Fall 2022

Applications of Machine Learning for Improved Patient Selection and Therapy Recommendations

Brendan Elochukwu Odigwe

APPLICATIONS OF MACHINE LEARNING FOR IMPROVED PATIENT SELECTION AND THERAPY RECOMMENDATIONS

by

Brendan Elochukwu Odigwe

Bachelor of Engineering
Covenant University, 2016

Master of Science
University of South Carolina, 2019

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Computer Science

College of Engineering and Computing

University of South Carolina

2022

Accepted by:

Homayoun Valafar, Major Professor

Marco Valtorta, Committee Member

Forest Agostinelli, Committee Member

Michael Huhns, Committee Member

Roy Mathew, Committee Member

Cheryl L. Addy, Interim Vice Provost and Dean of the Graduate School

DEDICATION

Dedicated to the Odigwe Clan. May we always find ourselves in the midst of noble people.

ACKNOWLEDGEMENTS

I would like to acknowledge the excellent professors at the University of South Carolina, especially my committee members, for their guidance throughout my graduate education.

I would like to express my deepest appreciation to my Advisor, Dr. Homayoun Valafar, for being the best mentor I could possibly ask for.

I could not have undertaken this journey without the unwavering support of my parents, Professor Clement Odigwe and Dr. Francisca Odigwe and the rest of my amazing family. They have been present with words of encouragements, every step of the way, and for this, I am eternally grateful.

Special thanks to my friends who have been there for me throughout this journey and to my colleagues and lab-mates whom I have had the pleasure of working and collaborating with, my sincere gratitude for all the years of support.

ABSTRACT

The public health domain continues to battle with illness and the growing need for continuous advancement in our approach to clinical care. Individuals experiencing certain conditions undergo tried and tested therapies and medications, practices that have become the mainstay and standard of care in clinical medicine. As with all therapies and medications, they don't always work the same way and do not work for everyone. Some Treatment regimens, like Hydroxyurea medication, which is commonly administered to Sickle cell anemia patients, come with some adverse side effects due to the chemotherapeutic nature of the drug. This would be particularly disappointing if the patients must be subjected to such medications without improving their health and quality of life. Some patients, like those battling chronic kidney disease face a more arduous healthcare journey due to the degenerative nature of their illness, coupled with the fact that there are limited tools to forecast the rate of progression of the disease. Asides from the physical toll patients could be subjected to; there is the matter of the economic impact of these therapies on the patients, their family members, insurance companies and even the government. Life-saving therapies like cardiac re-synchronization therapy are cost intensive in addition to requiring surgical procedures. It would be great if we had more ways of identifying patients that are most likely to receive significant benefits from recommended therapies before they are subjected to them.

We will employ a series of machine learning techniques to create models that can indicate a patient's response pattern to recommended therapy. To ensure that our approaches are widely applicable, we would be investigating multiple pressing healthcare problems, namely; Chronic Kidney Disease, Heart Failure, Sickle Cell Anemia, and Peripheral Arterial Disease.

PREFACE

Patients undergo therapies and take medications with the hope of substantial improvement in the conditions that ail them. We know they could benefit from said therapy when subjecting a patient to the tried and tested treatments that have become the standard of care. But it is impossible to tell if the treatment will solve their problems and improve their quality of life. These therapies have adverse side effects, whether physical, financial, social, or otherwise. Some of these side effects present physicians with the challenge of determining whether the therapy elicited benefits outweigh its detriments.

In some cases, patients have conditions that are degenerative in their presentation, i.e., illnesses, where patients are faced with progressive, often irreversible deterioration in the function of an organ or tissue. But, at what rate? The rate of degeneration often varies from patient to patient, and the approach to intervention and recommendations vary based on individual circumstances.

There is an obvious need for the development of more personalized and precise treatment strategies to improve and complement the approach to the patient selection and therapy recommendation currently in practice today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

In Medicine, there are a plethora of illnesses that continue to plague patients and their healthcare providers. This is primarily due to their complexities or the varying ways in which the symptoms manifest. Some diseases, particularly degenerative diseases, are a significant public health concern. Some patients are even put at higher risk of other adverse events they are exposed to due to their ailments. We know that certain medications and therapies alleviate patient symptoms and reduce the incidence of complications, but there is no way to be certain. Physicians often struggle with providing assurances that a patient will receive relief of their symptoms following certain therapies or medications. Some physicians also struggle with identifying the patient parameters most responsible or highly contributing to the nature of their patients' response to the medication or therapy.

With all types of disease treatment, from medication administration to device-driven therapy, a positive response is not guaranteed. Some patients respond positively, while other patients do not. A positive response is often depicted by a desirable change in the magnitude of some observable or measurable parameter.

Recommended therapies are usually tried and tested. Nevertheless, some therapies seem to do more harm than the expected benefits would suggest. Some procedures put patients' bodies under so much stress that it negatively impacts their quality of life for years to come. Some therapies, particularly those involving chemotherapeutic agents, elicit

negative side effects such as nausea, vomiting, acute pulmonary reactions, secondary complications, and illnesses etc.

Even when the physical side effects of therapies are not experienced, the economic cost of some procedures can be overwhelming and negatively impact the standard of living of entire families following the procedure. Some therapies also use up a lot of resources, be it human resources, time, equipment, etc. When the recommended therapy involves surgical procedures, they require a substantially larger amount of these resources, which are usually few and far between.

It is therefore beneficial to spare the non-responding patients from the unnecessary exposure to these side effects. The lack of integration of machine learning techniques in clinical decision systems and medical science is primarily due to the challenging nature of analyzing medical data and the absence of well-annotated and relevant data appropriate for machine learning endeavors.

We will be looking into the application of our proposal with regard to a few healthcare challenges and their corresponding methods of treatment—particularly Heart Failure, Sickle Cell Anemia, and Chronic Kidney Disease. We will also be exploring its potential applications with Peripheral Arterial Disease.

There are usually substantial impediments to which medically related experimentation is faced, e.g., relatively small data sizes, limited patient population diversity, incomplete or corrupted data, limited breadth of clinical data, distribution of data across many disconnected databases, lack of common dictionaries for integration of data, and absence of medical diagnosis or other needed annotations by the attending physicians etc. Even when a substantial amount of data is available, depending on the sizes and sources

of clinical data, some data come with their unique challenges. Retrieval, aggregation, pre-processing, filtering, and preparation of the data across patients with multiple clinical and genomic data points can constitute a considerable feat itself.

The potential for rich analytics is fundamentally dependent on the completeness and quality of the data. Acquisition and annotation of data are a requisite and one of the most cumbersome steps in the development of AI in any scientific endeavor. For instance, supervised training of an artificial neural network requires the annotation of data from every single patient to indicate their outcome. Despite the existence of a vast amount of medical data in electronic form, they are either unattainable or unusable by the community of machine learning and Artificial intelligence researchers.

