

**Multi-Dimensional Electrical Impedance Data
Analysis For Amyotrophic Lateral Sclerosis
Disease Diagnosis**

A thesis submitted to the University of Sheffield for the degree of Doctor of
Philosophy

Buket Sonbas

Department of Automatic Control and Systems Engineering

September 2020

To my past and my future

Acknowledgements

I would like to express my greatest appreciation to my supervisor Professor Visakan Kadirkamanathan. His knowledge, patience, and guidance helped me through the most challenging years of my academic life. I will take his teaching attitude as an example in my future teaching career.

I would like to offer my special thanks to my second supervisor Professor Lyudmila Mihaylova for her academic advice and moral support. I would like to show my gratefulness to Dr. James Alix MRCP PhD, for giving me the opportunity to work with him in this promising research project. He has been a great advisor throughout my work.

Advices given by Jamie Healey has been a great help in broadening the research questions. Dr. Parham Aram has been a great support during the first year of my project. I would like to thank him for introducing me to the Python programming language that helped me greatly with my analysis.

My special thanks to the Republic of Turkey, Ministry of National Education (MEB, MEBLEM) for their financial support and encouragement. Thank you very much for trusting me and for providing me with the scholarship and job opportunity in the city where I have always been dreaming of working.

Thanks to Signal Processing research group members between the years 2016-2020. My dear friend Raj, my friends with whom I shared four great years of companionship Uziel, Isaac, Oliver, Jinny, and Ph.D. Group 2016.

Finally, I am grateful to my mom Aysel, my dad Hamza, my sister Eda, and my precious nephew Uzay, for their guidance in every step of my life, their patience, emotional and financial support, and their trust in me. To my current and future family, Andrew, Rossie, and Jimmy, thank you for helping me through difficult times and helping me breathe deeper.

Abstract

Amyotrophic Lateral Sclerosis (ALS) is the most common type of motor neurone disease (MND), which affects the motor neurones that control the voluntary skeletal muscles. ALS patients with bulbar involvement experience weaknesses in muscles that control speech, swallowing, and breathing. Recent attempts in using Electrical Impedance (EI) techniques have shown promising results for early diagnosis of the bulbar involvement of the ALS disease.

A non-invasive, handheld EIM device is being developed as a potential biomarker for ALS bulbar involvement by a group in Sheffield. The device has a novel electrode configuration setting that employs a combination of eight electrodes used on both surfaces of the tongue to obtain the spectral EI measurements. This thesis proposes analysis of the acquired data using machine learning and system identification methods. The aim is to find patterns in data to assess bulbar dysfunction for diagnosis, and longitudinal assessment of the disease.

A data specific outlier removal algorithm was implemented as part of preliminary data analysis, and electrode configurations that did not comply with the EI characteristics of the tongue muscle along the full spectrum are eliminated. Classification of the patients and healthy volunteers demonstrated that the novel electrode configurations were capable of differentiating the two groups. Classification using the feature selection method not only gave better accuracy, but also revealed the most discriminatory frequency combination for disease diagnosis. Impedance data was modeled with an Integrated Fractional-Order Transfer Function. A model selection scheme over the model space was employed to avoid underestimation, overestimation, and over-parametrization. Estimated parameters were used to identify disease related patterns. Analysis shows the device's capacity to diagnose the disease and the non-parametric data classification has shown better accuracies than the parametric data classification.

Contents

List of Figures	vii
List of Tables	xi
1 Introduction	1
1.1 Motivation	2
1.2 Aims and Objectives	4
1.3 Overview of the Thesis	5
1.4 Publications	6
2 Electrical Impedance Data	7
2.1 Bioelectrical Impedance	8
2.2 Circuitry Modeling of the Bioelectrical Impedance Data	9
2.3 Electrical Impedance Myography	13
2.3.1 ALS Biomarkers	14
2.3.2 Disease-Related Muscle Groups	15
2.3.3 EIM Frequencies	17
2.3.4 Age and Gender-Related Studies	18
2.3.5 Longitudinal Analysis	19
2.4 Multi-dimensional Electrical Impedance Myography Device	20
2.4.1 Tongue Measurements	24
2.4.2 Patients and Data Collection	25

2.4.3	Multi-dimensional Electrical Impedance Myography Data . . .	26
2.5	Data Preprocessing	27
2.5.1	Outlier Detection	29
2.5.2	Preliminary Selection of Electrode Configurations	32
3	Non-Parametric Data Classification For Disease Diagnosis And Longitudinal Analysis	36
3.1	Supervised Classification	37
3.1.1	Dimensionality Reduction	38
3.1.2	Cross Validation	42
3.1.3	Classification Algorithms	43
3.1.4	Classification Results Evaluation	47
3.2	Non-Parametric Data Classification Using Feature Extraction	49
3.2.1	Preprocessing	49
3.2.2	Support Vector Classifier	52
3.2.3	k-Nearest Neighbour Classifier	54
3.3	Non-Parametric Data Classification Using Feature Selection	56
3.3.1	Feature Selection with Wrapper Method	57
3.3.2	k-NN Classifier with n-Fold Cross Validation	58
3.3.3	Pareto Dominance for Selecting Disease-Specific Frequencies	58
3.3.4	Diagnosis Using Multiple Electrode Configurations	61
3.4	Longitudinal Analysis Using Selected Features	63
3.5	Chapter Conclusion	65
4	Data Driven Modeling of Electrical Impedance Myography Measurements	67
4.1	Selecting A Model Structure For The EIM Data	68
4.2	Black Box Transfer Function Model	71
4.3	Rational Polynomial Transfer Function Model	74

4.4	Naive Least Squares Estimation	75
4.5	Factorized Transfer Function Model	78
4.5.1	Integral Fractional Order Transfer Function	79
4.5.2	Integral Fractional Order Transfer Function Parameters From Polar Components	81
4.6	Chapter Conclusion	83
5	Parameter Estimation and Model Selection	84
5.1	Parameter Estimation	85
5.1.1	Grid Search	86
5.1.2	Genetic Algorithm	90
5.1.3	Markov Chain Monte Carlo	96
5.2	Model Selection	102
5.2.1	Akaike Information Criterion	103
5.2.2	Bayesian information criterion	105
5.3	Chapter Conclusion	108
6	Parametric Data Classification For Disease Diagnosis and Longitudinal Analysis	110
6.1	Parameter Correlograms	111
6.2	Supervised Classification	113
6.2.1	Support Vector Classifier	113
6.2.2	k-Nearest Neighbour Classifier	117
6.3	Parametric and Non-Parametric Classification Results Comparison	120
6.4	Diagnosis Using Multiple Electrode Configurations	122
6.5	Longitudinal Analysis Using Selected Model	123
6.6	Chapter Conclusion	125

7 Conclusions	127
7.1 Summary and Conclusions	128
7.2 Future Work	130
Bibliography	133

List of Figures

2.1	Simple 3-Element [52] and advanced 5-Element [53] cellular circuit model. R_e is the resistance from extracellular fluids, R_i is the resistance from intracellular fluids, R_o is the resistance from organelle fluids, X_c is the reactance from cell membrane, X_o is the reactance from organelle membrane.	11
2.2	Electrical Impedance Spectrum visualisation with Cole-Cole plot (Nyquist plot) in complex domain [52]	12
2.3	Electrical Impedance characteristics of biological tissues along the frequency spectrum with representative visualisation of cell-frequency relation [110]. Important frequencies for ALS studies (50kHz, 100kHz) and frequencies that gives the relation in cell figures (1kHz, 10kHz, 1MHz) are noted on x-axis and Impedance readings are on y-axis.	18
2.4	Design of device for measuring tongue impedance. Four electrodes placed on the upper and lower plates. Current injected from the two red electrodes and the voltage measured from the other two blue electrodes (see Table 2.1). The figures are plotted by Dr. Jamie Healey.	22
2.5	Box-Whisker plot for complex impedance along the frequency spectrum. The box shows the interquartile range with lower quartile (25%), upper quartile (75%) and the median (50%). Whiskers show the minimum and maximum points of the data and without the outliers. The plots proves the data distribution is skewed.	28

2.6	Outlier removal algorithm result for twelve electrode configurations on Right-Intraoral placement	32
2.7	Outlier removed data for twelve electrode configurations on Center-Intraoral placement	33
2.8	Center-Intraoral and Center-Extraoral phase median along the spectrum for comparing the healthy volunteers and patients to visually identify the disease's effect on Intraoral and Extraoral measurements. The number of patients and volunteers involved in median statistics are noted in the legend, given in each axis.	34
3.1	A 2D SVC hyperplane with support vectors and margins.	44
3.2	A 2D k-NN plot depicting the effects of different number of neighbours. . . .	46
3.3	Explained Variance plot obtained with phase values. The electrode configurations exceed 90% variance with four principal components.	51
3.4	An example plot for first two Principal Components. The number of volunteers and patients are noted in each subfigure. The explained variance from each principal components are noted on axis labels. The electrode configuration initials are noted on the top of each axis.	52
3.5	Classification accuracy results heatmap for different electrode configurations and different classifier design parameters with SVC.	54
3.6	Classification accuracy results heatmap for different electrode configurations and different classifier design parameters with k-NN.	55
3.7	All sensitivity and specificity results obtained from the frequency subsets are given (blue stars) to show the variability of the classification results. Solutions on Pareto front (blue line) cannot be bettered in both the sensitivity and specificity by any other solution. The selected results (red points) are noted for each electrode configurations.	59

3.8	Percent change between baseline and 6th month visits for volunteers and patients in each electrode configurations to find the electrode configurations that can detect the patient EIM change.	63
4.1	System Identification diagram to represent the steps for identification process.	69
4.2	Black Box model representation of the system.	72
4.3	Bode Diagram for one impedance data set, given with Magnitude and Phase response plots.	78
4.4	Series black-box model for combining the instrument measurement model and the transfer function model of the tissue.	81
5.1	Model fit for one example data. The four model fit are plotted: 1-zero 1-pole, 1-zero 2-pole, 2-zero 2-pole, 1-zero 3-pole.	89
5.2	Genetic Algorithm flow chart with population, gene, chromosome terms and crossover types.	91
5.3	Model fit for one example data. The five model fit with different complexities are plotted: 1-zero 1-pole, 1-zero 2-pole, 2-zero 2-pole, 1-zero 3-pole, 2-zero 3-pole.	93
5.4	Frequency-related weighting factors for resistance and reactance.	94
5.5	Frequency-related weighted 2-zero, 3-pole model fit for one example data with parameters $\theta_{\alpha_0} = -3.42$, $\theta_{\alpha_1} = -0.004$, $\theta_K = 1.96$, $\theta_{i_p} = -3.22, -3.28, -7.29$, $\theta_{i_z} = -3.84, -3.91$	95
5.7	2-pole 1-zero model fit, credible and prediction intervals for one example data.	100
5.6	MCMC chain results with (a) Parameter density plot obtained from the chain after burnin is removed and (b) Pairwise parameter correlation plot to detect the correlation between pairs of parameters.	101

5.8	3-pole 2-zero model fit, credible and prediction intervals for one example data.	101
5.9	AIC model selection for all electrode configurations. The models are represented with n_p, n_z such as 21 is $n_p = 2$ and $n_z = 1$ for model $M_{\alpha 21}$	104
5.10	BIC model selection for all electrode configurations. The models are represented with n_p, n_z such as 21 is $n_p = 2$ and $n_z = 1$ for model $M_{\alpha 21}$	106
6.1	Correlogram for electrode configuration $[RIS]$ and model $M_{\alpha 22}$	112
6.2	Support Vector Classifier results for selected models and all electrode configurations using full parameter set.	114
6.3	Support Vector Classifier results for selected models and all electrode configurations using gain, pole and zero parameters.	115
6.4	Support Vector Classifier results for selected models and all electrode configurations using pole and zero parameters.	116
6.5	k-Nearest Neighbour Classifier results for selected models and all electrode configurations using full parameter set.	118
6.6	k-Nearest Neighbour Classifier results for selected models and all electrode configurations using gain, pole and zero parameters.	119
6.7	k-Nearest Neighbour Classifier results for selected models and all electrode configurations using pole and zero parameters.	120
6.8	Percent change between baseline and 6th month visits for volunteers and patients in each electrode configurations to detect the patient EIM change with model parameters; poles and zeros.	124

List of Tables

2.1	2-dimensional (2D) and 3-dimensional (3D) electrode configurations on representative tongue figure with current injection electrodes (red) and voltage sensing electrodes (blue) specified. 2D current electrodes and voltage sensing electrodes placed on same surface. 3D current electrodes and voltage sensing electrodes placed on opposing surfaces. Electrode names are initialized (please see Section 2.4.3). The electrode keys are obtained from Figure 2.4. The tongue figures are plotted by Dr. Jamie Healey.	23
2.2	Electrical Impedance phase data recorded for Centre Intraoral and Centre Extraoral measurements in each visits. Volunteers only have the first three measurements.	35
3.1	Confusion Matrix with four components; True Positive (TP) for correctly identified patients, True Negative (TN) for correctly identified healthy volunteers, False Positive (FP) for for volunteers identified as patient, False Negative (FN) for patients identified as volunteers.	47
3.2	Standard, Robust and Min-Max scaling equations comparison.	50
3.3	Selected frequencies for each electrode configurations with accuracy, sensitivity, and specificity values. The frequencies are represented as [f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14] which is [76.29395, 152.5879, 305.1758, 610.3516, 1220.703, 2441.406, 4882.813, 9765.625, 19531.25, 39062.5, 78125, 156250, 312500, 625000] Hz respectively.	60

3.4	Diagnosis outcomes across all electrode configurations. For each individual results are provided as a ratio of the number of configurations in which the correct classification was made to the total number of configurations that were included in the analysis. The red color indicates where the majority outcomes showed correct diagnosis while the blue indicates the alternative.	62
3.5	Total bulbar subscores of the ALSFRS of each patient for baseline and 6th Month visits.	64
5.1	Model fit transfer functions with poles and zeros for Figure 5.1	89
5.2	Model fit transfer functions with poles and zeros for Figure 5.3	93
5.3	AIC model selection rank based table for all electrode configurations.	105
5.4	BIC model selection rank based table for all electrode configurations.	107
6.1	Parametric and Non-parametric classification results comparison using 3-Nearest Neighbour classifier.	121
6.2	Diagnosis using all electrode configurations.	123

List of Algorithms

1	Unwanted and outlier data removal along the frequency spectrum . . .	30
2	Grid Search for estimating gain, fractional orders, poles and zeros with Integral Fractional Order Transfer Function	88

Nomenclature

V	voltage
I	current
Z	impedance
$ Z $	magnitude
Θ	phase
R	resistance
X	reactance
R_e	extracellular resistivity
R_i	intracellular resistivity
R_o	organelle resistivity
X_e	extracellular reactivity
X_i	intracellular reactivity
μ	mean value
σ	standard deviation

f_l	frequencies
σ_{jk}^2	variance
σ_{kk}^2	covariance
v	eigenvector
λ	scalar eigenvalues
Q	quartiles
y	output
u	input
M	model
N	zero parameters
D	pole parameters
$M_{\alpha n_p n_z}$	fractional order model
Y	output
X	information matrix
θ	parameters
z_{i_z}	zeros
p_{i_p}	poles
K	gain
α	fractional order

\in	element of
J	mean squared error
J_W	weighted mean squared error
n_θ	number of parameters
<i>argmin</i>	argument that minimizes

Acronyms

2D Two Dimensional

3D Three Dimensional

AC Alternating Current

ALS Amyotrophic Lateral Sclerosis

ALSFRS-R Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised

AUROC Area Under Receiver Operator Characteristics

DC Direct Current

DDM Data Driven Modeling

ECF Extracellular Fluids

ECS Extracellular Solids

EI Electrical Impedance

EIM Electrical Impedance Myography

EMG Electromyography

FA Factor Analysis

TN False Negative

TNR False Negative Rate

TP False Positive

TPR False Positive Rate

k-NN k-Nearest Neighbours

LDA Linear Discriminant Analysis

LOO CV Leave One Out Cross Validation

LpO CV Leave p Out Cross Validation

MAD Median Absolute Deviation

MCC Matthews Correlation Coefficient

MCMC Markov Chain Monte Carlo

MG Myasthenia Gravis

MH Metropolis-Hastings

MND Motor Neuron Disease

MRI Magnetic Resonance Imaging

N Negative

NHS National Health Service

NMD Neuromuscular Disease

NPV Negative Predictive Value

P Positive

PC Principal Component

PCA Principal Component Analysis

PPV Positive Predictive Value

RMSD Root Mean Square Deviation

ROC Receiver Operator Characteristics

SVC Support Vector Classifier

STH Sheffield Teaching Hospitals

TBW Total Body Water

TN True Negative

TNR True Negative Rate

TOC Total Operating Characteristics

TP True Positive

TPR True Positive Rate

UMN Upper Motor Neuron

Chapter 1

Introduction

Amyotrophic lateral sclerosis (ALS) is a type of neurodegenerative disease that causes muscle weakness. It is the most common form of motor neurone disease (MND) [1]. ALS is caused by progressive degeneration and loss of the motor neurones in both the brain (Upper Motor Neurones) and the spinal cord (Lower Motor Neurones). Symptoms include weakness of muscles that control limbs, speech, and swallowing. If the disease starts with difficulties in speech and swallowing, which are also known as bulbar features, the disease is called bulbar onset ALS [2]. The vast majority of the patients (around 80%) develop bulbar involvement, and they experience difficulties in swallowing which can cause aspiration. As a result bulbar involvement is associated with a poor prognosis. There are few tests available for the detection of bulbar involvement, either at the early or later stages of the disease [3], however there is still a need for a reliable biomarker for diagnosis and prognosis. With a reliable clinical device for detecting the disease, intervention is possible for drug trials and improving patient's quality of life.

The standard clinical investigation for ALS is needle electromyography (EMG) [4, 5]. This test is invasive and sometimes painful, particularly when examining the tongue, which is the most often tested bulbar muscle with EMG [6]. Recent advances with Electrical Impedance Myography (EIM), pioneered by Dr. Seward Rutkove have

shown that the non-invasive Electrical Impedance (EI) devices can be used to detect the disease [7, 8, 9, 10]. Numerous studies using EIM have been carried out in order to determine the disease's association to age and gender [11], muscle anisotropy [12], and the relevant frequencies [13, 14]. However, further investigation is required to detect the bulbar involvement and its progression using EIM.

This chapter begins with the motivation of the project, followed by aims and objectives in Section 1.2. The overview of the thesis outlines the rest of the chapters and the publication with relevance to this thesis is given in Section 1.4.

1.1 Motivation

The EIM technique consists of applying high frequency and low intensity Alternating Current (AC) with two electrodes ($\approx 5\mu A$) and measuring the output voltage with the same [15] or two other electrodes [16]. The data obtained from EIM devices is in complex numbers form, which represents the resistive and reactive characteristics of the tissue. The studies generally employ either resistance, reactance or the phase angle of the EI readings, to investigate the disease affect on the tissue composition [17, 18]. The phase angle is a product of the relation between the resistance and reactance. Most of the analysis confirmed the effectiveness of the phase angle in detecting disease-related patterns [13, 19, 20, 21].

Diagnosis [20, 14, 22] or longitudinal studies [19] using single-frequency measurements have returned reproducible results and show the technique's capability to detect disease measurements for clinical trials. Focus on a single-frequency limits the EIM's capacity to investigate the disease-related frequencies and spectral patterns. Moreover, the methods to compare the diseased tissue measurements to the healthy ones are simple statistical methods with mean and confidence interval [19, 17, 20], coefficient of variation [8] or the coefficient of determination [18] which may be in-

sufficient to find disease-specific patterns in data.

Different electrode settings and positioning are employed for measurements of both limbs and tongue [10, 8, 23, 16, 24] to investigate the disease effect on the anisotropic characteristics of the muscles [25]. Studies have shown that surface measurements (2-dimensional) are useful [26, 27, 28, 8, 29], however, due to the complex muscle structure of the tongue, it is essential to take measurements in cross-sectional plane (3-dimensional) to fully understand the disease's effect.

A non-invasive EIM method facilitates taking measurements quickly without pain or discomfort to the patients with the capacity to detect the muscle deterioration on both limb [10, 8] and tongue muscles [26, 27, 30, 25]. The technique is presented as a potential biomarker for ALS research, a disease for which there are still no validated biomarkers of bulbar onset. Quantitative assessment of the bulbar dysfunction with the EIM technique would help improve the diagnosis for the detection of the disease with clinical tests. Early identification could promote proactive clinical care rather than reactive care once the symptoms are established. EIM devices could serve as a longitudinal clinical assessment tool for drug trials and development. The hand-held, non-invasive EIM device developed by a group of researchers in Sheffield records measurements of the tongue with the 2-dimensional (2D) and 3-dimensional (3D) arrangement of the electrodes along the frequency spectrum. Analysis of new devices requires careful selection of methods. This thesis aimed at investigating the multi-dimensional data using machine learning methods to find disease-specific patterns for diagnosis and longitudinal analysis. It is also aimed to model the EI characteristics of the tissues using system identification methods.

1.2 Aims and Objectives

This thesis aims to find disease-specific patterns using machine learning and system identification methods for clinical diagnosis and longitudinal analysis of ALS disease with EIM data. According to the aims of the framework, the objectives of the project are given as follows:

- **Data Driven Disease Diagnosis:** Early diagnosis of diseases is a fundamental research problem in both engineering and medical fields. Classification is a commonly used machine learning method for diagnostic purposes [31, 32]. Supervised learning is a type of classification used to differentiate diseased and healthy individuals from a data set with known labels. The Sheffield device has multiple electrode configuration settings and takes measurements along the frequency spectrum to obtain the multi-dimensional characteristics of the diseased tissue. The aim is to find disease-related frequencies and electrode settings from the acquired data by using supervised classification algorithms together with feature selection method.
- **Data Driven Modeling:** The measurements are taken non-invasively from the surface of the tongue. The epithelial layer on both surfaces, muscle, and connective tissue between the muscle layers have an impact on EI readings. The device uncertainties, together with the different layers of tissues, make it challenging to model the system at a cellular level. The EIM measurements are considered as an input-output black box system (see Section 4.2) in order to build a mathematical model of the data over the spectra. The model is expected to give a spectral description of the underlying conditions. Classification using the model parameters would also help understand the model parameters role on the disease diagnosis.

- **Longitudinal Analysis for Disease Progression:** Disease progression has been a difficult problem for ALS studies due to the lack of a proper and convenient clinical prognostics biomarker. In this project, the patients have undergone the clinical measurements every three months for a longitudinal assessment of the disease. One challenge faced while studying longitudinal analysis is that the disease affects patients in different pace. Another challenge comes from the intervention for reducing the saliva effect [33]. The frequencies (features) with the most discriminatory information will be used in the analysis to find their role in longitudinal studies

1.3 Overview of the Thesis

The structure of this thesis is summarised as follows:

- **Chapter 2:** *Electrical Impedance Data* starts with brief information of the studies that gave rise to the Electrical Impedance technique. The type of data obtained from the devices is given with the introduction of the circuitry modeling of the cells. The EIM devices and related studies are presented as part of the literature review. The Sheffield EIM device, the data acquisition process, and patient groups used in this project is introduced in the following section. Data preprocessing of the raw EIM data is then given at the end of the chapter with an outlier detection algorithm developed for this data.
- **Chapter 3:** *Non-Parametric Data Classification For Disease Diagnosis And Longitudinal Analysis* implements the supervised classification for disease diagnosis. Different dimensionality reduction methods are focused upon, feature extraction and feature selection. The feature selection method together with Pareto Dominance aimed at finding the best performing frequency combinations. Longitudinal analysis is carried out with the selected features (frequencies)

cies) for each electrode configuration.

- **Chapter 4:** *Data Driven Modeling of EIM Measurements* introduces the integral fractional order transfer function modeling of the EIM data.
- **Chapter 5:** *Parameter Estimation and Model Selection* presents the statistical and non-statistical algorithms used to estimate the parameters of the model. A model space from simplest to most complicated models is created, and a model selection scheme is applied based on the results of one of the estimation method.
- **Chapter 6:** *Parametric Data Classification For Disease Diagnosis And Longitudinal Analysis* repeats the classification using the EIM model parameters as features and compares the results to the non-parametric data classification. Longitudinal analysis is completed using the model parameters.
- **Chapter 7:** *Conclusion and Future Work* gives an essence of the chapters and makes conclusions based on the results. A future direction to the research problems is also presented.

1.4 Publications

J. J. Alix, H. E. McDonough, B. Sonbas, S. J. French, D. G. Rao, V. Kadiramanathan, C. J. McDermott, T. J. Healey, and P. J. Shaw, "Multi-dimensional electrical impedance myography of the tongue as a potential biomarker for amyotrophic lateral sclerosis," *Clinical Neurophysiology*, vol. 131, no. 4, pp. 799–808, 2020.

Chapter 2

Electrical Impedance Data

Electrical Impedance (EI) is widely used measurement to assess biological tissues, and a variety of techniques have been developed such as EI Tomography, EI Spectroscopy, EI Myography, EI Mammography, EI Cardiography for measuring the EI data from the heart, muscles, brain, lungs, liver, for diagnosis and prognosis of diseases from cancer to heart conditions [34, 35, 36, 37, 38, 39]. One advantage that comes from using the EI is that the electrodes are connected to the tissue surface non-invasively. The measurements are in complex impedance form (see Section 2.2) over a range of frequencies. The use of the data in analysis with statistical, machine learning, and system identification methods help finding hidden patterns to reveal information related to diseases. The focus of the electrical impedance data analysis in this project is obtained from the tongue muscle of patients with ALS and healthy volunteers.

This chapter provides an overview of the key concepts of EI data, literature on EIM for ALS disease and details of the EIM data used in this research. Section 2.1 introduces the early concepts of electrical impedance data in the analysis of biological tissues. Section 2.2 covers electrical circuit based modeling which is an essential aspect of the electrical impedance studies. These techniques form the basis for developing the EIM devices as a biomarker for detecting ALS disease. Section 2.3 is a review on

ALS clinical testing, EIM data obtained from diseased muscles, frequencies, disease relation to age-gender and the longitudinal studies. Section 2.4 introduces the Multi-dimensional EIM Device, developed through a collaboration between the University of Sheffield and Sheffield Teaching Hospitals NHS Foundation Trust. Section 2.5 is dedicated to the preliminary analysis of EIM data with a data-specific outlier detection algorithm.

2.1 Bioelectrical Impedance

Electricity has been used in biomedical research for hundreds of years with early documented records going back to late 18th century with Galvani's frog test [40, 41]. It was followed by Vigouroux's studies to relate the electrical impedance measurements to diseases in the late 19th century [42]. In modern day biomedical studies, the use of electrical impedance varies from the analysis of blood cells [43] to the diagnosis of disease using single or multi-frequency EI measurements [34, 35, 36, 37, 38, 39].

Nyboer, in 1959 discovered that it is possible to use the volumetric relationship of resistance and conductance of the body to estimate the total body water (TBW). He used a 50 kHz frequency for his analysis [44, 45]. However, it was pointed out later that the single-frequency measurements are not sufficient for accurate estimation due to the non-uniformity of the human body. Thomasset in 1962 worked on this problem and expanded single frequency studies into multi-frequency level [46, 45]. He discovered the extracellular fluid characteristics at low frequencies. Following studies have shown that once the frequency is high enough to penetrate the cell membranes, it is possible to differentiate the extracellular and intracellular fluids [45]. These results lead to examining the tissues at the cellular and organelle level. The complex impedance is characterized by the cell and organelle membrane while real impedance is characterized by the extracellular and intracellular fluids of cells

and organelles. The EI measurement analysis uses this relation to investigate the underlying reasons for the deformation of the cell mass and fluids.

The simple idea behind the bioelectrical impedance is the human body is composed of 55–60% fluids that give the resistance characteristics and 40-45% solids that give the reactance characteristics to the impedance measurements. At a cellular level, the resistance behavior can be rationalized with the resistivity of the fluid to the flow of electrical current, and the reactance behavior can be rationalized with the shape of the cell membrane, which has two layers called the lipid bilayers. These layers look like the plates of a capacitor; hence the membranes return the capacitive reactance in the presence of an Alternating Current (AC). Capacitance causes the current to be shifted in time and results in a phase difference. The phase is expected to vary between 0° to -90° in an electrical impedance with capacitive reactance. Some studies employed the phase value [14, 47, 27] or resistance and reactance values [15, 48] or both complex components [28, 49, 50, 25], depending on the nature of the research.

2.2 Circuitry Modeling of the Bioelectrical Impedance

Data

Human body acts like an electrical circuit; when an alternating electrical current is applied, it returns a sinusoidal voltage reading. According to the Ohm's Law the proportion of the output voltage amplitude at a certain frequency (V) to the current at a certain frequency (I) returns a complex impedance reading when the signals are represented in the complex form arising from the phase shift.

$$Z = \frac{|V|e^{j(\omega t + \Theta_V)}}{|I|e^{j(\omega t + \Theta_I)}} (\Omega) = \frac{|V|}{|I|} e^{j(\Theta_V - \Theta_I)} (\Omega) \quad (2.2.1)$$

where $\Theta = \Theta_V - \Theta_I$. The real number of the impedance comes from the resistance (R) and the imaginary number comes from the capacitive reactance (X);

$$R = R/\underline{0^\circ} \quad (2.2.2)$$

$$X = \frac{1}{j\omega C} = \frac{1}{\omega C}/\underline{\Theta^\circ} \quad (2.2.3)$$

where C represents capacitance (F) and R represents resistance (Ω). The reactance readings of the biological tissues comes only from the capacitive reactance since the effect of the inductive reactance is infinitesimal compared to the capacitive reactance [51], therefore, for future references, the reactance refers to the capacitive reactance. Impedance representation can be both used in cartesian and polar coordinates with complex forms.

$$Z = R + jX \quad (2.2.4)$$

$$|Z|/\underline{\Theta} = |Z|e^{j\Theta} \quad (2.2.5)$$

Z , R , X , $|Z|$, $\underline{\Theta}$ are commonly used symbols for impedance, resistance, reactance, magnitude and phase respectively. For the Cartesian coordinate representation, the real unit resistance is placed on the x-axes and the imaginary unit reactance is placed on the y-axes. With Pythagoras theorem, the impedance magnitude and phase are calculated as follows:

$$|Z| = \sqrt{R^2 + X^2} \quad (2.2.6)$$

$$\Theta = \tan^{-1} \frac{X}{R} \quad (2.2.7)$$

These basic laws and equations form the basis of the EIM data analysis. As mentioned earlier, the human body acts like an electrical circuit, and cells, cell membranes, cellular fluids are the components of the electrical circuit. In the literature,

the cellular level electrical circuit representation of the impedance readings is given in two forms; simple 3-Element model [52] and a more advanced 5-Element model [53]. The 3-Element model captures the cell membrane, intracellular and extracellular fluids, and in the 5-Element model, in addition to these, the model captures the organelle fluids and organelle membranes. Circuitry representation with comparison to the cell compartments illustrated in Figure 2.1;

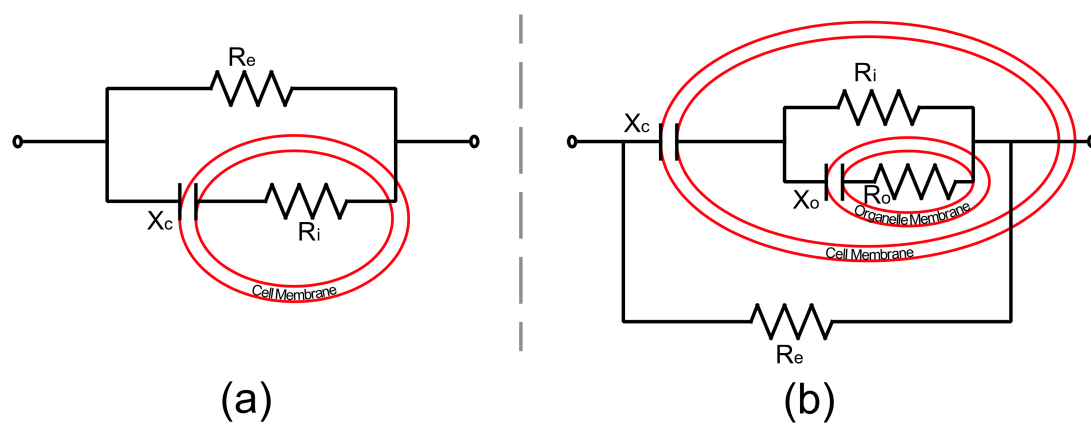


Figure 2.1: Simple 3-Element [52] and advanced 5-Element [53] cellular circuit model. R_e is the resistance from extracellular fluids, R_i is the resistance from intracellular fluids, R_o is the resistance from organelle fluids, X_c is the reactance from cell membrane, X_o is the reactance from organelle membrane.

These circuitry elements were useful in researches related to muscular disease with data obtained by the EIM devices [53, 54, 55] on different muscle group of interest. Studies also employed the Cole-Cole equations and Cole-Cole (Nyquist) plots that were introduced by K.S. Cole and R.H. Cole in 1941 [56] to interpret the impedance values of the tissues over the frequency spectrum [55, 57, 58, 59]. The plot is a product of Cole-Cole equations, and data points over the frequency spectrum are plotted on a semi-circle on the Cartesian coordinate. The real values (resistance) given in X-coordinate, and the imaginary values (reactance) is given in Y-coordinate. The

frequency components converted into the logarithmic scale; are the points over the semi-circle.

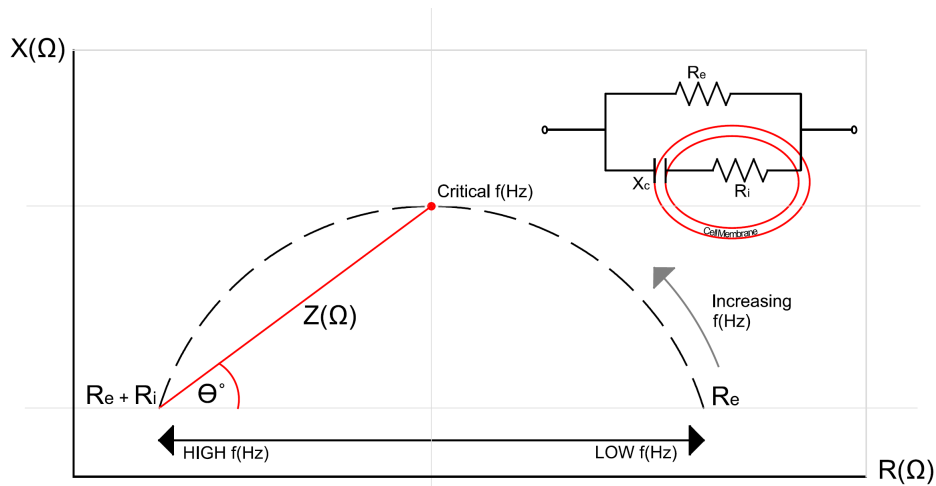


Figure 2.2: Electrical Impedance Spectrum visualisation with Cole-Cole plot (Nyquist plot) in complex domain [52]

The plots show the resistance and reactance changes along the frequency spectrum. It is assumed that at low frequencies, the current cannot go through the cell membrane and travels through the low resistivity path, which is the extracellular fluids between the cells and therefore returns high resistance readings [60, 44, 45]. As the frequency increases, the alternating current (AC) penetrates through the cell membranes, and the cell membranes act like a capacitor with its bilipid cell membranes [61, 62, 63, 45]. When the opposition to the AC is increased with the capacitive effect, a delay in time for the output voltage creates the phase angle (See Equation 2.2.3). The peak of the semi-circle in the Cole-Cole plot is called the critical frequency, where the reactance value reaches its highest value. The capacitance effect of the cells reduces as the frequency increases after the peak point, because the AC at high frequencies pierces through the cells. At this point, the resistance increases further due to the combination of intracellular and extracellular fluids.

Studies have employed the Cole-Cole plots to represent the electrical impedance over the frequency spectrum [64, 65, 66]. Some examples of these studies include disease-specific modeling with electrical impedance measurements [53], analysis of impedance characteristics of tissues [67, 68, 69], and progression analysis of the disease over time [70].

2.3 Electrical Impedance Myography

Bioelectrical impedance principles presented in Section 2.1 form the basis of electrical impedance technology. From single frequency [71] or painful needle electrodes [72] to advanced imaging devices [73], the electrical impedance technique is used to study neuromuscular diseases [74, 75, 76, 36]. EIM is the name of the technique that evaluates muscle or groups of muscles at a frequency spectrum. It can also be referred to as Electrical Impedance Spectroscopy in the sense that it is the EI that is applied to the muscles along the spectrum. It is widely used in studies for the diagnosis and assessment of the progression of neuromuscular diseases. In these types of diseases, the degeneration in the neurons that controls the motor function of the body results in muscle weaknesses [77]. Diagnosis and progression assessment is of great importance for early intervention and development of treatment for these diseases.

The EIM devices became popular after it was released as a potential biomarker for amyotrophic lateral sclerosis (ALS) disease, pioneered by Dr. Seward Rutkove from Harvard University [7, 78, 79, 19]. This technique has the potential for being an alternative to needle electromyography devices with its capacity to perform measurements on patients non-invasively [80]. There are different probe designs particular to the muscle group it is applied to, such as measurement of the muscle groups for bulbar involvement [30], and limb weaknesses [81]. The two electrode non-invasive

design [15] has shown promise for capturing the electrical impedance characteristics of the muscle. However, because the two electrodes are used both for applying the AC and measuring the output voltage, measurement errors are inevitable [82]. Four electrodes (also known as tetrapolar) probes overcome these errors by applying the AC with two electrodes and measuring the voltage with the other two electrodes across the muscle or muscle groups. The tetrapolar design avoids polarization effects and skin contact error [82]. Four electrode designs inspired the flexible [47] and rotational [83] electrode configurations that aim to deal with the anisotropy limitations of muscle measurements. Electrode sizes and array configuration plays an essential role in impedance measurements [84, 85]. Measurement of the small muscle groups highlighted the importance of electrode probes sizes [28]. A proposal was made for the regulation of electrodes [86] to improve the reliability of the technique as a biomarker and compare EIM studies.

2.3.1 ALS Biomarkers

The delay in diagnosis may cause a shorter life span or hardship to the patient due to lack of interventions to improve the quality of life and life span. Therefore, studies are focused on clinical diagnostic tests that are reliable and easy to perform. The El Escorial Revised [87] and the Awaji-Shima Criteria [88] are clinical tools to standardize the inclusion criteria, with former being related to the clinical examinations and latter being related to both clinical and electromyography (EMG) examinations. As the name suggests, EMG devices measure either the skeletal muscle's electrical activity with surface electrodes (non-invasive) [89] or the nerve that controls the particular muscle with needles (invasive) [90]. Magnetic Resonance Imaging (MRI) is used not as a diagnostic test but as a test to rule out other conditions [3], so that the diagnosis is focused more on the ALS disease. The clinician may advise muscle biopsy or blood and urine tests for eliminating diseases varying from myasthenia gravis (MG)

to spinal muscular dystrophy [3, 91, 92]. Patients' medical records and family history are taken as part of the clinical assessment in order to consider the environmental and genetic factors [93]. Respiratory tests may be necessary to check the bulbar involvement [94]. Tongue strength assessment [95] is another clinical test; however, it is inaccurate due to its dependency on patient effort. The ALS functional rating scale-revised (ALSF_{RS}-R) might not reflect the underlying progression of disease due to medical interventions that ease the disease effects [33, 96].

