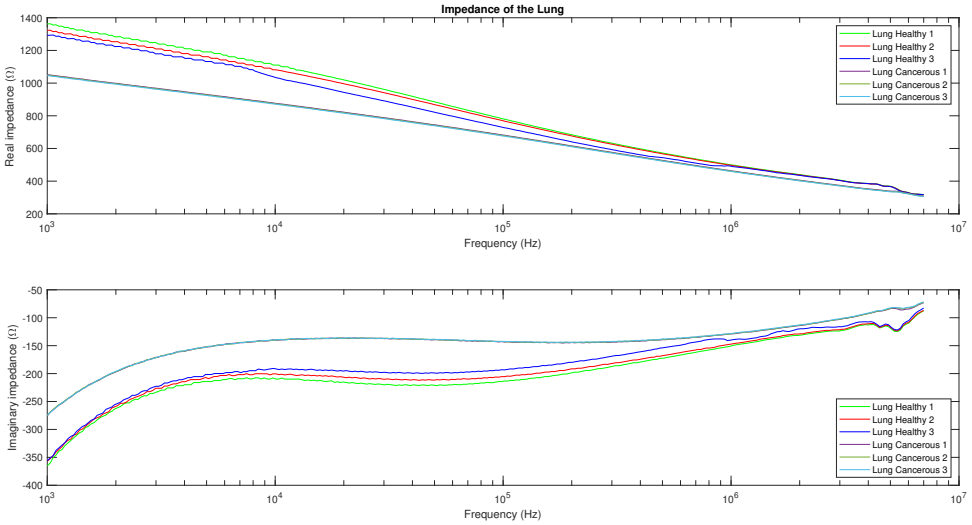


(a) The real part and the imaginary part of the resistance for the measured Pancreas.

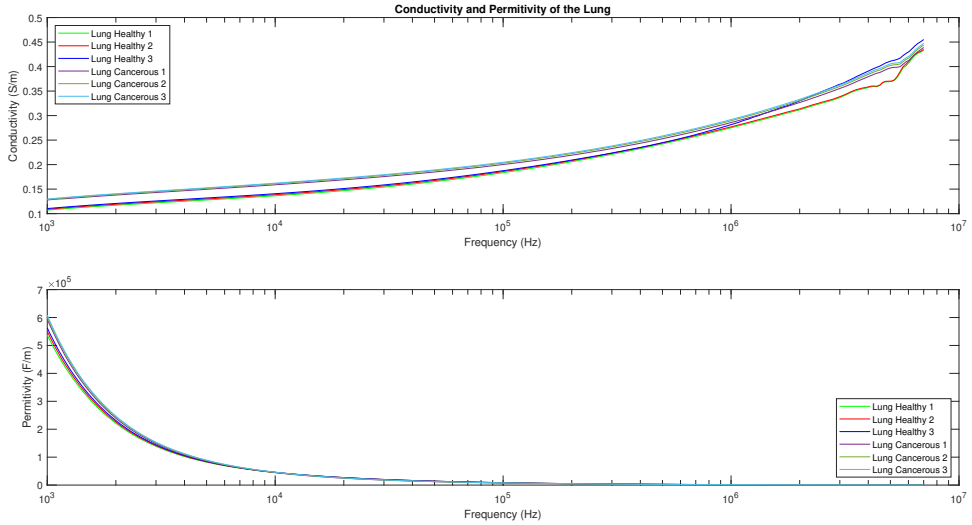


(b) The permittivity and conductivity for the measured Pancreas.

Figure A-13: The measured impedance of the Pancreas, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