Although various electronic health records, like the Veterans Affairs (VA) Electronic Health Record (EHR), have helped to remove some barriers, there remains a few technical difficulties in transforming the existing data into a useable set by the community of data analysts. In the course of this work, we have devised means of surmounting some of such challenges.

Another dimension of the proposed work is utilizing machine learning model e.g Artificial Neural Networks (Deep Neural Networks, DNN) to develop sophisticated prognostic models of disease patients. The approaches being proposed will be the first to utilize ML-based techniques, such as ANNs, in the context of chronic kidney disease and Heart Failure. Based on a recent literature search, this will be the first utilization of comprehensive integration of CKD data from a large multi-institute based EMR with a large number of subjects that have repeated measurements of structured and unstructured health information. The most novel aspect of our CKD study is the fact that all parameters

to be used during experimentation, though substantial in quantity, comprise those readily available from routine procedures in a detailed EMR.

It would also be the first to utilize deep learning to create dedicated models for Sickle cell anemia patients, i.e., computational models which are representative of patients, to serve as digital counterparts, with which patient response to parameter modifications and situational changes, purely for observational purposes, can be demonstrated.

In recent years, ANNs have demonstrated tremendous successes in accomplishing human-like recognition tasks. Some of these examples include real-time conversion of speech-to-text that has been implemented on our mobile phones [1], performing automated analysis of medical images [2][3], applications in autonomous vehicles, and predicting the mortality of patients within a year following some event [4]. Because of the successes in other domains of life, the extended use of ANNs in the medical domain is motivated and increased.

Artificial Intelligence (AI) and machine learning (ML) methods have been and continue to be utilized in the domain of healthcare in an academic capacity as topics of research. They have also been applied in areas such as medical fraud detection [5]. Pattern recognition and value prediction techniques rooted in machine learning theory have had great success in establishing complex relationships and identifying patterns between disparate parameters. The availability of predictive tools, such as the ones we are proposing, can be of paramount importance in the development and progression of personalized medicine. If the magnitude of therapy/medication elicited patient response can be predicted, it would aid in the identification of responders, non-responders, and non-compliant patients and an estimation of the extent of expected patient response under

certain, modifiable conditions. This will give physicians the incredibly important ability to predict, on a personalized level, whether a patient's manifested symptoms will be significantly reduced when subjected to the therapy/medication.

In our work, we explore and demonstrate the efficacy of AI/ML strategies in the domain of Medicine and Human Health. Our approach leverages the unique datasets made available to us through our existing collaborators. Using this data, we develop predictive models to develop pragmatic ways of developing alternative diagnostic and prognostic tools for several illnesses and develop an ability to predict those at varying levels of risk beyond what is currently in use in standard medical practice. We have created models capable of understanding patients at the individual level to aid their physicians in making personalized recommendations for their care. With this tool, the healthcare system will be able to effectively reduce the number of individuals adversely affected by unnecessarily being subjected to specific medical procedures, thus, generally improving patient outcomes and quality of life.

CHAPTER 2 MODELLING SICKLE CELL ANEMIA PATIENT RESPONSE TO HYDROXYUREA USING ARTIFICIAL NEURAL NETWORKS

2.1 INTRODUCTION

Sickle cell anemia is an inherited form of anemia, a condition in which there are not enough healthy red blood cells to carry adequate oxygen throughout the body. The disease is caused by defective hemoglobin, which is the oxygen-carrying protein in the red blood cell. Hemoglobins are made up of four peptide chains, each bound to the HEME group. Different Hemoglobins have different combinations of these chains. Adult Hemoglobin HbA, with 2 alpha globins and two beta globins, is the primary hemoglobin affected in sickle cell disease. Fetal Hemoglobin (HbF[1]) is another sub-unit, normally diminishingly expressed as we age but is not subject to the mutation that causes Sickle Cell Anemia (SCA). Hydroxyurea has been observed to increase the expression and, therefore, the amount of HbF in circulation. Most patients respond to Hydroxyurea (HU) therapy with an increase in the HbF concentration of blood [1], therefore reducing the incidents related to Sickle Cell Anemia. Although Hydroxyurea, in general, reduces the number of

[1] Brendan E. Odigwe, Homayoun Valafar. To be submitted to Blood – American Society of Hematology

complications of the SCA disease in most patients, it has numerous deleterious effects resulting from the fact that it is an agent used during chemotherapy. Therefore, the primary challenge facing a physician providing care to an SCA patient is whether the benefits of the HU outweigh its detriments. In the absence of any prior knowledge, a trial period of more than six months is initiated in order to determine a patient's response to potentially harmful treatment.

Pattern recognition and value prediction techniques based on artificial Intelligence or machine learning techniques have had great success, especially in establishing complex relationships and identifying patterns between disparate parameters such as age, gender, or genetic information. Artificial Intelligence tools, including Decision Trees, and Deep Neural Networks (DNN), have historically been used in the development of Clinical Decision Support [2], medical image processing [3][4], and prediction of patient outcomes [5][6].

The availability of predictive tools can be of paramount importance in the development of personalized medicine. More specific to the SCA, if the magnitude of the HU-elicited increase in the percentage of fetal hemoglobin in a patient with sickle cell anemia can be predicted, it will aid in the identification of responders and non-responders to therapy, non-compliant patients, and an estimation of the extent of the response for the identified responders [7]. This approach could also give physicians the incredibly important ability to predict whether a patient's sickle cell symptom(s) will be significantly reduced by the therapy. Having the ability to look at a patient on an individual basis and determine if the projected benefits of treatment are worth the deleterious side effects that could ensue will allow for a more specific patient-guided therapy[6].

Previous work [5] reported the utility of ANN in predicting the effectiveness of Hydroxyurea therapy for Sickle cell anemia patients. In that report, a shallow ANN was used to identify HU treatment responders using a limited definition of responders. Here we continue our previous work by establishing the use of machine learning techniques in a more pragmatic definition of responders as patients who more than doubled their initial %HbF. We also perform additional data analytics in order to expose the challenging nature of this dataset. We went further to explore the utilization of LSTM [8] neural networks to predict a patient's response to Hydroxyurea one month in advance.

With the information provided by the month-in-advance prediction of each patient, we were given the ability to visually observe the period during continuous therapy a previously responding patient began to regress. Various machine learning techniques, like the ones utilized in the paper, can exploit previously indiscernible relationships between parameters to produce results [2]. Finally, we attempted to observe the effect that modifications of one or more of the network parameters would have on patient response.

2.2 BACKGROUND

Sickle Cell Disease/Anemia is an autosomal recessive genetic disease where Red Blood Cells (RBCs) take the shape of a crescent/sickle. Sickle cell mutation is a non-conservative missense mutation where the Glutamate (hydrophilic) in the 6th amino acid of the beta-globin is substituted with Valine (hydrophobic). A hemoglobin tetramer, which is 2- alpha and 2-mutated beta, is called Sickle Hemoglobin (HbS). HbS is able to carry oxygen, but when deoxygenated, it changes its shape, making it capable of clustering with other HbS complexes that form polymers and distort the shape of red blood cells into a

crescent shape (Sickling) (see figure 1)[8]. This change allows for the easier destruction of erythrocytes, causing anemia, among other things [9].