2.3.2 Disease-Related Muscle Groups

There are more than 600 names for different skeletal muscles in the human body, and these muscles are made out of three different muscle fibers, that are: cardiac muscle fiber, skeletal muscle fiber, and smooth muscle fiber [97]. The limb muscles play an important role in the movement, skeletal support, and the spinal cord motor neurons that control them. The skeletal muscle fiber that exists on limbs are layers of clustered muscle fibers. These muscle fibers are called myocytes, and they are elongated multinuclear cells that vary in length from 1 mm to 30 cm. When the AC is applied to the muscle tissue, the current will return different voltage readings dependent on the frequency. The AC chooses the least resistive path along the muscle fibers rather than across the muscle fibers [98], and the dependence on the muscle fiber direction is called the anisotropic property of the muscles [99]. The mathematical modeling of the limb muscles with EI data may be simpler due to its more simplistic anisotropic characteristics [98].

The tongue muscles play an important role in swallowing, breathing, chewing, and speech, and four different cranial nerves (the hypoglossal nerve, the facial nerve, the glossopharyngeal nerve, the vagus nerve) control these factors. Tongue muscles are made of two muscle fibers; these are skeletal and smooth muscle fibers [100]. The muscles are classified with where they are originated, either within or outside

the tongue as intrinsic and extrinsic muscle fibers, respectively [101]. The intrinsic muscle fibers are layered in different directions, and they are named after their orientation, which is a superior longitudinal, inferior longitudinal, transverse, and vertical. Extrinsic muscles, unlike intrinsic muscles, are attached to the tongue, but they are originated from outside the tongue. These muscles help protruding-retracting, elevating-depressing the tongue, and they are named after where they are originated from; that are genio-glossus, hyo-glossus, stylo-glossus, and palato-glossus [100]. The muscle fibers in many parts of the tongue interweave, and this means that the anisotropic structure is more sophisticated than limb muscle groups.

EIM studies focused on different muscle groups [28, 49]. The limb muscle studies expanded on developing EIM devices with different types of electrode configurations or orientations such as the rotational EIM [99], electronically reconfigurable electrode array EIM [29] and other configurations by rotating the direction of the electrode array [47, 15, 48, 102, 103, 104] in order to investigate the anisotropic characteristics of the muscles. The motivation behind different electrode configurations is that the disease-related atrophy affects the anisotropy of the muscle [37]. Moreover, the impedance characteristics of different diseases show different anisotropic atrophies [83, 85].

ALS is known to change the tongue muscle structures and tongue thickness [105]. The bulbar involvement related studies mostly focused on the measurement of the tongue EI [26, 27, 25]. The electrodes are placed on the tongue in order to take the measurements. Studies also showed that a proper electrode array size is important to reduce the gag reflexes of the patients [26, 27]. The muscles evaluated using electrical impedance readings at 50kHz has shown the disease effects on contraction state [106]. This implies that the muscles that are active when the tongue is inside or outside of the mouth are different, and the relation between the intraoral and extraoral measurements is an important research matter that needs to be investigated.

2.3.3 EIM Frequencies

EIM devices rely on the injection of alternating current (AC) rather than the direct current (DC) to measure the impedance characteristics of the muscles in order to prevent polarization, therefore avoid tissue damages. AC reverses in direction periodically, and the waveform for biological tissues is usually a sine wave. The period in which the sine wave completes one cycle determines the frequency of the wave. This is determined by the number of cycles completed in one second. The unit is Hertz (Hz), which is equal to the cycles per second.

Research suggests that the frequency spectrum for biological tissue measurements can be divided into three bands; α , β , γ [107, 108] (see Figure 2.3). The α band is the frequencies below 100 Hz, and the γ band is the frequencies above 1GHz [107, 108]. The β band frequency readings of the tissues return impedance values with high resistance at the lower end [109]. Once the AC can penetrate through the cell membranes, impedance readings are affected by the intracellular fluids and organelles at the higher end of the β band (around 1MHz) [109]. Figure 2.3 illustrates the electrical impedance characteristics of muscle cells over the frequency range.

Early studies showed that the estimation of body composition gives the most accurate results at 50 kHz. Nyboer first proposed the hypothesis in 1959 [44, 52]. In this study, healthy skeletal muscle showed a peak in reactance values at this particular frequency. This knowledge led to studying different muscle groups with EIM at 50kHz [14, 49, 55, 74, 98]. Later studies on tongue muscle suggested that the phase values at 100kHz also show promise for the diagnosis of ALS with EIM [30]. High reactance readings are captured at the critical frequency (See Figure 2.2) at the β band where the reactance readings are maximal due to AC penetrating the cell membranes. The critical frequency may be different for diseased patients and volunteer subjects, considering the nature of the disease. With this in mind, ALS data analysis

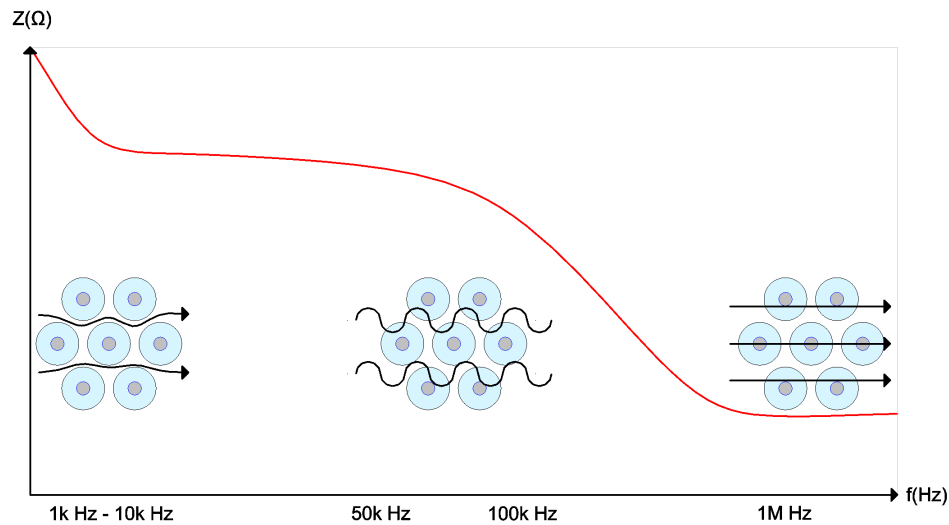


Figure 2.3: Electrical Impedance characteristics of biological tissues along the frequency spectrum with representative visualisation of cell-frequency relation [110]. Important frequencies for ALS studies (50kHz, 100kHz) and frequencies that gives the relation in cell figures (1kHz, 10kHz, 1MHz) are noted on x-axis and Impedance readings are on y-axis.

with multi-frequency EIM devices may capture valuable patterns on the spectral data rather than relying on a single frequency measurements [37, 111]. Another study on the contracted and relaxed skeletal muscle analysis with multi-frequency EIM technique [50] demonstrated that the biochemical activity of diseased muscle (as a result of force generated in the contracted muscle) is different to the healthy muscle. Results show that the relaxed muscle phase is lower than the contracted muscle phase at the frequency spectrum up to 2.5 kHz. At 2.5 kHz, both the relaxed and contracted muscle phase drops down to a minimum.

2.3.4 Age and Gender-Related Studies

ALS is categorized dependent on the background of the disease; whether it is genetic (familial) or not (sporadic). Depending on the background, the disease age-onset range is different, with 58 to 63 for sporadic and 47 to 52 is familial [112]. According to the statistics of ALS research from Europe, the disease is also gender-

biased, with most cases occurring in males [113]. These statistics suggest that the disease may be age and gender-biased.

A study has shown that the EIM mean phase readings of the limb muscles have dropped for older patients [11]. The study's reliability is questionable since the muscle composition of the patients in their youth is unknown. Another study compared two age groups: 19-50 and 60-85; mean phase readings did not show a difference between the groups; however, when looking at the resistance and reactance readings, a difference was more noticeable [114]. The same study also showed that the resistance and reactance readings revealed information towards gender differences. A comprehensive multi-frequency EIM study for different diseases that affects the muscle aimed at finding gender and age-related patterns, considering all impedance elements (resistance, reactance, magnitude, and phase) [111]. The results have shown random age and gender related changes for the healthy and diseased subjects using resistance, reactance and magnitude values, however, the changes were more visible using phase values. Another 50kHz EIM study on the tongue showed that the patients have smaller phase values and higher resistance values [8]. In this study, it is shown that age and gender factors did not play an important role in tongue measurements. Shellikeri et al. (2015) [26] inferred that age-related differences could be detected only over the age of 80 years old for bulbar involvement. With the controversies among the studies, it is not easy to conclude what effects age and gender have on the disease and whether or not it could be detected with EIM devices.

2.3.5 Longitudinal Analysis

The life span of the patients from the diagnosis date to the death varies depending on the bulbar involvement. Diagnosis to death time could be anywhere between 20 months to 4 years, and around 10% of the patients survive up to 10 years [115, 3, 116, 117, 118]. In extreme cases, patients can live up to 55 years which is a known

case of cosmologist Stephen Hawking, whom was diagnosed in 1963 [119]. With this variation in the life span, it is difficult to infer results from the longitudinal data for prognostics. ALS Functional Rating Scale-Revised (ALSFRS-R) [120, 121, 122, 123] is a 12-item questionnaire used by clinicians to assess disease progression, but it could be a biased assessment due to dependence on the patient's answers. Therefore, it is important to employ devices in clinical tests that can measure physical data and estimate the progression of the disease.

There are studies that observe the disease over a period of time in order to detect disease related changes [118, 124, 125, 126, 127]. A longitudinal study conducted with an EIM device for ALS compared the phase values from six and twelve months after diagnosis [8]. The study recorded a decrease in the mean phase and an increase in the steepness of phase slope as the disease progresses.

Production of affordable devices that are suitable for the use in inexperienced carers is important in longitudinal analysis. These devices may help anticipate the right time to intervene with respiratory and feeding devices in order to extend the patient's life as well as improve the quality of life. Collecting data for the progression analysis may also help improve the clinical testings for ALS sub-groups.

2.4 Multi-dimensional Electrical Impedance Myography Device

The EIM devices with different electrode sizes and orientation that uses different frequency spectrums, to study different muscle groups, showed that the EI is a reliable measure for identifying the muscular diseases, including ALS [111, 36, 30]. The optimal frequency-electrode combination is still unknown. Most studies focused on the limbs, but the disease is fatal when it is bulbar onset and once it paralyses the muscles that make you breath. The device design focuses on capturing the tongue's

EI characteristics from the surface measurements. However, the tongue's anisotropic characteristics have the most complicated muscle fiber arrangement, and the muscle fibers are mashed in longitudinal, transverse, and vertical directions [128, 129]. The previous studies of the tongue with 2D surface electrodes may not be sensitive enough to detect ALS bulbar involvement due to the anisotropic muscle structures, that would require more comprehensive measurements. Therefore the tongue organ requires measurement with more sophisticated electrode designs than those surface electrodes.

A group in Sheffield University Teaching Hospital designed a novel EIM device with the motivation to create an EIM device that is user and patient-friendly, and that can detect the ALS bulbar involvement, whether it is a bulbar onset or once the disease spread to the bulbar muscles. Dr. James J. Alix, a National Institute for Health Research clinical lecturer in neurophysiology and specialty registrar in clinical neurophysiology, coined the idea and Dr. Jamie Healey, a senior medical physicist with many years of experience in the development of impedance devices, designed the device. This comprehensive device has all the features that the EIM tongue research was interested in: that are multi-frequency measurements, multi-dimensional electrode configurations to investigate the disease effect on the anisotropic characteristics of the muscles, and data collection for longitudinal analysis. The device is patient-friendly, in that it reduced the gag reflexes caused by the big electrodes, and it is user friendly with high interrater and intrarater reliability [130].

The device is assembled with two hinged arms. A stop prevents the excess compression on the tongue and maintains the inter-electrode distances. At the end of each arm, there is an array with four gold electrodes. Electrodes are placed at 5mm distances with each other to form the corners of a cube with eight electrodes on two arms. The tetrapolar measurement idea is adopted in order to avoid electrode polarization. The two electrodes inject the AC at $5\mu A$, and the other two electrodes read

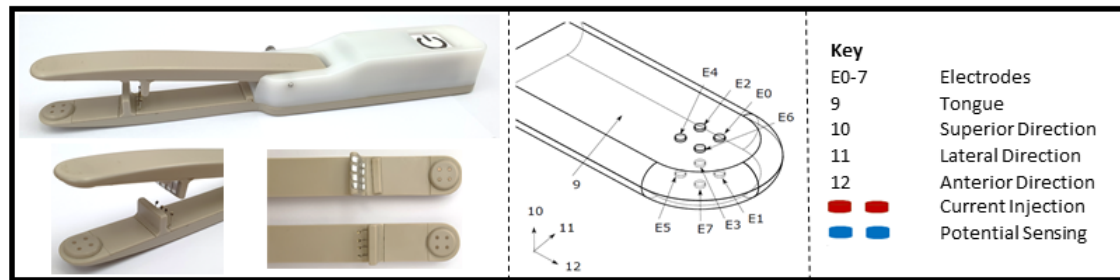






Figure 2.4: Design of device for measuring tongue impedance. Four electrodes placed on the upper and lower plates. Current injected from the two red electrodes and the voltage measured from the other two blue electrodes (see Table 2.1). The figures are plotted by Dr. Jamie Healey.

the output voltage across the fourteen frequencies while the other four electrodes are inactive. The injected current is below the threshold of electricity sensation; which makes the device non-invasive and painless. There are four settings of electrode configuration to obtain the surface measurements (2-dimensional (2D) measurements), and eight settings of electrode configuration to obtain cross-sectional measurements (3-dimensional (3D) measurements). The 3D measurements are designed to capture the anisotropic characteristics of the disease. This novel set is hypothesized to reveal more information about the disease effect on muscle EI characteristics.

The electrode array is handheld, lightweight, and wireless. At the end of each measurement recording, the impedance readings for each electrode configuration is visually inspected on custom software. This inspection enabled researchers to discard impractical data as a result of unwanted circumstances such as patient movement during measurements. The measurements were then repeated for reliability analysis. Measurements are downloaded into the computer via Bluetooth and sorted into a .csv file. EI readings in the complex form are converted to polar readings (see Equations 2.2.6, 2.2.7) with custom software. Test-retest reliability of 3D and 2D electrode configurations intraclass correlation (0.86 and 0.81 respectively), was consistent across visits for patients and volunteers [130]. The inter and intra-rater reliabili-

Electrode Names	TPA (2D)	TML (2D)	BPA (2D)	BML (2D)
Electrode Figures				
Current Injection	E2-E0	E6-E0	E3-E1	E7-E1
Voltage Sensing	E4-E6	E4-E2	E5-E7	E5-E3









Electrode Names	LPA (3D)	LIS (3D)	RPA (3D)	RIS (3D)	II (3D)	IM (3D)	SS (3D)	SM (3D)
Electrode Figures								
Current Injection	E4-E6	E6-E7	E2-E0	E0-E1	E4-E7	E4-E2	E5-E6	E5-E3
Voltage Sensing	E5-E7	E4-E5	E3-E1	E2-E3	E2-E1	E7-E1	E3-E0	E6-E0

Table 2.1: 2-dimensional (2D) and 3-dimensional (3D) electrode configurations on representative tongue figure with current injection electrodes (red) and voltage sensing electrodes (blue) specified. 2D current electrodes and voltage sensing electrodes placed on same surface. 3D current electrodes and voltage sensing electrodes placed on opposing surfaces. Electrode names are initialized (please see Section 2.4.3). The electrode keys are obtained from Figure 2.4. The tongue figures are plotted by Dr. Jamie Healey.

ties (with intraclass correlation of 0.788 and 0.854 respectively) proved that the device is reliable for both experienced and inexperienced hands [130].

The device measures the EI readings in 14 frequencies, starting with $\approx 76Hz$ that is in the α region, and the frequencies follow a doubling sequence with $\approx 76Hz * 2^j$ where $j \in \{0, 1, 2, \dots, 13\}$. The frequencies from $\approx 152Hz$ to $625kHz$ are in the β region. The data in raw form can reveal important frequency or frequency groups for the disease with machine learning methods. Modeling the data in full frequency spectrum may help explore the frequencies between $\approx 76Hz$ to $625kHz$. This spectrum would help investigate previously researched single and multi frequency bands.

2.4.1 Tongue Measurements

The bespoke handheld device ensures a fixed distance between the plates in order to measure a similar volume of tongue for each electrode configuration [130]. It does not cause gag reflexes to the patients with its small electrode array arrangements. In order to investigate the tongue EI characteristics with all perspectives from the literature, a comprehensive measurement setting is designed to measure different tongue placements for 2D and 3D electrode configurations on the multi-frequency device.

The measurement procedure starts with activating the handheld device with an on/off button. The device is then linked to the computer to record the measurements. Calibration and cleaning was ensured before each recording in order to avoid measurement errors. Once the device is placed on the tongue, a $\approx 5\mu A$ current is injected via the excitation electrodes. The device has been safety tested with standard international protocols for medical devices (ISO 13485) and with British Standards Institution guidelines (N606012-10). Different muscle groups are used when the tongue is protruded and retracted. Therefore patients were asked to sit upright and first protrude their tongue. In this position, the electrodes are placed on the centre, left and right sides of the tongue subsequently. The sensor configuration automatically activates to measure all twelve directions at one contact. The device beeped after each measurement is finished (about 10 to 30 seconds) to inform the observer to change the position of the device on the tongue. The same measurements are repeated when the tongue was retracted. This is equal to the recording of the each twelve electrode configurations six times which means measuring same muscle group in 72 different aspects. The measurements are repeated once more with the same observer or a different observer to calculate the intra or inter-rater reliability of the device. For each electrode configuration, each frequency is measured

eight times, and this is repeated until the standard deviation of the last set of eight measurements is less than 10% of the mean value of the eight measurements.

2.4.2 Patients and Data Collection

Patients were selected for data collection from the Sheffield ALS clinic between years 2015 and 2018. The inclusion criteria comprise the patients with a possible, probable, or definite ALS from the Ajwani Criteria [88], or patients with bulbar pathology from EMG examinations. Patients or volunteers with oral infection and diseases, tongue injury, tongue EMG appointment within a week, implanted medical electrical devices were excluded. The medical records of the patients include information regarding gender, age, site of disease onset, duration of symptoms, level of diagnostic certainty, ALS functional rating scale (ALSFR-S), tongue strength testing, clinical saliva score, and speech tests. ALS patients were assessed at three-month intervals for a period of up to a twelve months duration. Volunteers aged 40 or older were recruited from a local advertisement given in the Department of Clinical Neurophysiology, Royal Hallamshire Hospital, and the Sheffield Institute for Translational Neuroscience. Healthy patients were assessed twice at six-month intervals.

The study was approved by the National Research Ethics Service Committee Yorkshire & The Humber- South Yorkshire, and was approved by STH Trust Research and Development (Research Ethics Committee reference: 15/YH/0121, Integrated Research Application System project ID: 160983). Informed consent was obtained from the participants before recordings were undertaken. All the subjects were given the right to withdraw from the studies at any time. Each subject was given a code and the collected data was stored electronically and on paper with subject information kept in confidentiality.

Forty-one ALS patients (22 men, 19 women) and thirty healthy volunteers (14 men, 16 women; mean age 56) were recruited. The age criteria were compatible with

the statistics obtained from the ALS patient population in order to match the ages of the patients and volunteers to avoid any conflict of results [130]. At baseline, three patients were unable to perform the EIM recordings due to the severity of their disease. Seventeen patients had the bulbar-onset disease, twenty-four had the limb-onset disease.

2.4.3 Multi-dimensional Electrical Impedance Myography Data

The multi-dimensional data with 2D and 3D electrode configuration settings and different tongue placement is recorded into a .csv data file extension from the device to the computer via bluetooth. Python 2.7 and 3.4 programming language with Pandas library is used to read the data file.

Thirty-eight patients and thirty volunteers completed twelve electrode configurations with six tongue placements at fourteen frequencies within three months and six months of intervals. However, some recordings were not completed or not recorded due to errors in measurements. Each patients and volunteers completed different number of visits (varying from 1 to 6) in each electrode configuration. The full names of the electrode configurations come from the electrodes used in the configurations. The electrode configurations and the tongue placements are initialized as follows;

Tongue Location: Center = C, Left = L, Right = R

Tongue Position: Intra-Oral = I, Extra-Oral = E

2D Electrode Configurations:

- Top-Posterior-Anterior = TPA
- Top-Medial-Lateral = TML
- Bottom-Posterior-Anterior = BPA
- Bottom-Medial-Lateral = BML

3D Electrode Configurations:

- Left-Posterior-Anterior = LPA
- Left-Inferior-Superior = LIS
- Right-Posterior-Anterior = RPA
- Right-Inferior-Superior = RIS
- Infant-Infant = II
- Infant-Medial-Lateral = IM
- Supant-Supant = SS
- Supant-Medial-Lateral = SM

2.5 Data Preprocessing

Statistics are essential in biomedical data analysis because they help to explain the variability and randomness in the data. It also helps with the decision making of new data within a confidence interval. In medical studies, statistics are used to differentiate between healthy and patient groups [19, 17, 20], used to ascertain how effective a medication is against disease [131, 132], used to investigate the life expectancy of different genders with underlying conditions and many other researches [113, 133, 134, 135].

It is important to have a visual inspection of the data before applying a statistical method. The data is in the frequency domain. The seventy-two clinical tests for different patients can be pooled in order to study the statistical properties of the frequencies. Below, the box and whisker plot of data along the frequencies for real and imaginary impedance is given. This illustration contains information of the frequency components in terms of quartiles. Quartiles, as the names indicate, are the 25% intervals of the data. The first quartile is the lower quartile data up to 25%; the second quartile is the 50% of the data, the middle point of the distribution (also called

the median), and the third quartile is the upper quartile 75% of the data. The same method can be applied with the percentile statistics by setting the lower limit and upper limit to the confidence intervals with percentile values, where the 50% value is the median of the distribution.

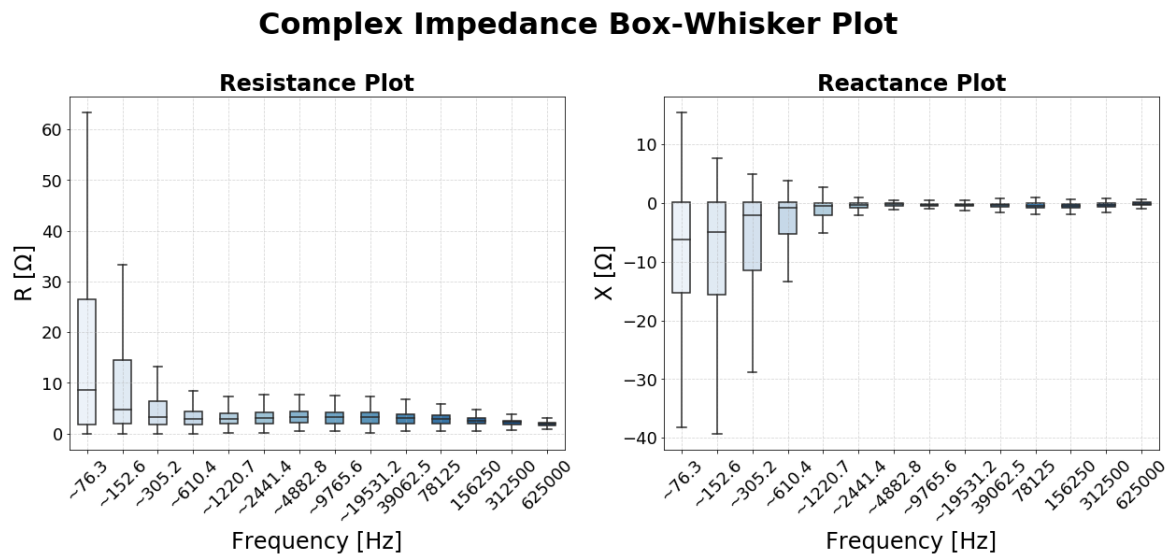


Figure 2.5: Box-Whisker plot for complex impedance along the frequency spectrum. The box shows the interquartile range with lower quartile (25%), upper quartile (75%) and the median (50%). Whiskers show the minimum and maximum points of the data and without the outliers. The plots proves the data distribution is skewed.

There are four moments in descriptive statistics. The first moment is called the mean that helps to understand the average outcome of a measurement. The second moment called the variance, which shows how much the data is spread. The third moment is called the skewness that helps to understand the distribution type of the data, and the fourth is called kurtosis [136]. If a data is normally distributed, then the normal distribution statistics with mean μ (Equation (2.5.1)) and standard deviation σ (Equation (2.5.2)) around the mean ($\mu \pm \sigma$) would be suitable. The box and whisker plots with the quartiles indicate the existence of the skewness of the distribution, therefore the median (50% percentile) and median absolute deviation (MAD) (Equa-

tion (2.5.3)) around the median value ($M \pm MAD$) [137] is a more robust statistical measure for representing the data. For a univariate data set: $X_i = X_1, X_2, \dots, X_n$;

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (2.5.1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n - 1}} \quad (2.5.2)$$

$$MAD = \text{median}(|X_i - \text{median}(X)|) \quad (2.5.3)$$

2.5.1 Outlier Detection

EIM data collected and stored into the .csv file from the device is in raw form. This means the saved data may contain some unwanted recordings or data that does not fit within the confidence intervals of the distribution of the whole population. Removal of the unwanted and outlier data preliminary data analysis is essential. Visual inspection of the raw data from the whole population by graphical representations may be informative. Once the unwanted data and outliers that conceals the true patterns are removed, the remaining data will be investigated further for finding discriminatory information of the healthy volunteers and patients.

The data is stored in complex impedance form. Prior to outlier removal, if a data set with negative resistance readings at any frequency is detected, the entire spectrum data is removed from the analysis. This is the first step in the outlier removal algorithm. The negative resistance is considered erroneous due to the biological nature of the data. There are two recordings of most measurements. The outlier and unwanted data removal are applied to the whole dataset; however, only the first available run is used in future analysis. If the initial run was removed from the analysis by the unwanted and outlier removal algorithm, then the second available run was used.

Algorithm 1 Unwanted and outlier data removal along the frequency spectrum

Input: Data set $D : D^i(f_l) = \{R^i(f_l), Z^i(f_l)\}$, for measurements $i, i \in \{1, 2, \dots, n\}$ and frequencies $f_l, l \in \{1, 2, \dots, 14\}$, with $R^i(f_l) = [R^i(f_1), R^i(f_2), \dots, R^i(f_{14})]$ and

Magnitude $Z^i(f_l) = [Z^i(f_1), Z^i(f_2), \dots, Z^i(f_{14})]$;

Output: Outliers removed data set $\tilde{D} : \tilde{D}^k(f_l)$, for measurements $k, k \in \{1, 2, \dots, K\}$;

// Remove negative resistances

- 1: $\phi = \{\}$ (create empty set of outliers)
- 2: **for** $i \leftarrow 1$ to total number of measurements n **do**
- 3: **for** $l \leftarrow 1$ to total number of frequencies 14 **do**
- 4: **if** $R^i(f_l) < 0$ (resistance is negative at frequency (f_l) for measurement i **then**
- 5: $\phi\{\phi, D^i(f_l)\}$ (include full spectral measurement i into outlier set ϕ)
- 6: **end if**
- 7: **end for**
- 8: **end for**
- 9: $\bar{D} = D - \phi$ (creates new dataset by removing data in outlier set ϕ from original data set)

// Detecting and removing outliers along the spectrum for each electrode configuration

- 10: **for** $c \leftarrow 1$ to total number of electrode configurations in $\bar{D}_c(f_l) = \{\bar{R}_c(f_l), \bar{Z}_c(f_l)\}$ **do**
- 11: $Q_a[\bar{Z}_c(f_l)] =$ quantile of variable $\bar{Z}_c(f_l)$ with pdf of $p[\bar{Z}_c(f_l)]$ along frequency spectrum (f_l) where $a = 1, 2, 3$ represents lower, median and upper quartiles respectively in each electrode configurations
- 12: **for** $k_c \leftarrow$ to total number of measurements K_c at each electrode configuration c in $\bar{D}_c(f_l)$ **do**
- 13: $J_c^{k_c} = \sqrt{\frac{1}{14} \sum_{l=1}^{14} \left(\frac{Z_c^{k_c}(f_l) - Q_2[\bar{Z}_c(f_l)]}{Q_3[\bar{Z}_c(f_l)] - Q_1[\bar{Z}_c(f_l)]} \right)^2}$ calculate the RMSD of each (robust scaled) full spectral measurements
- 14: **end for**
- 15: $Q_{c_1}[J_c], Q_{c_3}[J_c], MC[J_c] =$ lower quartile, upper quartile and MedCouple if variable $[J_c]$ with pdf of $p[J_c]$ for each electrode configuration respectively
- 16: $T_c = Q_{c_3}[J_c] + 3 * (Q_{c_3}[J_c] - Q_{c_1}[J_c]) * e^{4 * MC[J_c]}$
- 17: **for** $k_c \leftarrow k_c \in \{J_c\}$ the measurements k_c at each electrode configuration c **do**
- 18: **if** $J_c^{k_c} \geq T_c$ (RMSD greater than threshold) **then**
- 19: $\phi = \{\phi, \bar{D}_c^{k_c}(f_l)\}$ (include measurement k_c into outlier set)
- 20: **end if**
- 21: **end for**
- 22: **end for**
- 23: $\tilde{D} = \bar{D} - \phi$ (creates final dataset by removing data in outlier set ϕ from previous negative resistance removed data set)

The resistance and reactance measurements that make up the real and imaginary parts of the complex impedance are converted into the polar form as magnitude (Equation 2.2.6) and phase (Equation 2.2.7). From the equations, one can see that the calculated magnitude values are more sensitive to the measurement errors (with the squared sum of real and imaginary parts) than calculated phase values (where the proportion of the real and imaginary values tend to cancel out); therefore outlier detection is carried out based on the magnitude $|Z|$ values. Outlier detection is performed using the entire spectral data in order to accept or discard the full set of patient/volunteer measurements. Such a procedure requires ensuring that significant variations at individual frequencies, with potentially useful biomarker information, are not excluded owing to the small population size in our analysis. Each of the seventy-two tests (six tongue placements for each twelve electrode configurations) has its own impedance characteristics along the spectrum. With the observed spectral patterns showing distinct differences for each electrode configuration, the outlier detection was carried out separately for the different configurations.

Statistical analysis has shown that the data is non-normally distributed; therefore, in the Outlier Detection Algorithm, median and quartiles are preferred since these measures are not affected by the extremely low or extremely high values [138]. Feature scaling is another important pre-processing step in data analysis for data with widely varying features (in this instance frequencies). Impedance data has a high amplitude in low frequencies, and low amplitude in high frequencies and scaling is used to normalize data in different frequencies so that they contribute equally in the analysis. Robust scaling is an appropriate method for this non-normally distributed data [139].

High dimensional data is difficult to interpret. Root Mean Squared Deviation (RMSD) is a method to reduce the measurements into a single value along the frequency spectrum for easier interpretation of the statistical distribution of the data

set. The skewed distribution observed from the RMSD calculations suggests the use of Medcouple, which is a robust statistics measure that is more sensitive to the outliers [140]. This is, however, applied as a one-sided test owing to RMSD being a measure of difference. The outlier detection algorithm is broadly described in pseudocode, indicating the key steps in Algorithm 1 with respect to the methods justified above. The algorithm applied to data returned the results in Figure 2.6. The given Figure is for one of the six tongue placement figures that shows the outlier data was successfully removed from the raw data, and the processed data is ready for further analysis without the risk of outlier impact.

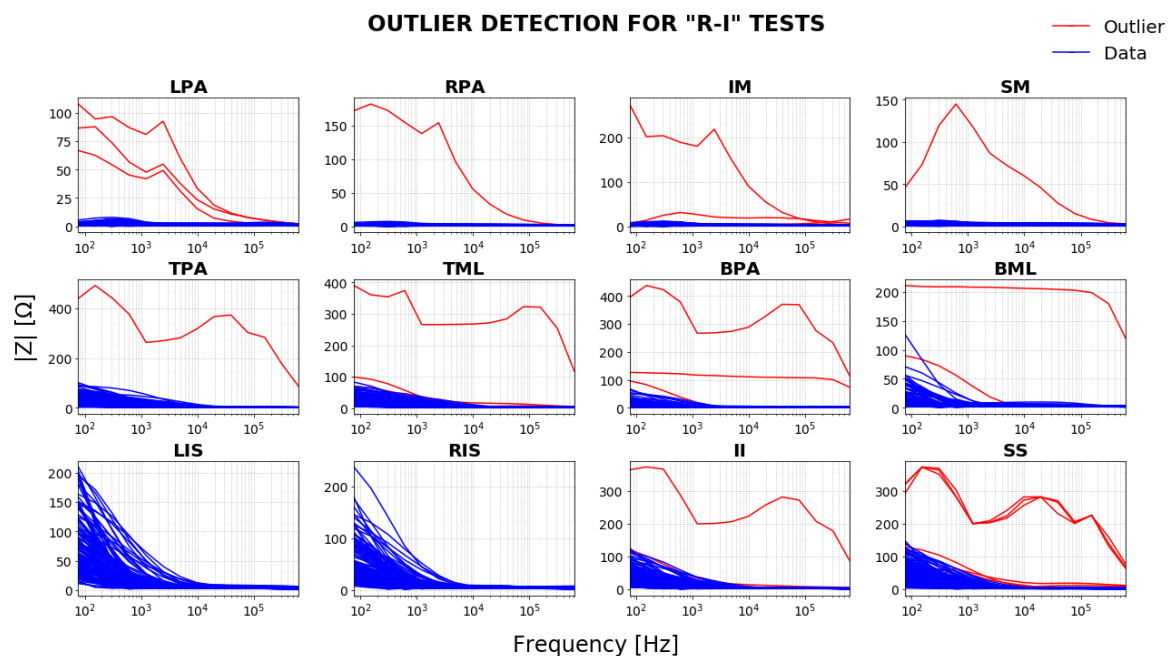


Figure 2.6: Outlier removal algorithm result for twelve electrode configurations on Right-Intraoral placement

2.5.2 Preliminary Selection of Electrode Configurations

Seventy-two test settings were studied as part of the preliminary data analysis. In early analysis of the project, a closer examination of four of the 3D electrode config-

urations shows that the electrical impedance measurements have a random shape, whereas the other electrode configurations seem to share a similar pattern among them along the frequency spectrum. Figure 2.7 is an example plot of the Center-Intraoral placement that shows this dissimilarity of the pattern. It is assumed that the electrode configurations with a random pattern may not reveal useful information for disease diagnosis, and it would be difficult to find a common mathematical representation for the tests. Moreover, these electrode configurations were not reproducible [130]. For this reason, the following chapters use four of the 2D electrode configurations: [TPA, TML, BPA, BML] and 3D electrode configurations: [LIS, RIS, II, SS].

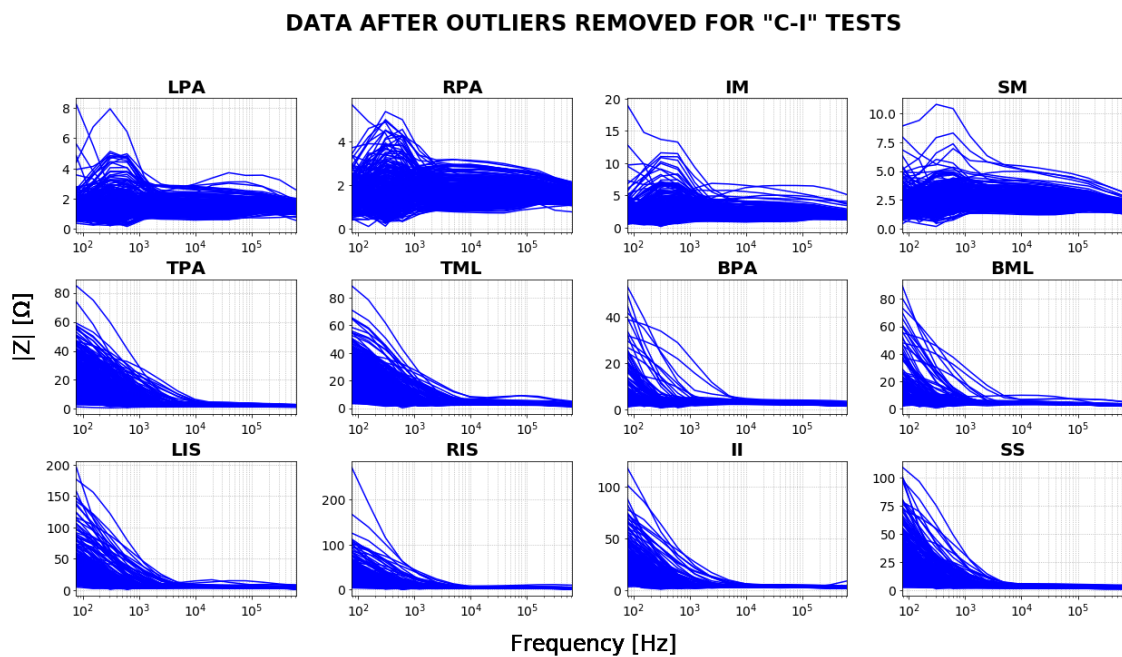


Figure 2.7: Outlier removed data for twelve electrode configurations on Center-Intraoral placement

This leaves us with the eight electrode configuration for six different tongue placements, which makes up to forty-eight clinical tests of the EIM device. Later in the project, concerns were raised about the left and right tongue location measurements.

The accuracy of the placement, particularly how close the electrodes might be to the edge of the tissue was identified as a potential problem. If the electrodes are close to the edge, the measurements could have a volume effect, i.e. less tongue tissue for current flow. However, centre measurements are not affected with this source of error. Therefore left and right placements are also removed from the future analysis in this project.

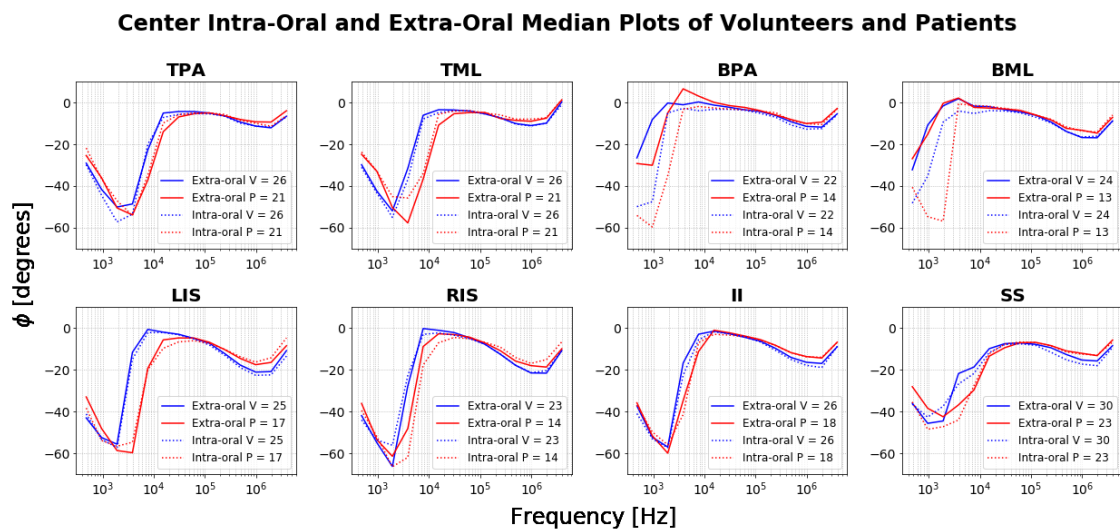


Figure 2.8: Center-Intraoral and Center-Extraoral phase median along the spectrum for comparing the healthy volunteers and patients to visually identify the disease's effect on Intraoral and Extraoral measurements. The number of patients and volunteers involved in median statistics are noted in the legend, given in each axis.