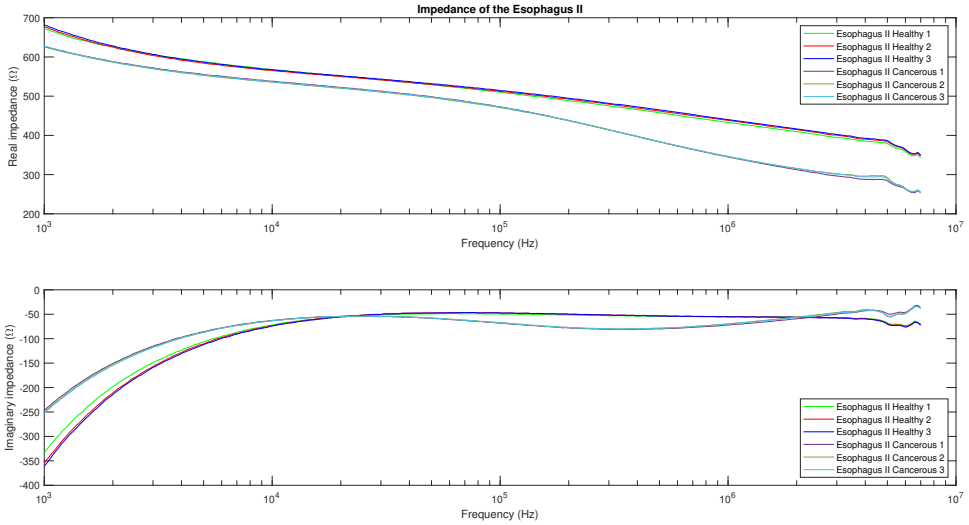


(a) The real part and the imaginary part of the resistance for the measured lung.

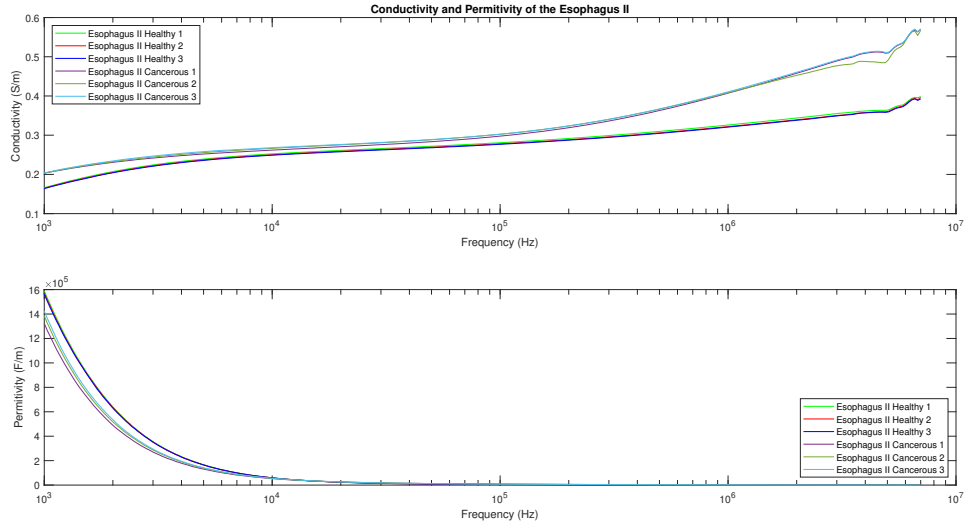


(b) The permittivity and conductivity for the measured lung.

Figure A-14: The measured impedance of the lung, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.



(a) The real part and the imaginary part of the resistance for the measured esophagus II.



(b) The permittivity and conductivity for the measured esophagus II.

Figure A-15: The measured impedance of the esophagus II, which shows the real impedance and negative imaginary part of the impedance, corresponding to capacitive reactance. And the measured permittivity and conductivity.

Appendix B

Trial Two Measurement

This appendix will cover all measurements conducted during the second measurement trial at the Erasmus MC in Rotterdam. This period started on 31-01-2023 and ended on 09-06-2023. We focused on esophagus, ileum, and colon tissue during this period. All tissues were measured within 24 hours of resection. It should be noted that, although we focused on these three tissues, no ileum specimens were available during this period.

During the trial period, impedance, conductivity, and capacitance measurements were made over 1 kHz to 7 MHz in 300 steps. We measured at least three places: A, B, and C. If we had enough time, we would include a fourth place: D. Locations A and B would always be non-cancerous tissue, whereas location C would always be cancerous tissue. Additionally, histological assessment would always verify the tissue at locations A, B, and C. Due to time and resource constraints, this was not done for location D. The measurement protocol for the second trial can be found in Figure B-1.

In Table B-1, one can find an overview of the measurements conducted during the trial period. Additionally, Table B-2 shows where the measurements were conducted and the results of the histological analysis performed in the Erasmus MC.