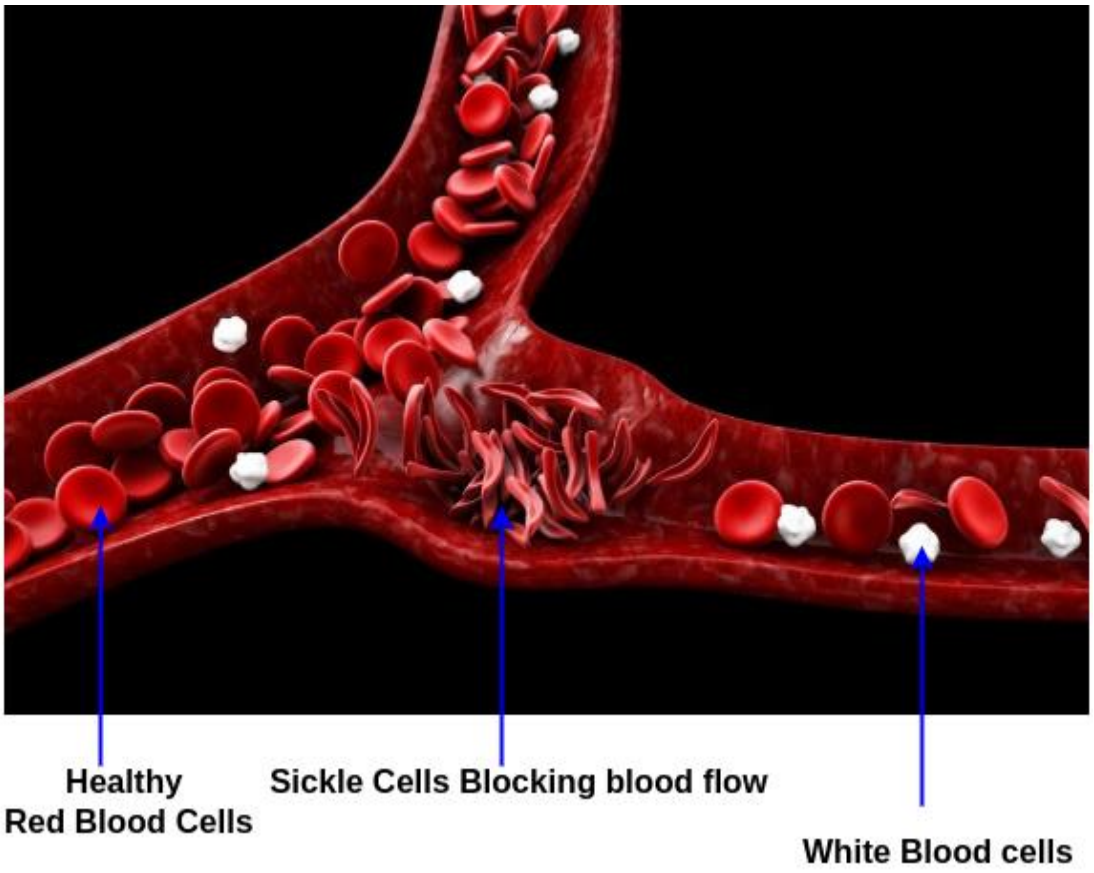


Figure 2.1 Blood Vessel showing the Sickling caused by defective hemoglobin in Sickle Cell Anemia Patients

Hydroxyurea (Hydrea or HU) is a preventive medication that reduces the complications associated with Sickle Cell Anemia (SCA). HU operates by increasing the amount of gamma-globin in the RBCs, which results in the expression of more fetal hemoglobin (HbF) that is not subject to Sickle Cell mutation. HbF is the primary Hemoglobin during the fetal development stages and at birth, which explains why Sickle cell symptoms do not appear until a few months of life (when adult Hb, containing the mutated beta globin, starts to dominate) [10].

As with all medications, a positive response is not guaranteed; some patients respond positively, as depicted by an increase in the concentration of their fetal hemoglobin, while other patients do not respond positively (see Figure 1). While Hydrea reduces the complications associated with SCA, it is a chemotherapeutic agent and may elicit the following side effects: nausea, vomiting, diarrhea or constipation, acute pulmonary reactions, genetic mutation, secondary leukemia, and hair loss, to name a few [11]]. It is therefore beneficial to spare the non-responding patients from unnecessary exposure to this drug's side effects.

In previous work [5], ANNs have been utilized to predict SCA patients' responsiveness by way of a classification model, which categorizes patients into responders and nonresponders. The criteria employed for the classification of responders consisted of patients whose HbF concentration exceeded the 15% threshold. Based on this definition of a responder, a person whose %HbF increased from an initial value of 4.0 to a final value of 4.7 is considered a responder. In contrast, a person whose %HbF increased from an initial value of 20.0 to a final value of 22.9 is considered a non-responder. Based on similar observations, the community has scrutinized the definition of a responder, and new definitions have been proposed. In this work, we explore the usability of various ML techniques as predictive tools for the early identification of responders based on different criteria. In addition, we present a new approach to utilizing the predictive power of Machine Learning techniques that allows us to remain agnostic to the definition of a responder.

2.3 MATERIALS AND/OR METHODS

2.3.1 Dataset

For this study, data from 122 sickle cell anemia patients were collected over a varying period, depending on the patient's response to the treatment and adherence to the study protocol. The cohort of patients were at least 16 years of age and were treated with a daily oral dose of Hydroxyurea, determined based on body weight. The HU dosage was subsequently increased if the HbF levels plateaued (stopped increasing) in successive monthly visits. Only patients who had received HU for 8 months or longer were included in the experiments

Blood samples were obtained and analyzed monthly. The values of 20 parameters [5] for each of the 122 patients were recorded for a minimum of 8 months and a maximum of 98 months. The parameters analyzed for each patient are those normally included in a Complete Blood Count (CBC), Chemistry Profile (SMA18), serial HbF levels (absolute and percent), DNA analyses for globin gene haplotype and the number of genes, treatment duration, weight, age, and gender. The values of these parameters were used to train ML models and for statistical analyses [5].

2.3.2 Statistical Analyses

We took steps to understand the complexity of this study. These steps consisted of performing several standard analytics of the data. In addition to performing the first and second-ranked statistical analyses, density estimation, principal component analysis, and correlation coefficient analysis were performed on the entire dataset.

Table 2.1 A Description of the 23 parameters for which data was obtained from the patients.