Analysis on skeletal muscles has shown the contracted and protruded muscles present different impedance characteristics along the frequency spectrum [50]. As explained in Section 2.3.2 different tongue muscles are activated when the tongue is retracted and protruded. Intra-oral and extra-oral measurements are taken in order to analyse the deterioration of the muscle in different muscle groups. Figure 2.8 shows the median values of the intra-oral and extra-oral measurements which appear as random for the different electrode configurations. Moreover, the Table 2.2

shows that the number of extra-oral measurements are much lower than the intra-oral measurements, which verified that obtaining extra-oral measurements are relatively difficult. For the reasons presented above, the chapters following contain analysis on centre intra-oral with four 2D electrode configurations and four 3D electrode configurations.

<i>Visit</i>	<i>Placement</i>	TPA	TML	BPA	BML	LIS	RIS	II	SS
1	<i>E</i>	v=27 p=21	v=26 p=22	v=26 p=17	v=29 p=15	v=26 p=21	v=23 p=14	v=26 p=20	v=30 p=23
	<i>I</i>	v=27 p=33	v=28 p=32	v=24 p=18	v=25 p=18	v=27 p=22	v=25 p=26	v=27 p=29	v=30 p=34
2	<i>E</i>	v=17 p=10	v=18 p=10	v=18 p=10	v=19 p=9	v=17 p=8	v=17 p=8	v=18 p=12	v=22 p=14
	<i>I</i>	v=21 v=23	v=20 p=23	v=19 p=14	v=19 p=18	v=19 p=21	v=18 p=18	v=20 p=23	v=23 p=26
3	<i>E</i>	p=8 p=10	v=7 p=10	v=8 p=9	v=8 p=9	v=6 p=7	v=6 p=8	v=6 p=11	v=8 p=12
	<i>I</i>	v=9 p=17	v=9 p=20	v=9 p=9	v=9 p=15	v=9 p=12	v=8 p=13	v=9 p=17	v=9 p=20
4	<i>E</i>	p=8 p=17	p=9 p=19	p=9 p=10	p=7 p=9	p=9 p=11	p=8 p=10	p=10 p=14	p=10 p=18
	<i>I</i>	p=2 p=9	p=2 p=9	p=2 p=4	p=2 p=7	p=2 p=7	p=2 p=4	p=2 p=9	p=2 p=10
5	<i>E</i>	—	—	—	—	—	—	—	—
	<i>I</i>	p=1	p=1	p=1	p=1	p=1	p=1	p=1	p=1
6	<i>E</i>	—	—	—	—	—	—	—	—
	<i>I</i>	p=1	p=1	p=1	p=1	p=1	p=1	p=1	p=1

Table 2.2: Electrical Impedance phase data recorded for Centre Intraoral and Centre Extraoral measurements in each visits. Volunteers only have the first three measurements.

Chapter 3

Non-Parametric Data Classification For Disease Diagnosis And Longitudinal Analysis

Clinical assessments are carried out by a medical team to evaluate the symptoms for diagnosis, in order to ensure patients receive the correct medical treatment. In this sense, the clinical assessments are a decision making tool for classifying the patient as diseased or healthy. The clinical assessment for diagnosis is the first step to treatment. Diagnosis starts with taking patient's clinical history and follows up with a series of appropriate physical and medical tests. Testing may involve using a medical device that measures the patients vital statistics, and creates a data pool to use in data analysis for diagnostic purposes.

Machine learning belongs to the family of artificial intelligence approaches and offers a wide variety of algorithms that can be used for cost-effective, automated diagnostic decision making of high-dimensional large datasets. Machine learning algorithms are capable of learning from observed data and applying it to the new data to make predictions. Machine learning methods are widely used in medicine for vari-

ous reasons such as diagnosis of the diseases [141, 142, 143], tracking the progression of the diseases [144, 145], or detecting impacts of the medicines used for treatment [146].

This project employs machine learning algorithms for the classification of patients and healthy volunteers data obtained from the novel Sheffield EIM device for investigating its potential for diagnosing the ALS bulbar involvement in the early stages. This chapter is presented in four main sections. Section 3.1 explains the steps to implement supervised classification. Different dimensionality reduction methods, cross-validation methods, and the two most common classification algorithms with classification evaluation metrics are discussed in this section. Section 3.2 focuses on non-parametric data classification using feature extraction and compares the classification algorithms. Section 3.3 focuses on non-parametric data classification approach with a feature selection algorithm. The classification algorithm that gave better discriminability in the section 3.2 is employed in this section. Pareto dominance is applied to the classification results with feature selection in the selection of the effective frequencies for each electrode configuration. Section 3.4 presents the longitudinal analysis with the selected frequencies for each electrode configuration. From herein and in the following chapters, the term clinical test and electrode configuration are used interchangeably. The phase readings are used in the classification analysis due to their capability in detecting the disease changes from EIM measurements [8, 30, 78]

3.1 Supervised Classification

Supervised classification is one of the three types of classification methods in machine learning. Supervised means that the algorithm learns from the labeled training data to develop a prediction model to which the new test data will fit. There are es-

quential steps to follow, and there are several different options used in each step that requires careful selection of appropriate methods for accurate classification of the labeled data. In machine learning, no one algorithm stands out in its ability to classify, but some algorithms are more widely used than the others.

The data obtained from the device was in raw form, and therefore a data pre-processing step (see Section 2.5) is essential in order to remove the outliers before classification. Dimensionality reduction is another step that removes unwanted and irrelevant information for greater accuracy, when mining data to find hidden patterns, especially for high dimensional data. The classification algorithms separate the data into a testing set and a training set, and use the selected/extracted features to obtain a model with the training data to evaluate it on the testing data.

The number of patients and volunteers recorded in each electrode configuration are different because some of the recordings were either faulty or removed from the preprocessing. Moreover, each electrode configuration aims at detecting disease effects on the different muscle groups. For this reason, each of electrode configuration is analysed separately in order to select the most informative electrode configurations paired with most discriminative frequency combinations. It is expected that the different electrode configurations will identify the different frequencies that are relevant to detect the disease effects on the anisotropic characteristics of the tongue.

3.1.1 Dimensionality Reduction

In Multi-Frequency EIM, impedance elements are measured at a range of frequencies to model the impedance characteristics of the tissue over the spectrum. This is creating a high dimensional data [147], which makes it challenging to perform computationally cost-effective classification. Moreover, high dimensional data causes a phenomenon called the curse of dimensionality, which is a result of the sparseness of the data in the high dimensional space, particularly with a low number

of data samples. Another challenge may arise from the overfitting, which is caused by using all features that create a model particular to that training data set, and this may lead to the poor fitting of the new data. For these reasons, dimensionality reduction is important for removing irrelevant and unnecessary features from the dataset that may obscure the discriminatory patterns.

There are methods used to reduce the high dimensional data into lower dimensions to outline the data by only considering the most important features. Two ways of reducing the dimension of the original dataset can be summarised as, feature extraction and feature selection. Feature extraction projects the features onto a new space with linear and non-linear methods to maximize the variability obtained by each feature. Feature selection is done either manually (if features are known to have discriminatory information according to the previous studies) or automatically with algorithms to create feature combinations and select the best performing feature set. Both feature selection [148, 149, 150] and feature extraction [151, 152, 153] are widely used in medical data analysis.

Feature Extraction

Feature extraction is used to reduce the dimensionality by performing a linear or non-linear transformation on the data [154, 155]. The most commonly used feature extraction method is a linear method called the Principal Component Analysis (PCA). PCA is preferred above others for its convenience in its simplicity as well as its widespread applicability [50, 156]. In PCA, the data is projected onto a lower-dimensional space with high variance features. The distribution characteristics and the feature labels are disregarded (unsupervised). In Linear Discriminant Analysis (LDA), the distributions of the features are labeled (supervised) and assumed to have a normal distribution. Factor Analysis (FA) is concerned about extracting latent variables of the data.

In PCA, the mean values of each feature are calculated to find the centre of the data in the high dimensional space. Then the centre of the data is moved to the origin of the high dimensional space, i.e., $[0, 0, 0, \dots]$; this is called the mean centering. Next, a line that has the minimum distances to the data points is fitted through the origin. This first line is called the first Principal Component (PC1). The unit vector obtained from PC1 contains information on how much each feature contributes to the variation; it is called the Eigenvector for PC1 and the method is called the Eigenvector decomposition. The next Principal Component (PC2) is the best fitting line that goes through the origin and is orthogonal to PC1. The PC3 is the best fitting line that goes through the origin and is orthogonal to PC1 and PC2. The process is repeated to find all principal components where the number of PCs is equal to the number of features. The percentile values of the variation can be calculated from the Eigenvalues, and this can then be plotted on a scree plot to select the number of PC's.

Mathematically, the above can be described as follows; consider the data given as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1l} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nl} \end{bmatrix} \quad (3.1.1)$$

where d is the dimension and n is the number of measurement. The covariance matrix of the data is:

$$\sum_{l \times l} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \dots & \sigma_{1l}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 & \dots & \sigma_{2l}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{l1}^2 & \sigma_{l2}^2 & \sigma_{l3}^2 & \dots & \sigma_{ll}^2 \end{bmatrix} \quad (3.1.2)$$

with

$$\sigma_{jk}^2 = \frac{1}{n-1} \sum_{i=1, j=1}^l (x_{ij} - \mu_j)(x_{ik} - \mu_k) \quad (3.1.3)$$

where μ_j and μ_k are the mean values, the σ_{kk}^2 is the variances and the σ_{jk}^2 is the covariance of the variables. Eigenvalues and eigenvectors can be calculated with the covariance matrix as follows;

$$\Sigma v = \lambda v \quad (3.1.4)$$

where the covariance matrix Σ in $l \times l$ form is multiplied with eigenvector v in $l \times 1$ form into scalar eigenvalues λ and eigenvectors. The sorted eigenvalues in descending order could be scaled into having percent based explanation, which is known as explained variance.

Feature Selection

Feature selection is useful in applications where the selected features can reveal some underlying characteristics of the problems or of the data. It could be done in two ways: one way is to perform manual feature selection and combine them with the classification algorithm if the analyst is familiar with the data features, another way is to use algorithms or visual tools to select relevant features. For low dimensional data, it is easier to combine features and plot one, two, or three-dimensional graphs for visual inspection of the discriminatory properties of the features. Heatmaps are also used for visual inspection of the correlation of features with each other. However, in high dimensional datasets, where the user has little to no knowledge about the background of the dataset, it is best to perform feature selection programmatically. In its most basic form, all features in the feature space are used to create feature combinations, and these combinations are then used in the classification algorithm to select the best performing feature subsets. In the literature, there are three classes of methods for selecting the best feature combination: wrappers, filters, and embedded methods.

- **Wrapper:** This class of methods considers all possible feature combinations, and performs an exhaustive search. Consider a dataset with l features; all pos-

sible combinations of features makes up to $2^l - 1$ different subsets. In the wrapper method, all subsets are used with the classification algorithm separately and result obtained with the evaluation metrics compared for selecting the feature subset with the highest classification score.

- **Filter:** This class of methods uses a more direct approach by selecting variables with proxy measures to eliminate the least interesting variables and then fitting the model. It is computationally more efficient than the wrapper method; however, it may not find the best performing feature subset for the dataset because of the elimination step. Several algorithms are available in the literature for implementing this method [157, 158, 159, 160].
- **Embedded:** This class of methods combines both filters and wrappers for the benefits of the filter's computational ease and the wrapper's exhaustive search. It adds and deletes best and worst features in each iteration, and requires manual input of the number of iterations, which could be a disadvantage for the method. LASSO algorithm is one of the examples for implementing this method [161].

3.1.2 Cross Validation

Model validation is important when the data set is used both for training and testing. It will give insight into how the classifier model will perform in predicting the new (unseen) data [162]. Cross-validation is useful in avoiding overfitting since it uses the training and test datasets with random splits repeatedly. The repetition is an important part of the procedure in order to reduce variability of scores in each test set.

A method called leave-p-out cross-validation (LpO CV) uses p number of data points from the whole dataset as test data, while the rest of the data is used as training

data. The procedure repeats until every possible subset of the p data points is used as the test data point. The averaged result from each test gives the performance of the model. Leave-one-out cross-validation (LOO CV) is a special case of LpO CV in that the p-value is equal to one [163].

Other two cross-validation are the Holdout method and the n-fold method. In the holdout method, the data is divided into training and test sets with a random number of data points in each sets. In n-fold cross-validation, the data set will be shuffled in order to split data into n groups randomly. The training data consist of $(n - 1)/n$ fraction of the data, and the remaining $1/n$ fraction of the data is used as the validation (test) set. The procedure will repeat n times using different training and test sets in each repetition. The repetition will allow each data to be used in both training ($n - 1$ times) and test (n times) sets. The score of each repetition is averaged to summarize the overall performance of the classification model. When n is equal to the number of data points, the method is equivalent to the LOO method [164].

3.1.3 Classification Algorithms

Classification is both a machine learning and statistics task that is used to classify different populations by extracting patterns that have discriminatory information in a dataset. In machine learning, the classification is used together with dimensionality reduction, cross-validation, and evaluation metrics in order to find the optimal classifier model for a dataset. The aim of performing classification in this project is to find the useful frequencies and best performing electrode configurations that will help with the diagnosis of the bulbar involvement of ALS with the Sheffield EIM device with novel electrode configurations. In order to do that, it is important to find a suitable classification algorithm and evaluation metric particular to the type of the analysis.

There are a large number of algorithms available in machine learning, and there

is no clear favorite. For this reason, it is crucial to carefully select a classification algorithm that suits well to the dataset and the goals of the research problems. Most of the classification algorithms are categorized as linear classifiers, and the name comes from the linear functions used in these algorithms. Among the linear classifiers, Fisher's Linear Discriminant, Logistic Regression, Naive Bayes Classifiers are some of the most commonly known. Support Vector Machines (this refers to Support Vector Classifier (SVC) in this work) and a kernel estimation method k-Nearest Neighbour (k-NN) are the interest of this research. This is due to the popularity of both methods and the appropriateness of the algorithms to the dataset. In the following sections, these two algorithms will be further detailed, and they will then be employed in Section 3.2 for performance comparison.

Support Vector Classifiers

Support Vector Machines (SVM) is a supervised machine learning algorithm that is used for many purposes, including classification, regression, outlier detection. Support Vector Classifier (SVC) is the SVM algorithm used to perform classification. Support Vector Classifiers have been used in medical diagnosis studies from cancer researches [165, 166] to electrical impedance studies [167].

Support Vector Classifiers starts with mapping each data in the training dataset into the high dimensional feature space with selected features. It draws a line between the classes in this plane with

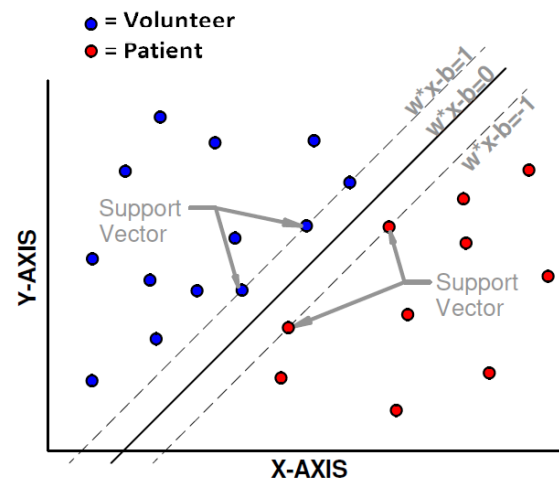


Figure 3.1: A 2D SVC hyperplane with support vectors and margins.

a maximum distance (called the hyperplane) that ensures a good class separation. The hyperplane has $l - 1$ dimensions with a dataset with l dimensions where l is the number of features after dimensionality reduction where necessary. The hyperplane can be written as a linear problem in its simplest form that makes the classifier a linear classifier. There are other kernel functions for separating the data to perform non-linear classification, but for a simplified explanation of the hyperplanes, a linear example is used in Figure 3.1. The linear relation can be expressed as:

$$\vec{\omega} \cdot \vec{x} - b = 0 \quad (3.1.5)$$

with $\vec{\omega}$ normal vector and \vec{x} is the set of data points. The hard margins created to separate classes from each other with $\vec{\omega} \cdot \vec{x} - b = v$ border belongs to volunteer class and $\vec{\omega} \cdot \vec{x} - b = p$ border belongs to patient class. The points at the borders are called the support vectors that help identify the borders of the hyperplane.

k-Nearest Neighbours

k-Nearest Neighbours (k-NN) is a frequently used classification algorithm in medical studies [168, 169, 170]. k-Nearest Neighbours is a classification algorithm that works only with labels and features. It is a supervised, non-parametric and a lazy learning algorithm [171]. It is a supervised algorithm because it learns from the labelled input data to estimate the label or class of the newly introduced unlabelled data. It is a lazy learning (or instance based) algorithm because it does not have a training step and the nearest neighbour in terms of feature similarities is searched from the whole data set each time for each prediction. It is non-parametric because it does not rely on data distribution assumptions, which makes it more practical for real-world applications.

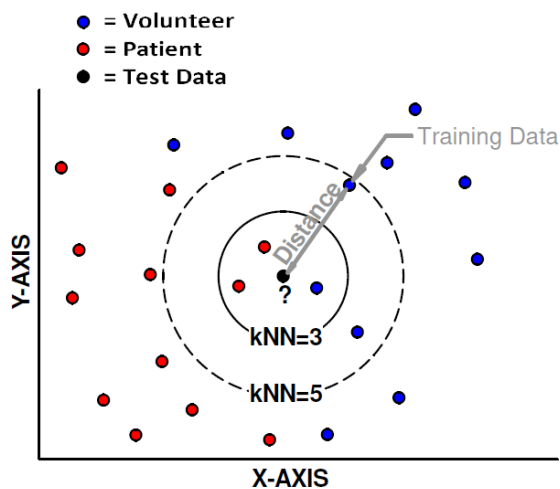


Figure 3.2: A 2D k-NN plot depicting the effects of different number of neighbours.

As the name suggests, k is the number of nearest neighbours that the algorithm tries to get votes from [172]. When a new data is introduced to the trained data space, a distance metric is used to calculate the distance to determine the k closest neighbours [172]. Different distance metrics can be used

in the distance calculation that are Manhattan (L1 norm), Euclidean (L2 norm), Chebyshev (maximum distance), Minkowski, and Mahalanobis. Votes from the surrounding neighbours are calculated, and the data point is labeled as the same label that has the highest votes amongst the nearest neighbour data points. This shows that the number of neighbours, k is a crucial factor for labeling the new data point. The number of neighbours should not be too small in order to avoid bias towards closest data points, and it should not be too high in order to avoid the costly label selection process.

There are two metrics to focus on when using a k-NN algorithm, that is the distance metric and k , the number of neighbours. It is important to know that the number of neighbours should be odd for two-class classification because if it is an even number, the number of closest neighbors may be the same, and therefore a decision could not be made. If the number of the nearest neighbour is high, the selection is less biased and less vulnerable to noise; however, the boundaries become less flexible, especially for datasets with a small number of data points.

3.1.4 Classification Results Evaluation

Classifier performance measurement is an essential part of comparing different methods to select the one that performs best for the purpose of the research. Both classifier and the evaluation metric depends on the nature of the study. For instance, this study employs supervised classification in order to identify the electrode configurations and the frequencies that shows the most discrimination between the patients and volunteers. This statement also means that the patients should be identified correctly, while a false alarm is avoided in the clinical assessments.

	<i>Positive Test</i>	<i>Negative test</i>
<i>Disease Present</i>	True Positive (Hit)	False Negative (Miss) (Type II Error)
<i>Disease Absent</i>	False Positive (False Alarm) (Type I Error)	True Negative (Correct Rejection)

Table 3.1: Confusion Matrix with four components; True Positive (TP) for correctly identified patients, True Negative (TN) for correctly identified healthy volunteers, False Positive (FP) for for volunteers identified as patient, False Negative (FN) for patients identified as volunteers.

One way to score the classification results for disease diagnosis is to calculate the confusion matrix that returns the proportion of accurately and inaccurately detected healthy volunteers and patients [173]. A confusion matrix (also called the error matrix) for a two-class classification problem has four components; two of them are the correctly identified classes, and two of them are the number of incorrectly identified classes. From these four numbers, several evaluation metrics can be calculated. Disease diagnosis studies mostly focus on the three out of all metrics, which are sen-

sitivity, specificity, and accuracy. The confusion matrix and a short description of evaluation metrics obtained from the confusion matrix components are presented below for two-class classification with healthy and diseased individuals.

From Table 3.1, the True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Positive ($P = TP + FN$), Negative ($N = TN + FP$) values help to calculate the following evaluation metrics;

Accuracy is a balanced representation of the Sensitivity and Specificity. It is the accuracy of how well the classifier can identify classes correctly. It is calculated as follows;

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (3.1.6)$$

Sensitivity is the sensitivity towards detecting the patients correctly. Also known as True Positive Rate (TPR), hit rate, recall is calculated as follows;

$$TPR = \frac{TP}{TP + FN} = 1 - FNR \quad (3.1.7)$$

Specificity is the specificity towards detecting the healthy individuals correctly. Also known as True Negative Rate (TNR), is calculated as follows;

$$TNR = \frac{TN}{TN + FP} = 1 - FPR \quad (3.1.8)$$

Other Confusion Matrix Metrics:

- Precision (Positive Predictive Value): $PPV = \frac{TP}{TP + FP}$
- Negative Predictive Value: $NPV = \frac{TN}{TN + FN}$
- False Negative Rate (miss rate): $FNR = \frac{FN}{FN + TP} = 1 - TPR$
- False Positive Rate (fall-out rate): $FPR = \frac{FP}{FP + TN} = 1 - TNR$

- F1 Score: $2 * \frac{PPV * TPR}{PPV + TPR} = \frac{2 * TP}{2 * TP + FP + FN}$
- Matthews Correlation Coefficient: $MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

3.2 Non-Parametric Data Classification Using Feature Extraction

Early diagnosis of patients without evidence of the bulbar involvement is important. From a machine learning perspective, classification of data that is collected from diseased and healthy individuals could help with diagnosing the disease in the early stages. Careful selection of methods is key to achieving the most accurate results in identifying the disease clinically with the EIM technique. In this section, the aim is to select methods that will help most accurately identify bulbar involvement of ALS disease on patients with EI data obtained from the Sheffield EIM device.

The classification requires the dimensionality reduction prior to implementing the algorithm. Commonly used feature extraction with PCA is implemented on EI data, and extracted features are tested on two different classification algorithms, namely k-NN and SVC. The results evaluation is validated with both exhaustive and non-exhaustive cross-validation methods, namely LOO and n-fold CV.

3.2.1 Preprocessing

The feature extraction method PCA is an orthogonal linear transformation that uses data points in its original feature space and projects it onto a new feature space where the maximum variance is preserved with the first few PC. Steps followed to obtain the first few PCs are listed below.

- **Feature Scaling:** Scaling normalizes the features into a range so that they all contribute the same amount to the calculations. Some of the most used scal-

ing methods are Standard Scaling, Robust Scaling, and Min-Max Scaling. Standard scaling uses mean and standard deviation moments of the normal distribution; therefore, it is not suitable for our dataset (see Section 2.5). Min-Max scaling subtracts the minimum value of the distribution from all data and divides them into the difference between the minimum and maximum values. Robust scaling uses 25% (Q1), 50% (Q2 = median), and 75% (Q3) for a more robust scaling of the features in a skewed data distribution like our data (see Section 2.5).

<i>Standard Scaling</i>	<i>Robust Scaling</i>	<i>Min-Max Scaling</i>
$\frac{x-\bar{x}}{\sigma}$	$\frac{x-Q_2}{Q_3-Q_1}$	$\frac{x-\min(x)}{\max(x)-\min(x)}$

Table 3.2: Standard, Robust and Min-Max scaling equations comparison.

- **Explained Variance:** An important step of feature extraction is to select how many components are necessary to explain enough variation on data for good discrimination between groups with the minimum number of components. Explained variance (cumulative variance) is a visual tool plotted using the same rationale as a scree plot, with an ascending percentage on the y-axis and component number on the x-axis. This visual tool will help identify the percentage of the explained variance, coupled with the number of principal components. An explained variance plot with selected principal components is plotted in Figure 3.3 using the phase readings of the EI measurements. There are no rules on the amount of variation that should be included in the classification; however, principal components that add up to above 90% is usually considered a good choice for differentiating the classes [174]. In this dataset, the electrode configurations have commonly shown a single number of principal components that exceed 90% variation between classes. For this reason, dimension-

ality was successfully reduced from fourteen into a maximum of four for all electrode configurations.

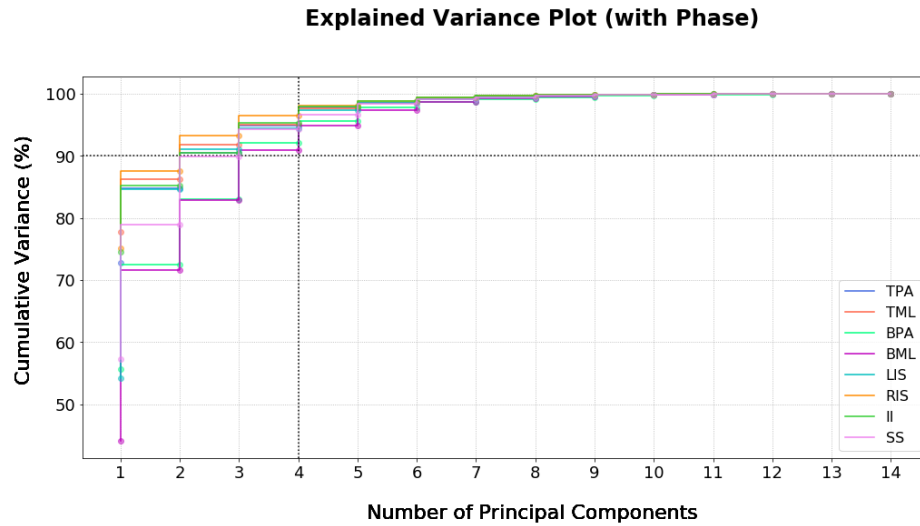


Figure 3.3: Explained Variance plot obtained with phase values. The electrode configurations exceed 90% variance with four principal components.

- **Feature Projection:** The first two principal components in Figure 3.4 have shown a $\approx 72\%$ to $\approx 87\%$ variation among the electrode configurations between the classes. A plot of the data in two-dimensions with first and second principal components on x-axis and y-axis respectively may give insight into how well the classification accuracy is. An example plot of this first two principal components is given in Figure 3.4.

The preprocessing analysis has shown that the dimensionality of the data is reduced successfully using feature extraction. However, it is not possible to isolate which frequencies are essential to reduce the dimensionality. In this study, the frequencies that give better discrimination are considered an important research question as much as the electrode configurations that give better discrimination.

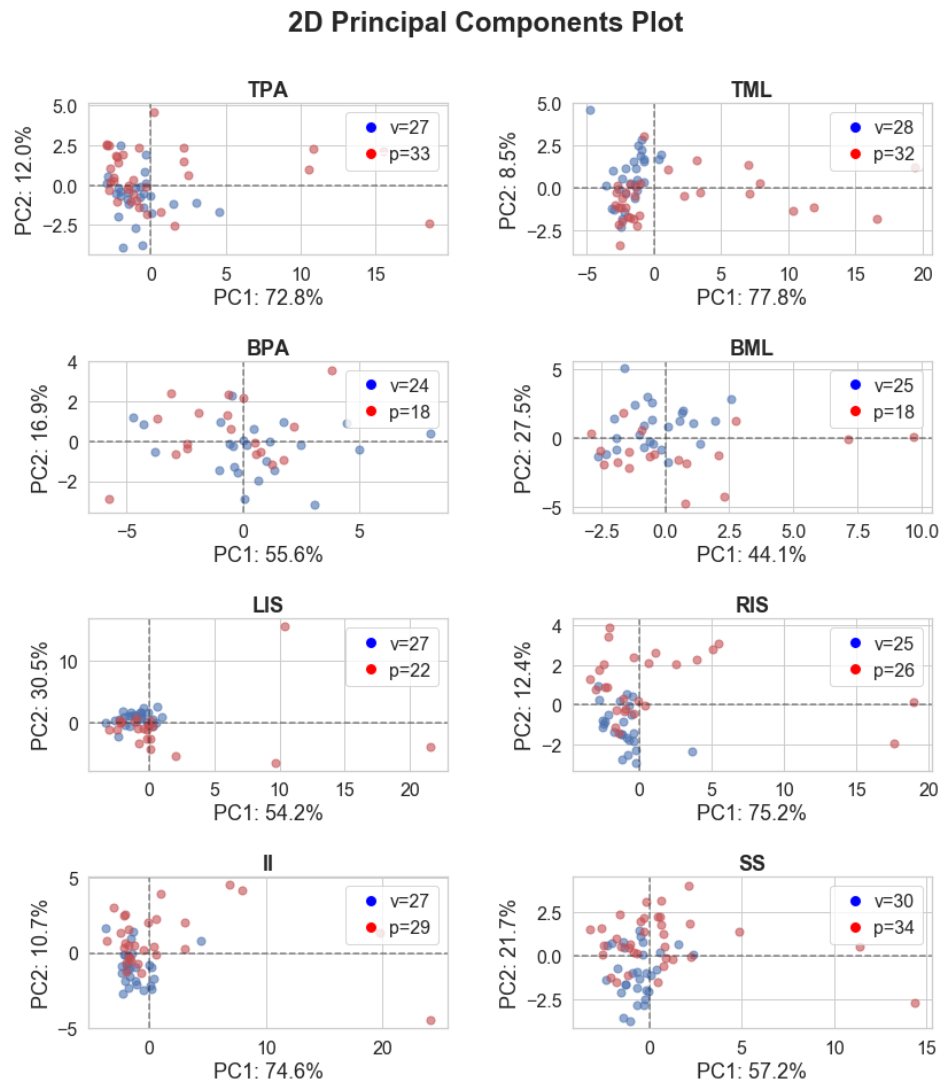


Figure 3.4: An example plot for first two Principal Components. The number of volunteers and patients are noted in each subfigure. The explained variance from each principal components are noted on axis labels. The electrode configuration initials are noted on the top of each axis.

3.2.2 Support Vector Classifier

Support Vector Classifiers avoid overfitting with high dimensional data, are memory-efficient with support vectors, and are versatile with built-in or custom made kernel

functions. The real-world data problems may not have a linear separation of measured features, and in order to separate non-linear data in high dimensional space, and to find the hyperplanes for a good separation of classes, one can use different kernel functions. Computer programming languages like Matlab and Python have libraries that help with implementing SVC for various research questions. There are three built-in kernel functions used with Python's SVC library, which are linear, polynomial and radial basis function.

Kernel functions $k(\vec{x}_i, \vec{x}_j)$ transforms the non-linearly separable feature space into a linearly separable feature space using scalar products of pairs of data vectors $(\vec{x}_i), (\vec{x}_j)$ to fit the hyperplane. The linear kernel function is defined as $k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j)$, polynomial kernel function is defined as $k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d$ where the d value is user defined degree of the polynomial and the radial basis function is defined as $k(\vec{x}_i, \vec{x}_j) = e^{-\gamma \|\vec{x}_i - \vec{x}_j\|^2}$ with γ being a positive user defined influence of a single training data point in the feature space.

The results were plotted using heatmaps with the classification accuracy metric, computed from the confusion matrix to determine the diagnostic accuracy of the device. The same visual inspection on the heatmap is done for sensitivity and specificity metrics of the classification, and the conclusions presented at the end persist.

Figure 3.5 shows the summary performance of the classification with SVC. Firstly, different kernel methods were used for comparing the results obtained for each electrode configuration, and on average, the poly kernel has lower and sigmoid kernel has higher accuracy results. Secondly, the 3D electrode configurations gives better accuracies compared to the 2D electrode configurations. Thirdly, electrode configurations seem to have similar accuracy results with a different number of n for cross-validation. It is also visible that 2-fold and LOO CV does not give good classification accuracies. However, there is no single good parameter set to satisfy the best classification accuracy for all electrode configurations.

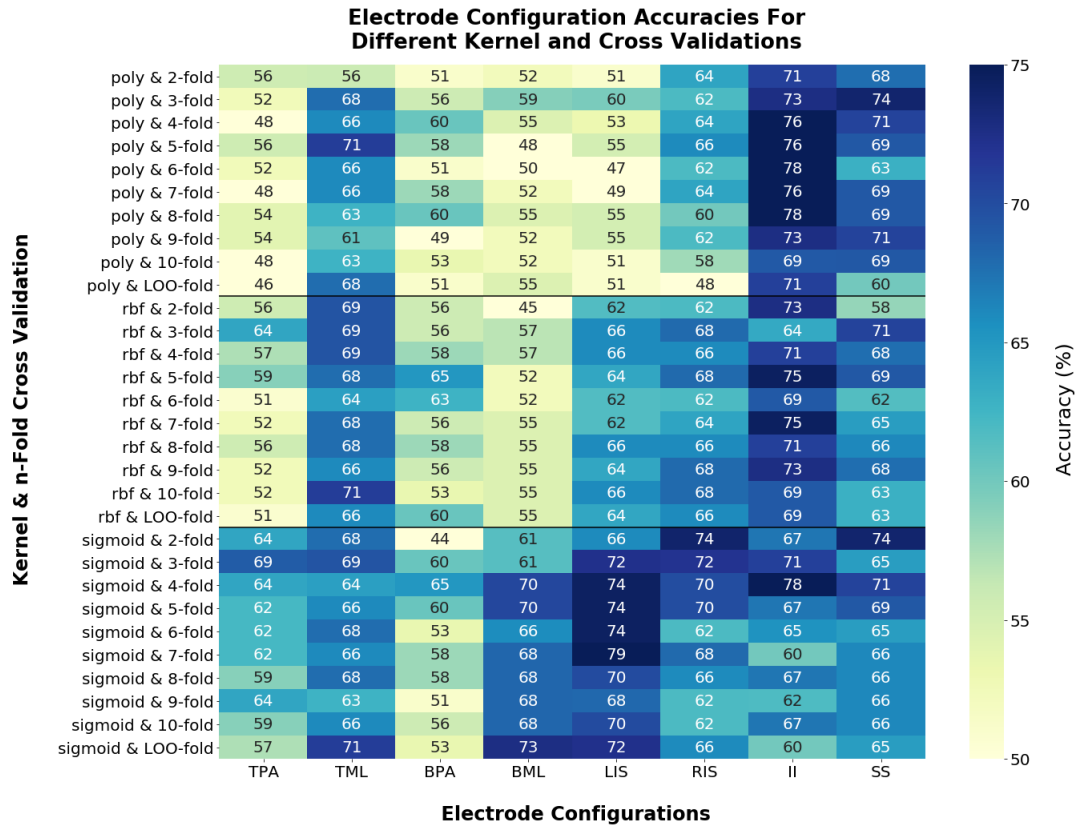


Figure 3.5: Classification accuracy results heatmap for different electrode configurations and different classifier design parameters with SVC.

3.2.3 k-Nearest Neighbour Classifier

k-Nearest Neighbours algorithm uses the training data points with the extracted features and class labels to calculate the k-nearest neighbours of a single test data point each time to assign the data point to the class of the closest neighbours. In this framework, the algorithm requires two user inputs that would affect the classification accuracy; that is, the k value for the nearest neighbours and the distance metric to calculate the distances to the closest neighbours. Due to a low number of data points, and the concerns related to even number of neighbours discussed under k-Nearest Neighbours title, $k = 3$ and $k = 5$ nearest neighbours together with commonly used

L1 (Manhattan) and L2 (Euclidean) distance metrics were tested with both LOO and n-fold cross-validation to select the optimal parameters for the user input variables. The value of n varies from a minimum of $n = 2$ to a maximum of most commonly used $n = 10$ fold.

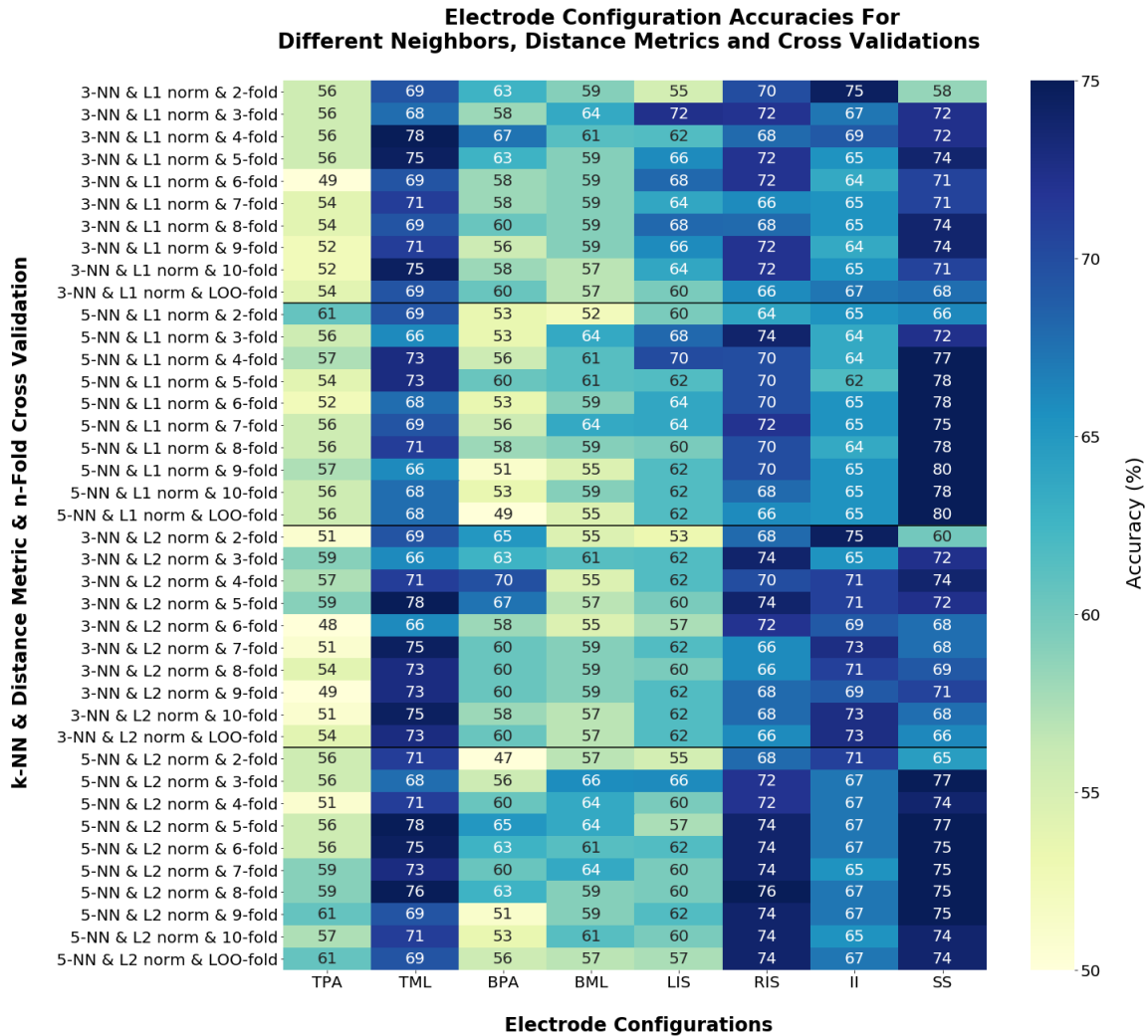


Figure 3.6: Classification accuracy results heatmap for different electrode configurations and different classifier design parameters with k-NN.