#	Date of Measurement	Gender (M/F)	Year of Birth	Tissue
1	31-01-2023	F	1979	Colon
2	28-03-2023	M	1960	Colon
3	09-06-2023	M	1943	Esophagus

Table B-1: The year of birth, gender, and date of the measurement per patient and the corresponding tissue the measurements were conducted on during the second trial measurement period in the Erasmus MC.

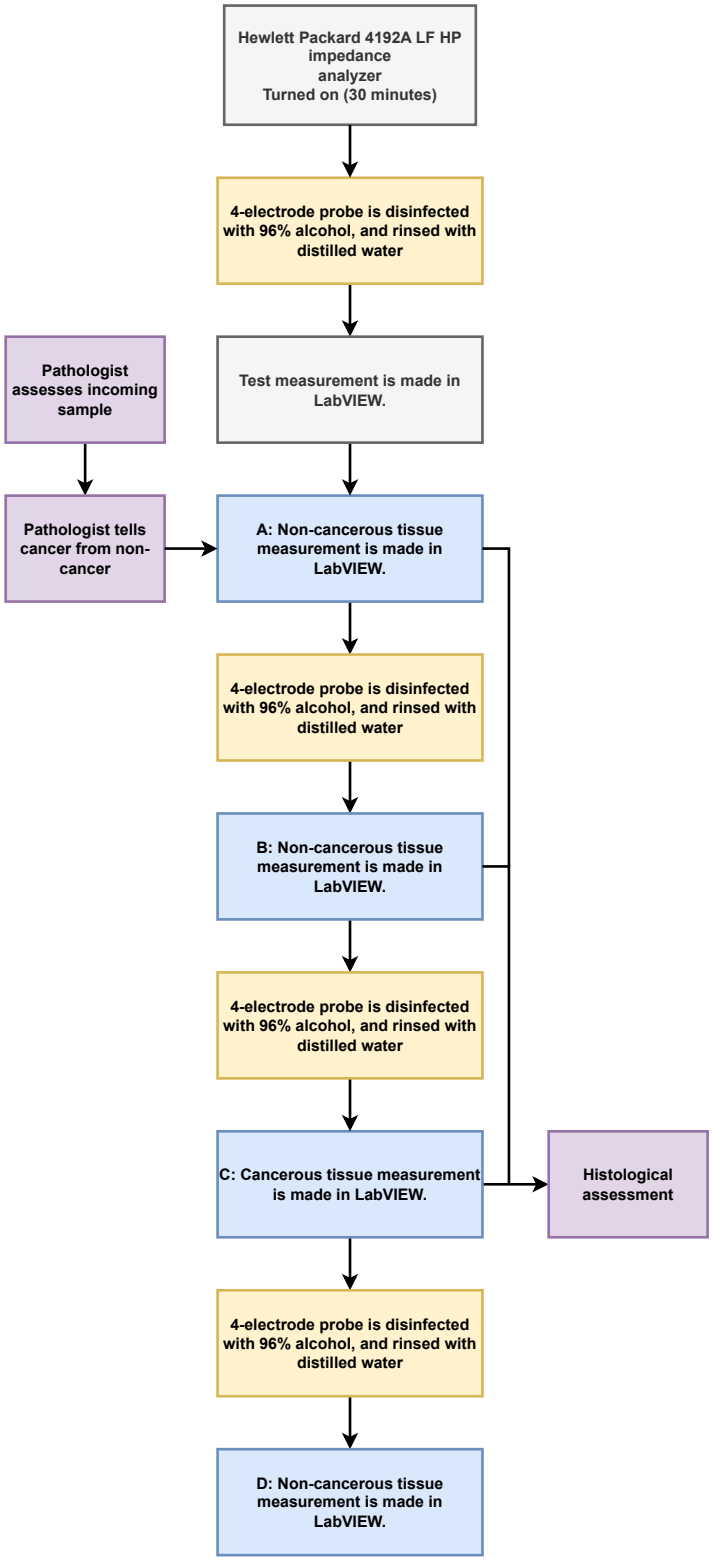


Figure B-1: The measurement protocol for the second trial that was conducted at the Erasmus MC.