KEY	Parameters	Description	Units
X.17	Age	Age of patient at the time of analysis	Days
X.18	Sex	Male/Female	F=1, M=2
X.19	NAGG	Globin gene number	None
X.3	BAN	Number of BAN haplotypes	1,2,or 3
X.20	BEN	Number of BEN haplotypes	1,2,or 3
X.15	Bili	Bilirubin Concentration in blood	mg/dl
X.1	WGT	Patient weight	kg
N/A	% HbF	Fetal hemoglobin as % of total	None
X.4	HbF	Absolute value for fetal hemoglobin	g/l of blood
X.5	Hb	Total hemoglobin concentration	g/dl of blood
X.6	RBC	Red blood cell count	x 10^12

X.8	PCV	Packed Cell Volume	l/l
X.7	RDW	% Variation in the size of red blood cells	None
X.9	Retic	Reticulocytes	x 10^3
X.10	MCV	Mean Cell Volume	Femtolitres
X.11	MCH	Mean Cell Hemoglobin	Picograms
x.13	Polys	Polymorphonuclear leukocytes	x 10^9/l
X.12	WBC	White Cell Count	x 10^9/l
X.14	Plats	Platelet Count	x 10^9/l
X.2	Dos	HU Dosage Administered	g
x.16	NRBC	Nucleated Red Blood cells	Number/WB

Density estimation [12] aims to visualize the effects of HU on patients who undergo the therapy by observing the distribution of HbF before and after the therapy across the entire cohort of participants.

Principal component analysis [13] aims to explore meaningful strategies in performing dimension-reduction in the principal component space to remove the unnecessary complexities of the problem. The correlation coefficient [13], on the other

hand, will be useful in reducing the dimensionality of the problem in the original parameter space by identifying the features with a degree of correlation.

2.3.3 Computational Approaches

We explored computational methods, particularly Decision Trees and ANNs, for categorizing sickle cell anemia patients based on their functional outcomes to HU. We also employed Linear regression and ANNs to predict the HU elicited fetal hemoglobin increase expected after HU administration. Our choice of machine learning techniques is in line with following the minimalist approach of employing a simple model suitable for the problem requirements and extending efforts when necessary.

2.3.3.1 Decision Trees

Decision trees (DT) are a reliable and effective decision-making technique that provides high classification accuracy with a simple representation of gathered knowledge. They have been used in different areas of medical decision-making [7]. They learn from the provided data to produce an approximation with a set of if-then-else rules. The deeper the tree, the more complex the decision rules, and the fitter the model. DT breaks down a dataset into increasingly smaller subsets while, at the same time, an associated decision tree is incrementally developed. Decision trees use different object attributes to classify different subsets of objects. Their great advantage is that they do not use a fixed number of predetermined features. In the decision tree approach, the members of a set of objects are classified as either positive or negative instances (in our case, patients that responded positively to CRT and patients that did not). Candidate attributes that may describe the concept are then outlined [8].

A decision tree construction tool uses outlined attributes to formulate the appropriate decision tree that identifies all positive instances of the underlying concept according to subjects with known classification. The first set of objects used for tree generation is usually called the training set. This initial decision tree becomes a basis for:

1. Forecasting whether a new subject (previously unseen) is a positive or negative instance of the concept being modeled.
2. Investigating the hierarchical representation of the most critical attributes. [9]

As the DT is incrementally developed and the dataset is broken down, class entropy is calculated at each decision to determine if maximal information has been gained from the decisions made on our subjects. The subjects, in this case, are the patients. Our primary interest is to predetermine patients' responses and to identify the contributing attributes supporting each decision. A restriction on the minimum number of patients at a decision node was implemented to prevent over-training. Parameters used for this investigation included patient attributes such as weight, dosage, age, etc.

2.3.3.2 Linear Regression

Linear Regression is a machine learning approach that assumes a linear relationship between the input variables and a single output variable, more specifically, that the output can be calculated from a linear combination of the inputs. It attempts to model the relationship between parameters by fitting a linear equation to observed data. The linear equation assigns one scale factor to each input value, called coefficients. One additional coefficient is also added, giving the line an additional degree of freedom, with which to move up and down on a two-dimensional plot[10].

The least square error is used to derive a function to predict the linear relationship between variables. Suppose, for a dependent variable y and one or more independent variables X, we want to find β that gives us the minimum error following the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

From data, we have observations $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$, and the variables $X_{n\times(p+1)}$ that can be written as follows

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

Therefore, the expectation function is

$$Y' = X\ B$$

where $B = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$.

Now the sum of the squared error is

$$SSE = (Y - Y')^2 = (Y - XB)^2 = \sum_{n=1}^{N} \left(y_n - \sum_{p=1}^{P+1} x_{np}\beta_p \right)^2$$

SSE can be minimized by taking its derivitive and solving it for B which will lead to optimum coefficients B [11].

$$\frac{\partial SSE}{\partial \beta_p} = 0 \ => \ B = (X^T X)^{-1} X^T Y$$

Learning a linear regression model means estimating the values of the coefficients used in the representation with the data that we have available. Given the representation of a linear equation, making predictions is as simple as solving the equation for a specific set of inputs.

2.3.3.3 Artificial Neural Networks

Our investigations utilized Artificial Neural Networks (ANNs) as another predictive machine learning approach. ANNs are brain-inspired systems, which are intended to replicate the way that humans learn. Neural networks consist of input and output layers and (in most cases) a hidden layer consisting of units that transform the input into complex abstracted information that the output layer can use. They are excellent tools for finding patterns that are far too complex or numerous for a human analyst to extract. We utilized several ANN architectures with varying complexities, including the legacy shallow ANNs. Our developments included neural network models using MATLAB [15], TensorFlow [16] and Keras [17].

Within the dataset, some patients did not have complete values for all 22 parameters. In this work, the missing data were substituted with a zero so that it has no contribution to the updated weights during the learning process of the back-propagation algorithm. A Shift of one was added to all data with a legitimate value of 0 (e.g. number of BAN). In addition, the dataset contained data with substantially different ranges. For instance, the age of the patient is measured in days (range was in thousands), while other parameters such as haplotypes units have single-digit values. This disparity in the range of

data is automatically resolved by the normalization of data during the training/testing processes of the models.

Due to the limited data, we employed an n-fold cross-validation approach; train with *n-1*, test with *1*, and repeat *n* times, where *n* is the total number of patients in the dataset. This approach was applied to all experiments conducted with ANNs in this study. The details related to the optimal network architecture and the training-testing processes are discussed later in this chapter.

2.3.4 Experimentation

We developed a shallow two-stage fully connected feed-forward artificial neural network topology with a backpropagation learning algorithm for training and testing. Patients whose post-therapy HbF experienced a 15% increase or more were classified as responders and anyone whose values did not meet that threshold were classified as non-responders. This was performed as a repetition of earlier work[18], [19] to establish comparability in results and further ensure the integrity of results. 72 of a total of 122 patients with the highest number of observations were used for this experiment.