Figure 3.6 shows the summary performance of the classification with k-NN. Firstly, higher accuracies are achieved by the 3D electrode configurations [RIS, II, SS]. This shows that the 3D electrode configurations are able to identify the disease degener-

ation effects on tongue muscles better. Secondly, different neighbours; $k = 3, 5$ used with different distance metrics; Manhattan (L1 Norm) and Euclidean (L2 Norm) were combined (such as $k = 3$ with L1, $k = 3$ with L2, $k = 5$ with L1, $k = 5$ with L2) to calculate the nearest neighbours. There is no particular set of parameters that have a better accuracy pattern for all electrode configurations, and therefore, it is not possible to say that one of the combinations of k -neighbours and distance metric gives more accurate results for the classification. Moreover, the classification using k -NN overall has better accuracies than classification using SVC.

3.3 Non-Parametric Data Classification Using Feature Selection

In the analysis in Section 3.2 the dimensionality of the data was reduced down to four Principal Components while preserving the 90% variance in the data. However, the information on the individual frequencies that gives good discriminability of the volunteers and patients were not obtained with this dimensionality reduction method. One of the critical research questions of the ALS studies is to find the frequencies that give the most discriminatory information of the two groups (see Section 2.3.3). For that reason, to reduce the dimensionality of the data for computational efficiency and to remove redundant and irrelevant frequencies from the classification analysis, as well as obtaining the discriminatory frequencies, feature selection is considered here.

This section introduces the classification with the feature selection based dimensionality reduction for identifying the discriminatory frequencies for each electrode configuration. Pareto dominance was used to select the best performing frequency combination among all possible combinations. This section also looks into the correctly and incorrectly identified patients and healthy volunteers utilizing all electrode

configurations on the disease diagnosis with majority voting method.

3.3.1 Feature Selection with Wrapper Method

Differentiating between patients and healthy volunteers might not necessarily require the whole impedance spectrum, with some frequencies being redundant or irrelevant [175]. For this reason, dimensionality reduction is essential in order to reduce the computational complexity of the learning process associated with high dimensional data. In real-world data applications, useful features cannot be hand-picked due to unknown structures in the data. The identification of important frequencies is a feature selection problem with dimensionality reduction. Feature selection reduces the computational cost of classification, simplifies data for easier interpretation, reduces overfitting, and helps to avoid the curse of dimensionality and improve classification accuracy [176].

The two main steps in feature selection are subset generation and subset evaluation. There are three main subset evaluation algorithms for feature selection: wrappers, filters, and hybrids (embedded) [177]. In this study, the wrapper algorithm was employed, which, although computationally expensive, is the most reliable among all three [157]. It trains and tests the model for each new subset with the learning algorithm and selects the best subset depending on their performances.

There are several subset generation algorithms to choose the subset for a wrapper algorithm, including the genetic algorithm, simulated annealing, and several complete or heuristic search methods [178]. Subset generation is carried out by the evaluation of each possible subset, which requires $2^l - 1$ repetitions of the subset evaluation algorithm. This is referred to as a complete (or exhaustive) search. Some other widely used heuristic algorithms for subset searches include greedy backward elimination or forward selection [176] and genetic algorithms [157]. In this study, a fourteen-dimensional search will only cost $2^{14} - 1$ repetitions of the selected sub-

set evaluation and will give a complete understanding of all feature combinations. Therefore, a complete search with the wrapper evaluation algorithm is suitable for our purpose.

3.3.2 k-NN Classifier with n-Fold Cross Validation

The two classification algorithms were compared in Section 3.2; k-NN and SVC. The SVC has performed better with a sigmoid kernel for all electrode configurations; however, when it was compared to the k-NN, SVC has shown lower classification accuracies overall. For that reason, this section considers using k-NN for further analysis. Classification with k-NN has shown that the 3D electrode configurations *[RIS, II, SS]* performed better than other electrode configurations with all of the parameter combinations. The heatmap also has shown that there is no single common parameter combination that is best for all electrode configurations. Therefore the choice of the parameters are selected as; k=3 with L2 norm and 4-fold cross-validation for this dataset.

The wrapper method, with an exhaustive search for feature selection, created a list of 16383 ($2^{14} - 1$) accuracy, sensitivity, and specificity results for each electrode configuration. The choice of useful frequency combination must not be reliant on a single performance metric like accuracy. For this reason, the Pareto Dominance method is considered for the selection of best-performing frequencies for each electrode configuration. The results are presented in the following section.

3.3.3 Pareto Dominance for Selecting Disease-Specific Frequencies

Classification with feature selection returned $2^{14} - 1$ classification scores for each electrode configuration. The ranking of these results is necessary in order to find the best performance of each electrode configuration with their selected features.

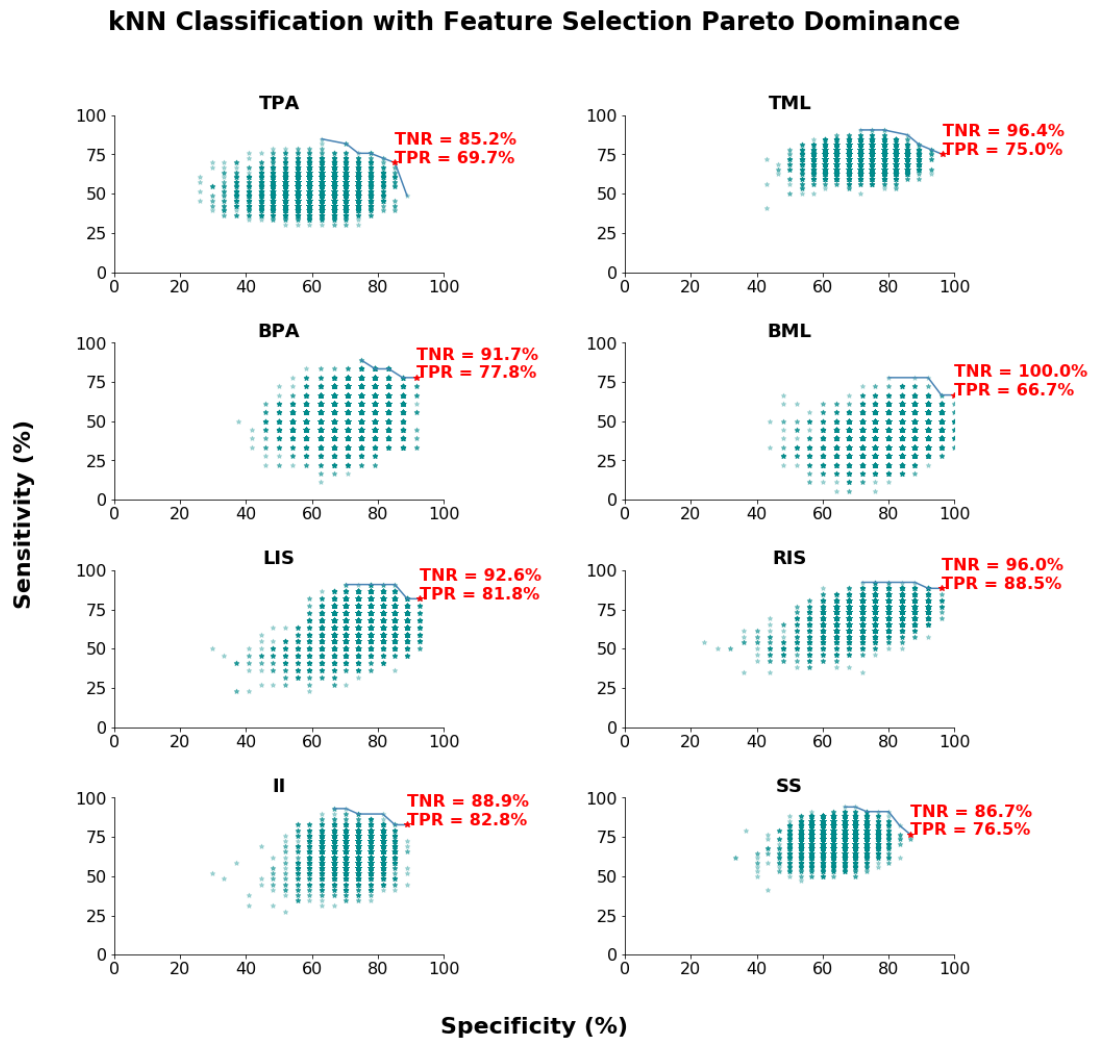


Figure 3.7: All sensitivity and specificity results obtained from the frequency subsets are given (blue stars) to show the variability of the classification results. Solutions on Pareto front (blue line) cannot be bettered in both the sensitivity and specificity by any other solution. The selected results (red points) are noted for each electrode configurations.

Electrode configurations were evaluated by their sensitivity and specificities, using a multi-objective optimization method called the Pareto ranking [179]. The robustness of classification accuracy depends on sensitivity and specificity and hence it is important to identify the subsets of frequencies that are important for both good sensitivity and specificity. By choosing the sensitivity and specificity values as two

objective functions, the performance of each configuration can be mapped into this objective space. A Pareto optimal solution in the objective space is one that is not bettered by another solution in both the sensitivity and specificity. Such Pareto optimal solutions define the Pareto front, and all solutions that are not on the Pareto front are not optimal solutions to the classification task.

Electrode Configurations	Features (Frequencies)	Accuracy	Sensitivity	Specificity
TPA (2D)	f1, f11, f12, f13	76.7%	69.7%	85.2%
TML (2D)	f2, f5, f6, f7, f8, f10, f11, f13	85.0%	75.0%	96.4%
BPA (2D)	f1, f2, f3, f5, f9, f12, f13, f14	85.7%	77.8%	91.7%
BML (2D)	f1, f3, f4, f8, f12, f14	86.0%	66.7%	100.0%
LIS (3D)	f1, f3, f4, f6, f9, f10, f11, f14	87.8%	81.8%	92.6%
RIS (3D)	f2, f3, f4, f8, f9, f10, f11, f12	92.2%	88.5%	96.0%
II (3D)	f1, f2, f5, f10, f12, f13	85.7%	82.8%	88.9%
SS (3D)	f6, f7, f8, f13, f14	81.3%	76.5%	86.7%

Table 3.3: Selected frequencies for each electrode configurations with accuracy, sensitivity, and specificity values. The frequencies are represented as [f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14] which is [76.29395, 152.5879, 305.1758, 610.3516, 1220.703, 2441.406, 4882.813, 9765.625, 19531.25, 39062.5, 78125, 156250, 312500, 625000] Hz respectively.

In this work, the solutions on the Pareto front were ranked with descending specificities to form a ranked list of electrode configurations. This is because it is more important to avoid misdiagnosing healthy volunteers as diseased in clinical assessments. The choice of evaluating the results with the best accuracy, sensitivity, or specificity usually depends on the type of medical tests and the concept of the study. Given the different feature subsets for each electrode configuration, similar feature

configurations may appear on the Pareto front. Only the highest-ranked solution with the minimum number of features for each electrode configuration is retained in the list to avoid the inclusion of redundant features.

Selected sensitivity-specificity pairs are indicated in Figure 3.7 in red. The results show that the electrode configurations are achieving the highest specificity (TPR) of 100% and the highest sensitivity (TNR) of 88.5%. Table 3.3 summarizes the chosen results for each configuration with the identified frequencies.

3.3.4 Diagnosis Using Multiple Electrode Configurations

Electrode configurations achieve sensitivities varying between $\approx 70\%$ to $\approx 89\%$ and specificities varying between $\approx 85\%$ to 100%. These results indicate that the electrode configurations are highly capable of avoiding misdiagnosis and also capable of detecting the disease. It also means that some of the patients and healthy volunteers are misclassified in some of the electrode configurations. All the measurements are taken with a single placement of the device at the center of the tongue with intra-oral placement. Across all electrode configurations, measurements give more comprehensive measurement of the characteristics of the tongue to make a better diagnostic decision. Therefore the diagnosis method should employ all electrode configurations with the outcome of the analysis given in Table 3.4. The majority voting method was used to combine the outcomes from all the electrode configurations.

Table 3.4 is presented for volunteers and patients diagnostics ratio for all electrode configurations. There are eight electrode configurations available, but measurements are not captured in some of the electrode configurations for some patients and volunteers. The ratio represents the proportion of the number of electrode configurations that correctly identified the patient/volunteer to the total number of electrode configurations that the measurement was performed. The blue colours represent the low ratios, where less than or equal to 50% of the electrode configurations

correctly identified the patient/volunteer. Converting red and blue color codes into confusion matrix table, true positives and true negatives are represented with red colour where false positives and false negatives are represented with blue colour.

<i>Patient</i>	<i>Rate</i>	<i>Patient</i>	<i>Rate</i>	<i>Patient</i>	<i>Rate</i>	<i>Patient</i>	<i>Rate</i>
p001	8/8	p012	4/8	p023	6/8	p034	5/6
p002	5/8	p013	5/7	p024	6/8	p035	8/8
p003	6/6	p015	0/1	p026	3/5	p036	6/6
p004	5/8	p016	3/4	p027	1/1	p037	3/8
p006	6/6	p017	0/0	p028	1/1	p038	8/8
p007	3/5	p018	5/5	p029	2/4	p039	4/4
p008	5/7	p019	1/1	p030	3/3	p040	2/3
p009	5/8	p020	5/5	p031	5/6	p041	7/8
p010	5/5	p021	6/8	p032	5/5	Blue	Red
p011	3/8	p022	5/8	p033	4/4	Low	High
<i>Volunteer</i>	<i>Rate</i>	<i>Volunteer</i>	<i>Rate</i>	<i>Volunteer</i>	<i>Rate</i>	<i>Volunteer</i>	<i>Rate</i>
v001	8/8	v009	3/3	v017	8/8	v025	8/8
v002	7/7	v010	7/7	v018	6/7	v026	5/8
v003	8/8	v011	4/4	v019	8/8	v027	7/7
v004	8/8	v012	7/8	v020	7/8	v028	8/8
v005	8/8	v013	8/8	v021	3/4	v029	8/8
v006	7/8	v014	7/8	v022	2/4	v030	3/5
v007	4/6	v015	8/8	v023	8/8	Blue	Red
v008	7/8	v016	8/8	v024	6/7	Low	High

Table 3.4: Diagnosis outcomes across all electrode configurations. For each individual results are provided as a ratio of the number of configurations in which the correct classification was made to the total number of configurations that were included in the analysis. The red color indicates where the majority outcomes showed correct diagnosis while the blue indicates the alternative.

The results show that only five of the patients and one healthy volunteer was misdiagnosed based on the 50% rule. In confusion matrix, this translates to true positive value of 32, true negative value of 29, false positive value of 1 and false negative value of 5. This is equivalent to 86% sensitivity, 96.7% specificity, and 91.0% accuracy. The results have similar scores to the best performing 3D electrode configuration [RIS], but the scores are higher than the rest of the electrode configurations. The results are

also more reliable in comparison to depending on the analysis from a single electrode configuration.

3.4 Longitudinal Analysis Using Selected Features

In order to examine the relationship between patient symptoms and the EIM measurements, classification with feature selection followed by Pareto ranking was undertaken. The selected frequencies for each electrode configurations are employed in the longitudinal analysis to examine whether the EIM can detect the change over time.

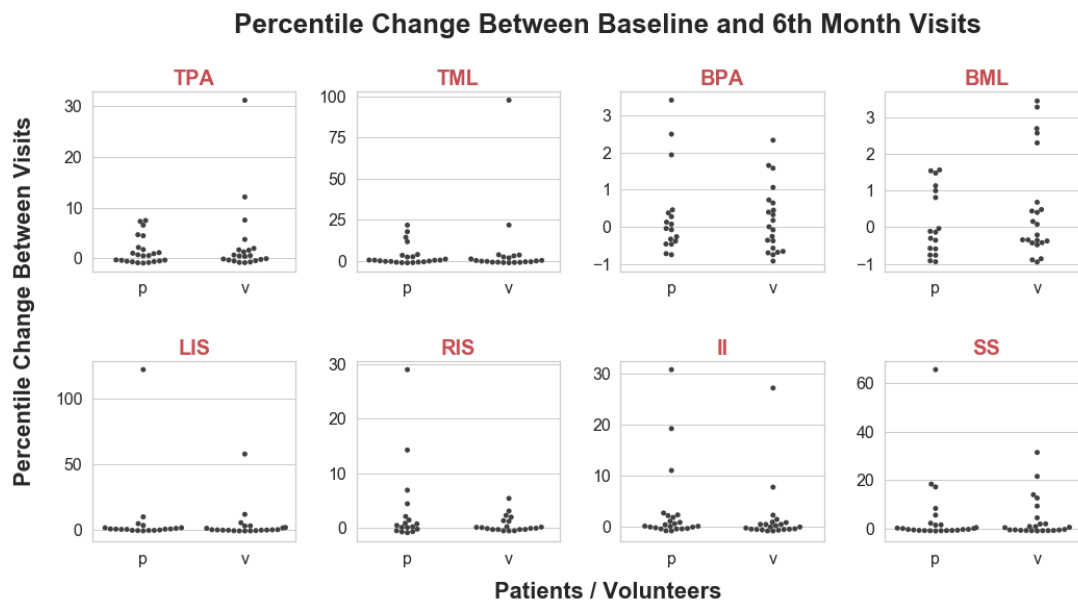


Figure 3.8: Percent change between baseline and 6th month visits for volunteers and patients in each electrode configurations to find the electrode configurations that can detect the patient EIM change.

The measurements with the EIM device are taken with three months intervals for patients and six months intervals for volunteers. In order to make the intervals equal, patients' baseline and sixth-month visits are used in the analysis. Phase angle values in the new feature space for each electrode configuration were then summarized with

a single score for ease of interpretation by calculating the L2 (Euclidean) norm for each visit per patient and healthy volunteers. The percentile change between the baseline visit L2 norm and the sixth-month visit L2 norm is an appropriate choice to detect the changes for the two groups.

Figure 3.8 presents the swarm plots of the percentile change in the L2 norm between visits. The results indicate that it is difficult to detect the longitudinal changes for patients with EIM readings. Only the *[RIS]* electrode configuration appeared to show some discrimination between volunteers and patients compared to the other electrode configuration. Further results are reported in [130].

<i>Patient</i>	<i>Baseline Score</i>	<i>6th Month Score</i>	<i>Patient</i>	<i>Baseline Score</i>	<i>6th Month Score</i>	<i>Patient</i>	<i>Baseline Score</i>	<i>6th Month Score</i>
p001	4	3	p020	9	10	p033	12	8
p003	9	10	p021	8	3	p035	1	1
p004	12	11	p022	8	10	p037	10	10
p008	7	6	p023	9	10	p038	12	11
p009	10	10	p024	9	8	p039	3	1
p013	9	9	p026	12	11	p040	12	12
p016	8	10	p028	7	8	p041	12	12
p017	10	7	p029	11	11	blue	green	red
p019	8	10	p030	10	8	Stable	Increase	Decrease

Table 3.5: Total bulbar subscores of the ALSFRS of each patient for baseline and 6th Month visits.

Clinically, the longitudinal disease progression is usually carried out from the bulbar score. Table 3.5 shows the total bulbar subscores of the ALSFRS of each patient in each visit. It was expected that there will be a drop in the bulbar scores as the disease progresses; however, some of the patient's bulbar scores are unchanged, and some even had an increase in bulbar subscore. A few of the patients have shown mild reduction, and just three patients have shown severe reduction on the tongue bulbar score. This provides a hint that the tongue structure for the majority of patients have not changed over the six months, showing the difficulty of the longitudinal data anal-

ysis.

3.5 Chapter Conclusion

In this chapter, the classification of healthy volunteers and ALS patients using novel Sheffield EIM device's phase readings was proposed. Two of the popular classification methods, k-NN and SVC were compared with feature extraction. Feature selection was implemented with k-NN classification, and the best performing frequency subset for each electrode configuration was selected after a Pareto Dominance analysis. Diagnosis with multiple electrode configurations was implemented. Longitudinal analysis was completed in the new feature space formed from every electrode configuration.

Dimensionality reduction with feature extraction was implemented using PCA. Preprocessing analysis has shown that all electrode configurations have above 90% variability with only four PC. The projection of the first two PCs in two-dimensional space has revealed that there is discriminatory information between the two groups. Classification using the four PCs with SVC algorithm has determined that the 3D electrode configurations perform better compared to the 2D electrode configurations. k-NN classification has achieved better accuracies than the SVC and confirmed that the 3D electrode configurations perform better compared to the 2D electrode configurations. Overall accuracies were varied among the different electrode configurations, reaching up to 80%. Classification with feature extraction showed sufficient discriminability of groups with EIM data.

Dimensionality reduction with feature selection is implemented using the exhaustive feature subset search and k-NN classification algorithm. Pareto dominance is investigated for the selection of the best performing frequency subsets. The 3D electrode configuration *[RIS]* gave the most accurate results among all. The diagnos-

tic ability of the device is improved with diagnosis using multiple electrode configurations. A sensitivity of 86.5%, a specificity of 97.7% is achieved using all electrode configurations.

Longitudinal analysis was completed using the new feature subset for each electrode configuration obtained by the feature selection. Results have shown that the 3D electrode configuration *[RIS]* performed better than the rest of the electrode configurations in detecting the percentile change of the EIM phase readings.

Chapter 4

Data Driven Modeling of Electrical Impedance Myography Measurements

Advances in technology facilitated the collection and storage of large amounts of data obtained from experimental sources. Growing computational efficiencies helped improve methods that serve as tools to mathematically model the observations in order to relate it to real-world applications. Data-driven modeling (DDM) benefits from the advantages of the computational advances to model the obtained data without any prior knowledge on the governing physical laws of the data source or experimental settings.

Data-driven modeling has been the focus of the studies where there is little or no knowledge about the mathematical or physical principles that describe the systems' characteristics. The models that are hypothesized by the physical laws and existing mathematical equations often use simplifications and idealizations that do not capture the relevant complexity in the actual system. This can lead to a bias in the model that is chosen to represent the system. As opposed to knowledge-based modeling, DDM tries to create mathematical models that are unbiased, and that can provide the flexibility to finding new patterns and mathematical laws from the observations.

In DDM, firstly, a model choice is made based on some assumptions and the measurement data that is available from the system. And secondly, a model estimation is performed by defining a loss function to quantify the fit of the model to the data. Some models can be used to give the required flexibility while also retaining some physical interpretability.

The previous chapter focused on machine learning methods to tackle difficult, high-dimensional data problems with limited input-output readings to find patterns relevant to the research question. This chapter focuses on the DDM with unknown system dynamics for identifying model related parameters that may have a physical meaning. Relating the EIM data to the physiology of the tongue through mathematical models is challenging. Some electrical circuit models have been proposed in the literature, but they lack the flexibility that may be required to fit to the data. The model structure that is chosen here belongs to a form that can include some of the electrical circuit representations. These are introduced in the following sections.

4.1 Selecting A Model Structure For The EIM Data

Model construction from the observed data has applications in industry [180] and natural sciences [181]. System Identification is a field of research within control systems engineering that tries to build a proxy mathematical model from observed input/output data. It is a method of model reduction by selecting models with only the dominant characteristics that describe the system. Compared to modeling implemented by using machine learning methods in the previous chapter, System Identification has been around longer [182, 183]. For a successful system identification implementation, there are steps to follow. First, identify the model structure and parametrize the model that is relevant to the data based on the assumptions made about the system. Second, infer a system description that summarizes the data con-

cerning the model, complexity levels, and underlying principles. Third, select algorithms that are suitable for estimating the numerical parameters of the model. Fourth and last, validate the model on new data. The steps to a system model identification process are given in the following diagram:

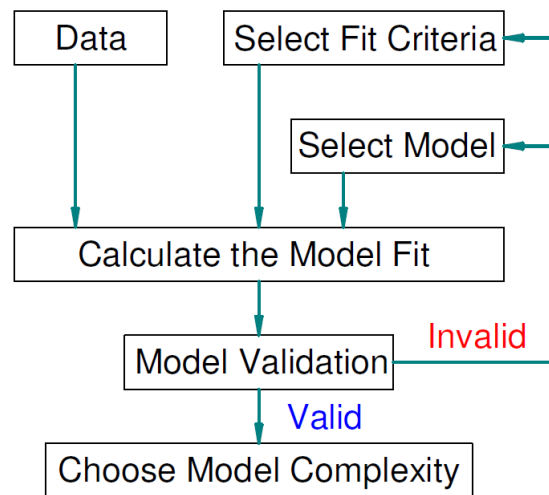


Figure 4.1: System Identification diagram to represent the steps for identification process.

The selected model and the cost function, together with the data, contribute to the estimation of the parameters. It is important to remember one fact that the 'Nature is simple' [184], and the model complexity should be assessed in order to obtain the model that best represents the system without the redundant and irrelevant parameters. Moreover, reducing model complexity will reduce the computational complexity. However, the balance between the model complexity and the fit should be carefully analysed. For this reason, a model space consisting of different model orders is essential. There are model selection schemes that trade off the improved model fit while penalizing the model complexity.

Systems can be described either non-parametrically or by a parametric model with linear or nonlinear representations. The essential characteristics of a model can be summarized by parameters that make the model a parametric model. Parameters

with a linear relation between them make the model linear in parameters. Real-life models are nonlinear; however, if a linearized representation can sufficiently approximate the system characteristics, then the system will be represented with a linear model. This approach will reduce the model complexity and computational cost. System identification is mostly built around the linear models, but nonlinear system identification is an open and growing field [185]. The objective of using a loss (cost) function is to evaluate the model's performance and thereby to estimate the optimal parameters giving best fit.

Systems are represented with block diagrams to show the input-output relation. These diagrams are a useful tool for the construction of the model with simpler building blocks. The diagrams' complexity depends on whether the system is measured in a well-controlled environment with each variable known to affect the output(s). The diagrams are named with the grey colour scale. As the shades of grey get darker and lighter, the background knowledge or the knowledge on the experimental setup gets darker and lighter. Modeling with the white box models requires complete understanding and background knowledge of the system from which the observations were obtained. The model has applications in literature [186]; however, it is not widely used since it is not possible to understand the whole system and know its physical laws from the basic principles. Therefore, grey box models are used to overcome the complicated structuring of the white-box models. The grey shade comes from the modeling with some unknown parameters or unidentified relationships between the variables. With the grey-box model, the observer has some idea about the actual model but not the entire settings of the experiments. Grey-box models are a more realistic model structure where the observer can depict the model with some of the straightforward physical laws of the experiment settings. There are different shades of grey, described by L. Ljung, in his summary of system identification methods and directions for future works [184]. These shades are; off-white, smoke-grey, steel-grey,

slate-grey models in which the examples and works are more focused on the nonlinear system identification. The darkest shade of grey, the black box is the most flexible model. It is dependent on the inputs/outputs of the system without knowledge of how the system works.

The data obtained from the EIM device is a frequency response data that will draw our focus on frequency domain parametric linear system identification. Since there is a need to select the model structure and parameters, the black-box model will be used for the flexibility required for modeling the data and to extract the underlying structures of the ALS disease. The parametrized transfer function form permits the use of identified parameters to classify the patients and volunteers for the diagnosis of the disease.

4.2 Black Box Transfer Function Model

The literature suggest that the electrical circuitry representation of the cell attributes with 3-element [52] or 5-element [53] models could help represent the bio-electrical impedance data [187, 188, 189]. The 3-element model consists of the cell wall mass X_c that acts as reactance, intracellular R_i , and extracellular R_e fluid that act as resistances. In 5-element model, in addition to these 3-elements, there are organelle resistance R_o and reactance X_o . In simple terms, the measurement process consists of applying a non-invasive alternating current to the tissue and recording the output voltage. Ohm's Law states that the proportion of the current and voltage returns the impedance readings. In these terms, the suggested circuitries in the literature rely on the physical model of a cell; however, the tissue level observations have a more complicated structure than a single cell. In a sample tissue that is measured with electrical impedance spectroscopy, there are influences of the layers of tissue with different components such as connective tissue, muscle tissue, nervous tissue,

and fat tissue. Moreover, the attachment of the electrodes to the tissue will have additional influences on the measurement. Modeling such complex systems with very little knowledge of the effects of these components is insufficient. Therefore, modeling the tissue without relying on the background information will give the freedom necessary to identify the disease characteristics with the parameters.

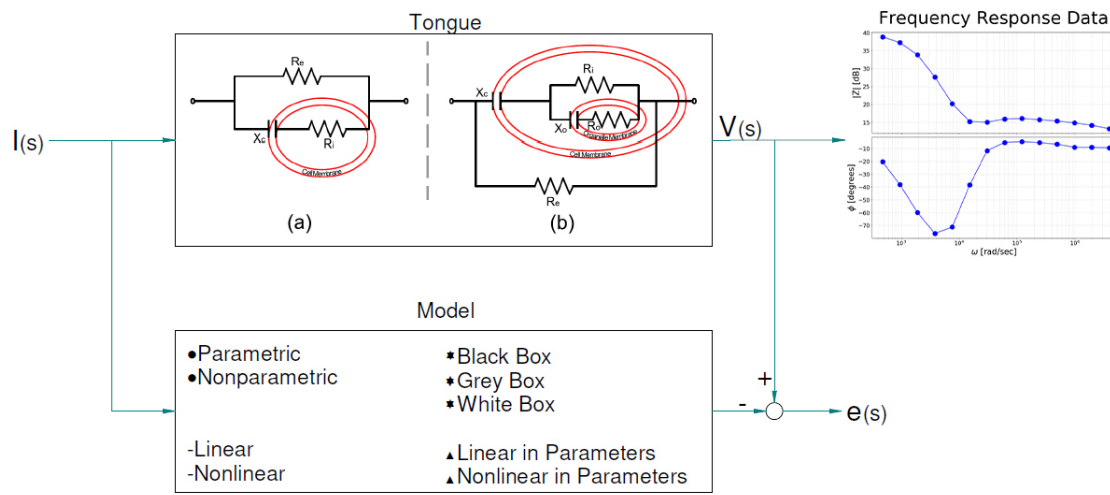


Figure 4.2: Black Box model representation of the system.

Linear black-box models described with the input $u(s)$ and output $y(s)$ is given as follows;

$$A(s)y(s) = \frac{B(s)}{D(s)}u(s) + \frac{C(s)}{F(s)}e(s) \tag{4.2.1}$$

This model is a representation for both the deterministic and stochastic parts of a system. There are parameters used to describe the input-output characteristics as well as the error characteristics. Using all parameters in this representation is computationally expensive, and it may not best represent the system. For this reason, reduced complexity models are available. Transfer function models are used to describe the

deterministic part of the system as

$$y(s) = \frac{B(s)}{D(s)} u(s) \quad (4.2.2)$$

In Control Systems Engineering, the proportion of the output-input of a linear time-invariant system in mathematical function form is called its transfer function. For frequency domain data, the transfer functions are represented in s domain with Laplace transform. The visualization of the bioelectrical impedance system's transfer function is possible using the independent frequencies and the dependent impedance output in a Bode plot or a Cole-Cole plot. Bode plots with its magnitude and phase values are useful in representing the impedance readings in polar form, and Cole-Cole plots are useful in representing the impedance in complex form. The transfer functions are given in a mathematical form of rational polynomials with numerator and denominator functions. The roots of the polynomial function in the numerator give the zeros, and those in the denominator give the poles. A transfer function model in the Laplace domain (s -domain) is as follows;

$$M(s) = \frac{N(s)}{D(s)} \quad (4.2.3)$$

where order of polynomial $N(s)$ is less than or equal to the order of polynomial $D(s)$. The s -domain model can be related to the frequency response function of the system where $s = j\omega$ where $\omega = 2\pi f$ and the frequency response function is given by

$$M(j\omega) = \frac{N(j\omega)}{D(j\omega)} \quad (4.2.4)$$

A linear black-box transfer function with an unknown number of poles and zeros allows for the type of flexibility necessary for selecting a suitable mathematical structure for the Electrical Impedance measurement from the tongue. The estima-

tion process in terms of black-box modeling includes exploration of different complexity models and selection of a suitable model that best represent the data. The aim is to find the set of poles and zeros that would best represent the disease characteristics while a good fit to the data is obtained.

4.3 Rational Polynomial Transfer Function Model

Transfer functions can be formulated in polynomial and factorized form. The first approach to formulating the linear frequency response data transfer function is given in the following general form as;

$$M_{n_p, n_z}(s) = \frac{N_0 + N_1 s + N_2 s^2 + \dots + N_{n_z} s^{n_z}}{D_0 + D_1 s + D_2 s^2 + \dots + D_{n_p} s^{n_p}} \quad (4.3.1)$$

The electrical impedance data are in the form of $R(\omega_l) + jX(\omega_l)$ with $s = j\omega$, where $l = 1, 2, \dots, 14$. The order of poles and zeros $n_z \leq n_p$, and the order of the system is determined by the order of the poles where $n_p < 14$. The D_0 coefficient is constrained to be 1, $D_0 = 1$, to ensure that the mathematical model has the uniqueness. The polynomial functions in the numerator and denominator are solved after complex components are separated and the difference between the data real and imaginary part (given by $R(\omega_l) + jX(\omega_l)$) with the model real and imaginary part from $M_{n_p, n_z}(j\omega_l)$ is minimized. The following section shows the steps for converting the equation into naive least squares for a closed form solution, which is one of the most common estimator for linear systems. With naive least squares, the parameter vector θ is estimated by minimizing the sum squared error.

4.4 Naive Least Squares Estimation

Following the Transfer Function Equation 4.3.1, we have;

$$M_{n_p, n_z}(s) + M_{n_p, n_z}(s)(D_1 s + D_2 s^2 + \dots + D_{n_p} s^{n_p}) = (N_0 + N_1 s + \dots + N_{n_z} s^{n_z}) \quad (4.4.1)$$

In matrix form;

$$M_{n_p, n_z}(s) + M_{n_p, n_z}(s) \begin{bmatrix} D_1 & D_2 & \dots & D_{n_p} \end{bmatrix} \begin{bmatrix} s \\ s^2 \\ \dots \\ s^{n_p} \end{bmatrix} = \begin{bmatrix} N_0 & N_1 & \dots & N_{n_z} \end{bmatrix} \begin{bmatrix} 1 \\ s \\ \dots \\ s^{n_z} \end{bmatrix} \quad (4.4.2)$$

For calculating the real and imaginary roots, $s = j\omega_l$, with $\omega_l = 2\pi f_l$ is an array from the frequencies with $l = 1, 2, \dots, L$ and $L = 14$

$$\begin{bmatrix} s^0 \\ s^1 \\ s^2 \\ s^3 \\ s^4 \\ s^5 \\ s^6 \\ s^7 \\ \dots \end{bmatrix} = \begin{bmatrix} 1 \\ j\omega \\ -\omega^2 \\ -j\omega^3 \\ \omega^4 \\ j\omega^5 \\ -\omega^6 \\ -j\omega^7 \\ \dots \end{bmatrix} = \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \\ \dots \end{bmatrix} \odot \begin{bmatrix} 1 \\ j \\ -1 \\ -j \\ 1 \\ j \\ -1 \\ -j \\ \dots \end{bmatrix} \quad (4.4.3)$$

From Equation 4.4.3, the even roots are imaginary and odd roots are real. This helps

formulate the numerator and denominator values as follows;

$$\mathbf{N}_{even}^{real} = \mathbf{N}_E = [N_0 \ N_2 \ N_4 \ \dots], \mathbf{N}_{odd}^{imag} = \mathbf{N}_O = [N_1 \ N_3 \ N_5 \ \dots], \quad (4.4.4)$$

$$\mathbf{D}_{even}^{real} = \mathbf{D}_E = [D_2 \ D_4 \ D_6 \ \dots], \mathbf{D}_{odd}^{imag} = \mathbf{D}_O = [D_1 \ D_3 \ D_5 \ \dots]$$

and

$$\mathbf{q}_{N_E} = \begin{bmatrix} \omega^0 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \dots \end{bmatrix} \odot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \dots \end{bmatrix}, \mathbf{q}_{N_O} = \begin{bmatrix} \omega^1 \\ \omega^3 \\ \omega^5 \\ \omega^7 \\ \dots \end{bmatrix} \odot \begin{bmatrix} j \\ -j \\ j \\ -j \\ \dots \end{bmatrix}, \mathbf{q}_{D_E} = \begin{bmatrix} \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \dots \end{bmatrix} \odot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \dots \end{bmatrix}, \mathbf{q}_{D_O} = \begin{bmatrix} \omega^1 \\ \omega^3 \\ \omega^5 \\ \omega^7 \\ \dots \end{bmatrix} \odot \begin{bmatrix} j \\ -j \\ j \\ -j \\ \dots \end{bmatrix} \quad (4.4.5)$$

From Equations 4.4.4 and 4.4.5

$$\begin{bmatrix} D_1 & D_2 & \dots & D_{n_p} \end{bmatrix} \begin{bmatrix} \omega \\ \omega^2 \\ \dots \\ \omega^{n_p} \end{bmatrix} = D_E q_{D_E} + j D_O q_{D_O} \quad (4.4.6)$$

$$\begin{bmatrix} N_0 & N_1 & \dots & N_{n_z} \end{bmatrix} \begin{bmatrix} 1 \\ \omega \\ \dots \\ \omega^{n_z} \end{bmatrix} = N_E q_{N_E} + j N_O q_{N_O}$$

Leaving $M_{n_p, n_z}(s)$ alone on the left hand side, and inserting Equation 4.4.6 in Equation 4.4.2 :

$$Y_r + jY_x = N_E q_{N_E} + j N_O q_{N_O} - D_E q_{D_E} (Y_r + jY_x) - j D_O q_{D_O} (Y_r + jY_x) \quad (4.4.7)$$

Therefore, the real part (Y_r) and imaginary part (Y_x) from the model output is calculated as follows;

$$Y_r = \begin{bmatrix} q_{N_E} & -q_{D_E} * Y_r & q_{D_O} * Y_x \end{bmatrix} \begin{bmatrix} N_E \\ D_E \\ D_O \end{bmatrix}, Y_x = \begin{bmatrix} q_{N_O} & -q_{D_E} * Y_x & q_{D_O} * Y_r \end{bmatrix} \begin{bmatrix} N_O \\ D_E \\ D_O \end{bmatrix} \quad (4.4.8)$$

In matrix form;

$$\begin{bmatrix} Y_r \\ Y_x \end{bmatrix} = \begin{bmatrix} q_{N_E} & 0 & -q_{D_E} * Y_r & q_{D_O} * Y_x \\ 0 & q_{N_O} & -q_{D_E} * Y_x & q_{D_O} * Y_r \end{bmatrix} \begin{bmatrix} N_E \\ N_O \\ D_E \\ D_O \end{bmatrix} \quad (4.4.9)$$

$$Y = X\theta \quad (4.4.10)$$

where Y is output, X is information matrix and θ is the parameters. Coefficients in the transfer function can be obtained in closed form from naive least squares with

$$\tilde{\theta} = (X^T X)^{-1} X^T Y \quad (4.4.11)$$

Inserting Equation 4.4.11 into 4.4.10, we get:

$$\tilde{Y} = X\tilde{\theta} \quad (4.4.12)$$

The coefficients can then be mapped into the frequency response function form to obtain the model representation. The naive least squares estimation provides a

mathematically convenient form of model identification. However, the direct implementation of the estimator resulted in a model with unstable poles and non-minimum phase zeros. The main contributing factor appeared to be the choice of frequencies that increased in log-scale.

4.5 Factorized Transfer Function Model

The closed-form linear least-squares approach formulated in the polynomial transfer function has shown that the numerical issues that arise from the frequency data should be addressed. The factorized transfer function is a simpler and numerically stable form that aims to resolve the numerical instability problem. This form only allows for real poles and zeros. Magnitude plots, in particular, are useful in determining the number of poles and zeros of the impedance data.