#	Tissue	Measurement Location	Histological Analysis
1	Colon	A, B: Normal C: Tumor	A: Normal B, C: Tumor
2	Colon	A, B: Normal C: Tumor	A, B: Normal C: Normal
3	Esophagus	A, B: Normal C: Tumor	A, B: Normal C: Normal

Table B-2: The measurement tissue and the measured placed with the results from the histological results per patient.

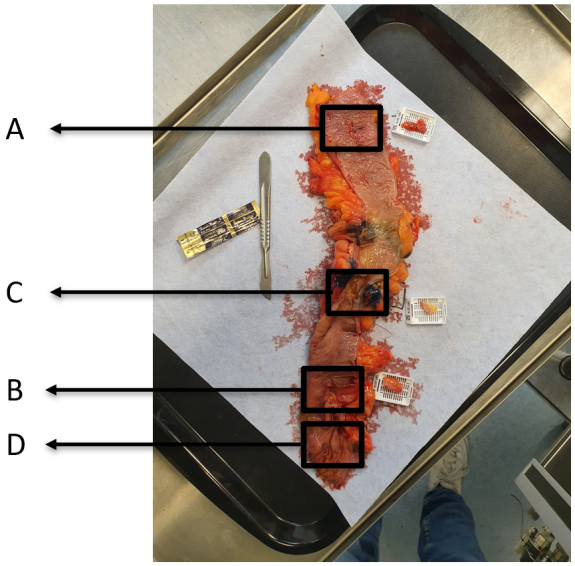


Figure B-2: Different measurement locations of the first patient of the second trial

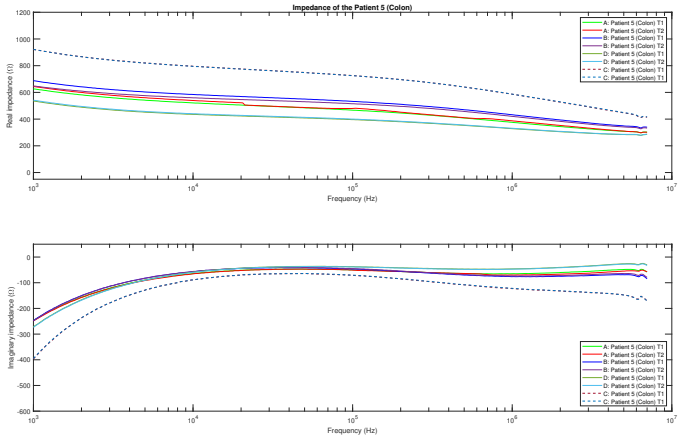


Figure B-3: Impedance Patient 1 trial 2 (Colon).

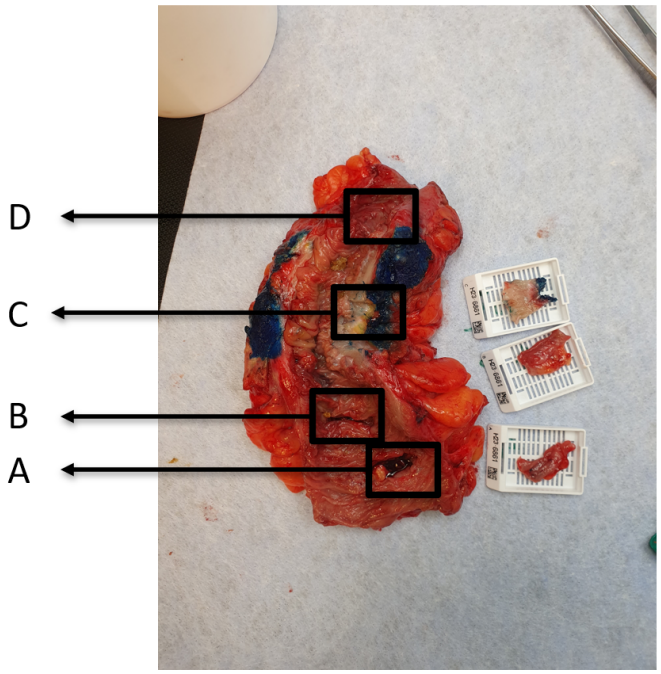


Figure B-4: Different measurement locations of the second patient of the second trial