Due to the lack of general agreement on the specific percentage regarded as responding, we established a threshold for a positive response as patients who experience a HbF increase of 100% (double the initial HbF) or more because of its popularity among the community of Sickle Cell Anemia (SCA) researchers [20].

After the initial results obtained from implementing various Predictive models for classifying responders and non-responders, the next step is to attempt predicting a patient's final fetal hemoglobin value after undergoing Hydroxyurea therapy. This is to predict the

highest %HbF observed over the course of medication administration. Again, due to the lack of conclusive agreement on how percentage increase would constitute positive response, it would be of great value to arm the physicians with actual values with which they can determine, on a patient-by-patient basis, if they would consider that to positive response or not. This value would be the highest HbF expected, that the subject will encounter during their course of therapy. We have approached this experiment using both Linear regression and an Artificial neural network.

Although it is of theoretical benefit to gain prior knowledge of a patient's maximum response to HU therapy, such information would be of little practical use without knowing when the effect will be observed (in 6 months or 6 years of therapy). To resolve this limitation, we have explored a new approach where we can predict a patient's HbF one month in advance as a preliminary step toward predicting the time when the full effect of the medication will be observed. To that end, we implemented a recurrent neural network (RNN) with Long short-term memory (LSTM) for the time series prediction.

A recurrent neural network is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. LSTM networks are well-suited to making predictions based on time series data since there can be lags of unknown duration between important events in a time series, as is the case with the problem at hand [21], [22].

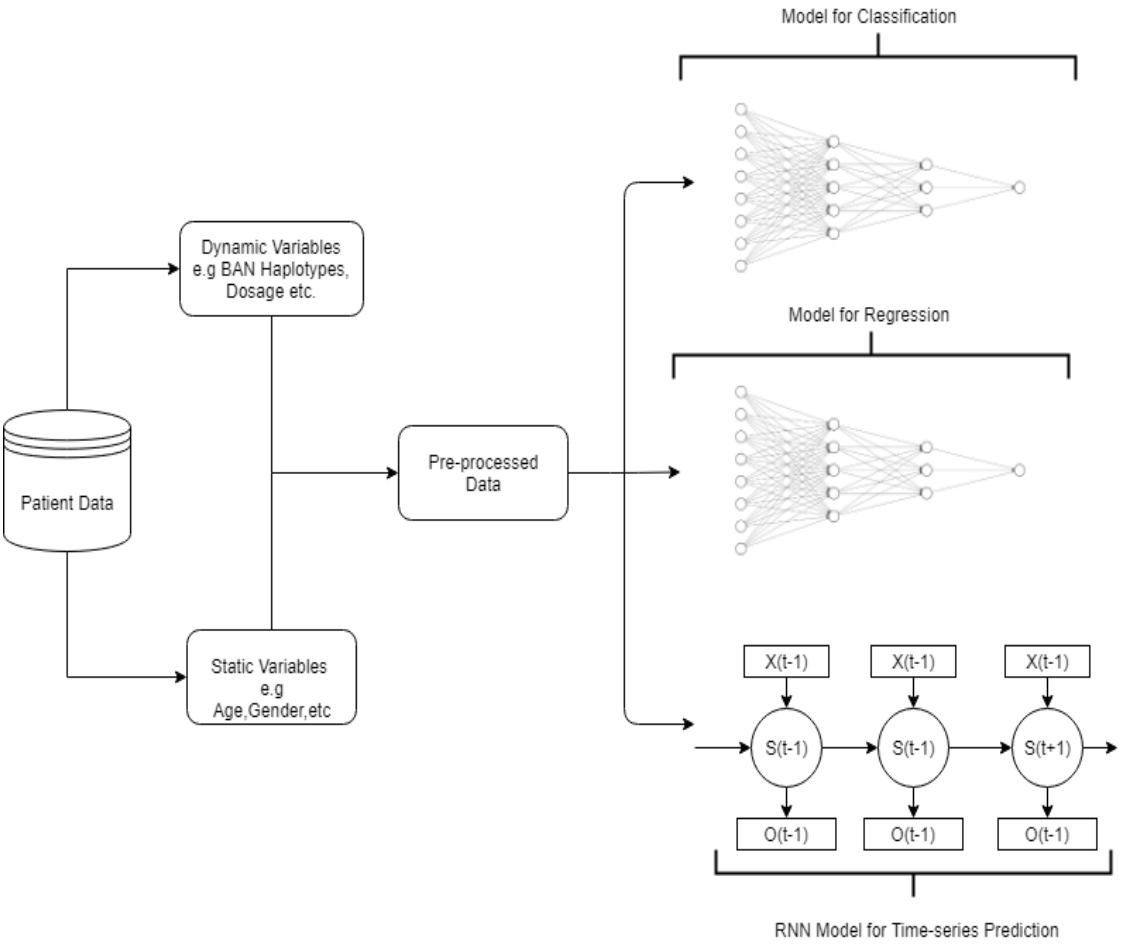


Figure 2.2 Diagram of the Neural Network implementations

During experimentation, we believed that it would be of importance to observe the impact certain parameters have on the prediction task(s). Such information is used to determine the highest contributing parameter and evaluate the effect that changes in a certain parameter could have on the resultant prediction. One of the input parameters used with the neural networks for computation is the hydroxyurea dosage received by the patient. Certain patients were believed (based on response observation) to have failed to respond to medication over the period it was administered. With this, we can tentatively simulate the HbF response the patient could have, should their dosage be modified in the

same way. Information like this could prove useful for designing new studies into SCA patient response.

Regular feed-forward and recurrent deep neural networks were developed using Tensorflow (an interface for expressing and executing machine learning algorithms) and Keras (Python machine learning application programming interface, which allows you to easily run neural networks). Training and testing were performed using the patient data previously obtained after pre-processing. The neural network architecture was regularly modified to obtain the optimal model.

The neural network model used for the final HbF prediction was a deep model constituting 3 hidden layers. 21 input neurons, 12 neurons for the first hidden layer, 8 neurons for the second hidden layer, 4 neurons for the third hidden layer and then a single neuron in the output layer (21-12-8-4-1) since just the single continuous value prediction is expected.

We made use of the Adam optimizer for the iterative update of the network weights and for each layer within the model, we made use of the rectified linear unit (ReLU) activation function with exception to the final layer which has a linear activation function because the output is expected to be a continuous value. This network architecture was used to predict the final fetal hemoglobin value after treatment.