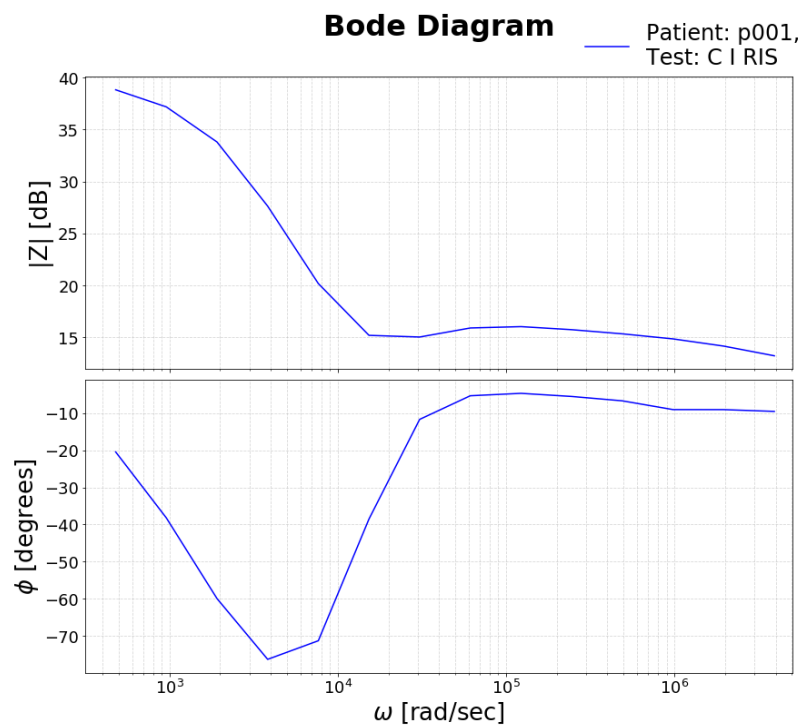


Figure 4.3: Bode Diagram for one impedance data set, given with Magnitude and Phase response plots.

The magnitude response of one data set in full spectrum is shown in Figure 4.3 for a visual inspection. When the numerator of the transfer function approaches to zero, then the frequency point approaches a zero point; in Bode plot magnitude response, this will add a slope of +20dB/dec. When the denominator approaches to zero, then the frequency point approaches to a pole point; in Bode plot magnitude response, this will add a slope of -20dB/dec. From this perspective, a transfer function equation with varying poles and zeros in factorized form is given as follows;

$$M_{n_p, n_z}(s) = K \frac{\prod_{i_z=0}^{n_z} (s + z_{i_z})}{\prod_{i_p=0}^{n_p} (s + p_{i_p})} \quad (4.5.1)$$

where K is the gain, $s = j\omega_l$ is the frequency inputs of the observation, z_{i_z} and p_{i_p} are the zeros and poles respectively.

4.5.1 Integral Fractional Order Transfer Function

The integer-order transfer function in the factorized form in Equation 4.5.1 will help obtain the poles and zeros directly from the denominator and numerator. In Figure 4.3, it is visible that the slopes at the high frequency spectrum are smaller than $\pm 20dB/dec$, and in order to capture these slopes that may reveal important information relating to the data, the order of s in Equation 4.5.1 needs to be revised.

Representing systems by fractional order rather than with integer order has proven to be useful in real world applications [190]. In Cole-Cole plots, fractional powers have been used to identify the models from EI data [191]. Fractional powers have been proven to be useful in modeling the tissue level biological electrical impedance measurements [192, 193, 194, 195]. Looking into the mathematical function in Equation 4.5.1, the $\pm 20dB/dec$ slope is a result of computing the poles and zeros with integer-order. One method to achieve a lower than $\pm 20dB/dec$ slope is by the choice of the order of s being less than 1. Replacing the integer orders of the poles and ze-

ros with the fractional-order may improve the fit to the system's anomalous behavior. Fractional order systems can be formulated in two different ways; commensurate order, where $\alpha_{p_1} = \alpha_{p_2} = \dots = \alpha_{i_p} = \alpha_{z_1} = \alpha_{z_2} = \dots = \alpha_{i_z} \equiv \alpha$ and noncommensurate order where $\alpha_{p_1} \neq \alpha_{p_2} \neq \dots \neq \alpha_{i_p} \neq \alpha_{z_1} \neq \alpha_{z_2} \neq \dots \neq \alpha_{i_z}$ in Equation 4.5.2 [196]. In this project, the commensurate fractional order transfer function is chosen because of the limited amount of impedance data that is available to fit the model.

$$M_{\alpha n_p n_z}(s) = K \frac{\prod_{i_z=0}^{n_z} (s^\alpha + z_{i_z})}{\prod_{i_p=0}^{n_p} (s^\alpha + p_{i_p})} \quad (4.5.2)$$

There is also a gentle slope that appears at the low frequency of the magnitude plot. From a physical point of view, in cell level measurements, the identical cells would be expected to have the same model, but in tissue level, there are different cells in the system. In this data, they can be used to explain the integration effect and the process of cell continuum approximation respectively. In literature, there are examples of this integration effect coming from the electrode-tissue interface model [197]. This effect may also come from the tissue layers of the tongue. An epithelial layer on top of tongue may cause interference that needs to be taken into account. Epithelial acting like a capacitance could also explain modeling of $M_0(s) = \frac{1}{s^{\alpha_0}}$ integration. In order to account for the gentle slope that appears at the low frequency of the magnitude plot, a different fractional order of an integrator is chosen to represent that part. Schematically this can be shown via two separate processes, one associated with tongue, and the other fractional integration potentially associated with the measurement process.

The second block $M_0(s)$ on the block diagram above represents the fractional integration that may come from instrument process or some other affect on the measurements and the first block $M_{\alpha_1 n_p n_z}(s)$ represents the tongue process. This series block diagram returns the Equation (4.5.3) for the integral fractional order transfer

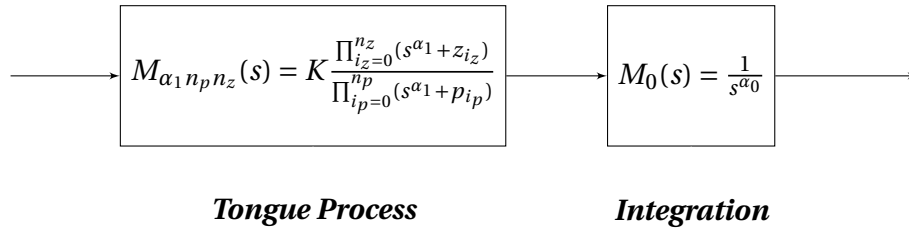


Figure 4.4: Series black-box model for combining the instrument measurement model and the transfer function model of the tissue.

function used here.

$$M_{\alpha_1 n_p n_z}(s) = K \frac{1}{s^{\alpha_0}} \frac{\prod_{i_z=0}^{n_z} (s^{\alpha_1} + z_{i_z})}{\prod_{i_p=0}^{n_p} (s^{\alpha_1} + p_{i_p})} \quad (4.5.3)$$

4.5.2 Integral Fractional Order Transfer Function Parameters From Polar Components

In this section, the calculations will be explained by partitioning the Equation 4.5.3. These calculations are used in the estimation process and it explains how the parameters of the integral fractional order transfer function is obtained. Consider one pole fractional order system:

$$D(s, \alpha_1, p_1) = \frac{1}{s^{\alpha_1} + p_1} \quad (4.5.4)$$

where $s = j\omega_l$ and $T_{p_1} = \frac{1}{p_1}$ is the equivalent of the time constant and the equation becomes:

$$D(\omega_l, \alpha_1, p_1) = \frac{1}{T_{p_1} (j\omega_l)^{\alpha_1} + 1} \quad (4.5.5)$$

For $(j\omega_l)^{\alpha_1} = \omega_l^{\alpha_1} e^{j\frac{\pi}{2}\alpha_1}$ where $e^{j\frac{\pi}{2}\alpha_1} = \cos(\frac{\pi}{2}\alpha_1) + j \sin(\frac{\pi}{2}\alpha_1)$, we get:

$$D(\omega_l, \alpha_1, p_1) = \frac{1}{T_{p_1} \omega_l^{\alpha_1} \cos(\frac{\pi}{2}\alpha_1) + j T_{p_1} \omega_l^{\alpha_1} \sin(\frac{\pi}{2}\alpha_1) + 1} \quad (4.5.6)$$

For Magnitude

$$|D(\omega_l, \alpha_1, p_1)| = \frac{1}{\sqrt{T_{p_1}^2 \omega_l^{2\alpha_1} + 2 * T_{p_1} \omega_l^{\alpha_1} \cos(\frac{\pi}{2} \alpha_1) + 1}} \quad (4.5.7)$$

and Phase

$$\angle D(\omega_l, \alpha_1, p_1) = \tan^{-1} \left(\frac{-T_{p_1} \omega_l^{\alpha_1} \sin(\frac{\pi}{2} \alpha_1)}{T_{p_1} \omega_l^{\alpha_1} \cos(\frac{\pi}{2} \alpha_1) + 1} \right) \quad (4.5.8)$$

The transfer function part associated with the zeros can be calculated the same way:

$$|N(\omega_l, \alpha_1, z_1)| = \sqrt{T_{z_1}^2 \omega_l^{2\alpha_1} + 2 * T_{z_1} \omega_l^{\alpha_1} \cos(\frac{\pi}{2} \alpha_1) + 1} \quad (4.5.9)$$

and Phase

$$\angle N(\omega_l, \alpha_1, z_1) = \tan^{-1} \left(\frac{-T_{z_1} \omega_l^{\alpha_1} \sin(\frac{\pi}{2} \alpha_1)}{T_{z_1} \omega_l^{\alpha_1} \cos(\frac{\pi}{2} \alpha_1) + 1} \right) \quad (4.5.10)$$

Same way, the fractional integration $D(s, \alpha_0) = \frac{1}{s^{\alpha_0}}$ Magnitude and Phase is;

$$|D(\omega_l, \alpha_0)| = \omega^{\alpha_0} \quad (4.5.11)$$

$$\angle D(\omega_l, \alpha_0) = \frac{-\pi * \alpha_0}{2} \quad (4.5.12)$$

The equations above returns the magnitude and phase of the whole model is as follows;

$$|M_{\alpha n_p n_z}(\omega_l)| = K \frac{\prod_{i_z=0}^{n_z} |N(\omega_l, \alpha_1, z_{i_z})|}{|D(\omega_l, \alpha_0)| \prod_{i_p=0}^{n_p} |D(\omega_l, \alpha_1, p_{i_p})|} \quad (4.5.13)$$

$$\angle M_{\alpha n_p n_z}(j\omega_l) = \angle D(\omega_l, \alpha_0) + \sum_{i_z=0}^{n_z} \angle N(\omega_l, \alpha_1, z_{i_z}) - \sum_{i_p=0}^{n_p} \angle D(\omega_l, \alpha_1, p_{i_p}) \quad (4.5.14)$$

These formulas permit us to calculate the gain and phase of fractional order transfer functions of varying model orders.

4.6 Chapter Conclusion

In this chapter, the aim was to create a generic model with sufficient flexibility for identifying an appropriate model structure that can fit the observed impedance data. The model based approach was to consider a transfer function model to represent the underlying tongue impedance characteristics. The transfer function model with the choice of polynomial form combined with the naive least squares estimation did not yield satisfactory results. In fact it showed numerical instability in the calculation of the parameters even though a numerically stable version had been proposed in the literature, they had been applied to linearly varying frequencies and did not provide an adequate solution for the problem considered here. A more restricted transfer function model with real valued poles and zeros were considered. A parametrization that was linear in the log scale was considered as a means to overcome the numerical instability. However this model also could not account for specific patterns in the impedance data, most notably the slope at the low frequency end. The final choice of the model was the fractional transfer function form. Following chapter will present the results to model parameter estimation and model selection schemes using the final choice of model.

Chapter 5

Parameter Estimation and Model Selection

In the previous chapter, the first challenge of DDM, choosing a suitable model structure for EIM tongue data was addressed. In this chapter, the focus is on the estimation of the model parameters. In knowledge-based modeling, it may be possible to obtain the parameters from the experimental setup or background information; however, in data-driven modeling, the parameters are estimated from the chosen model and the observed data.

The modeling process presented in Section 4.3 has shown that the linear optimization methods are incapable of calculating stable poles and zeros. For this reason, this chapter considers the estimation of parameters with nonlinear optimization algorithms in an iterative search scheme. Among several optimization algorithms, three commonly used ones are: exhaustive search [198], metaheuristic algorithms [199, 200] and probabilistic methods [201]. Section 5.1 compares three algorithms; each belonging to one of these families for parameter estimation of the tongue EIM data model.

Model complexity selection is the last step in identifying a suitable model for the

observed data. The Integral Fractional Order Transfer Function model described in Section 4.5.2 is flexible for creating a model space with a set of candidate models from simplest to most complex form. A model selection criteria will be used to select a model from the model space to represent the system in its optimal form. In this framework, modeling is a type of dimensionality reduction that aims at selecting a model that represents the system in its simplest form. Section 5.2, compares two of the commonly used model selection criteria. The parameters of the selected model will be considered the features of the data that may have disease-related patterns.

5.1 Parameter Estimation

The model parameters that return values for the gain, poles, and zeros, which may contain discriminatory information between the healthy volunteers and patients, require reliable parameter estimation methods. There are conventional methods that are easy to implement, but they suffer from finding solutions that are not the global optimum. For linear optimization, the global optimum is reached when the optimal solution is obtained. In contrast, nonlinear optimization may have several local optimum among all other optimal solutions. In computational biology, there are examples of using Markov Chain Monte Carlo as a probabilistic approach and Genetic Algorithm as a heuristic method to find the best set of parameters while fitting the model to the observed data [202].

The fitting process incorporates a cost function to find parameters that minimize the error between the model and the observed data. In optimization problems, the cost function is a part of the process for estimating the model parameters commonly in an iterative scheme. The observed impedance data and the model output is given in complex form with $R(\omega_l) + jX(\omega_l)$ and $\tilde{R}(\omega_l) + j\tilde{X}(\omega_l)$ respectively. Consider parametrized model given in Equation 4.5.3: $M_{\alpha n_p n_z}$, we can write the function that

minimizes the Mean Squared Error (MSE) between the observations and the model as;

$$J_{\alpha n_p n_z}(\theta, M_{\alpha n_p n_z}) = \frac{1}{L} \sum_{l=1}^L [(R(\omega_l) - \tilde{R}(\omega_l))^2 + (X(\omega_l) - \tilde{X}(\omega_l))^2] \quad (5.1.1)$$

with

$$\tilde{\theta}(M_{\alpha n_p n_z}) = \underset{\theta}{\operatorname{argmin}} J_{\alpha n_p n_z}(\theta, M_{\alpha n_p n_z}) \quad (5.1.2)$$

where α, n_p, n_z represents fractional order, poles and zeros parameters respectively. $\tilde{\theta}(M_{\alpha n_p n_z})$ is the parameters that minimizes the error between the model and the observed data.

The parameters associated with the pole and zero locations will be changing in logarithmic scale, because the frequencies are increasing in a logarithmic scale. Hence the parameters have to be modeled as linearly varying in the logarithmic space. Parameters related to gain, fractional orders, poles and zeros are given as $\theta \in [\theta_K, \theta_{\alpha_0}, \theta_{\alpha_1}, \theta_{p_{i_p}}, \theta_{z_{i_z}}]$ where $i_p = 0, \dots, n_p$ and $i_z = 0, \dots, n_z$.

5.1.1 Grid Search

Grid Search algorithm is widely used for its simplicity and convenience to find an optimal solution to a given function. As the name suggests, the search space is divided into grids of given intervals, and each point on the grid is searched to find the parameters that gives the minimum error. The algorithm is implemented iteratively, and in each search, the size of grid interval is reduced. Grid search is an exhaustive search, with the flexibility of selecting the grid sizes, intervals and iterations. The algorithm is suitable for nonlinear parameter optimization problems.

In this project, Grid search is used to test the general model and model complexity while reaching to an optimal solution for the nonlinear parameter estimation. This method is considered simple and easy to implement to obtain a first cut model for the tongue EIM data. This section presents the algorithm implemented for this dataset

and the results of the model fit.

Algorithm

Grid Search algorithm requires parameter interval, grid sizes, model, cost function, data, and the number of iterations. The EIM data is measured in frequencies that vary between $\approx 76Hz$ to $625kHz$; this range is important for identifying the poles and zeros search interval. The poles and zeros have an inverse relation with the parameters (see Section 4.5.2), and the frequencies range between $[10^1, 10^6]$. In order to keep the range larger to identify the poles and zeros in a more flexible search space and considering the inverse relation, the search interval for the poles and zeros are set to be $[10^{-9}, 10^0]$. The value of gain K is likely to be indicative of the value of $\omega = 10^0$ assuming that the poles and zeros are placed much further than this frequency. Since the value of the magnitude around this frequency ranges between ≈ 0 and ≈ 200 , the search space for K is set to be $[10^{-2}, 10^3]$. To capture the slowly decaying/rising slopes in the magnitude plots, the fractional order parameters of the integration blocks (see Figure 4.4) are searched over the region $[10^{-4}, 10^0]$.

Step sizes are also a critical user-defined variable, which affects the grid sizes. Step size is set to 10; this means that each search space is going to be divided into 10 grids. A stopping criteria is defined to terminate the algorithm as the model fit is improved. The stopping criteria will avoid unnecessary iterations once the appropriate grid resolution is reached. The pseudo-code for the Grid Search method is given in Algorithm 2.

Results

The results of the Grid Search Algorithm applied to one of the patient's data is presented in Figure 5.1 and the parameters are given in Table 5.1. The poles, zeros, gain, and fractional-order results are presented with the parameters θ . The imple-

Algorithm 2 Grid Search for estimating gain, fractional orders, poles and zeros with Integral Fractional Order Transfer Function

Input: Angular frequency: $\omega_l = j2\pi f_l$
 Complex Resistance and Reactance data: $R(\omega_l) + jX(\omega_l)$
 Model: $M_{\alpha n_p n_z}(s) = K \frac{1}{s^{\alpha_0}} \frac{\prod_{i_z=0}^{n_z} (s^{\alpha_1} + z_{i_z})}{\prod_{i_p=0}^{n_p} (s^{\alpha_1} + p_{i_p})}$
 Parameters: $\theta \in [\theta_K, \theta_{\alpha_0}, \theta_{\alpha_1}, \theta_{p_{i_p}}, \theta_{z_{i_z}}]$
 Search intervals: $S_\theta = S_{\theta_K} \times S_{\theta_{\alpha_0}} \times S_{\theta_{\alpha_1}} \times S_{\theta_{p_{i_p}}} \times S_{\theta_{z_{i_z}}}$,
 for $S_{\theta_K} = [-2, 3]$, $S_{\theta_{\alpha_0}} = [-4, 0]$, $S_{\theta_{\alpha_1}} = [-4, 0]$, $S_{\theta_{p_{i_p}}} = [-9, 0]$, $S_{\theta_{z_{i_z}}} = [-9, 0]$
 Step Size = $G = 10$
 Stopping Criteria = $\delta = 0.1$: *If the error does not improve more than the stopping criteria, the algorithm terminates*

Output: Estimated Resistance and Reactance: $\tilde{R}(\omega_l), \tilde{X}(\omega_l)$

Parameters: $\tilde{\theta} \in [\tilde{\theta}_K, \tilde{\theta}_{\alpha_0}, \tilde{\theta}_{\alpha_1}, \tilde{\theta}_{p_{i_p}}, \tilde{\theta}_{z_{i_z}}]$

Initialize: $S_{\theta_{K_{low}}}, S_{\theta_{K_{high}}} = [-2, 3]$; $S_{\theta_{\alpha_0_{low}}}, S_{\theta_{\alpha_0_{high}}} = [-4, 0]$; $S_{\theta_{\alpha_1_{low}}}, S_{\theta_{\alpha_1_{high}}} = [-4, 0]$;

$S_{\theta_{p_{i_p}_{low}}}, S_{\theta_{p_{i_p}_{high}}} = [-9, 0]$; $S_{\theta_{z_{i_z}_{low}}}, S_{\theta_{z_{i_z}_{high}}} = [-9, 0]$;

$\Delta_K = \frac{S_{\theta_{K_{high}}} - S_{\theta_{K_{low}}}}{G}$; $\Delta_{\alpha_0} = \frac{S_{\theta_{\alpha_0_{high}}} - S_{\theta_{\alpha_0_{low}}}}{G}$; $\Delta_{\alpha_1} = \frac{S_{\theta_{\alpha_1_{high}}} - S_{\theta_{\alpha_1_{low}}}}{G}$;

$\Delta_{p_{i_p}} = \frac{S_{\theta_{p_{i_p}_{high}}} - S_{\theta_{p_{i_p}_{low}}}}{G}$; $\Delta_{z_{i_z}} = \frac{S_{\theta_{z_{i_z}_{high}}} - S_{\theta_{z_{i_z}_{low}}}}{G}$

while $J_{\alpha, n_p, n_z}(\tilde{\theta}, M_{\alpha n_p n_z})[t] - J_{\alpha n_p n_z}(\tilde{\theta}, M_{\alpha n_p n_z})[t-1] > \delta$ **do**

for K within range $[S_{\theta_{K_{low}}} - S_{\theta_{K_{high}}}]$ **do**

for p_1 within range $[S_{\theta_{p_{1_{low}}}} - S_{\theta_{p_{1_{high}}}}]$ **do**

 ... number of for loops depends on the number of poles

for z_1 within range $[S_{\theta_{z_{1_{low}}}} - S_{\theta_{z_{1_{high}}}}]$ **do**

 ... number of for loops depends on the number of zeros

for α_0 within range $[S_{\theta_{\alpha_0_{low}}} - S_{\theta_{\alpha_0_{high}}}]$ **do**

for α_1 within range $[S_{\theta_{\alpha_1_{low}}} - S_{\theta_{\alpha_1_{high}}}]$ **do**

 Calculate $J_{\alpha n_p n_z}(\tilde{\theta}, M_{\alpha n_p n_z})[t]$ cost function in Equation 5.1.1

if $J_{\alpha n_p n_z}(\tilde{\theta}, M_{\alpha n_p n_z})[t] - J_{\alpha n_p n_z}(\tilde{\theta}, M_{\alpha n_p n_z})[t-1] < \delta$ **then**

return $\tilde{R}(\omega_l) + j\tilde{X}(\omega_l)$

else Recalculate the grid sizes from the new grid intervals

end if

end for

end for

end for

end for

end for

end while

mentation process deals with several problems. One problem arises when the number of parameters increases, the computational cost increases exponentially, and to compensate for this, the number of iterations that are needed to reach an optimal solution also increases. Most importantly, the algorithm is able to find an optimal solution; however, this is not guaranteed to be a global optimum.

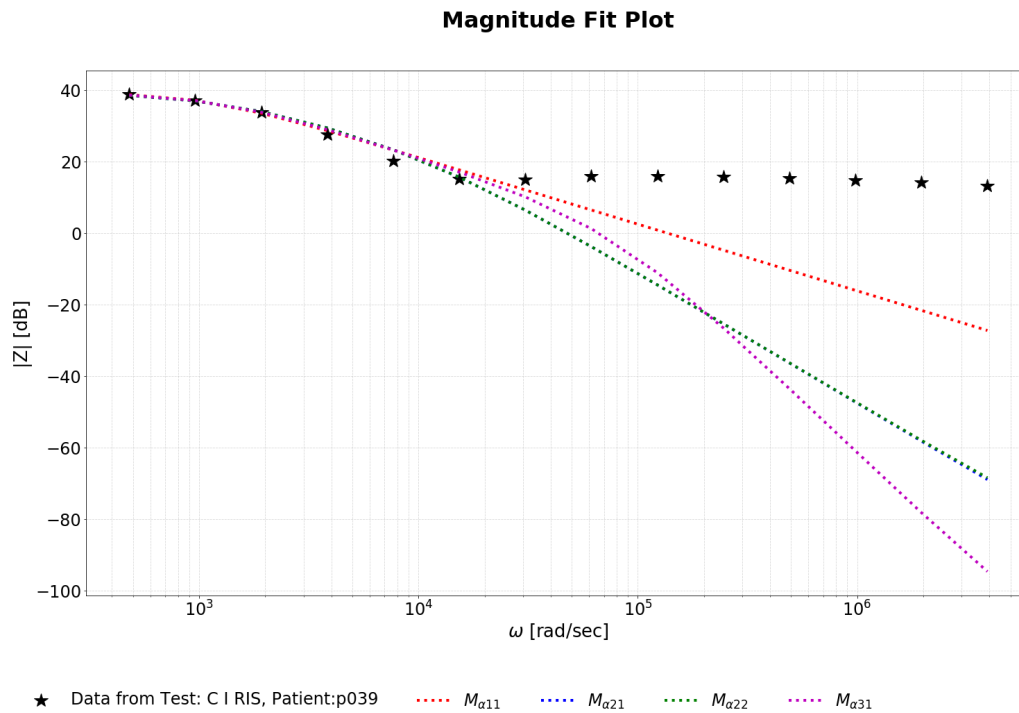


Figure 5.1: Model fit for one example data. The four model fit are plotted: 1-zero 1-pole, 1-zero 2-pole, 2-zero 2-pole, 1-zero 3-pole.

	θ_{α_0}	θ_{α_1}	θ_K	θ_{i_p}	θ_{i_z}
$M_{\alpha 11}$	-2.16	-0.04	2.30	-2.81	-9.90
$M_{\alpha 21}$	-3.17	-0.08	2.30	-2.61, -3.19	-3.51
$M_{\alpha 22}$	-4.99	-0.08	2.29	-2.61, -3.26	-3.88, -3.88
$M_{\alpha 31}$	-1.64	-0.17	2.39	-2.41, -2.41, -3.72	-6.60

Table 5.1: Model fit transfer functions with poles and zeros for Figure 5.1

As the model complexity increases, we will be looking into estimating parameters

with Grid Search with a grid of g equally spaced values and the number of parameters of n_θ making up to $(g + 1)^{n_\theta}$ search space of possible combinations of parameters. This search space causes a high computational cost and therefore makes the method intractable for complex models. For this reason the analysis investigated the alternatives to grid search especially algorithms that can find global optimum with metaheuristic methods such as Genetic Algorithm and probabilistic methods such as Markov Chain Monte Carlo.

5.1.2 Genetic Algorithm

Genetic Algorithm is a widely used global optimization method in parameter estimation for nonlinear optimization problems [203]. The algorithm is inspired by the survival of the fittest process in Darwin's evolution theory, and it is a member of the evolutionary algorithms family [204]. The algorithm selects the best-fitted parameter sets and mates them to create offsprings for the next generation. This is iterated through generations until a stopping condition is met and the final parameter set is then selected from the best fit model of the last iteration. In this respect, the number of iterations plays an essential role in reaching the fittest parameter set (elitist strategy). This section presents the details of the algorithm, and the results obtained from the implementation of the algorithm to the EIM data. A modification to this algorithm is also suggested.

Algorithm

The algorithmic steps are derived from the genetic splicing in nature, and each term has an equivalent term to the model parameter estimation. Figure 5.2 illustrates the steps of the algorithm in a flow diagram, and the terms are expanded within the subfigures. In further detailing the relationship of the terms to the model estimation, the chromosome is the parameter set, the gene is the individual parameters in

a parameter set, the population is the collection of chromosomes created with the random values selected within the interval of each parameter. The fitness of each chromosome is the quality of the model fit to data and is calculated with Equation 5.1.1. A number of best-fitted chromosomes are selected and placed into a mating pool. In the mating process, the selected parents in the mating pool are employed for crossover and mutation of the genes. The offsprings are created as a result of crossover and mutation, and new offsprings are created in each iteration (generation). When the set number of iterations are reached the last offspring is selected as a final parameter set.

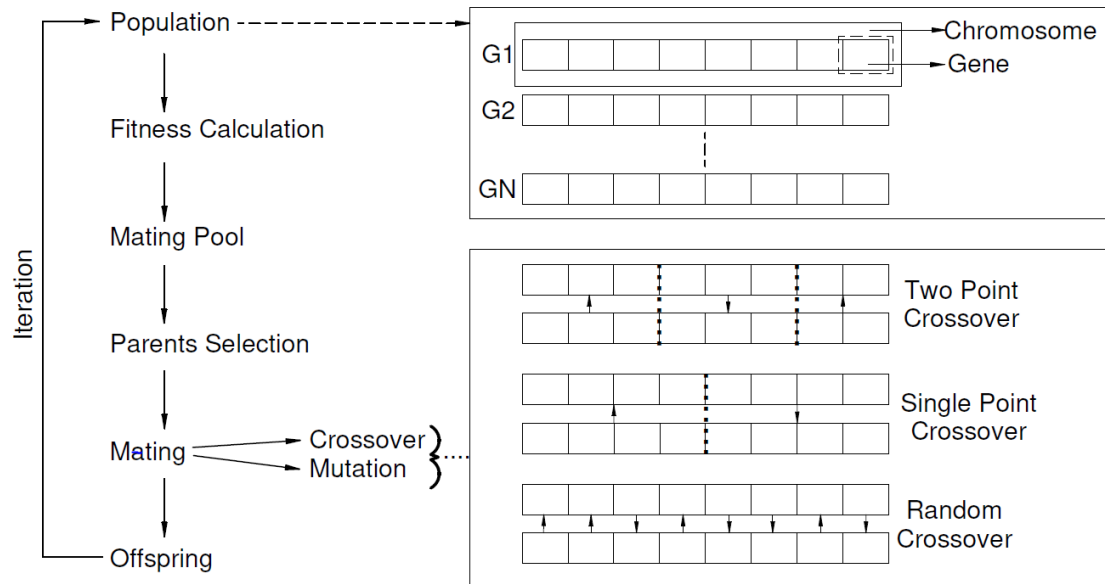


Figure 5.2: Genetic Algorithm flow chart with population, gene, chromosome terms and crossover types.

Fine-tuning of the variants for the GA has resulted in selecting the population size of 10000, a mating pool of 1000. There are different types of crossover in this algorithm; single point, two-point, and random. Single point crossover is implemented by dividing the chromosome into two and swapping the two parts between the par-

ents. Two-point crossover (or multi-point) is implemented by dividing the chromosome into three (or more) and swapping these parts between the parents. For a random crossover, the genes are crossed over randomly, without following a pattern. In this implementation, random crossover is used. The mutation is the process of altering the offsprings with random values; hence, a safety measure is taken to keep the parameters within range by checking whether the parameter is exceeding the initial limits.

Results

The parameters of the models are estimated using the Genetic Algorithm described above by minimizing the error function $J_{\alpha n_p n_z}(\theta, M_{\alpha n_p n_z})$ given in Equation 5.1.2. Search intervals were used according to 2 with respect to the discussion given in Section 5.1.1. Results for different models are presented in Figure 5.3 and the parameters are given in Table 5.2. The results indicate that as the model complexity increases (as the number of poles and zeros increase), the error reduces. However, for the 3-pole 1-zero model, the error appears to increase as the model has difficulty fitting the data at high frequencies. The θ_{α_1} is close to zero for most models which indicates there is no fractional order associated with the poles and zeros. Parameters shared from models of different complexity have similar values which suggests that the estimation is reliable.

The advantage of the Genetic algorithm over the conventional algorithms is that it creates a population of the fittest parameter sets and creates generations based on this pool rather than creating a single solution in each iteration. This helps with approaching the global minimum while avoiding the local minimum problem. However, each parameter's final values are the single best solution and that does not give information on the uncertainty of estimated parameters.

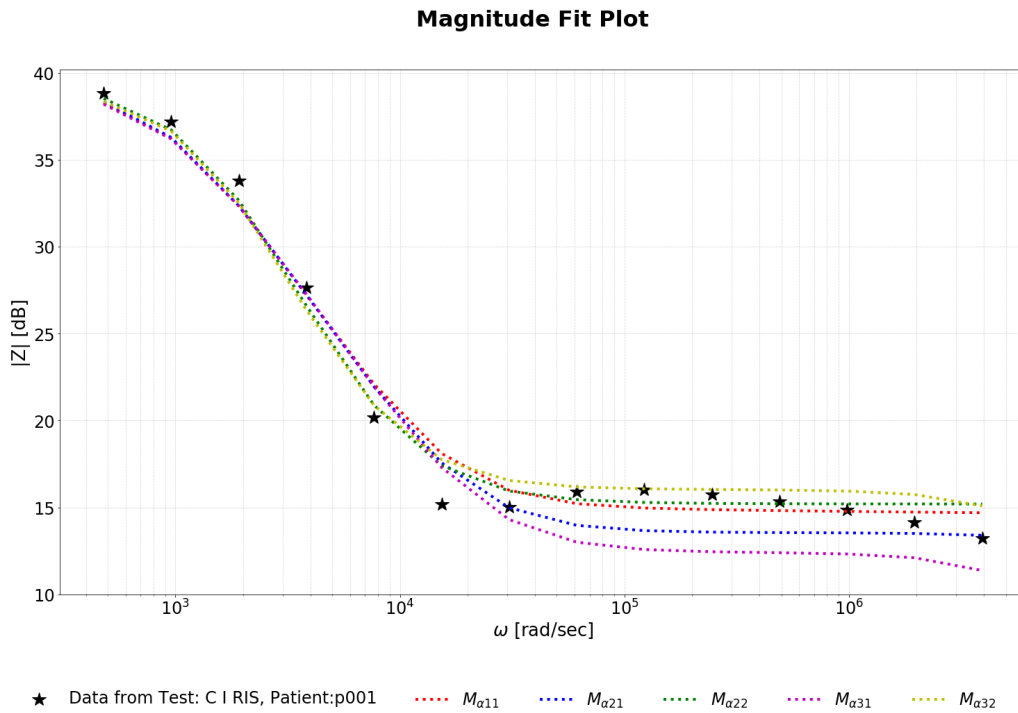


Figure 5.3: Model fit for one example data. The five model fit with different complexities are plotted: 1-zero 1-pole, 1-zero 2-pole, 2-zero 2-pole, 1-zero 3-pole, 2-zero 3-pole.

	θ_{α_0}	θ_{α_1}	θ_K	θ_{i_p}	θ_{i_z}
$M_{\alpha 11}$	-2.18	-0.002	1.98	-2.98	-4.18
$M_{\alpha 21}$	-3.61	-0.003	1.97	-2.96, -7.31	-4.25
$M_{\alpha 22}$	-4.17	-0.015	1.97	-3.01, -3.23	-3.59, -3.86
$M_{\alpha 31}$	-3.70	-0.005	1.96	-2.95, -6.90, -7.12	-4.29
$M_{\alpha 32}$	-3.08	-0.005	1.96	-3.12, -3.24, -6.85	-3.57, -3.94

Table 5.2: Model fit transfer functions with poles and zeros for Figure 5.3

Frequency-Related Weighting Factor

The magnitude of frequency response function in Figure 4.3, measurements at the high frequency data have low magnitude relative to the rest of the data, and these measurements may be more informative compared to the rest of the measurements as this region is in the β band (see Section 2.3.3). Moreover, the model fit results in

Figure 5.3 has shown that the models have difficulty fitting to high frequency data. Applying weighting to the measurements in this band may help provide a better fit to this part of the data. For this reason, the weighting factors are calculated from the data as follows;

$$\mu_R = \frac{1}{\sum_{l=1}^L (R(\omega_l))} \quad (5.1.3)$$

$$\mu_X = \frac{1}{\sum_{l=1}^L (X(\omega_l))} \quad (5.1.4)$$

$$W(\omega_l) = \frac{\log_{10} \omega_l}{\sum_{l=1}^L (\log_{10} \omega_l)} \quad (5.1.5)$$

where μ_R and μ_X have an inverse relation with the mean values of resistance and reactance readings of the whole data along the spectrum respectively. $W(\omega_l)$ has the property of $0 < W(\omega_l) < 1$ and $\sum_{l=1}^L W(\omega_l) = 1$. The weights of the resistance and reactance are calculated from the combination of these values as;

$$W_R(\omega_l) = \mu_R * W(\omega_l) \quad (5.1.6)$$

$$W_X(\omega_l) = \mu_X * W(\omega_l) \quad (5.1.7)$$

Weighting Factors

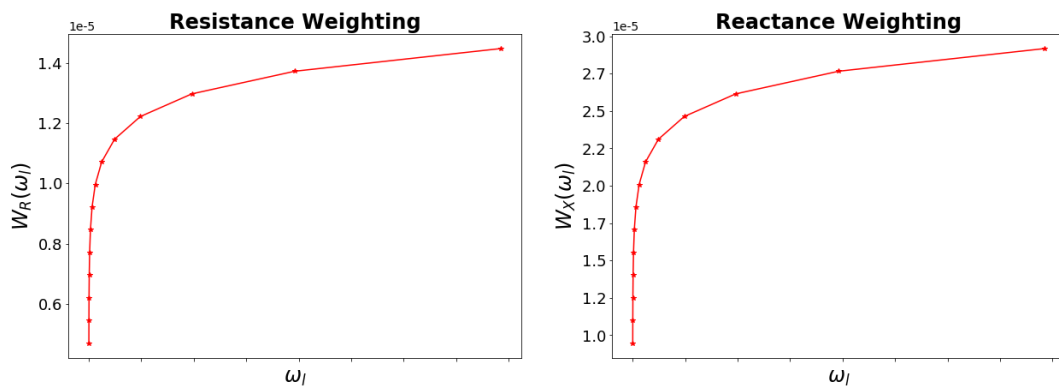


Figure 5.4: Frequency-related weighting factors for resistance and reactance.

The weightings $W_R(\omega_l)$ and $W_X(\omega_l)$ are given in Figure 5.4. The reason for different weightings used for resistance and reactance is because the range of resistance and reactance are different. The purpose of using μ_R and μ_X is to get a balance between resistance and reactance readings as the order of resistance is greater than order of reactance. The idea behind the frequency-related weighting is to sensitize the high frequency to capture poles and zeros from the model. From the above equations, the weighted cost function is calculated as follows;

$$J_W(\theta) = \sum_{l=1}^L [W_R(\omega_l) * (R(\omega_l) - \widetilde{R}(\omega_l))^2 + W_X(\omega_l) * (X(\omega_l) - \widetilde{X}(\omega_l))^2] \quad (5.1.8)$$

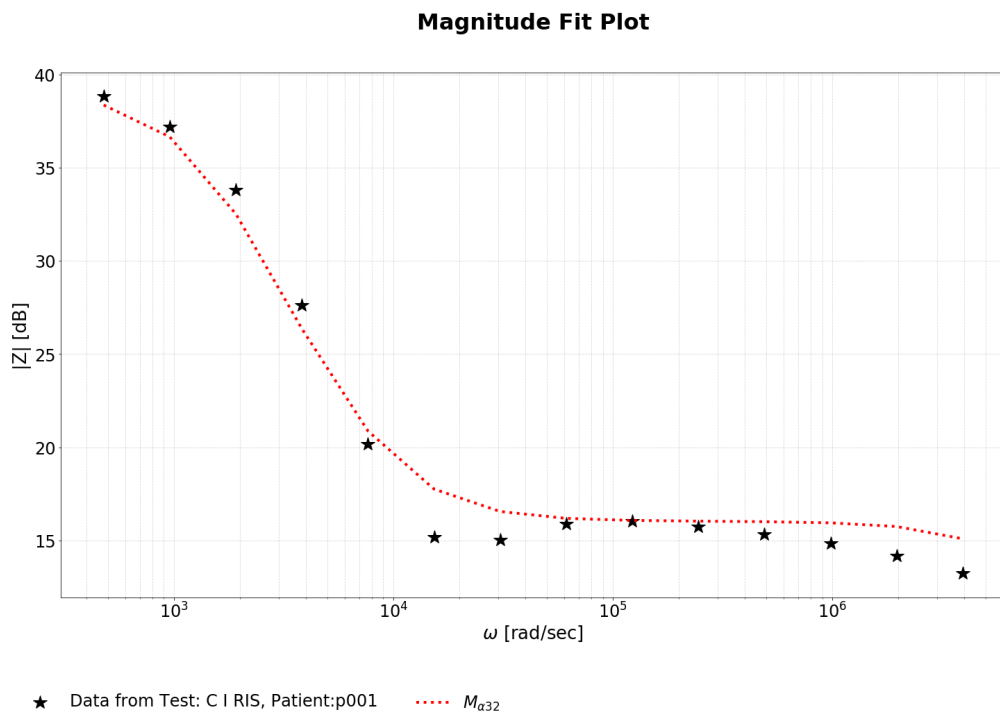


Figure 5.5: Frequency-related weighted 2-zero, 3-pole model fit for one example data with parameters $\theta_{\alpha_0} = -3.42$, $\theta_{\alpha_1} = -0.004$, $\theta_K = 1.96$, $\theta_{i_p} = -3.22, -3.28, -7.29$, $\theta_{i_z} = -3.84, -3.91$.