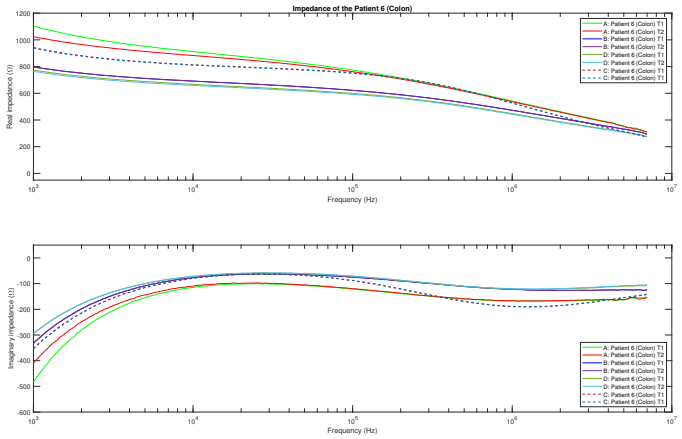


Figure B-5: Impedance Patient 2 trail 2 (Colon).

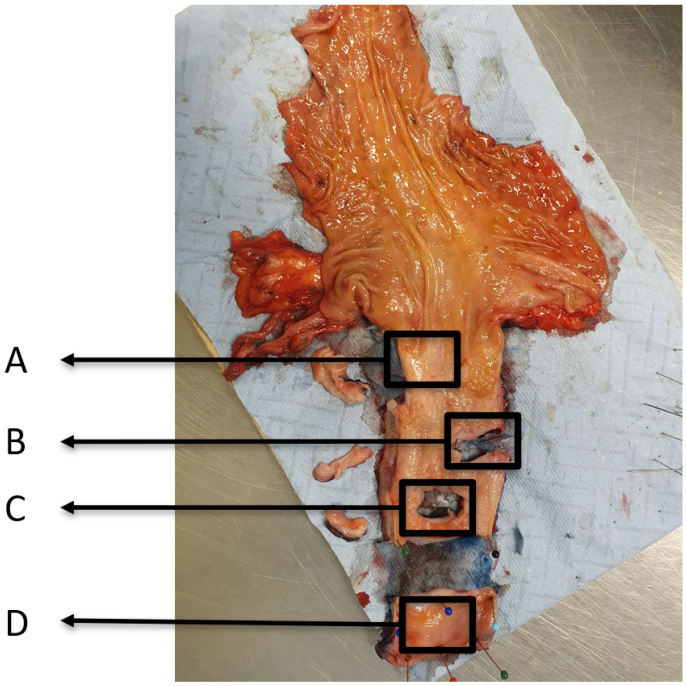


Figure B-6: Different measurement locations of the third patient of the second trial

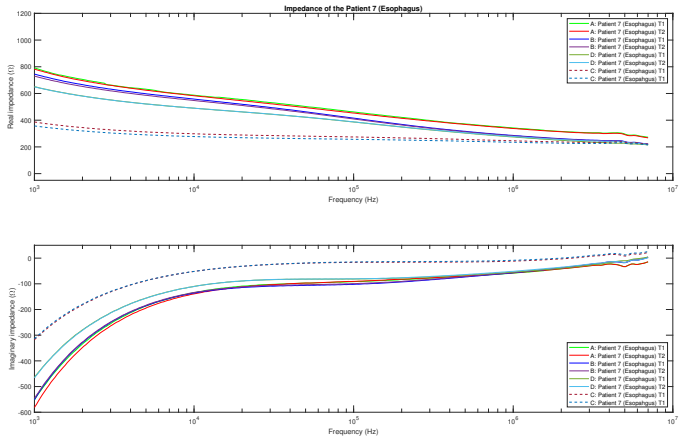


Figure B-7: Impedance Patient 3 trial 2 (Esophagus).

Bibliography

[1] Hervé Abdi and Lynne J Williams. "Principal component analysis". In: *Wiley interdisciplinary reviews: computational statistics* 2.4 (2010), pp. 433–459.