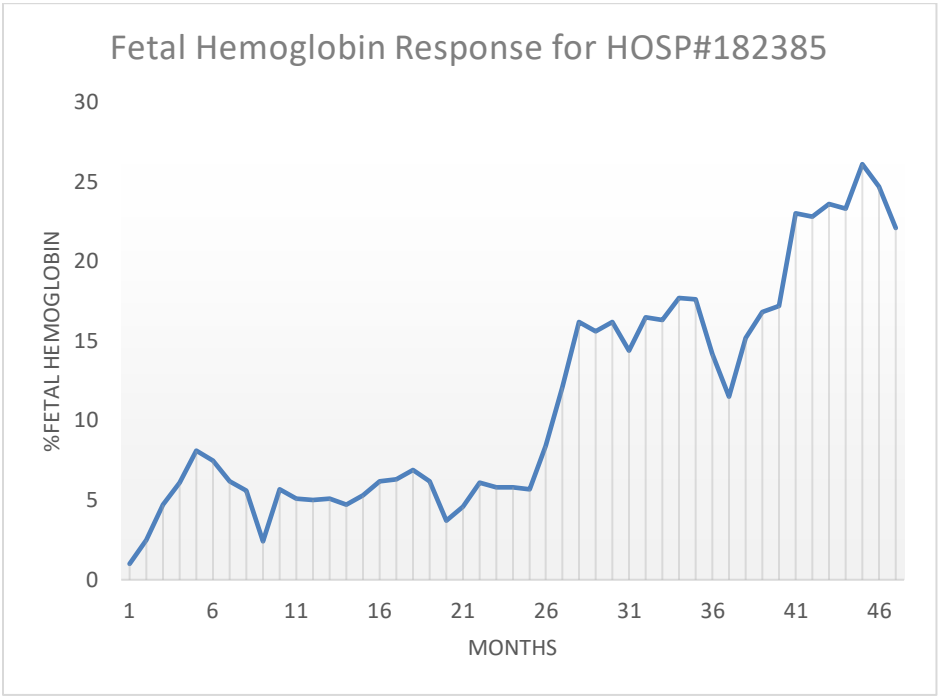
The neural network model used for the periodic prediction was a recurrent LSTM model. The data was prepared by normalizing the features and framing the dataset as a supervised learning problem as predicting the HbF response at the current month given the response and patient biological data from the previous month.

2.4 RESULTS AND DISCUSSION

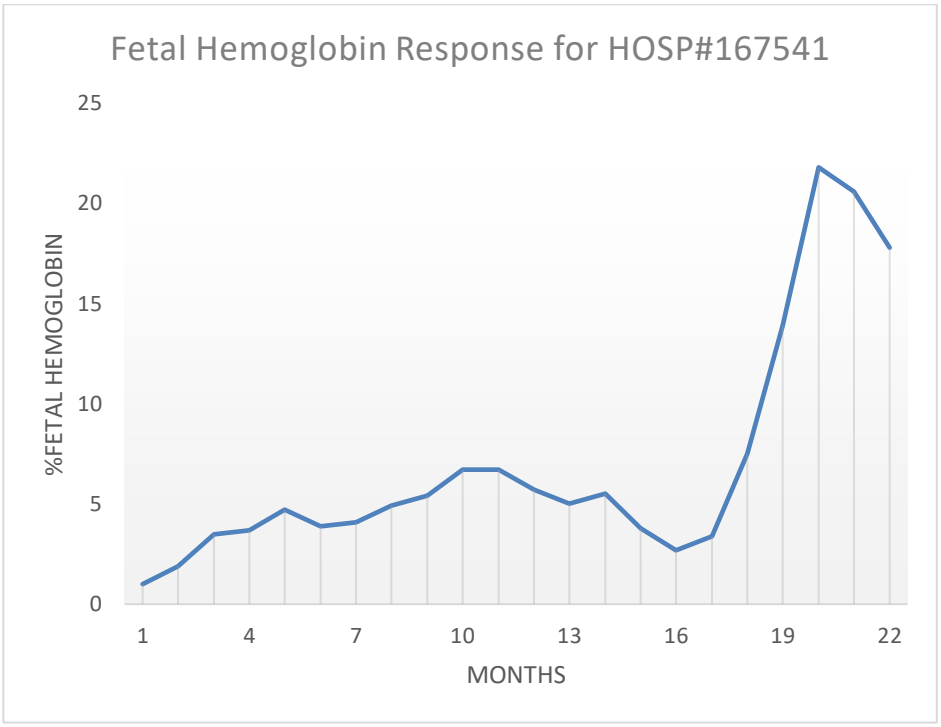
As a first step in conveying the complexity of the problem, Figure 3 illustrates some examples of patient responses to the HU treatment. As illustrated in these figures, not all patients' response to HU results in monotonically increasing levels of HbF. Furthermore, it is evident that in some instances, an initial period of response followed by a decline in the HbF levels. Visual inspection of these response profiles demonstrates the complexity of predicting a patient's response to HU treatment.

In predictive modelling, Traditional statistical methods break down partly because of the increase in the number of observations, but mostly because of the increase in the number of variables associated with each observation. The dimension of the data is the number of variables that are measured on each observation. High-dimensional datasets present many mathematical challenges as well as opportunities.

One of the problems with high-dimensional datasets is that, in many cases, not all the measured variables are "important" for understanding the underlying phenomena being examined. While certain computationally expensive novel methods can construct predictive models with high accuracy from high-dimensional data, it is still of interest in many applications to reduce the dimension of the original data prior to any modeling of the data. [23] [24]

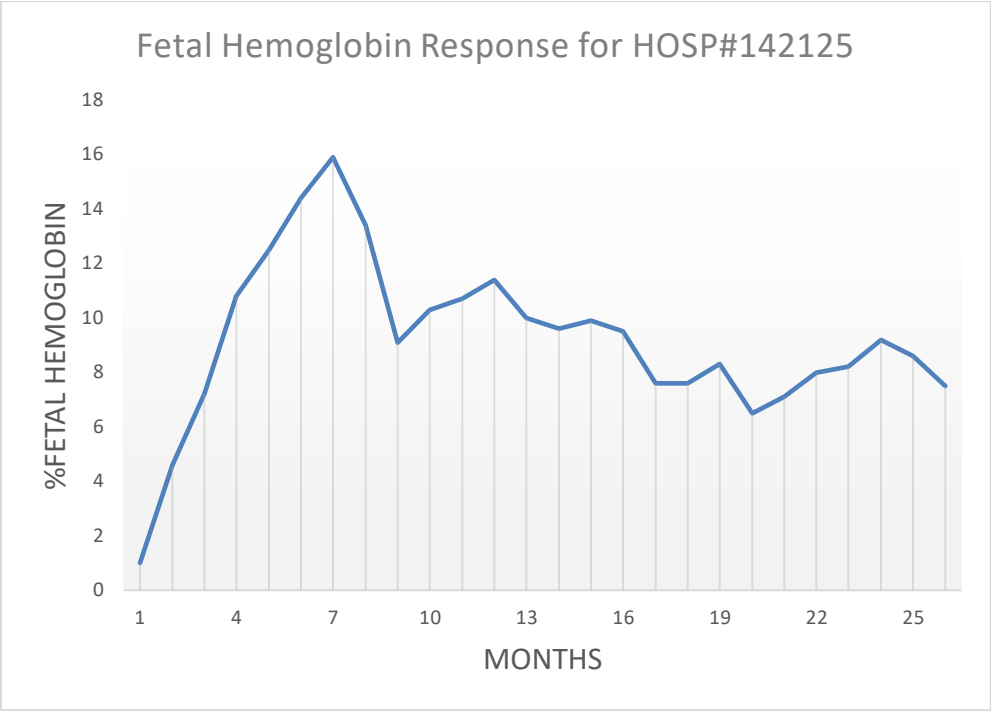

Fetal Hemoglobin Response for HOSP#182385
%FETAL HEMOGLOBIN
30
25
20
15
10
5
0
1
6
11
16
21
26
31
36
41
46
MONTHS