The fit results are presented on one dataset with resistance and reactance. Compared to the unweighted fit presented in page 92 from genetic algorithm results, the weighting has not improved the fit. In fact it reduced the variations in magnitude needed by the model at the higher frequencies. The weighting is not taken any further for this research because the important poles and zeros that have disease-related patterns are unknown and likely located at the high frequency data. These details may be captured with a better model estimation instead of a weighted estimation approach.

5.1.3 Markov Chain Monte Carlo

Genetic Algorithms have many advantages over conventional methods. For example, the mutation stage helps with expanding the search space for exploring nearby options to reach an optimal solution and is applicable to a wide range of optimization problems. On the contrary, the number of iterations and population size may result in difficulty in convergence. Moreover, the obtained parameters in Genetic Algorithms are single best results that do not show the estimation uncertainty.

Bayesian inference represents the unknown model parameters θ with a posterior probability distribution. This distribution captures the uncertainties in the model. The estimation of the posterior distribution of the model requires integration over the multiple parameters. This calculation may be easy for simple models, but as the model complexity increases, the computational complexity increases. Markov Chain Monte Carlo (MCMC) is a class of algorithms that can provide a solution to the complex models that requires integration over multiple parameters.

MCMC method refers to the memory-less stochastic process that uses random samples from a conditional posterior distribution to calculate the model output with the given input and parameter space. One of the advantages of the method comes from its computational convenience, where the next estimation process with the ran-

domly chosen parameters does not rely on past information, and it only depends on the current settings. The process of the next estimation, depending on the current state, is called the Markov Chain, and the process of repetitive random sampling is called the Monte Carlo. The following sections introduce the algorithm used to estimate the parameters of the model and the estimation results.

Algorithm

The implementation of the MCMC method is made possible with various algorithms. Some of the commonly used ones are the Metropolis-Hastings (MH) algorithm, Gibbs Sampler, and Hamiltonian [205, 206]. MH algorithm is the simplest one that relies on the accept/reject principle. Hamiltonian (aka Hybrid) is computationally more expensive but more efficient considering the large jumps between the steps, generates fewer rejection samples, and therefore converges to the posterior distribution faster [206]. In Gibbs Sampling, the parameter space is split into its parameters, and each parameter is sampled separately, while the rest of the samples remain fixed to their current values. In most references, in its simplest form, Gibbs Sampling is considered a particular case of the MH algorithm [207]. This similarity is due to the same random walk rules applied to the Gibbs Sampling. The only difference comes from the sampling of each parameter in turn in a cycle within the parameter space and the conditional posterior distribution caused by this cycle.

The initialization step of the MH algorithm requires a parameter distribution. This distribution comes from the assumption made based on the visual inspection of the data set. Initial values for the parameters (also known as proposal distribution or a transition model) are selected from this distribution. The current state parameters return a model fit error. Another set of parameters are drawn from the distribution and will be used to calculate the model fit error for this set. The comparison of the current state's errors and the new random state helps decide which set of parameters

become the next state. If the error of the proposed set is higher than the current state, then the proposed set is rejected, else the proposed set is accepted as the new current state. The algorithm will repeat proposing a new set of parameters from the current distribution until the specified number of iterations are completed. The resulting chain provides the distribution for the parameters. The algorithm steps can be expressed in terms of Bayesian Inference, and the following mathematical expressions show this relation. In Bayesian Inference;

$$P(\theta/D) = \frac{P(D/\theta)P(\theta)}{P(D)} \quad (5.1.9)$$

where; $P(\theta/D)$ is the posterior distribution that we wish to compute, $P(D/\theta)$ is the likelihood, $P(\theta)$ is the prior and $P(D)$ is the evidence. The selection process of the new set of parameters involves taking the proportion of the new and current posterior distribution. The proportion is given by:

$$\frac{P(\theta'/D)}{P(\theta/D)} = \frac{\frac{P(D/\theta')*P(\theta')}{P(D)}}{\frac{P(D/\theta)*P(\theta)}{P(D)}} \quad (5.1.10)$$

that cancels out the evidence $P(D)$. Canceling the evidence in this proportion is one of the convenience that comes from the MCMC methods because the evidence term involves the integration over the parameter space, and as the model complexity increases, the computational complexity of the evidence term increases. The proportion expresses that if the likelihood of the new proposal is higher than the current, the proposal $P(\theta'/D)$ is accepted, else it is rejected. The likelihood of each proposed sample of parameters is calculated by a function f . Including function f into the Eq.(5.1.10) returns;

$$\frac{P(D/\theta') * P(\theta')}{P(D/\theta) * P(\theta)} = \frac{\prod_i^n f(d_i/\theta') * P(\theta')}{\prod_i^n f(d_i/\theta) * P(\theta)} \quad (5.1.11)$$

Results

The uncertainty in data is not considered in methods like Genetic Algorithm, and Grid search. The uncertainty may arise from the lack of knowledge of the physical background knowledge of the system or noise in the measurement process. The results of an MCMC algorithm returns the model parameters with the underlying uncertainty.

One data set from [C I RIS] electrode configuration and patient *p039* is used for estimation of model parameters using MCMC. From Bayesian Inference, the posterior density for the probability distribution of the parameter values is calculated with the likelihood and the prior information about the assumptions made on the visual inspection on the parameters. Search intervals were used according to Algorithm 2 with respect to the discussion given in Section 5.1.1. In the implementation, the iterations will repeat the process until the desired number of realizations is completed. The results are presented after considering the burnin of the chain. Burnin in MCMC is used to remove the first part of the chain, where the samples has not been drawn from the stationary posterior distribution. The burnin considered a wrong term for the MCMC in many resources [208] [209] since it is a memoryless random walk. However, as long as the chain is not infinite, the approximation at the beginning of the chain will have an adverse effect on the distribution and therefore needs to be removed for the samples to be more associated with the real distribution. There are no real rules to select this value; therefore, selecting samples only from the last portion of the iteration to represent the whole distribution is reasonable. The output results present the model parameters' statistical properties and the credible interval of the model to represent the model uncertainty. The distribution of the parameters after the burnin iterations are removed, are given in Figure 5.6a. In this figure it is shown that the parameters related to the gain K , fractional order α_1 , pole and zero p_2, z_1 are clearly identified however there are ambiguities in parameters related

to pole p_1 and fractional order α_0 . In Figure 5.6b, the correlation between the parameters is plotted from the chain results in order to see if there is any correlation between two parameters. A high correlation between the two parameters may mean that one parameter may be determined from the paired parameter, and this is undesired because it shows that the correlating parameters are unnecessary. A correlation is visible between α_1 and p_1 parameters for this model.

The Figure 5.7 shows the credible and prediction intervals for $M_{\alpha 21}$ model. The uncertainty of the parameters obtained from the chain is reflected in the credible intervals. The prediction intervals represent the parameter uncertainty, together with the observation uncertainty, to depict the future predictions. The results of the credible intervals are compared to another model in order to see the effects of the parameter uncertainty on the model uncertainty. As seen in the $M_{\alpha 21}$ model, the credible interval is high in the high frequency data, and this uncertainty is reduced in the $M_{\alpha 32}$ model. This indicate that the appropriate model complexity may reduce the uncertainty in the high frequency data.

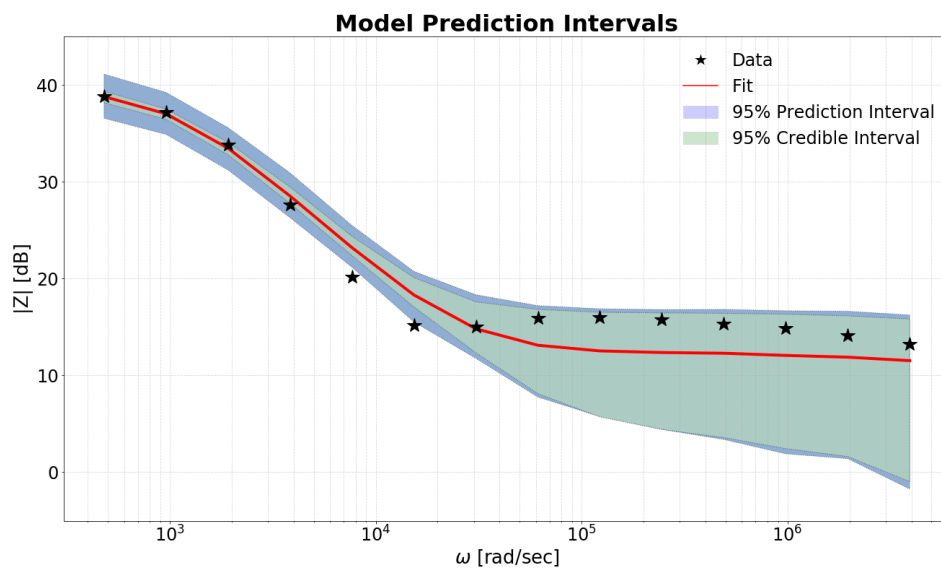


Figure 5.7: 2-pole 1-zero model fit, credible and prediction intervals for one example data.

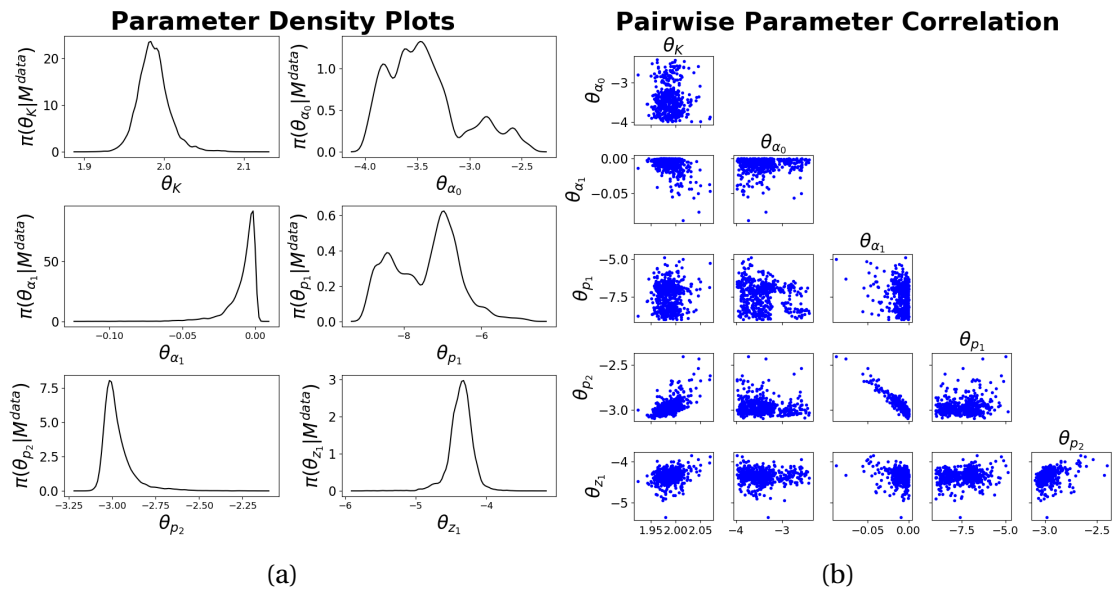


Figure 5.6: MCMC chain results with (a) Parameter density plot obtained from the chain after burnin is removed and (b) Pairwise parameter correlation plot to detect the correlation between pairs of parameters.

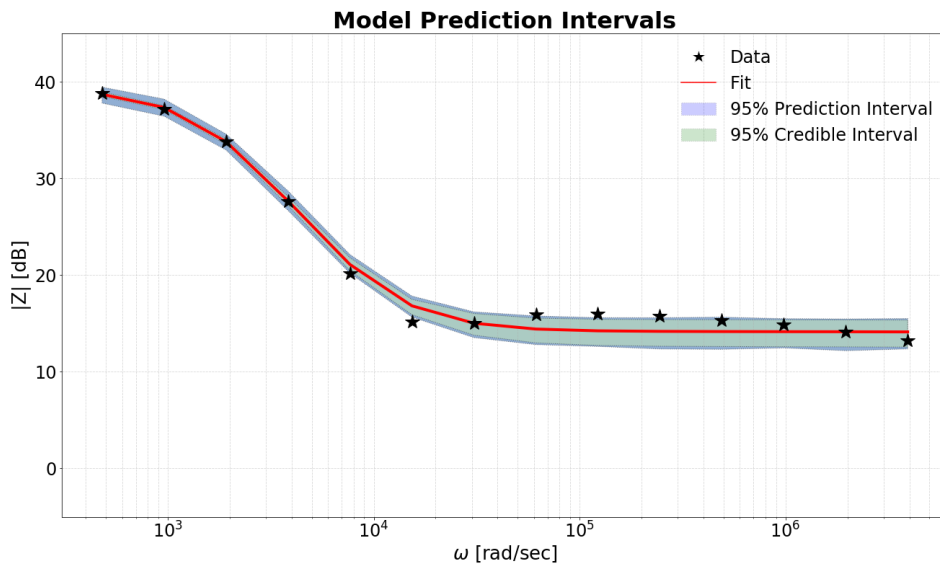


Figure 5.8: 3-pole 2-zero model fit, credible and prediction intervals for one example data.

5.2 Model Selection

Black-box modeling deals with several uncertainties, such as the observation noise, the construction of the model, the model parameters, or the initialization of the parameter space. A model selection procedure is essential to avoid overfitting and under-fitting that comes from these uncertainties. Section 4.5 on estimation has shown that the fractional-order transfer function model is suitable for this data. It is essential to acknowledge that increasing the model complexity may improve the model fit; however, complex models tend to capture noise, whereas simpler models tend to capture essential characteristics of the data. For this reason, a model space consisted of several candidate fractional-order transfer function model is used in the model selection procedure.

Visual inspection suggested the physical background of the EIS data can be explained with a minimum of one pole, one zero, and the fractional parameters as well as the gain for the measurement and the instrument processes. However, the appropriate model complexity cannot be found without a full investigation of all model orders. The $M_{\alpha 00}$ model has only the fractional integration part of the system together with gain, given in $\frac{K}{s^{\alpha_0}}$ with two parameters related to K, α_0 . The model $M_{\alpha 10} = \frac{K}{s^{\alpha_0} * (s^{\alpha_1} + p_1)}$ with four parameters related to $K, \alpha_0, \alpha_1, p_1$. Here we can conclude that the models with minimum of one pole has the parameter space dimension of $n_{\theta} = n_p + n_z + 3$ and the model order with no poles and zeros has the dimension of $n_o = 2$. The input-output data is has 14 data points and model parameters cannot exceed this number in order to avoid both the overfitting and over-parametrization. The model with $n_p = 5, n_z = 5$, we have parameter space with the size of 13. The fifth order model $M_{\alpha 55}$ helps explore full fifth order model space, however higher-order model spaces will exceed the number of data points. The estimated models will start from $M_{\alpha 00}$ and the maximum number of parameters will be achieved in $M_{\alpha 55}$. The

anticipated model space with candidate model is $M_{\alpha i_p i_z} = [M_{\alpha 00}, M_{\alpha 10}, M_{\alpha 11}, M_{\alpha 20}, M_{\alpha 21}, M_{\alpha 22}, M_{\alpha 30}, M_{\alpha 31}, M_{\alpha 32}, M_{\alpha 33}, M_{\alpha 40}, M_{\alpha 41}, M_{\alpha 42}, M_{\alpha 43}, M_{\alpha 44}, M_{\alpha 50}, M_{\alpha 51}, M_{\alpha 52}, M_{\alpha 53}, M_{\alpha 54}, M_{\alpha 55}]$. Here the i_p is the number of poles and determines the model order, and i_z is the number of zeros. The model space has models with the number of poles greater than or equal to the number of zeros $n_p \geq n_z$ models for proper systems. The number of parameters for the chosen models are $n_{M_{\alpha i_p i_z}} = [2, 4, 5, 5, 6, 7, 6, 7, 8, 9, 7, 8, 9, 10, 11, 8, 9, 10, 11, 12, 13]$.

5.2.1 Akaike Information Criterion

Akaike Information Criterion (AIC) is one of the most commonly used model selection method. The method is designed to make a trade off between the goodness of model and the simplicity of the model. This means that the model is useful for dealing with two of the model selection problems; overfitting and underfitting. The method does that by calculating a value for each model with the following equation;

$$AIC_{M_{\alpha n_p n_z}} = 2n_{\theta} - 2\ln(L) \quad (5.2.1)$$

where n_{θ} is the number of the parameter in the parameter space of that particular model and L is the likelihood function for the same model [210]. The equation indicates a reward for the goodness of fit with the likelihood function and a penalty for the number of parameters. For this reason, the AIC model selection is more about the quality of the model relative to the quality of the other models in the model space. Figure 5.9 compares the models for all electrode configurations from the Genetic Algorithm results.

AIC Model Comparison Plot

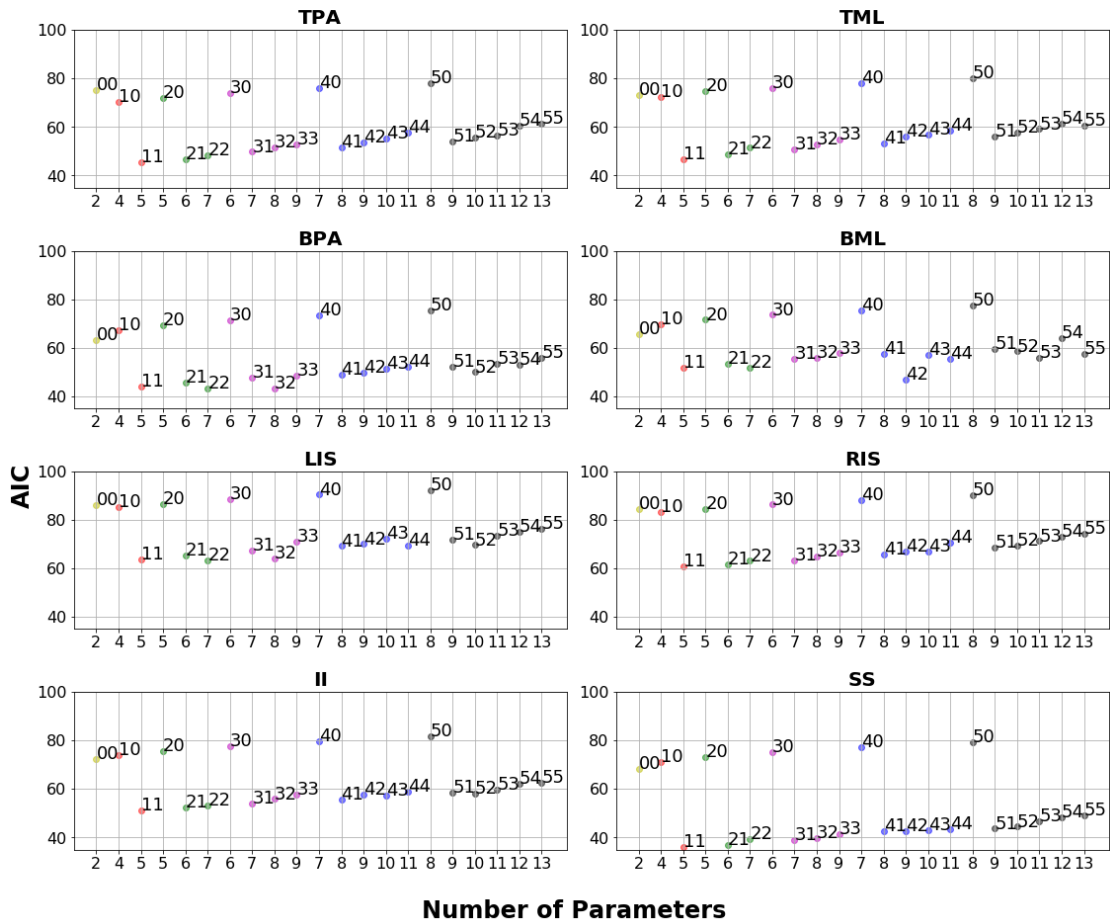


Figure 5.9: AIC model selection for all electrode configurations. The models are represented with n_p, n_z such as 21 is $n_p = 2$ and $n_z = 1$ for model $M_{\alpha 21}$.

Figure 5.9 is converted to a rank based table in order to select the best performing models. For this table, the models are ranked for each electrode configuration, with the smallest AIC value having the lowest rank. The ranks are then averaged for each model and the new rank is identified. Table 5.3 gives this new rank in the last column and indicates that $M_{\alpha 11}, M_{\alpha 21}, M_{\alpha 22}, M_{\alpha 31}, M_{\alpha 32}$ models has the minimum AIC values, with $M_{\alpha 11}$ model having the top rank and the $M_{\alpha 21}, M_{\alpha 22}$ models having the same rank value. The models with no zeros has the highest AIC values that puts them

lowest in the ranking, which is expected considering the frequency response function given in Figure 4.3. For this reason, in order to make a conclusion on the discriminatory characteristics of the parameters, the models with best ranked AIC values will be considered for further analysis.

	<i>BPA</i>	<i>BML</i>	<i>TPA</i>	<i>TML</i>	<i>LIS</i>	<i>RIS</i>	<i>II</i>	<i>SS</i>	AVG. SCORE	NEW RANK
00	16	16	19	17	17	18	16	16	16.9	17
10	17	17	16	16	16	16	17	17	16.5	16
11	3	2	1	1	2	1	1	1	1.5	1
20	18	18	17	18	18	17	18	18	17.8	18
21	4	4	2	2	4	2	2	2	2.8	2
22	1	3	3	4	1	3	3	4	2.8	2
30	19	19	18	19	19	19	19	19	18.9	19
31	5	5	4	3	5	4	4	3	4.1	4
32	2	8	5	5	3	5	6	5	4.9	5
33	6	12	7	7	10	7	8	6	7.9	8
40	20	20	20	20	20	20	20	20	20	20
41	7	11	6	6	6	6	5	7	6.8	6
42	8	1	8	8	9	9	9	8	7.5	7
43	10	9	10	10	12	8	7	9	9.4	9
44	11	6	13	12	7	12	12	10	10.4	10
50	21	21	21	21	21	21	21	21	21	21
51	12	14	9	9	11	10	11	11	10.9	12
52	9	13	11	11	8	11	10	12	10.6	11
53	14	7	12	13	13	13	13	13	12.3	13
54	13	15	14	15	14	14	14	14	14.1	14
55	15	10	15	14	15	15	15	15	14.3	15

Table 5.3: AIC model selection rank based table for all electrode configurations.

5.2.2 Bayesian information criterion

Bayesian information criterion (BIC) is another commonly used model selection method that penalizes the model complexity more heavily than the AIC. This means that more complex models will have a higher BIC;

$$BIC_{M_{\alpha p n z}} = \log(N)n_{\theta} - 2\ln(L) \quad (5.2.2)$$

where L is the likelihood function, n_θ is the number of parameters in model [210]. Compared to AIC, the penalty value is $2n_\theta$ for AIC and $\log(N)n_\theta$ for BIC where N is the number of data points. This indicates that the BIC method penalizes model complexity dependent on the number of data points. The Figure 5.10 compares the models for [RIS] electrode configuration from the Genetic Algorithm results.

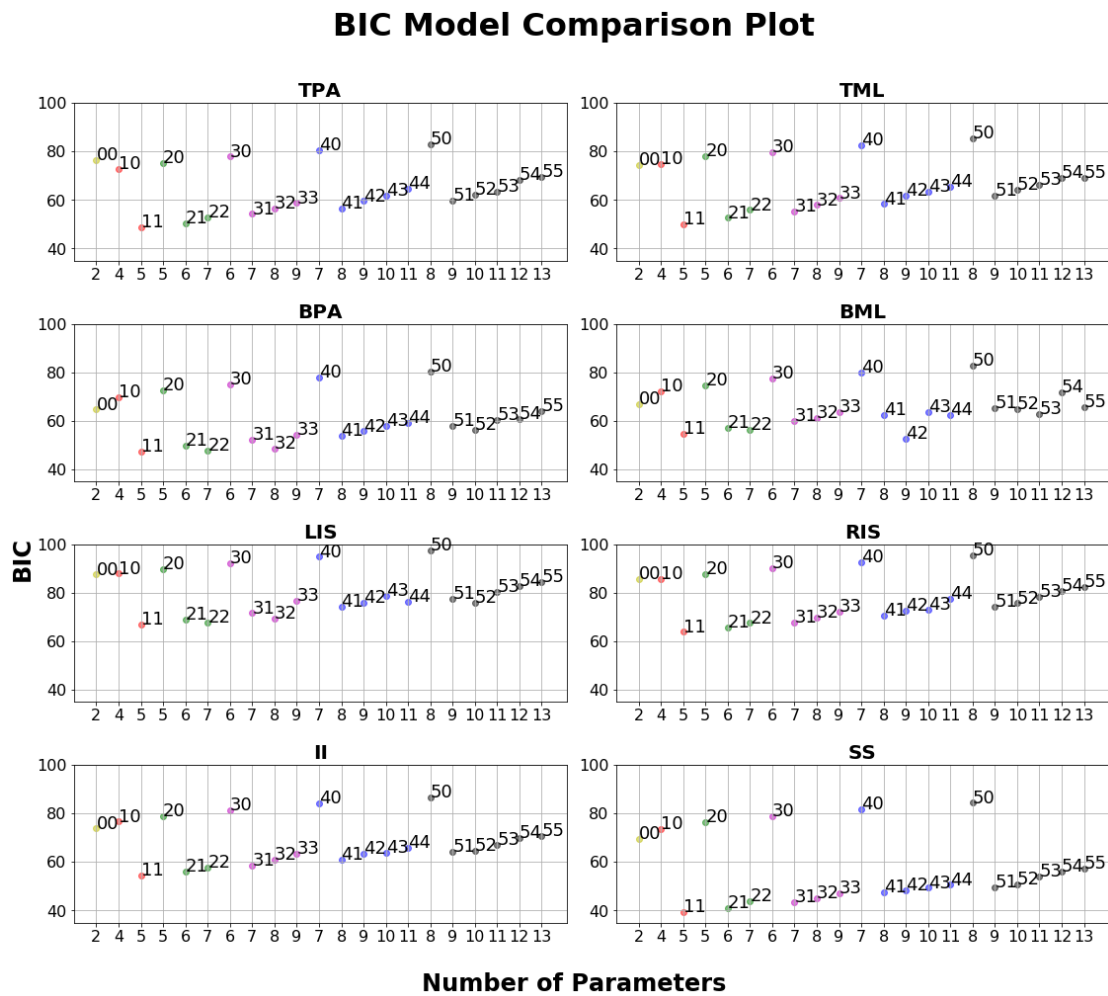


Figure 5.10: BIC model selection for all electrode configurations. The models are represented with n_p, n_z such as 21 is $n_p = 2$ and $n_z = 1$ for model $M_{\alpha 21}$.

The Figure 5.10 is converted to a rank based table in order to select the best per-

forming models. The Table 5.4 indicates that $M_{\alpha 11}, M_{\alpha 21}, M_{\alpha 22}, M_{\alpha 31}, M_{\alpha 32}$ models has the minimum BIC values, with $M_{\alpha 11}$ model having the top rank. The models with best ranked BIC values will be considered in further analysis. These results agrees with AIC results except BML electrode configuration which shows the minimum BIC value at $M_{\alpha 42}$ model while the overall ranking is around 8 for other electrode configurations. Moreover, models $M_{\alpha 00}, M_{\alpha 10}$ and $M_{\alpha 44}, M_{\alpha 51}$ changes in ranking due to extra penalty applied by BIC for number of parameters.

	<i>BPA</i>	<i>BML</i>	<i>TPA</i>	<i>TML</i>	<i>LIS</i>	<i>RIS</i>	<i>II</i>	<i>SS</i>	AVG. SCORE	NEW RANK
00	16	15	18	16	16	17	16	16	16.3	16
10	17	17	16	17	17	16	17	17	16.8	17
11	1	2	1	1	1	1	1	1	1.1	1
20	18	18	17	18	18	18	18	18	17.9	18
21	4	4	2	2	3	2	2	2	2.6	2
22	2	3	3	4	2	3	3	4	3	3
30	19	19	19	19	19	19	19	19	19	19
31	5	5	4	3	5	4	4	3	4.1	4
32	3	6	5	5	4	5	6	5	4.9	5
33	7	10	7	7	10	7	7	6	7.6	8
40	20	20	20	20	20	20	20	20	20	20
41	6	7	6	6	6	6	5	7	6.1	6
42	8	1	8	8	7	8	8	8	7	7
43	10	11	10	10	12	9	9	9	10	9
44	12	8	13	12	9	12	12	11	11.1	12
50	21	21	21	21	21	21	21	21	21	21
51	11	13	9	9	11	10	10	10	10.4	10
52	9	12	11	11	8	11	11	12	10.6	11
53	13	9	12	13	13	13	13	13	12.4	13
54	14	16	14	15	14	14	14	14	14.4	14
55	15	14	15	14	15	15	15	15	14.8	15

Table 5.4: BIC model selection rank based table for all electrode configurations.

5.3 Chapter Conclusion

This chapter aimed at estimating model parameters given in general model in Equation 4.5.3 and selecting a suitable model that can represent the tongue EIM data along the spectrum. The chapter is divided into two main sections in order to present the analysis results for parameter estimation and model selection. The key reason for doing the modeling of the tongue EIM data is, if a fairly good fit captures the parameters correctly, the model has the potential to link to the electrical characteristics of the muscle to the model parameters.

Section 5.1 presents three different algorithms that are widely used for parameter estimation. A grid search algorithm is employed to get an estimate of the simple models. This method is convenient in its easy implementation and exhaustive search processes. Results have shown that the method has the disadvantage of approximating to a local minimum value. The method also struggles with the increased computational cost with the increased model complexity. Genetic Algorithms are capable of reducing the computational complexity as the model complexity increases. The method is inspired by nature's survival of the fittest hypothesis. This method also has the capacity to approximate the global optimum with the mutation embedded in the algorithm that allows searching nearby optimal solutions as opposed to Grid Search. The algorithm results have shown the efficiency of the estimation of the model parameters. A frequency-related weighting factor is implemented due to the increased error of the fit in the high frequency region. However, the weighted estimation idea was abandoned due to the ambiguity of the importance of poles and zeros that may give better discriminatory information for the volunteers and patients. Markov Chain Monte Carlo method is also used to show the estimation uncertainty rather than a single best solution for each parameter. Results have shown that as the model complexity is increased, the credible intervals at the higher frequency end narrow down.

Section 5.2 is prepared to select a suitable model from the model space that avoids under or over parametrization. This is because an overcomplicated model with a better fit is not necessarily a good model if the model parameters do not show disease-related patterns. The selected model is going to be used in the classification of the patients and healthy volunteers to see the discriminatory capability of the parameters in the selected model.

Chapter 6

Parametric Data Classification For Disease Diagnosis and Longitudinal Analysis

Chapters 4 and 5 focused on modeling the EIM data taken from the tongue to identify the effects of the ALS disease on the muscle structure using model parameters. It is assumed that the electrical impedance measurements capture the changes of the muscle structure, where the coupling of the resistance and reactance influences the location of the poles and zeros of a transfer function model. The data has been carefully analysed in Chapter 4, and an integral fractional order transfer function model was deemed. This model was proven to be ideal with the estimation and model selection of the parameters. Chapter 5 has addressed the parameter estimation and model selection of this fractional order transfer function model.

The optimal models over the spectra obtained after a model selection scheme in Section 5.2 are expected to exhibit a complete representation of the tissue's impedance characteristic. The models capture the essential characteristics of the tissue with its fractional order, poles, zeros, and gain parameters. These parameters of the candi-

date models are going to be used as the data features, and classification methods will be employed to obtain the prediction accuracies for each of the electrode configurations. From the selected candidate models, the best performing model, according to the classification accuracies, is going to be used for the longitudinal analysis.

This chapter is similar to Chapter 3 in the sense that both chapters are focused on disease diagnosis and longitudinal analysis of the ALS bulbar involvement from the EIM data. One main difference is that Chapter 3 focused on selecting the most informative frequencies by using the frequencies as the data features. In contrast, this chapter focuses on the whole spectrum by using the selected model parameters as the data features.

6.1 Parameter Correlograms

Correlograms are the plots that show the relationship between the pairs of parameters and the distribution of each parameter. Distribution on the correlogram shows each of the parameter's capacity to discriminate between the patients and healthy volunteers. Scatter plots on the correlogram shows the variation across the pairs of parameters between the patients and healthy volunteers. The correlograms in this chapter are plotted to see the discriminability of each parameter and pairs of parameters in the classification task. The plots are also useful for visualizing the parameter correlation to check pole-zero cancellation (where pole-zero is on the $x=y$ line). Pole-zero cancellation means there are extra poles and zeros in the model to represent the data. Figure 6.1 is an example plot of $[RIS]$ electrode configuration and $M_{\alpha 22}$ model that shows the correlation of the parameters as well as the discriminability of the patients and healthy volunteers.

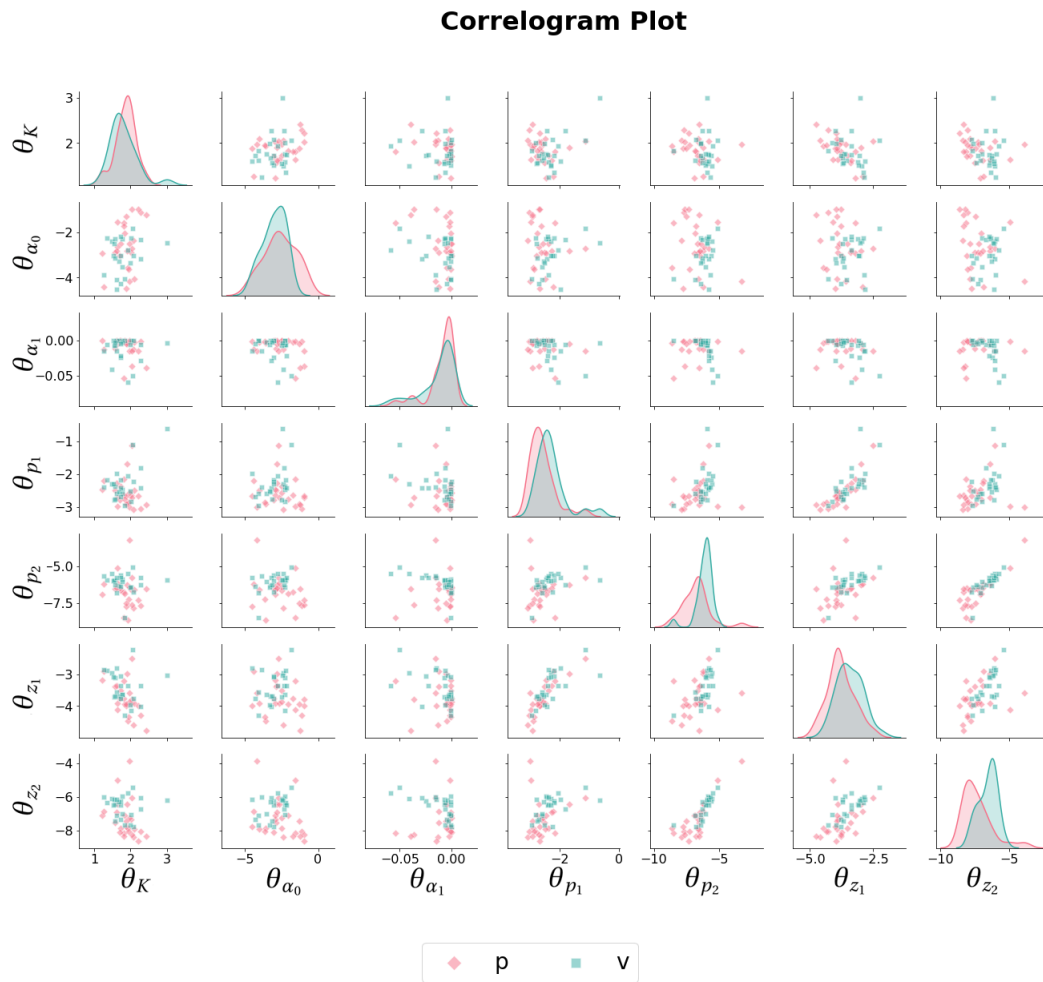


Figure 6.1: Correlogram for electrode configuration $[RIS]$ and model M_{a22} .

Figure 6.1 represents all electrode configurations, and the conclusion made for this figure matches with the rest of the electrode configurations. Other models also have similar characteristics to the model given in the above figure. The distributions indicate that the gain and fractional orders do not contribute to the discrimination between the patients and healthy volunteers. Poles and zeros show more discriminability between the two groups. The gain and fractional orders are important parameters to fit the model to the data; however, the generic tongue-related parameters are the poles and zeros that contribute to the cutoff of the magnitude of the

data. Scatter plots show significant overlap with some ability for classification and it is possible that the full dimensional features may show better discriminability that can be captured by the classification process. The following sections present the results to classification with different models and classifiers to find a suitable classifier and model for representing the data.

6.2 Supervised Classification

The supervised classification implemented in this chapter employs the features (model parameters) generated with the Genetic Algorithm for the selected models. The two classification algorithms, k-Nearest Neighbours and Support Vector Classifiers used in Chapter 3 will be used in this chapter to compare the results of different models for different classifiers. Classification results are expected to show the joint variations of the model parameters.

6.2.1 Support Vector Classifier

The Support Vector Classifier used in Chapter 3 has shown that different kernel methods and various values for n variable in n -fold cross-validation do not have much effect on the classification accuracy. It was also shown that the sigmoid kernel returned better discrimination between the classes. 4-fold cross-validation with sigmoid kernel is used as classifier parameters according to the results shown in Section 3.2.2. The features are scaled with robust scaling as part of the preprocessing so that each feature contributes proportionally. As discussed under Feature Selection title, it is possible to perform manual feature selection as part of dimensionality reduction. The manual feature selection is possible in this section by using the knowledge and familiarity with the features (model parameters). The following sections contain analysis with classification using a full parameter set as well as analysis excluding

some of the common parameters in the parameter set.

Classification Using Full Parameter Set

The full parameter set obtained from Genetic Algorithm results is used as the features of the data set. The classification results for all electrode configurations and selected models (see Section 5.2.2) are given as a heatmap for comparison.