[2] S Abdul et al. "The use of electrical impedance spectroscopy in the detection of cervical intraepithelial neoplasia". In: *International Journal of Gynecologic Cancer* 16.5 (2006).

[3] Peter Åberg. *Skin cancer as seen by electrical impedance*. Institutionen för laboratoriemedicin/Department of Laboratory Medicine, 2004.

[4] James J Ackmann and Martin A Seitz. "Methods of complex impedance measurements in biologic tissue." In: *Critical reviews in biomedical engineering* 11.4 (1984), pp. 281–311.

[5] Rasool Baghbani et al. "Bioimpedance spectroscopy measurement and classification of lung tissue to identify pulmonary nodules". In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–7.

[6] Kenneth S Cole. *Membranes, ions and impulses: a chapter of classical biophysics*. Vol. 1. Univ of California Press, 1972.

[7] Shekh Emran et al. "Bioimpedance spectroscopy and spectral camera techniques in detection of oral mucosal diseases: a narrative review of the state-of-the-art". In: *Journal of medical engineering & technology* 43.8 (2019), pp. 474–491.

[8] John WL Fielding et al. "The Anatomy and physiology of the Oesophagus". In: *Upper Gastrointestinal Surgery* (2005), pp. 1–15.

[9] Camelia Gabriel, Sami Gabriel, and y E Corthout. "The dielectric properties of biological tissues: I. Literature survey". In: *Physics in medicine & biology* 41.11 (1996), p. 2231.

[10] Sverre Grimnes and Orjan G Martinsen. "Cole electrical impedance model-a critique and an alternative". In: *IEEE transactions on biomedical engineering* 52.1 (2004), pp. 132–135.

[11] Igor Griva, Stephen G Nash, and Ariela Sofer. *Linear and Nonlinear Optimization 2nd Edition*. SIAM, 2008.

[12] Oliver Heaviside. *Electromagnetic waves*. Cambridge University Press, 2011.

[13] T Ikyumbur et al. "Optimization in the computation of dielectric constant of methanol using Debye relaxation method". In: *British Journal of Applied Science & Technology* 19.1 (2017), pp. 1–10.

[14] Paul Ben Ishai et al. "Electrode polarization in dielectric measurements: a review". In: *Measurement science and technology* 24.10 (2013), p. 102001.

[15] Antoni Ivorra et al. "Bioimpedance dispersion width as a parameter to monitor living tissues". In: *Physiological measurement* 26.2 (2005), S165.

[16] Håvard Kalvøy et al. "New method for separation of electrode polarization impedance from measured tissue impedance". In: *The open biomedical engineering journal* 5 (2011), p. 8.

[17] Shadab Khan et al. "Prostate cancer detection using composite impedance metric". In: *IEEE transactions on medical imaging* 35.12 (2016), pp. 2513–2523.

[18] Suguru Kimura et al. "Application of electrical impedance analysis for diagnosis of a pulmonary mass". In: *Chest* 105.6 (1994), pp. 1679–1682.

[19] FOSTER KR. "Dielectrical properties of tissues and biological materials. a critical review". In: *Crit Rev Biomed Eng.* 17 (1989), pp. 25–104.

[20] Rashmi Krishnappa et al. "A study on histopathological spectrum of upper gastrointestinal tract endoscopic biopsies". In: *International Journal of Medical Research & Health Sciences* 2.3 (2013), pp. 418–424.

[21] David C Lay. *Linear algebra and its applications*. Pearson Education India, 2003.

[22] Philine Luik. "Design of a cardiac phantom for development of electrophysiology sudies". In: (2020).

[23] Digby D Macdonald. "Reflections on the history of electrochemical impedance spectroscopy". In: *Electrochimica Acta* 51.8-9 (2006), pp. 1376–1388.

[24] Elaine Nicpon Marieb and Katja Hoehn. *Human anatomy & physiology*. Pearson education, 2007.

[25] Vaishali Mishra et al. "Electrical impedance spectroscopy for prostate cancer diagnosis". In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2012, pp. 3258–3261.

[26] Sepideh Mohammadi Moqadam et al. "Cancer detection based on electrical impedance spectroscopy: A clinical study". In: *Journal of Electrical Bioimpedance* 9.1 (2018), pp. 17–23.