A)

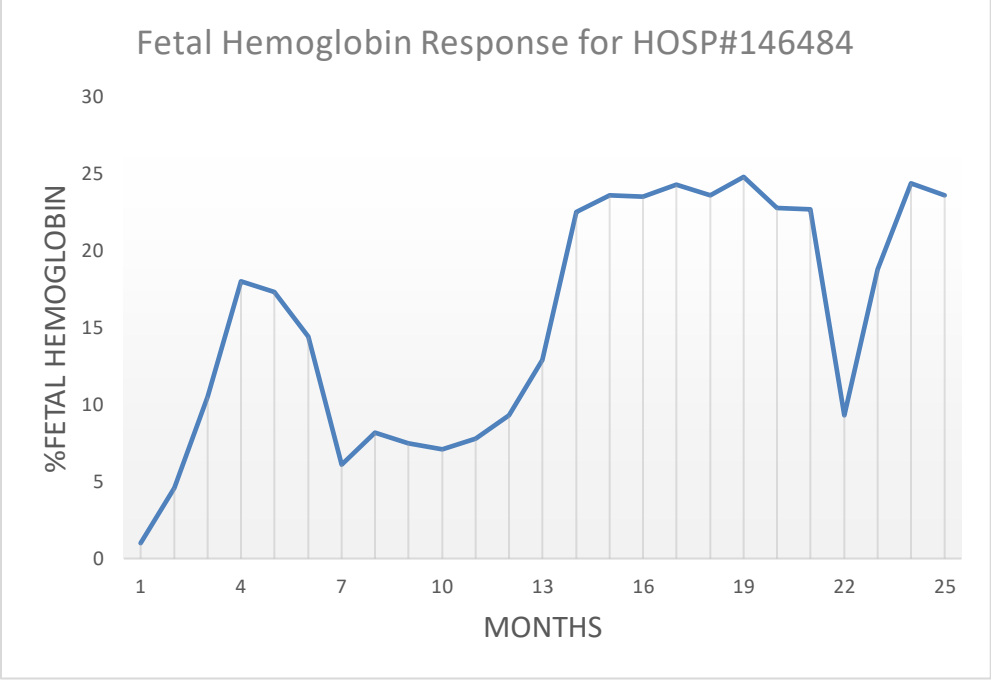

Fetal Hemoglobin Response for HOSP#167541
%FETAL HEMOGLOBIN
25
20
15
10
5
0
1
4
7
10
13
16
19
22
MONTHS

B)

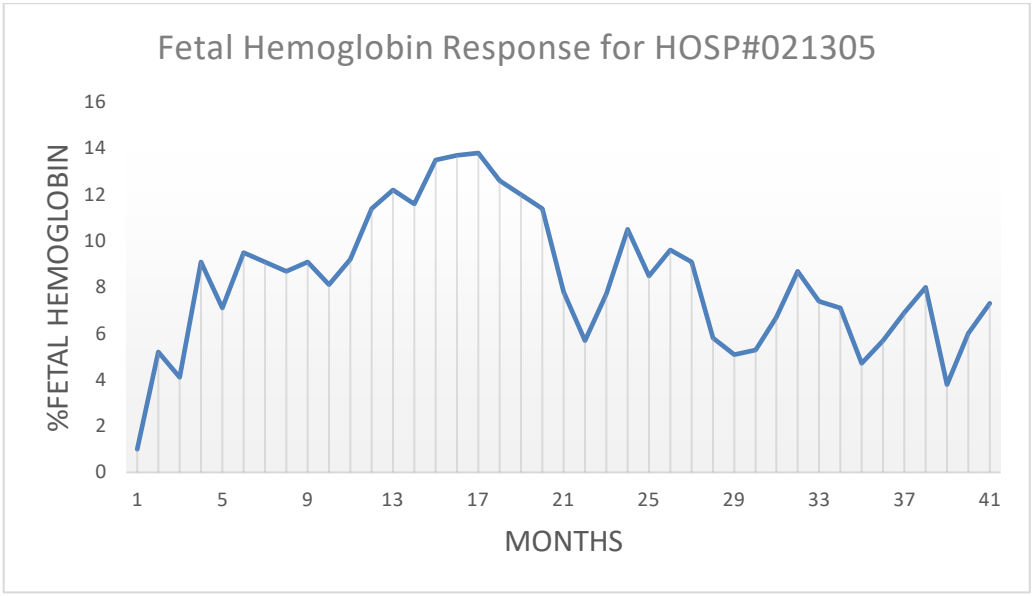
C)


Fetal Hemoglobin Response for HOSP#142125
%FETAL HEMOGLOBIN
0
2
4
6
8
10
12
14
16
18
1
4
7
10
13
16
19
22
25
MONTHS

D)


Fetal Hemoglobin Response for HOSP#146484
%FETAL HEMOGLOBIN
0
5
10
15
20
25
30
1
4
7
10
13
16
19
22
25
MONTHS

E)



F)

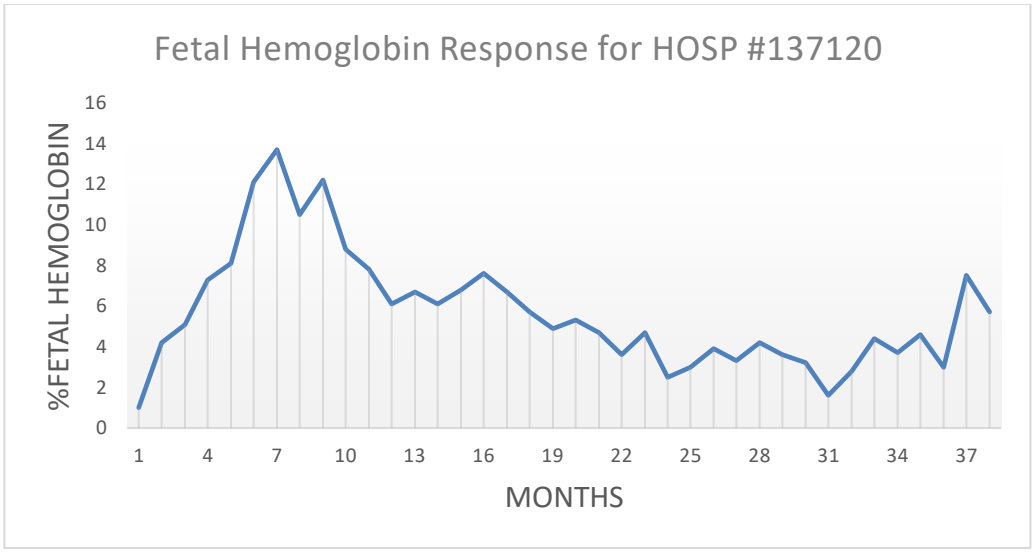


Figure 2.3 (A-F) Plots showing the varying response of different patient's Fetal Hemoglobin to Hydroxyurea

Dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables. The main objective of

dimensionality reduction is to find a low-dimensional representation of the data that retains as much information as possible.