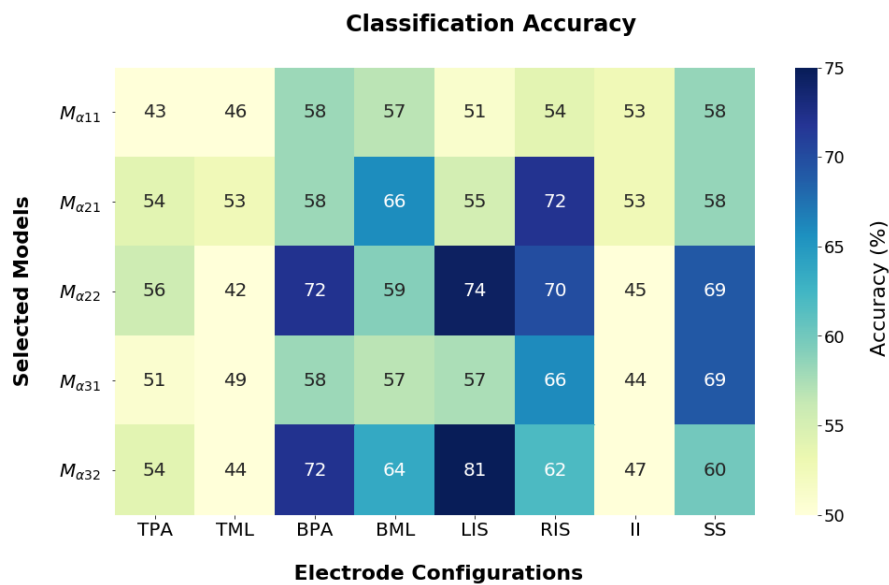


Figure 6.2: Support Vector Classifier results for selected models and all electrode configurations using full parameter set.

Figure 6.2 shows that the overall classification accuracies are low compared to the non-parametric classification results. The models with two zeros, namely $M_{\alpha 22}$ and $M_{\alpha 32}$, show higher accuracy than the rest of the models. It is visible that three of the 3D electrode configurations; [*LIS*, *RIS*, *SS*] and one of the 2D electrode configuration; [*BPA*] has higher accuracies compared to the rest of the electrode configurations with $M_{\alpha 22}$ and $M_{\alpha 32}$ models. This confirms the hypothesis that the 3D electrode configurations capture disease-related characteristics better than 2D surface electrode configurations.

Classification Using Poles, Zeros And Gain Parameters

The fractional-order in the model is used to capture the slow decline and increases on the slopes in the impedance spectrum because of the location of poles and zeros. This section investigates whether the fractional orders contribute to the discrimination of healthy volunteers and patients. The classification results for all electrode configurations and selected models (see Section 5.2.2) are given as a heatmap for comparison.

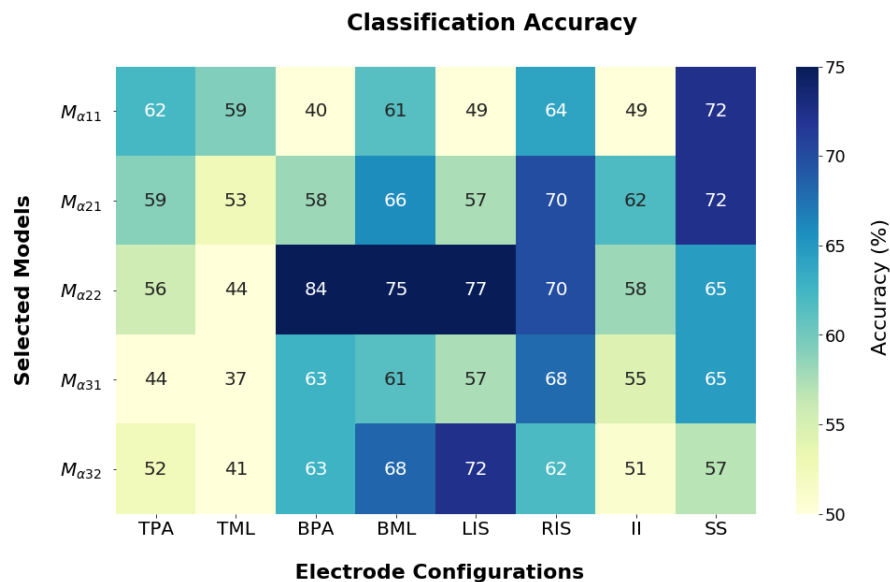


Figure 6.3: Support Vector Classifier results for selected models and all electrode configurations using gain, pole and zero parameters.

Figure 6.3 shows that the overall classification accuracy is higher than the classification results with the full parameter set. This result indicates that the fractional orders are not contributing to discriminating the two groups. There is an overall increase in classification with M_{22} model, but this increase does not apply to $M_{\alpha 32}$ model. Two of the 2D electrode configurations [*BPA*, *BML*] and three of the 3D electrode configurations [*LIS*, *RIS*, *SS*] with $M_{\alpha 22}$ model performed better compared to

the rest of the electrode configurations.

Classification Using Poles And Zeros Parameters

The gain value in the model determines where the magnitude value starts in the frequency response function. This section investigates whether the low frequency (at 76Hz) impedance reading contributes to the discrimination of the healthy volunteers and patients. Both the fractional-order α and the gain are removed from the feature space. The classification results for all electrode configurations and selected models (see Section 5.2.2) are given as a heatmap for comparison.

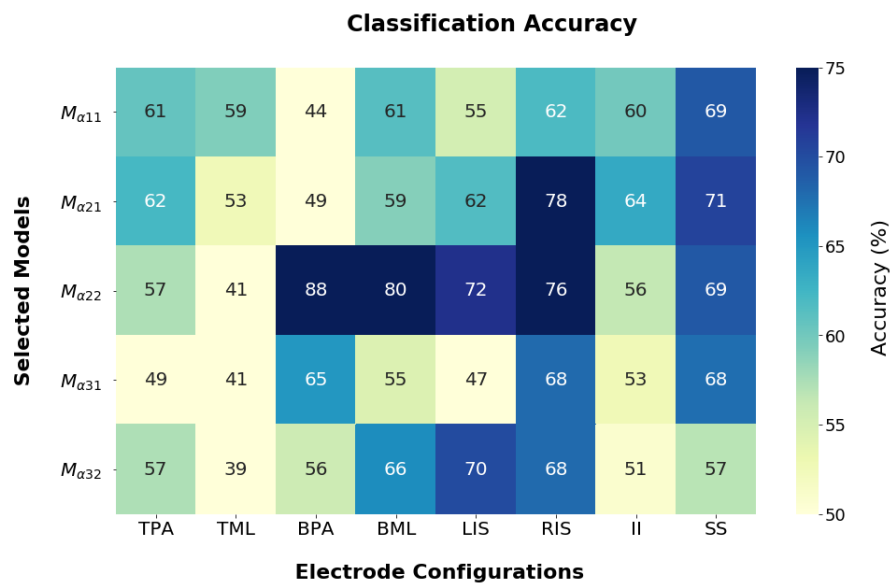


Figure 6.4: Support Vector Classifier results for selected models and all electrode configurations using pole and zero parameters.

Figure 6.4 shows that the overall classification accuracies are higher compared to the classification results with features that contain gain value. This is an indication of measurement captured with the first frequency not contributing to discriminating between the two groups. This could be due to this particular frequency impedance belonging in the low-frequency band (see Section 2.3.3). There is an overall increase

in classification with $M_{\alpha 22}$ model, but this increase does not apply to $M_{\alpha 32}$ model. Two of the 2D electrode configurations [*BPA*, *BML*] and three of the 3D electrode configurations [*LIS*, *RIS*, *SS*] with $M_{\alpha 22}$ model performed better compared to the rest of the electrode configurations.

6.2.2 k-Nearest Neighbour Classifier

k-Nearest Neighbour Classifier used in Chapter 3 has shown that different distance metrics, various values for k, n variables in k-nearest neighbours and n-fold cross validation does not have much effect on the variations of the results. Due to the results in Section 3.2.3, the classification used in this chapter have 3-nearest neighbours for k-value, 4-fold cross-validation for n-value. The features are scaled with robust scaling as part of the preprocessing so that each feature contributes proportionally. The Euclidean distance metric is used for calculation of the object's distance with scaled features to the neighbours in the classifier. The features will be selected with a similar approach to the classification with support vector classifier.

Classification Using Full Parameter Set

The full parameter set obtained from Genetic Algorithm results is used as the features of the data set. The classification results for all electrode configurations and selected models (see Section 5.2.2) are given as a heatmap below for comparison.

Figure 6.5 shows that the overall classification accuracies are low compared to the non-parametric classification results. The model with two-zeros and two-poles, namely $M_{\alpha 22}$ shows the higher accuracies compared to the rest of the models. This result is similar to that observed in SVC using full parameter set results. It is visible that three of the 3D electrode configurations; [*LIS*, *RIS*, *II*] and one of the 2D electrode configurations; [*BPA*] has higher accuracies compared to the rest of the electrode configurations with $M_{\alpha 22}$ model. This confirms the hypothesis that the 3D electrode

configurations perform better in detecting the disease. Comparing the overall results as well as the $M_{\alpha 22}$ results to SVC classifier performance, the k-NN classifier achieves better accuracies, and therefore, k-NN classifier results will be used for further analysis.

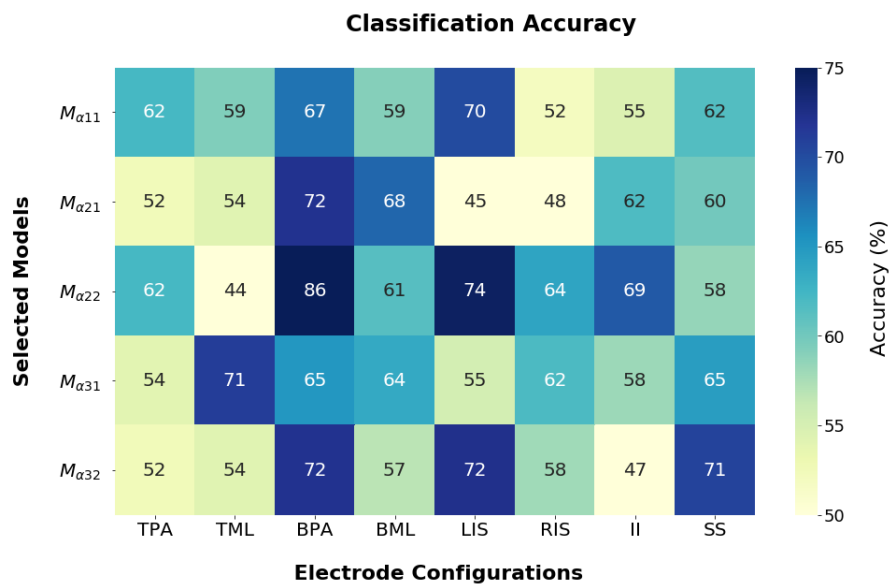


Figure 6.5: k-Nearest Neighbour Classifier results for selected models and all electrode configurations using full parameter set.

Classification Using Poles, Zeros And Gain Parameters

Similar to the analysis in classification using poles, zeros and gain parameters for SVC, this section investigates whether the slopes in the impedance spectrum because of the location of poles and zeros contribute to discrimination between the healthy volunteers and patients. The classification results for all electrode configurations and selected models (see Section 5.2.2) are given as a heatmap for comparison.

Figure 6.6 shows that the overall classification accuracies are not improved compared to the classification results with the full parameter set. Two of the 2D electrode configurations [*BPA*, *BML*] and three of the 3D electrode configurations [*LIS*, *RIS*, *SS*] with $M_{\alpha 22}$ model performed better compared to the rest of the electrode configura-

tions.

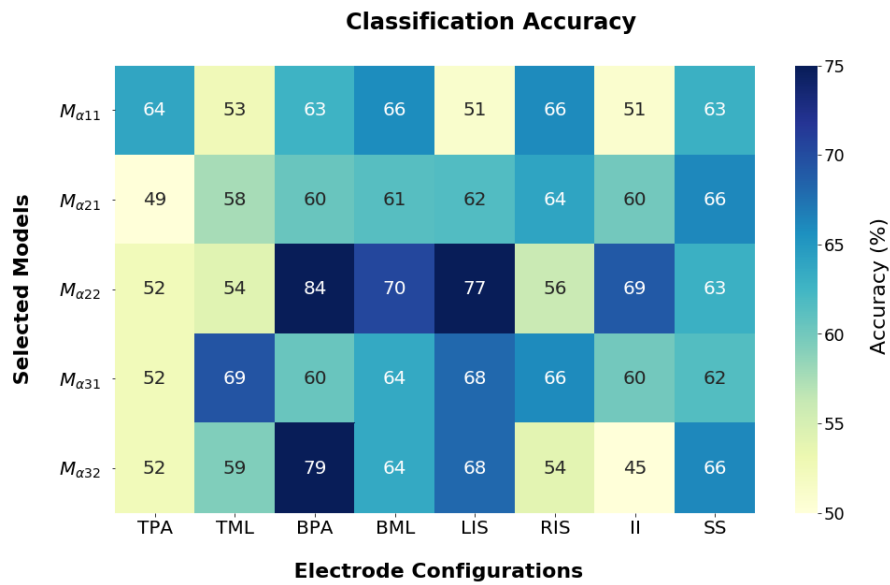


Figure 6.6: k-Nearest Neighbour Classifier results for selected models and all electrode configurations using gain, pole and zero parameters.

Classification Using Poles And Zeros Parameters

Similar to the analysis in classification using poles and zeros parameters for SVC, this section investigates whether the low frequency (at $76Hz$) impedance measurement contributes to the discrimination between the healthy volunteers and patients. Both the fractional-order α and the gain are removed from the feature space. The classification results for all electrode configurations and selected models (see Section 5.2.2) are given as a heatmap for comparison.

Figure 6.7 shows that the overall classification accuracies are higher compared to the classification results with features that contain the full parameter set and features that contain gain value. There is an overall increase in classification with $M_{\alpha 32}$ model, but this increase does not apply to $M_{\alpha 22}$ model.

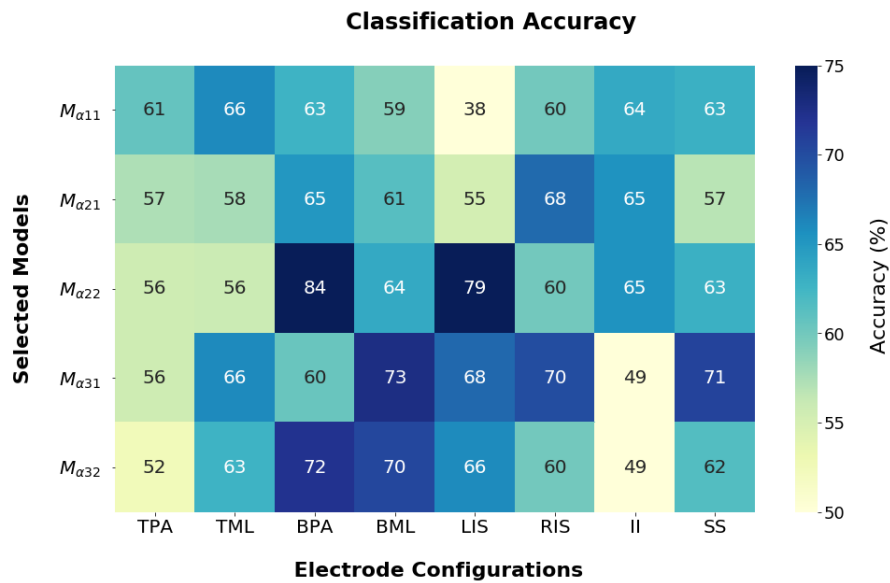


Figure 6.7: k-Nearest Neighbour Classifier results for selected models and all electrode configurations using pole and zero parameters.

The above analysis shows that the k-NN classifier is able to capture the discriminatory characteristics of the disease better than the support vector classifier. The heatmaps also confirm an overall increase in accuracy results when using only the poles and zeros parameters. However, the classification results for $M_{\alpha 22}$ model does not change significantly compared to the other results with different feature spaces. Considering the overall increase in classification accuracies with feature space using only poles and zeros, and better performance of $M_{\alpha 22}$ model, k-NN classification with $M_{\alpha 22}$ model will be used for subsequent analysis.

6.3 Parametric and Non-Parametric Classification Results Comparison

In this section we provide a comparison between parametric and nonparametric classification. The performances of the parametric and non-parametric classifica-

tion are presented in Table 6.1.

<i>Electrode Configurations</i>	<i>Non-Parametric Classification</i>			<i>Parametric Classification</i>		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
TPA (2D)	76.6 %	69.7 %	85.2 %	55.7 %	48.5 %	64.3 %
TML (2D)	85.0 %	75.0 %	96.4 %	55.9 %	71.9 %	37.0 %
BPA (2D)	85.7 %	77.8 %	91.7 %	83.7 %	83.3 %	84.0 %
BML (2D)	86.0 %	66.7 %	100.0 %	63.6.4 %	47.4 %	76.0 %
LIS (3D)	87.8 %	81.8 %	92.6 %	78.7 %	59.1 %	96.0 %
RIS (3D)	92.2 %	88.5 %	96.0 %	60.0 %	58.8 %	66.7 %
II (3D)	85.7 %	82.8 %	88.9 %	65.5 %	65.5 %	65.4 %
SS (3D)	81.3 %	76.5 %	86.7 %	63.1 %	71.4 %	53.3 %

Table 6.1: Parametric and Non-parametric classification results comparison using 3-Nearest Neighbour classifier.

The classification accuracies for the parametric model is lower than that of non-parametric model. Appropriate dimensionality reductions applied to non-parametric data prior to classification, whereas the dimensionality reduction is achieved with modeling in parametric classification. The results are different from each other because the non-parametric classification has less constraints in features since the features are obtained directly from individual frequency measurement, whereas the features are restricted to the models' parameters with parametric classification. The key difference between the two classification is that the features used in non-parametric classification is less interpretable whereas the features used in parametric classification has information relevant to the properties associated with muscle.

Results indicate that the electrode configurations have lower accuracy, sensitivity, and specificity with parametric classification. The classifier improved the sensitivity

results for [BPA] electrode configuration. Three of the 3D electrode configurations [LIS, RIS, II] and one of the 2D electrode configuration [BPA] has better accuracy results than the rest of the electrode configurations. The classifier significantly dropped the [TML] electrode configuration in detecting both healthy volunteers and patients.

6.4 Diagnosis Using Multiple Electrode Configurations

Disease diagnosis in Section 3.3.4 employed results from the non-parametric classification using multiple electrode configurations. This section considers employing $M_{\alpha 22}$ model parameter classification results to improve classification reliability and accuracy by combining all electrode configurations.

Table 6.2 is presented for volunteers and patients diagnostics ratio for all electrode configurations. The majority voting method was used to combine the outcomes from all the electrode configurations. There are eight electrode configurations available, but measurements were not captured in some of the electrode configurations for some patients and volunteers. The ratio represents the proportion of the number of electrode configurations that correctly identified the patient/volunteer to the total number of electrode configurations that measurements were obtained. The blue colours represent the low ratios, where 50% or less of the electrode configurations correctly identified the patient/volunteer.

The results show that 7 volunteers and 9 patients scored worse and $p011$, $p029$, $v022$ scored better compared to the non-parametric results. Blue colours are considered misdiagnosed, and red colours are considered the correct diagnosis. This means that the blue colors in patient group return *False Negative Rate* and volunteer group return *False Positive Rate*. Similarly, red colours in patient group return *True Positive Rate* and volunteer group return *True Negative Rate*. In confusion matrix, this table translates to true positive rate of 23, true negative rate of 23, false positive

rate of 7 and false negative rate of 12. Converting these values from confusion matrix (see Equations 3.1.6, 3.1.7, 3.1.8), the sensitivity value is 67.6% and the specificity value is 76.7%, which returns an accuracy value of 71.6% is obtained. The combined electrode configuration classification scores are higher than single electrode configurations: $[TPA, TML, BML, RIS, II, SS]$.

<i>Patient</i>	<i>Rate</i>	<i>Patient</i>	<i>Rate</i>	<i>Patient</i>	<i>Rate</i>	<i>Patient</i>	<i>Rate</i>
p001	5/8	p012	4/8	p023	3/8	p034	3/6
p002	6/8	p013	5/7	p024	6/8	p035	4/8
p003	5/6	p015	0/1	p026	1/5	p036	5/6
p004	6/8	p016	3/4	p027	0/1	p037	2/8
p006	4/6	p017	0/0	p028	1/1	p038	5/8
p007	3/5	p018	4/5	p029	4/4	p039	3/4
p008	5/7	p019	1/1	p030	3/3	p040	1/3
p009	6/8	p020	4/5	p031	4/6	p041	5/8
p010	2/5	p021	5/8	p032	4/5	Blue	Red
p011	5/8	p022	4/8	p033	2/4	Low	High
<i>Volunteer</i>	<i>Rate</i>	<i>Volunteer</i>	<i>Rate</i>	<i>Volunteer</i>	<i>Rate</i>	<i>Volunteer</i>	<i>Rate</i>
v001	5/8	v009	3/3	v017	6/8	v025	2/8
v002	4/7	v010	7/7	v018	4/7	v026	8/8
v003	6/8	v011	1/4	v019	5/8	v027	4/7
v004	5/8	v012	5/8	v020	7/8	v028	4/8
v005	7/8	v013	4/8	v021	2/4	v029	6/8
v006	5/8	v014	6/8	v022	3/4	v030	2/5
v007	4/6	v015	7/8	v023	8/8	Blue	Red
v008	4/8	v016	6/8	v024	4/7	Low	High

Table 6.2: Diagnosis using all electrode configurations.

6.5 Longitudinal Analysis Using Selected Model

Model parameter based classification has shown promising results for detecting disease with the highest accuracy of 83.7% with electrode configuration $[BPA]$ and accuracy of 71.6% with all electrode configurations combined. Chapter 3 considered using the selected frequencies with Phase angle data to show the efficiency of the

obtained data for longitudinal analysis. The longitudinal analysis is studied to show whether the EIM data can reveal disease-related changes in muscle structures as the disease progresses. As explained in Section 3.4, in order to make the measurement intervals equal for both healthy volunteers and patients, two measurements with six months intervals are used in the analysis. The selected model parameters with poles and zeros are used with parametric longitudinal analysis instead of phase readings at selected frequencies with non-parametric longitudinal analysis. The selected model parameter values were summarized with a single score for ease of interpretation by calculating the L2 (Euclidean) norm for each visit per patient and healthy volunteers. The parameters were normalized with robust scaling prior to calculating the L2 norm. The percentile change between the two visit's L2 norm is calculated and presented in Figure 6.8 with swarm plots. The results indicate that it is difficult to detect the longitudinal changes for patients with EIM readings from model parameters.

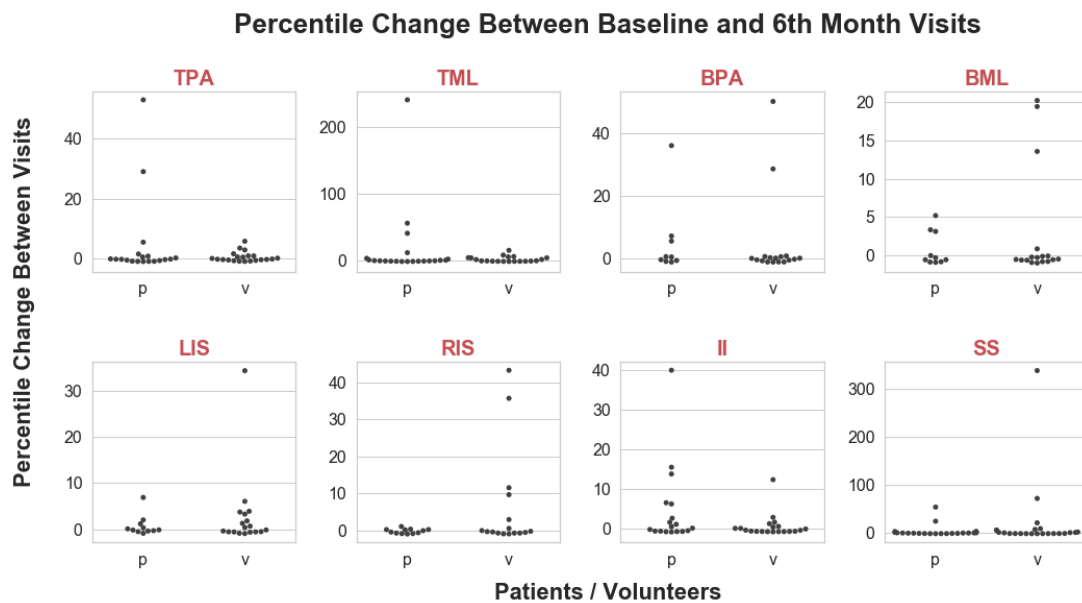


Figure 6.8: Percent change between baseline and 6th month visits for volunteers and patients in each electrode configurations to detect the patient EIM change with model parameters; poles and zeros.

6.6 Chapter Conclusion

Modeling of EIM data with integral fractional-order transfer function has shown that models; $M_{\alpha 11}$, $M_{\alpha 21}$, $M_{\alpha 22}$, $M_{\alpha 31}$, $M_{\alpha 32}$ have a good fit for the data. The selected models were employed for classification analysis in this chapter. Parameter correlograms were plotted in order to identify the pole-zero cancellation, variation across the pairs of parameter as well as the discriminatory characteristics of each parameter with distribution plots. The idea behind plotting the correlograms is that if there is a visible separation between the two group in these plots, then a proper classifier may be able to separate in higher dimensions. Overall, poles and zeros has shown better discriminatory characteristics in distribution plots.

Parametric classification employed the two commonly used classifiers, that are the k-Nearest Neighbour and the Support Vector Classifiers. A subset of parameters were chosen to evaluate a classifier in the reduced dimensionality space. Familiarity with features enabled the selection of features from the classification results. Support Vector Classifier has shown that the classification using poles and zeros has shown better discrimination than using full parameter space or parameter space that contains the gain parameter. Results also show that the $M_{\alpha 22}$ and $M_{\alpha 32}$ models have shown better discrimination than the rest of the selected models in the previous chapter. Similar conclusions were obtained from the analysis with k-NN classifier. Moreover, the k-NN classifier distinguished the two groups better than the SVC. For this reason, further analysis is produced using the k-NN classifier and the $M_{\alpha 22}$ model.

The parametric classification is different from non-parametric classification in that the features are the parameters of the model and not the raw data points from the full frequency spectrum. This means that the non-parametric approach has less constraints and is likely to find better discriminatory information. However, non-

parametric approach may not reveal any properties associated with the muscles in the way that the parametric approach does with poles and zeros.

Diagnosis using all electrode configurations has shown good discriminability between the two groups with a sensitivity of 67.6% and a specificity of 76.7%, which returns an accuracy of 71.6%. The results show that the model parameters, poles and zeros have some ability to discriminate between volunteers and patients. Longitudinal analyses were produced using the selected model parameters. Longitudinal analysis did not show any detectability of disease progression.

Chapter 7

Conclusions

Early stage diagnosis of bulbar disease in ALS is a challenging task. Discovery of EIM as a biomarker of motor neuron disease has opened new possibilities in the diagnosis and assessment of this disease. The development of EIM devices for accurate and reliable assessment has received considerable attention in recent years. The success of data obtained from such devices is highly dependent on the careful selection of methods for data analysis.

Statistical analysis of the data obtained from patients with EIM devices revealed fundamental aspects relating to the disease effect on the muscle tissue's electrical properties [36]. Various frequency electrical signals with different electrode configurations were tested in data acquisition. The overall conclusion of these researches indicates that there is still a need for more advanced analysis for detecting useful frequency characteristics associated with the disease and suitable electrode configuration associated with muscle anisotropy.

The Sheffield team has designed an EIM device that takes measurements along the desired frequency range (in both α and β regions) to measure the tongue muscles. The tongue comprises a complex muscle structure. Electrical properties of the muscle structure gives different electrical characteristics of the muscles (anisotropic characteristics). In order to capture the anisotropy, different electrode configurations

are necessary. For this reason, the device is designed to capture the tongue's electrical impedance characteristics with both *2D* and *3D* electrode configurations. Analysis of such high dimensional data was addressed in this dissertation.

7.1 Summary and Conclusions

The Sheffield device obtained measurements from both the patients with bulbar involvement and healthy volunteers. Measurements were taken from the left, right, and centre, when the tongue is relaxed and protruded, to investigate the electrical impedance characteristics of the contracted and relaxed tongue muscles. The preliminary analysis comprises the outlier detection of the raw data and the identification of the reliable measurements. A data specific outlier removal algorithm successfully removed the outliers along the spectrum. Further visual inspection pointed out the inconsistency of four of the *2D* electrode configurations, and these electrode configurations were removed from further analysis. Concerns were raised about the tissue volume captured with the hinged arms, which may affect the reliability of the measurements. For this reason, further analysis was completed using only the centre measurements. The contracted and relaxed muscle electrical impedance characteristics were analysed using phase values of the impedance spectra, and the two conditions did not show any discernible difference. For the reasons mentioned above, four *2D* and four *3D* electrode configurations in center location when the tongue is relaxed were used in further analysis. Removing the unwanted and unnecessary data brought the seventy-two (three locations * two tongue postures * twelve electrode configurations) measurements down to eight electrode configuration measurements, which emphasize the importance of the preliminary analysis.

As part of first objective of the project, classification methods were compared, and a suitable classifier was used with both the feature extraction and feature selec-

tion for dimensionally reduced classification. Classification with feature extraction has shown the device's diagnostic capacity. However, dimensionality reduction using feature extraction failed to answer one of the important research questions of this project: a consistent set of frequencies that have a better capacity to discriminate between volunteers and patients. Classification was applied to each electrode configurations separately because different electrode configurations measure different muscle groups. The results have shown that different frequency subsets show more discrimination between the two groups with different electrode configurations. This result relates to the anisotropic characteristics of the muscle groups. Diagnosis using *3D* electrode configurations were more successful compared to *2D* electrode configurations. Analysis was continued by combining the classification results of all electrode configurations to validate the existence of disease without relying on a single electrode configuration. As part of third objective of the project, longitudinal analysis was followed with the selected frequencies for each electrode configuration. Due to the course of the disease (see Section 2.3.5) and the medical interventions administered to ease disease-related difficulties, the findings did not support the device's capacity to serve as a tool to detect disease change over time.

As part of second objective of the project, data-driven modeling was studied to find the disease-related characteristics from the electrical impedance data with model parameters. The first attempt to model the data with closed-form linear least-squares approach had numerical issues, and an integral fractional-order transfer function was proposed to deal with these issues. In this model, the slow-decaying and increasing slopes of impedance spectral function were captured with fractional orders, and an integral pole with different fractional-order was used to represent the measurement process. The tongue measurements were represented with the gain, poles, and zeros. Parameter estimation was implemented using three different algorithms, from the most straightforward Grid Search Algorithm to nature-inspired Genetic Al-

gorithm and Bayesian Inference related Markov Chain Monte Carlo Algorithm. A model space was proposed with respect to the data, and suitable models were selected with model selection schemes. Selected models were used to implement parametric data classification in the last chapter. Selection of features with interpretability was employed in order to identify the model parameters that provide the most discriminatory information. Compared to non-parametric data classification accuracies were lower. As part of third objective of the project, longitudinal analysis was followed with the model and parameters selected by the classification. The conclusion was similar to Chapter 3 that the findings did not support the device's capacity to serve as a tool to detect disease change over time.

In conclusion, the device has shown a good diagnostic capacity for ALS disease bulbar involvement. Useful frequencies and electrode configurations were identified and it was shown that the *3D* electrode configurations outperformed the *2D* electrode configurations. The results indicate that the widely used surface electrodes in ALS researches were less discriminatory than the multi-dimensional electrodes developed by the Sheffield team for investigating the anisotropy characteristics. Parametric modeling of such data has never been done in ALS research to date to the author's knowledge. The transfer function modeling of the data will enable interpretation of the impedance characteristic changes in a more holistic way.

7.2 Future Work

Detection of ALS disease with electrical impedance data has taken a new direction in this dissertation. Statistical methods used to date are replaced with more advanced machine learning and data modeling studies. Classification using machine learning methods has shown better discrimination between healthy and diseased individuals. Generic transfer function modeling of the data replaced the 3 and 5 el-

ement cell electrical circuit models. Machine learning methods using the parameterized model have shown that the selected model can detect diseased data. The research also encouraged further analysis of this study field. These analyses are outlined as follows:

- The classification analysis given in Chapter 3 has shown results related to classification using machine learning methods. The dimensionality reduction is handled with two methods, one related to feature extraction and the other related to feature selection. The results produced with feature extraction employed PCA, which is a linear method. The alternative results may be obtained with nonlinear methods and compared to the presented results. The feature selection is achieved using an exhaustive search wrapper method, and the useful frequencies are selected after employing Pareto Dominance. The alternative non-exhaustive hybrid (embedded) method may be employed in order to reduce computational complexity.
- The amount of data used for this research is limited to the measurements obtained with the new Sheffield EIM data. Further analysis may be repeated with more data obtained with this device to confirm the reliability of outlier detection, classification, and modeling analysis.
- Longitudinal analysis suggested in Chapter 3 and repeated in Chapter 6 used the $L2$ norm percentile change of phase data. Results have shown that the devices capacity to detect the disease change over time remains an open question with the current methods to address this particular issue to make a more clear conclusion. This is due to varying change of health status of patients. Further analysis may be repeated with patients with increased disease severity. The study is open to new methods for analysis.
- The model proposed in Chapter 4 is capable of dealing with numerical insta-

bilities. However, the model is in the factorized form that does not allow for complex poles and zeros in the transfer function model. The capacity of complex roots for improving the model fit and detection of disease-related characteristics could be further investigated.

- The model selection schemes employed in Chapter 5 are related to the analysis obtained from Genetic Algorithm results. A section on Markov Chain Monte Carlo for parameter estimation is proposed in this chapter in order to show the advances of this method. The MCMC results can be produced to use a different model selection scheme called the Bayesian Model Selection.

The impedance spectroscopy based ALS disease diagnosis is entering an exciting area and this thesis is taking the first step towards providing a set of methods to facilitate this journey.

Bibliography

- [1] S. Boillée, C. V. Velde, and D. W. Cleveland, “ALS: a disease of motor neurons and their nonneuronal neighbors,” *Neuron*, vol. 52, no. 1, pp. 39–59, 2006.
- [2] A. D. Hillel and R. Miller, “Bulbar amyotrophic lateral sclerosis: patterns of progression and clinical management,” *Head & neck*, vol. 11, no. 1, pp. 51–59, 1989.
- [3] N. I. of Neurological Disorders and Stroke, *Amyotrophic Lateral Sclerosis (ALS) Fact Sheet*, Jun 2013. [Online]. <https://www.ninds.nih.gov/disorders/patient-caregiver-education/fact-sheets/amyotrophic-lateral-sclerosis-als-fact-sheet>.
- [4] J. Finsterer, M. Erdorf, B. Mamoli, and A. Fuglsang-Frederiksen, “Needle electromyography of bulbar muscles in patients with amyotrophic lateral sclerosis: evidence of subclinical involvement,” *Neurology*, vol. 51, no. 5, pp. 1417–1422, 1998.
- [5] K. W. Lee and C. H. Yun, “Value of needle electromyography for the diagnosis of amyotrophic lateral sclerosis,” *J Korean Neurol Assoc*, vol. 16, no. 1, p. 42, 1998.
- [6] T. M. Jenkins, J. J. Alix, R. H. Kandler, P. J. Shaw, and C. J. McDermott, “The role of cranial and thoracic electromyography within diagnostic criteria for amyotrophic lateral sclerosis,” *Muscle & nerve*, vol. 54, no. 3, pp. 378–385, 2016.
- [7] J. Talan, “A million dollar idea—a potential biomarker for ALS,” *Neurology Today*, vol. 11, no. 6, pp. 1–7, 2011.

- [8] S. B. Rutkove, J. B. Caress, M. S. Cartwright, T. M. Burns, J. Warder, W. S. David, *et al.*, “Electrical impedance myography as a biomarker to assess ALS progression,” *Amyotroph. Lateral Scler.*, vol. 13, no. 5, pp. 439–445, 2012.
- [9] M. R. Turner, R. Bowser, L. Bruijn, L. Dupuis, A. Ludolph, M. McGrath, *et al.*, “Mechanisms, models and biomarkers in amyotrophic lateral sclerosis,” *Amyotrophic Lateral Sclerosis and Fronto-temporal Degeneration*, vol. 14, no. sup1, pp. 19–32, 2013.
- [10] S. B. Rutkove, H. Zhang, D. A. Schoenfeld, E. M. Raynor, J. M. Shefner, M. E. Cudkowicz, *et al.*, “Electrical impedance myography to assess outcome in amyotrophic lateral sclerosis clinical trials,” *Clin. Neurophysiol.*, vol. 118, no. 11, pp. 2413–2418, 2007.
- [11] R. Aaron, G. J. Esper, C. A. Shiffman, K. Bradonjic, K. S. Lee, and S. B. Rutkove, “Effects of age on muscle as measured by electrical impedance myography,” *Physiol. Meas.*, vol. 27, pp. 953–959, jul 2006.
- [12] S. Budrewicz, P. Szewczyk, J. Bładowska, R. Podemski, E. Kozirowska-Gawron, M. Ejma, *et al.*, “The possible meaning of fractional anisotropy measurement of the cervical spinal cord in correct diagnosis of amyotrophic lateral sclerosis,” *Neurol. Sci.*, vol. 37, no. 3, pp. 417–421, 2016.
- [13] A. Tarulli, C. Shiffman, R. Aaron, A. Chin, and S. Rutkove, “Multifrequency electrical impedance myography in amyotrophic lateral sclerosis,” 2007.
- [14] S. B. Rutkove, P. M. Fogerson, L. P. Garmirian, and A. W. Tarulli, “Reference values for 50-khz electrical impedance myography,” *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 38, no. 3, pp. 1128–1132, 2008.

- [15] F. Hart, N. Berner, and R. McMillen, "Modelling the anisotropic electrical properties of skeletal muscle," *Physics in Medicine & Biology*, vol. 44, no. 2, p. 413, 1999.
- [16] B. Sanchez and S. B. Rutkove, "Electrical impedance myography and its applications in neuromuscular disorders," *Neurotherapeutics*, vol. 14, no. 1, pp. 107–118, 2017.
- [17] A. W. Tarulli, L. P. Garmirian, P. M. Fogerson, and S. B. Rutkove, "Localized muscle impedance abnormalities in amyotrophic lateral sclerosis," *Journal of clinical neuromuscular disease*, vol. 10, no. 3, p. 90, 2009.
- [18] M. Sung, A. J. Spieker, P. Narayanaswami, and S. B. Rutkove, "The effect of subcutaneous fat on electrical impedance myography when using a hand-held electrode array: the case for measuring reactance," *Clin. Neurophysiol.*, vol. 124, no. 2, pp. 400–404, 2013.
- [19] J. M. Shefner, S. B. Rutkove, J. B. Caress, M. Benatar, W. S. David, M. S. Cartwright, *et al.*, "Reducing sample size requirements for future ALS clinical trials with a dedicated electrical impedance myography system," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 19, no. 7-8, pp. 555–561, 2018.
- [20] J. Li, A. Pacheck, B. Sanchez, and S. B. Rutkove, "Single and modeled multifrequency electrical impedance myography parameters and their relationship to force production in the ALS SOD1G93A mouse," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 17, no. 5-6, pp. 397–403, 2016.
- [21] J. C. Desport, B. Marin, B. Funalot, P. M. Preux, and P. Couratier, "Phase angle is a prognostic factor for survival in amyotrophic lateral sclerosis," *Amyotroph. Lateral Scler.*, vol. 9, no. 5, pp. 273–278, 2008.