[27] Larissa Mulder. "Development of a parametric model to distinguish between specific tissue types for the application in a radiofrequency ablation catheter". In: *TU Delft Repository* (2022).

[28] Angela Pathiraja et al. "Detecting colorectal cancer using electrical impedance spectroscopy: an ex vivo feasibility study". In: *Physiological measurement* 38.6 (2017), p. 1278.

[29] Angela A Pathiraja et al. "The clinical application of electrical impedance technology in the detection of malignant neoplasms: a systematic review". In: *Journal of Translational Medicine* 18.1 (2020), pp. 1–11.

[30] Armando Peixoto et al. "Biopsies in gastrointestinal endoscopy: when and how". In: *GE Portuguese journal of gastroenterology* 23.1 (2016), pp. 19–27.

[31] David R Perera, Wilfred M Weinstein, and Cyrus E Rubin. "Small intestinal biopsy". In: *Human pathology* 6.2 (1975), pp. 157–217.

[32] Medline Plus. *Small intestine biopsy*. URL: https://medlineplus.gov/. (accessed: 4.4.2023).

[33] S Poompavai and V Gowri Sree. "Dielectric property measurement of breast—Tumor phantom model under pulsed electric field treatment". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 2.6 (2018), pp. 608–617.

[34] Rob Reilink, Stefano Stramigioli, and Sarthak Misra. "Image-based flexible endoscope steering". In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2010, pp. 2339–2344.

[35] Evan Remmerswaal. "Electrical Impedance Spectroscopy for Real-Time Cancer Detection". In: (2023).

[36] Hugo Sanabria and John H Miller Jr. "Relaxation processes due to the electrode-electrolyte interface in ionic solutions". In: *Physical Review E* 74.5 (2006), p. 051505.

[37] Herman P Schwan. "Electrical properties of tissue and cell suspensions". In: *Advances in biological and medical physics*. Vol. 5. Elsevier, 1957, pp. 147–209.

[38] Mount Sinai. *Enteroscopy*. URL: https://www.mountsinai.org/health-library/tests/enteroscopy. (accessed: 4.4.2023).

[39] Gilbert Strang. *Linear algebra and its applications*. Belmont, CA: Thomson, Brooks/-Cole, 2006.

[40] Daniil D Stupin et al. "Bioimpedance spectroscopy: basics and applications". In: *ACS Biomaterials Science & Engineering* 7.6 (2021), pp. 1962–1986.

[41] Oakleaf Surgical. *COLON CANCER MORE COMMON THAN YOU THINK!* URL: https://www.oakleafsurgical.com/hv/2012_wi/wi2012_coloncancer.php. (accessed: 4.4.2023).

[42] Agilent Technologies. "Agilent Impedance Measurement Handbook A guide to measurement technology and techniques". In: (2009).

[43] JA Tidy et al. "Accuracy of detection of high-grade cervical intraepithelial neoplasia using electrical impedance spectroscopy with colposcopy". In: *BJOG: An International Journal of Obstetrics & Gynaecology* 120.4 (2013), pp. 400–411.

[44] Valery Vodovozov. *Introduction to electronic engineering*. Bookboon, 2010.

[45] Helen H Wang and Gamze Ayata. "Gastrointestinal tract". In: *Cytology E-Book: Diagnostic Principles and Clinical Correlates* (2019), p. 214.

[46] Jian Wu et al. "A new method to rapidly identify benign and malignant breast lumps through bioelectrical impedance spectroscopy". In: *Medical physics* 46.5 (2019), pp. 2522–2525.

[47] Takahiko Yamamoto and Kohji Koshiji. "Electromagnetic high-hydrous gel phantom at a low-frequency band-Improvement in the electrical characteristics by using a carbon microcoil and investigation of its mechanism". In: *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*. Springer. 2015, pp. 1383–1386.

[48] LTD. Yogogawa-Hewlett-Packard. *Operation and Service Manual: Model 4192A LP Impedance Analyzer*. 1984.

[49] HM El-Zimaity, Kemal Itani, and David Y Graham. "Early diagnosis of signet ring cell carcinoma of the stomach: role of the Genta stain." In: *Journal of clinical pathology* 50.10 (1997), pp. 867–868.