- [22] M. de Carvalho and M. Swash, "Nerve conduction studies in amyotrophic lateral sclerosis," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 23, no. 3, pp. 344–352, 2000.
- [23] C. McIlduff, *Electrical Impedance Myography (EIM) and Quantitative Ultrasonography (QUS) Measurements of the Tongue: Biomarkers of Bulbar Dysfunction*. Master's thesis, Harvard University, Cambridge, MA, United States, 2015.
- [24] S. B. Rutkove and J. L. Dawson, "Hand-held device for electrical impedance myography," May 22 2018. US Patent 9,974,463.
- [25] A. Pacheck, A. Mijailovic, S. Yim, J. Li, J. R. Green, C. E. McIlduff, *et al.*, "Tongue electrical impedance in amyotrophic lateral sclerosis modeled using the finite element method," *Clin. Neurophysiol.*, vol. 127, no. 3, pp. 1886–1890, 2016.
- [26] S. Shellikeri, Y. Yunusova, J. R. Green, G. L. Pattee, J. D. Berry, S. B. Rutkove, and L. Zinman, "Electrical impedance myography in the evaluation of the tongue musculature in amyotrophic lateral sclerosis," *Muscle & nerve*, vol. 52, no. 4, pp. 584–591, 2015.
- [27] C. McIlduff, S. Yim, A. Pacheck, T. Geisbush, A. Mijailovic, and S. B. Rutkove, "An improved electrical impedance myography (EIM) tongue array for use in clinical trials," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 932–935, 2016.
- [28] Z. Li, L. Chen, Y. Zhu, Q. Wei, W. Liu, D. Tian, *et al.*, "Handheld electrical impedance myography probe for assessing carpal tunnel syndrome," *Ann. Biomed. Eng.*, vol. 45, no. 6, pp. 1572–1580, 2017.
- [29] O. T. Ogunnika, M. Scharfstein, R. C. Cooper, H. Ma, J. L. Dawson, and S. B. Rutkove, "A handheld electrical impedance myography probe for the assessment of neuromuscular disease," in *2008 30th Annual International Conference*

- of the IEEE Engineering in Medicine and Biology Society*, pp. 3566–3569, IEEE, 2008.
- [30] C. E. Mcilduff, S. J. Yim, A. K. Pacheck, and S. B. Rutkove, “Optimizing electrical impedance myography of the tongue in amyotrophic lateral sclerosis,” *Muscle & nerve*, vol. 55, no. 4, pp. 539–543, 2017.
- [31] T. Srivastava, B. T. Darras, J. S. Wu, and S. B. Rutkove, “Machine learning algorithms to classify spinal muscular atrophy subtypes,” *Neurology*, vol. 79, no. 4, pp. 358–364, 2012.
- [32] M. López *et al.*, “Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer’s disease,” *Neurocomputing*, vol. 74, no. 8, pp. 1260 – 1271, 2011. Selected Papers from the 3rd International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC 2009).
- [33] S. Pinto, M. Gromicho, and M. de Carvalho, “Sialorrhoea and reversals in ALS functional rating scale,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 88, no. 2, pp. 187–188, 2017.
- [34] R. H. Bayford, “Bioimpedance tomography (electrical impedance tomography),” *Annu. Rev. Biomed. Eng.*, vol. 8, pp. 63–91, 2006.
- [35] D. Dean, T. Ramanathan, D. Machado, and R. Sundararajan, “Electrical impedance spectroscopy study of biological tissues,” *J. Electrostat.*, vol. 66, no. 3-4, pp. 165–177, 2008.
- [36] S. Rutkove, “Electrical impedance myography as a biomarker for ALS,” *The Lancet Neurology*, vol. 8, no. 3, p. 226, 2009.

- [37] S. B. Rutkove, "Electrical impedance myography: background, current state, and future directions," *Muscle & nerve*, vol. 40, no. 6, pp. 936–946, 2009.
- [38] O. Trokhanova, M. Okhapkin, and A. Korjenevsky, "Dual-frequency electrical impedance mammography for the diagnosis of non-malignant breast disease," *Physiol. Meas.*, vol. 29, no. 6, p. S331, 2008.
- [39] I. B. Salah, R. De la Rosa, K. Ouni, and R. B. Salah, "Automatic diagnosis of valvular heart diseases by impedance cardiography signal processing," *Biomed. Signal Process. Control*, vol. 57, p. 101758, 2020.
- [40] M. Piccolino, "Luigi galvani and animal electricity: two centuries after the foundation of electrophysiology," *Trends Neurosci.*, vol. 20, no. 10, pp. 443–448, 1997.
- [41] M. Pera, *The ambiguous frog: the Galvani-Volta controversy on animal electricity*. Princeton University Press, 2014.
- [42] H. C. Lukaski, "Bioimpedance analysis," in *Quality of the Body Cell Mass*, pp. 76–88, Springer, 2000.
- [43] T. X. Zhao and B. Jacobson, "Quantitative correlations among fibrinogen concentration, sedimentation rate and electrical impedance of blood," *Med. Biol. Eng. Comput.*, vol. 35, no. 3, pp. 181–185, 1997.
- [44] J. Nyboer, "Electrical impedance plethysmography," *Springfield-344*, 1959.
- [45] W. C. Chumlea and S. S. Guo, "Bioelectrical impedance: a history, research issues, and recent consensus," *Emerging technologies for nutrition research. Potential for assessing military performance capability*, pp. 169–192, 1997.

- [46] A. Thomasset, "Bioelectric properties of tissue. impedance measurement in clinical medicine. significance of curves obtained," *Lyon Med.*, vol. 28, pp. 107–109, 1962.
- [47] J. Li and S. Rutkove, "A flexible electrode array for muscle impedance measurements in the mouse hind limb: A tool to speed research in neuromuscular disease," in *Journal of Physics: Conference Series*, vol. 434, p. 012021, IOP Publishing, 2013.
- [48] C. Shiffman and R. Aaron, "Angular dependence of resistance in non-invasive electrical measurements of human muscle: the tensor model," *Physics in Medicine & Biology*, vol. 43, no. 5, p. 1317, 1998.
- [49] C. Nichols, M. S. Jain, K. G. Meilleur, T. Wu, J. Collins, M. R. Waite, *et al.*, "Electrical impedance myography in individuals with collagen 6 and laminin α -2 congenital muscular dystrophy: a cross-sectional and 2-year analysis," *Muscle & nerve*, vol. 57, no. 1, pp. 54–60, 2018.
- [50] T. Zagar and D. Krizaj, "Multivariate analysis of electrical impedance spectra for relaxed and contracted skeletal muscle," *Physiol. Meas.*, vol. 29, no. 6, p. S365, 2008.
- [51] W. Y. Du, *Resistive, capacitive, inductive, and magnetic sensor technologies*. CRC Press, 2014.
- [52] W. C. Chumlea and S. S. Guo, "Bioelectrical impedance and body composition: present status and future directions," *Nutr. Rev.*, vol. 52, no. 4, pp. 123–131, 1994.
- [53] C. Shiffman and S. Rutkove, "Circuit modeling of the electrical impedance: I. neuromuscular disease," *Physiol. Meas.*, vol. 34, no. 2, p. 203, 2013.

- [54] C. Shiffman and S. Rutkove, "Circuit modeling of the electrical impedance: Ii. normal subjects and system reproducibility," *Physiol. Meas.*, vol. 34, no. 2, p. 223, 2013.
- [55] B. Sanchez, S. R. Iyer, J. Li, K. Kapur, S. Xu, S. B. Rutkove, *et al.*, "Non-invasive assessment of muscle injury in healthy and dystrophic animals with electrical impedance myography," *Muscle & nerve*, vol. 56, no. 6, pp. E85–E94, 2017.
- [56] K. S. Cole and R. H. Cole, "Dispersion and absorption in dielectrics i. alternating current characteristics," *The Journal of chemical physics*, vol. 9, no. 4, pp. 341–351, 1941.
- [57] C. A. Shiffman, H. Kashuri, and R. Aaron, "Electrical impedance myography at frequencies up to 2 mhz," *Physiol. Meas.*, vol. 29, no. 6, p. S345, 2008.
- [58] B. Sanchez, J. Li, S. Yim, A. Pacheck, J. J. Widrick, and S. B. Rutkove, "Evaluation of electrical impedance as a biomarker of myostatin inhibition in wild type and muscular dystrophy mice," *PLoS One*, vol. 10, no. 10, 2015.
- [59] C. T. S. Ching, Y. C. Chen, L. H. Lu, P. F. Hsieh, C. S. Hsiao, T. P. Sun, *et al.*, "Characterization of the muscle electrical properties in low back pain patients by electrical impedance myography," *PLoS One*, vol. 8, no. 4, 2013.
- [60] H. Schwan, C. Kay, P. Bothwell, E. Foltz, *et al.*, "Electrical resistivity of living body tissues at low frequencies," in *Federation Proc*, vol. 13, p. 131, 1954.
- [61] L. E. Baker, "Principles of the impedance technique," *IEEE Eng. Med. Biol. Mag.*, vol. 8, no. 1, pp. 11–15, 1989.
- [62] A. Barnett and S. Bagnò, "The physiological mechanisms involved in the clinical measure of phase angle: Circuits simulating the phase angle properties of the body; correlations with experimental findings in normal and pathological

- states," *American Journal of Physiology-Legacy Content*, vol. 114, no. 2, pp. 366–382, 1935.
- [63] H. Schwan and C. Kay, "The conductivity of living tissues," *Ann. N.Y. Acad. Sci.*, vol. 65, no. 6, pp. 1007–1013, 1957.
- [64] S. Laufer, S. B. Solomon, and B. Rubinsky, "Tissue characterization using electrical impedance spectroscopy data: a linear algebra approach," *Physiol. Meas.*, vol. 33, no. 6, p. 997, 2012.
- [65] M. Singh, S. Semwal, A. Kumar, and S. Singh, "Impedance modeling for classification of flavored green teas," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 23, no. Sup. 1, pp. 2208–2214, 2015.
- [66] O. Casas, R. Bragos, P. Riu, J. Rosell, M. Tresanchez, M. Warren, *et al.*, "In vivo and in situ ischemic tissue characterization using electrical impedance spectroscopy," *Ann. N.Y. Acad. Sci.*, vol. 873, no. 1, pp. 51–58, 1999.
- [67] E. Mylott, E. Kutschera, and R. Widenhorn, "Bioelectrical impedance analysis as a laboratory activity: At the interface of physics and the body," *Am. J. Phys.*, vol. 82, no. 5, pp. 521–528, 2014.
- [68] R. Valdiosera, C. Clausen, and R. Eisenberg, "Circuit models of the passive electrical properties of frog skeletal muscle fibers," *The Journal of general physiology*, vol. 63, no. 4, pp. 432–459, 1974.
- [69] G. Falk and P. Fatt, "Linear electrical properties of striated muscle fibres observed with intracellular electrodes," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 160, no. 978, pp. 69–123, 1964.

- [70] B. Brown, D. Barber, W. Wang, L. Lu, A. Leathard, R. Smallwood, *et al.*, “Multi-frequency imaging and modelling of respiratory related electrical impedance changes,” *Physiol. Meas.*, vol. 15, no. 2A, p. A1, 1994.
- [71] S. B. Rutkove, G. J. Esper, K. S. Lee, R. Aaron, and C. A. Shiffman, “Electrical impedance myography in the detection of radiculopathy,” *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 32, no. 3, pp. 335–341, 2005.
- [72] H. Kwon, S. Rutkove, and B. Sanchez, “Recording characteristics of electrical impedance myography needle electrodes,” *Physiol. Meas.*, vol. 38, no. 9, p. 1748, 2017.
- [73] M. Cheney and D. Isaacson, “Issues in electrical impedance imaging,” *IEEE Computational Science and Engineering*, vol. 2, no. 4, pp. 53–62, 1995.
- [74] A. Tarulli, G. Esper, K. Lee, R. Aaron, C. Shiffman, and S. Rutkove, “Electrical impedance myography in the bedside assessment of inflammatory myopathy,” *Neurology*, vol. 65, no. 3, pp. 451–452, 2005.
- [75] S. B. Rutkove, J. M. Shefner, M. Gregas, H. Butler, J. Caracciolo, C. Lin, *et al.*, “Characterizing spinal muscular atrophy with electrical impedance myography,” *Muscle & nerve*, vol. 42, no. 6, pp. 915–921, 2010.
- [76] C. M. Zaidman, L. L. Wang, A. M. Connolly, J. Florence, B. L. Wong, J. A. Parsons, *et al.*, “Electrical impedance myography in duchenne muscular dystrophy and healthy controls: A multicenter study of reliability and validity,” *Muscle & nerve*, vol. 52, no. 4, pp. 592–597, 2015.
- [77] A. E. Emery, “Population frequencies of inherited neuromuscular diseases—a world survey,” *Neuromuscul. Disord.*, vol. 1, no. 1, pp. 19–29, 1991.

- [78] S. B. Rutkove, J. B. Caress, M. S. Cartwright, T. M. Burns, J. Warder, W. S. David, *et al.*, “Electrical impedance myography correlates with standard measures of ALS severity,” *Muscle & nerve*, vol. 49, no. 3, pp. 441–443, 2014.
- [79] B. Sanchez and S. B. Rutkove, “Present uses, future applications, and technical underpinnings of electrical impedance myography,” *Curr. Neurol. Neurosci. Rep.*, vol. 17, no. 11, p. 86, 2017.
- [80] N. C. Joyce and G. T. Carter, “Electrodiagnosis in persons with amyotrophic lateral sclerosis,” *PM&R*, vol. 5, no. 5, pp. S89–S95, 2013.
- [81] S. B. Rutkove, M. C. Gregas, and B. T. Darras, “Electrical impedance myography in spinal muscular atrophy: a longitudinal study,” *Muscle & nerve*, vol. 45, no. 5, pp. 642–647, 2012.
- [82] U. Birgersson, *Electrical impedance of human skin and tissue alterations: Mathematical modeling and measurements*. Karolinska Institutet, 2012.
- [83] O. Ogunnika, S. Rutkove, H. Ma, P. Fogerson, M. Scharfstein, R. Cooper, and J. Dawson, “A portable system for the assessment of neuromuscular diseases with electrical impedance myography,” *Journal of medical engineering & technology*, vol. 34, no. 7-8, pp. 377–385, 2010.
- [84] L. A. Bolgla and T. L. Uhl, “Reliability of electromyographic normalization methods for evaluating the hip musculature,” *J. Electromyogr. Kinesiol.*, vol. 17, no. 1, pp. 102–111, 2007.
- [85] A. W. Tarulli, A. B. Chin, R. A. Partida, and S. B. Rutkove, “Electrical impedance in bovine skeletal muscle as a model for the study of neuromuscular disease,” *Physiol. Meas.*, vol. 27, no. 12, p. 1269, 2006.

- [86] B. Sanchez, A. Pacheck, and S. B. Rutkove, "Guidelines to electrode positioning for human and animal electrical impedance myography research," *Sci. Rep.*, vol. 6, 2016.
- [87] B. R. Brooks, R. G. Miller, M. Swash, and T. L. Munsat, "El escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis," *Amyotroph. Lateral Scler. Other Motor Neuron Disord.*, vol. 1, no. 5, pp. 293–299, 2000.
- [88] M. de Carvalho, R. Dengler, A. Eisen, J. D. England, R. Kaji, J. Kimura, *et al.*, "Electrodiagnostic criteria for diagnosis of ALS," *Clin. Neurophysiol.*, vol. 119, no. 3, pp. 497–503, 2008.
- [89] B. U. Kleine, W. A. Boekestein, I. M. Arts, M. J. Zwarts, H. J. Schelhaas, and D. F. Stegeman, "Fasciculations and their f-response revisited: high-density surface EMG in ALS and benign fasciculations," *Clin. Neurophysiol.*, vol. 123, no. 2, pp. 399–405, 2012.
- [90] M. A. de Carvalho, S. Pinto, and M. Swash, "Paraspinal and limb motor neuron involvement within homologous spinal segments in ALS," *Clin. Neurophysiol.*, vol. 119, no. 7, pp. 1607–1613, 2008.
- [91] C. Arsac, C. Raymond, N. Martin-Moutot, B. Dargent, M. Seagar, F. Couraud, *et al.*, "Immunoassays fail to detect antibodies against neuronal calcium channels in amyotrophic lateral sclerosis serum," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 40, no. 5, pp. 695–700, 1996.
- [92] C. Dettmers, D. Fatepour, H. Faust, and F. Jerusalem, "Sympathetic skin response abnormalities in amyotrophic lateral sclerosis," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 16, no. 9, pp. 930–934, 1993.

- [93] T. S. Wingo, D. J. Cutler, N. Yarab, C. M. Kelly, and J. D. Glass, "The heritability of amyotrophic lateral sclerosis in a clinically ascertained united states research registry," *PLoS One*, vol. 6, no. 11, 2011.
- [94] P. J. Shaw and C. Wood-Allum, "Motor neurone disease: a practical update on diagnosis and management," *Clin. Med.*, vol. 10, no. 3, p. 252, 2010.
- [95] J. Weikamp, H. Schelhaas, J. Hendriks, B. de Swart, and A. Geurts, "Prognostic value of decreased tongue strength on survival time in patients with amyotrophic lateral sclerosis," *J. Neurol.*, vol. 259, no. 11, pp. 2360–2365, 2012.
- [96] R. Giess, M. Naumann, E. Werner, R. Riemann, M. Beck, I. Puls, *et al.*, "Injections of botulinum toxin a into the salivary glands improve sialorrhoea in amyotrophic lateral sclerosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 69, no. 1, pp. 121–123, 2000.
- [97] J. Glaser and M. Suzuki, "Skeletal muscle fiber types in neuromuscular diseases," *Muscle Cell and Tissue: Current Status of Research Field*, p. 65, 2018.
- [98] S. B. Rutkove, R. Aaron, and C. A. Shiffman, "Localized bioimpedance analysis in the evaluation of neuromuscular disease," *Muscle & nerve*, vol. 25, no. 3, pp. 390–397, 2002.
- [99] S. Rutkove, A. Chin, C. Shiffman, and R. Aaron, "P22.9 measurement of muscle anisotropy using electrical impedance myography," *Clin. Neurophysiol.*, vol. 117, no. Supplement 1, p. 102, 2006.
- [100] A. W. V. Drake, Richard and A. W. Mitchell., *Gray's Anatomy for Students*. Elsevier Health Sciences, 2009.
- [101] M. Wu, J. Han, O. Röhrle, W. Thorpe, and P. Nielsen, "Using three-dimensional finite element models and principles of active muscle contraction to analyse

- the movement of the tongue,” in *11th Australian International Conference on Speech Science & Technology*, pp. 354–359, 2006.
- [102] A. B. Chin, L. P. Garmirian, R. Nie, and S. B. Rutkove, “Optimizing measurement of the electrical anisotropy of muscle,” *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 37, no. 5, pp. 560–565, 2008.
- [103] J. Li and S. Rutkove, “Alteration in surface muscle electrical anisotropy in the rat sod1 model of amyotrophic lateral sclerosis,” *Clin. Neurophysiol.*, vol. 123, no. 1, pp. 206–210, 2012.
- [104] L. P. Garmirian, A. B. Chin, and S. B. Rutkove, “Discriminating neurogenic from myopathic disease via measurement of muscle anisotropy,” *Muscle & Nerve*, vol. 39, no. 1, pp. 16–24, 2009.
- [105] M. Nakamori, N. Hosomi, S. Takaki, M. Oda, A. Hiraoka, M. Yoshikawa, *et al.*, “Tongue thickness evaluation using ultrasonography can predict swallowing function in amyotrophic lateral sclerosis patients,” *Clin. Neurophysiol.*, vol. 127, no. 2, pp. 1669–1674, 2016.
- [106] C. A. Shiffman, R. Aaron, and S. B. Rutkove, “Electrical impedance of muscle during isometric contraction,” *Physiol. Meas.*, vol. 24, no. 1, p. 213, 2003.
- [107] B. H. Brown, J. A. Tidy, K. Boston, A. D. Blackett, R. H. Smallwood, and F. Sharp, “Relation between tissue structure and imposed electrical current flow in cervical neoplasia,” *The Lancet*, vol. 355, no. 9207, pp. 892–895, 2000.
- [108] H. P. Schwan, “Electrical properties of tissues and cell suspensions: mechanisms and models,” in *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. A70–A71, IEEE, 1994.

- [109] R. Buendia, F. Seoane, K. Lindecrantz, I. Bosaeus, R. Gil-Pita, G. Johannsson, *et al.*, “Estimation of body fluids with bioimpedance spectroscopy: state of the art methods and proposal of novel methods,” *Physiol. Meas.*, vol. 36, no. 10, p. 2171, 2015.
- [110] R. Harder, A. Diedrich, J. S. Whitfield, M. S. Buchowski, J. B. Pietsch, and F. J. Baudenbacher, “Smart multi-frequency bioelectrical impedance spectrometer for bia and biva applications,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 4, pp. 912–919, 2016.
- [111] G. J. Esper, C. A. Shiffman, R. Aaron, K. S. Lee, and S. B. Rutkove, “Assessing neuromuscular disease with multifrequency electrical impedance myography,” *Muscle & nerve*, vol. 34, no. 5, pp. 595–602, 2006.
- [112] M. C. Kiernan, S. Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, *et al.*, “Amyotrophic lateral sclerosis,” *The lancet*, vol. 377, no. 9769, pp. 942–955, 2011.
- [113] G. Logroscino, B. J. Traynor, O. Hardiman, A. Chiò, D. Mitchell, R. J. Swingler, *et al.*, “Incidence of amyotrophic lateral sclerosis in Europe,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 81, no. 4, pp. 385–390, 2010.
- [114] H. G. Kortman *et al.*, “Age and gender associated differences in electrical impedance values of skeletal muscle,” *Physiol. Meas.*, vol. 34, no. 12, p. 1611, 2013.
- [115] E. V. Hobson and C. J. McDermott, “Supportive and symptomatic management of amyotrophic lateral sclerosis,” *Nature Reviews Neurology*, vol. 12, no. 9, p. 526, 2016.
- [116] B. Swinnen and W. Robberecht, “The phenotypic variability of amyotrophic lateral sclerosis,” *Nature Reviews Neurology*, vol. 10, no. 11, p. 661, 2014.

- [117] K. Talbot, "Motor neuron disease: the bare essentials," *Practical neurology*, vol. 9, no. 5, pp. 303–309, 2009.
- [118] A. Chio, G. Logroscino, O. Hardiman, R. Swingler, D. Mitchell, E. Beghi, *et al.*, "Prognostic factors in ALS: a critical review," *Amyotroph. Lateral Scler.*, vol. 10, no. 5-6, pp. 310–323, 2009.
- [119] E. B. Kelly, *Encyclopedia of human genetics and disease*, vol. 1. ABC-CLIO, 2013.
- [120] A. Chiò, A. Calvo, C. Moglia, L. Mazzini, and G. Mora, "Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 82, no. 7, pp. 740–746, 2011.
- [121] P. H. Gordon, R. G. Miller, and D. H. Moore, "ALSFRS-R," *Amyotroph. Lateral Scler. Other Motor Neuron Disord.*, vol. 5, no. sup1, pp. 90–93, 2004.
- [122] H. Creemers, H. Grupstra, F. Nollet, L. H. van den Berg, and A. Beelen, "Prognostic factors for the course of functional status of patients with ALS: a systematic review," *J. Neurol.*, vol. 262, no. 6, pp. 1407–1423, 2015.
- [123] C. Castrillo-Viguera, D. L. Grasso, E. Simpson, J. Shefner, and M. E. Cudkovicz, "Clinical significance in the change of decline in ALSFRS-R," *Amyotroph. Lateral Scler.*, vol. 11, no. 1-2, pp. 178–180, 2010.
- [124] M. R. Turner, J. Scaber, J. A. Goodfellow, M. E. Lord, R. Marsden, and K. Talbot, "The diagnostic pathway and prognosis in bulbar-onset amyotrophic lateral sclerosis," *J. Neurol. Sci.*, vol. 294, no. 1-2, pp. 81–85, 2010.
- [125] P. H. Gordon, F. Salachas, L. Lacomblez, N. Le Forestier, P.-F. Pradat, G. Bruneteau, *et al.*, "Predicting survival of patients with amyotrophic lateral sclerosis at presentation: a 15-year experience," *Neurodegenerative Diseases*, vol. 12, no. 2, pp. 81–90, 2013.

- [126] J. Labra, P. Menon, K. Byth, S. Morrison, and S. Vucic, "Rate of disease progression: a prognostic biomarker in ALS," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 87, no. 6, pp. 628–632, 2016.
- [127] M. R. Turner, A. Brockington, J. Scaber, H. Hollinger, R. Marsden, P. J. Shaw, *et al.*, "Pattern of spread and prognosis in lower limb-onset ALS," *Amyotroph. Lateral Scler.*, vol. 11, no. 4, pp. 369–373, 2010.
- [128] R. J. Gilbert and V. J. Napadow, "Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion tensor magnetic resonance imaging," *Dysphagia*, vol. 20, no. 1, pp. 1–7, 2005.
- [129] T. A. Gaige, T. Benner, R. Wang, V. J. Wedeen, and R. J. Gilbert, "Three dimensional myoarchitecture of the human tongue determined in vivo by diffusion tensor imaging with tractography," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 26, no. 3, pp. 654–661, 2007.
- [130] J. J. Alix, H. E. McDonough, B. Sonbas, S. J. French, D. G. Rao, V. Kadiramanathan, *et al.*, "Multi-dimensional electrical impedance myography of the tongue as a potential biomarker for amyotrophic lateral sclerosis," *Clin. Neurophysiol.*, vol. 131, no. 4, pp. 799–808, 2020.
- [131] J. Hallas and A. Nissen, "Individualized drug utilization statistics," *Eur. J. Clin. Pharmacol.*, vol. 47, no. 4, pp. 367 – 372, 1994.
- [132] H. Støvring, "New statistical methods for estimation of prevalence, incidence and mortality based on pharmacoepidemiological and other health-related databases.," *Dan. Med. Bull.*, vol. 50, 2003.
- [133] D. C. Malins, K. M. Anderson, N. L. Polissar, G. K. Ostrander, E. T. Knobbe, V. M. Green, *et al.*, "Models of granulocyte dna structure are highly predictive of

- myelodysplastic syndrome,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 14, pp. 5008–5011, 2004.
- [134] L. Fratiglioni, M. Grut, Y. Forsell, M. Viitanen, and B. Winblad, “Clinical Diagnosis of Alzheimer’s Disease and Other Dementias in a Population Survey: Agreement and Causes of Disagreement in Applying Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition, Criteria,” *Arch. Neurol.*, vol. 49, pp. 927–932, 09 1992.
- [135] S. K. Van Den Eeden, C. M. Tanner, A. L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch, *et al.*, “Incidence of Parkinson’s Disease: Variation by Age, Gender, and Race/Ethnicity,” *Am. J. Epidemiol.*, vol. 157, pp. 1015–1022, 06 2003.
- [136] F. K. Onu, D. O. Ikoru, U. C. Emeruem, and D. O. Inyang, “Discrimination of depositional environment using statistical parameters of conglomeratic outcrop sections, obubra southeastern nigeria,” *British Journal of Applied Science & Technology*, 2014.
- [137] T. Pham-Gia and T. Hung, “The mean and median absolute deviations,” *Math. Comput. Modell.*, vol. 34, no. 7-8, pp. 921–936, 2001.
- [138] M. Bland, *An introduction to medical statistics*. Oxford University Press (UK), 2015.
- [139] G. Brys, M. Hubert, and A. Struyf, “A robust measure of skewness,” *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 996 – 1017, 2004.
- [140] G. Hackeling, *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd., 2017.

- [141] X. Zhang, P. E. Barkhaus, W. Z. Rymer, and P. Zhou, "Machine learning for supporting diagnosis of amyotrophic lateral sclerosis using surface electromyogram," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 96–103, 2013.
- [142] F. Aydin and Z. Aslan, "Classification of neurodegenerative diseases using machine learning methods," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 1, no. 5, pp. 1–9, 2017.
- [143] M. S. Bereman, J. Beri, J. R. Enders, and T. Nash, "Machine learning reveals protein signatures in csf and plasma fluids of clinical value for ALS," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018.
- [144] T. Hothorn and H. H. Jung, "Randomforest4life: a random forest for predicting ALS disease progression," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 15, no. 5-6, pp. 444–452, 2014.
- [145] R. Küffner, N. Zach, R. Norel, J. Hawe, D. Schoenfeld, L. Wang, *et al.*, "Crowd-sourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression," *Nat. Biotechnol.*, vol. 33, no. 1, p. 51, 2015.
- [146] N. Savage, "Calculating disease," *Nature*, vol. 550, no. 7676, pp. S115–S117, 2017.
- [147] J. S. Shyam Mohan, P. Shanmugapriya, and N. Kumaran, "Data reduction techniques for high dimensional biological data," *International Journal of Research in Engineering and Technology*, 2016.
- [148] T. H. Cheng, C. P. Wei, and V. S. Tseng, "Feature selection for medical data mining: Comparisons of expert judgment and automatic approaches," in *19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)*, pp. 165–170, IEEE, 2006.

- [149] C. J. Tu, L. Y. Chuang, J. Y. Chang, and C. H. Yang, "Feature selection using PSO-SVM," *International Journal of Computer Science*, 2007.
- [150] K. K. Gandhi and N. B. Prajapati, "Diabetes prediction using feature selection and classification," *International journal of advance Engineering and Research Development*, vol. 1, no. 05, 2014.
- [151] S. E. Umbaugh, Y. S. Wei, and M. Zuke, "Feature extraction in image analysis. a program for facilitating data reduction in medical image classification," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 4, pp. 62–73, 1997.
- [152] M. L. Raymer, W. F. Punch, E. D. Goodman, P. C. Sanschagrin, and L. A. Kuhn, "Simultaneous feature extraction and selection using a masking genetic algorithm," in *Proceedings of the 7th International Conference on Genetic Algorithms*, pp. 561–567, Citeseer, 1997.
- [153] D. C. Li, C. W. Liu, and S. C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets," *Artif. Intell. Med.*, vol. 52, no. 1, pp. 45–52, 2011.
- [154] C. Ding, X. He, H. Zha, and H. D. Simon, "Adaptive dimension reduction for clustering high dimensional data," in *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, pp. 147–154, IEEE, 2002.
- [155] A. Widodo and B.-S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 241–250, 2007.
- [156] A. Miszczyk and K. Darowicki, "Multivariate analysis of impedance data obtained for coating systems of varying thickness applied on steel," *Prog. Org. Coat.*, vol. 77, no. 12, pp. 2000–2006, 2014.

- [157] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [158] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Icml*, vol. 97, p. 35, Nashville, TN, USA, 1997.
- [159] H. Yuan, S. Tseng, Shian, W. Gangshan, and Z. Fuyan, "A two-phase feature selection method using both filter and wrapper," in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 2, pp. 132–136, IEEE, 1999.
- [160] P. E. Jouve and N. Nicoloyannis, "A filter feature selection method for clustering," in *International Symposium on Methodologies for Intelligent Systems*, pp. 583–593, Springer, 2005.
- [161] F. R. Bach, "Bolasso: model consistent lasso estimation through the bootstrap," in *Proceedings of the 25th international conference on Machine learning*, pp. 33–40, 2008.
- [162] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010.
- [163] Q. Lin and Y. Chen, "Application of machine learning method in simulation model validation," in *2018 3rd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2018)*, Atlantis Press, 2018.
- [164] H. Cheng, D. J. Garrick, and R. L. Fernando, "Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction," *Journal of animal science and biotechnology*, vol. 8, no. 1, p. 38, 2017.

- [165] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.
- [166] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," *Digital Signal Process.*, vol. 17, no. 4, pp. 694–701, 2007.
- [167] P. K. Grewal and F. Golnaraghi, "Pilot study: electrical impedance based tissue classification using support vector machine classifier," *IET Science, Measurement & Technology*, vol. 8, no. 6, pp. 579–587, 2014.
- [168] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of heart disease using k-nearest neighbor and genetic algorithm," *arXiv preprint arXiv:1508.02061*, 2015.
- [169] I. Babaoğlu, M. S. Kıran, E. Ülker, and M. Gündüz, "Diagnosis of coronary artery disease using artificial bee colony and k-nearest neighbor algorithms," *International Journal of Computer and Communication Engineering*, vol. 2, no. 1, pp. 56–59, 2013.
- [170] K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, and S. Mukherjee, "A comparative study on thyroid disease detection using k-nearest neighbor and naive bayes classification techniques," *CSI transactions on ICT*, vol. 4, no. 2-4, pp. 313–319, 2016.
- [171] L. Y. Hu, M. W. Huang, S. W. Ke, and C. F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," *SpringerPlus*, vol. 5, no. 1, pp. 1–9, 2016.
- [172] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers," *arXiv preprint arXiv:2004.04523*, 2020.

- [173] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J Inf Eng Appl*, vol. 3, no. 10, 2013.
- [174] J. Fu, C. Huang, J. Xing, and J. Zheng, "Pattern classification using an olfactory model with pca feature selection in electronic noses: Study and application," *Sensors*, vol. 12, no. 3, pp. 2818–2830, 2012.
- [175] L. Wang, N. Zhou, and F. Chu, "A general wrapper approach to selection of class-dependent features," *IEEE Trans. Neural Netw.*, vol. 19, no. 7, pp. 1267–1278, 2008.
- [176] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data classification: Algorithms and applications*, p. 37, 2014.
- [177] H. H. Hsu, C. W. Hsieh, and M. D. Lu, "Hybrid feature selection by combining filters and wrappers," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8144–8150, 2011.
- [178] A. G. Karegowda, M. Jayaram, and A. Manjunath, "Feature subset selection problem using wrapper approach in supervised learning," *International journal of Computer applications*, vol. 1, no. 7, pp. 13–17, 2010.
- [179] M. S. Sánchez, M. C. Ortiz, L. A. Sarabia, and R. Lletí, "On pareto-optimal fronts for deciding about sensitivity and specificity in class-modelling problems," *Anal. Chim. Acta*, vol. 544, no. 1-2, pp. 236–245, 2005.
- [180] K. Mitra and M. Ghivari, "Modeling of an industrial wet grinding operation using data-driven techniques," *Computers & chemical engineering*, vol. 30, no. 3, pp. 508–520, 2006.
- [181] J. B. Tylcz, K. El Alaoui Lasmali, E. H. Djermoune, N. Thomas, B. Faivre, and T. Bastogne, "Data-driven modeling and characterization of anti-angiogenic

- molecule effects on tumoral vascular density,” *Biomed. Signal Process. Control*, vol. 20, pp. 52–60, 2015.
- [182] N. J. Nilsson, “Learning machines,” 1965.
- [183] L. Zadeh, “On the identification problem,” *IRE Trans. Circuit Theory*, vol. 3, no. 4, pp. 277–281, 1956.
- [184] L. Ljung, “Perspectives on system identification,” *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.
- [185] J. Schoukens and L. Ljung, “Nonlinear system identification: A user-oriented road map,” *IEEE Control Syst. Mag.*, vol. 39, no. 6, pp. 28–99, 2019.
- [186] L. V. Kalmykov and V. L. Kalmykov, “A white-box model of s-shaped and double s-shaped single-species population growth,” *PeerJ*, vol. 3, p. e948, 2015.
- [187] B. Cornish, B. Thomas, and L. Ward, “Improved prediction of extracellular and total body water using impedance loci generated by multiple frequency bio-electrical impedance analysis,” *Physics in Medicine & Biology*, vol. 38, no. 3, p. 337, 1993.
- [188] L. Kalakutskiy and S. Akulov, “Modeling of the bioelectrical impedance of blood by synthesis of the equivalent electrical circuits,” in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pp. 575–577, Springer, 2009.
- [189] S. K. Dhar and Q. D. Hossain, “Non-invasive bio-impedance measurement using voltage-current pulse technique,” in *Int. Conf. on Electrical, Electronics and Biomedical Engineering*, pp. 70-74, 2012.
- [190] J. Sabatier, O. P. Agrawal, and J. T. Machado, *Advances in fractional calculus*, vol. 4. Springer, 2007.

- [191] D. Copot, R. De Keyser, and C. Ionescu, "In vitro glucose concentration estimation by means of fractional order impedance models," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 002711–002716, IEEE, 2016.
- [192] R. L. Magin, *Fractional calculus in bioengineering*, vol. 2. Begell House Redding, 2006.
- [193] R. Magin and M. Oviaia, "Modeling the cardiac tissue electrode interface using fractional calculus," *J. Vib. Control*, vol. 14, no. 9-10, pp. 1431–1442, 2008.
- [194] R. L. Magin, "Fractional calculus models of complex dynamics in biological tissues," *Computers & Mathematics with Applications*, vol. 59, no. 5, pp. 1586–1593, 2010.
- [195] O. G. Martinsen and S. Grimnes, *Bioimpedance and bioelectricity basics*. Academic press, 2011.
- [196] K. Hcheichi and F. Bouani, "Comparison between commensurate and non-commensurate fractional systems," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 9, no. 11, 2018.
- [197] T. J. Freeborn, "A survey of fractional-order circuit models for biology and biomedicine," *IEEE Journal on emerging and selected topics in circuits and systems*, vol. 3, no. 3, pp. 416–424, 2013.
- [198] V. Kotu and B. Deshpande, *Data Science: Concepts and Practice*. Morgan Kaufmann, 2018.
- [199] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," *Computer*, vol. 27, no. 6, pp. 17–26, 1994.

- [200] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods," *Genome Res.*, vol. 13, no. 11, pp. 2467–2474, 2003.
- [201] V. Plagnol and S. Tavaré, "Approximate bayesian computation and mcmc," in *Monte Carlo and Quasi-Monte Carlo Methods 2002*, pp. 99–113, Springer, 2004.
- [202] F. Reali, C. Priami, and L. Marchetti, "Optimization algorithms for computational systems biology," *Frontiers in Applied Mathematics and Statistics*, vol. 3, p. 6, 2017.
- [203] H. Li, P. E. Love, and S. Ogunlana, "Genetic algorithm compared to nonlinear optimization for labour and equipment assignment," *Building Research & Information*, vol. 26, no. 6, pp. 322–329, 1998.
- [204] J. Sadeghi, S. Sadeghi, and S. T. A. Niaki, "Optimizing a hybrid vendor-managed inventory and transportation problem with fuzzy demand: an improved particle swarm optimization algorithm," *Information Sciences*, vol. 272, pp. 126–144, 2014.
- [205] B. Sengupta, K. J. Friston, and W. D. Penny, "Gradient-free mcmc methods for dynamic causal modelling," *Neuroimage*, vol. 112, pp. 375–381, 2015.
- [206] P. J. Green, K. Łatuszyński, M. Pereyra, and C. P. Robert, "Bayesian computation: a summary of the current state, and samples backwards and forwards," *Statistics and Computing*, vol. 25, no. 4, pp. 835–862, 2015.
- [207] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The american statistician*, vol. 49, no. 4, pp. 327–335, 1995.
- [208] S. N. MacEachern and L. M. Berliner, "Subsampling the gibbs sampler," *The American Statistician*, vol. 48, no. 3, pp. 188–190, 1994.

- [209] S. Brooks, A. Gelman, G. Jones, and X. L. Meng, *Handbook of markov chain monte carlo*. CRC press, 2011.
- [210] T. Haslwanter, “An introduction to statistics with python,” *With applications in the life sciences. Switzerland: Springer International Publishing*, 2016.