When there is a discernible relationship between two or more variables in the dataset, those features/variables are said to have covariance. The higher the covariance, the more substantial the correlation between these variables and that implies redundancy in the dataset since there is more information than is actually needed. Correlation coefficient analysis was performed on the dataset used for these experiments. The correlation coefficient analysis results are shown in the heat map (shown in Figure 5). As shown in this figure, most of the 20 input parameters of this problem show correlation values within the mid-range of -0.5 to 0.5, indicating that there is no substantial correlation between any two parameters in the dataset to assist with dimensional reduction.

X10 is the Mean Cell Volume which is an indicator for the size of the red blood cell and X11 is the Mean Cell Hemoglobin which indicates the amount of hemoglobin the cell carries. It makes sense that they are highly correlated but cannot be substituted for each other because they are both important measurements towards the physicians understand of anemia.

Finally, the results of the spectral analysis obtained from the principal component analysis are summarized in Figure 5. Based on the gradual decrease in the spectral analysis, it can be concluded that dimensional reduction in the principal component space is not a very effective approach.

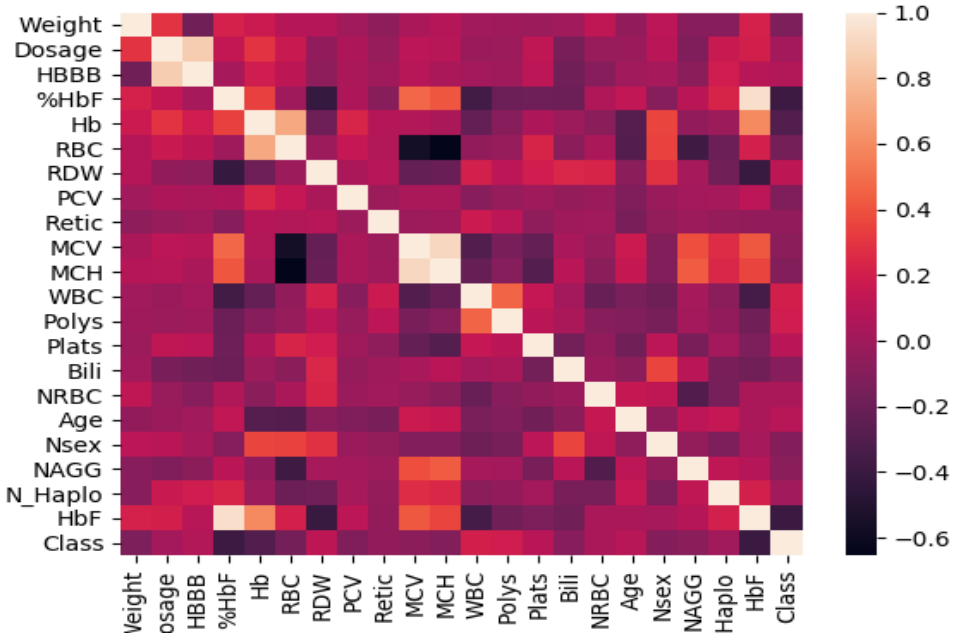


Figure 2.4 Correlation coefficient heatmap of all parameters in the patient dataset

Table 2.2 A Ranking of the 10 most correlated parameters (Use Key from Table 2.1 above)

S/N	Parameters		Corr. Coefficient
1	X.10	X.11	0.914
4	X.13	X.12	0.461
5	X.10	X.4	0.460
6	X.11	X.19	0.430
7	X.11	X.4	0.405

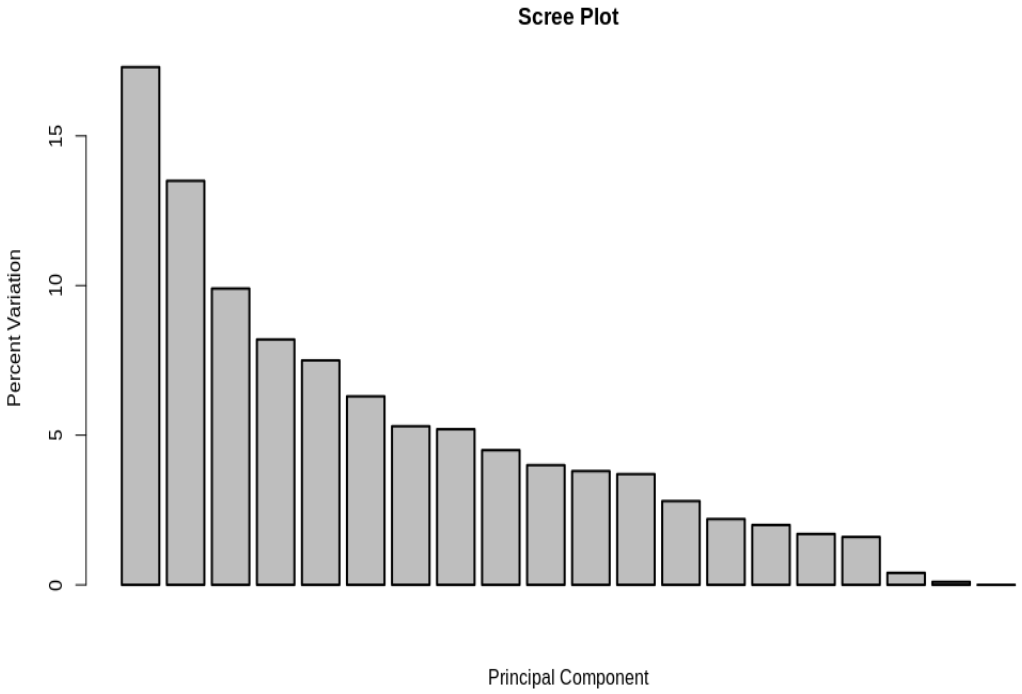


Figure 2.5 Spectral analysis from Principal Component Analysis conducted on the patient dataset

In summary, the collective results of the statistical analyses expose the complexity of the problem statement and suggest that the problem could benefit from the application of artificial neural networks.

2.4.1 Results for Classification Tree Experiments

The HU patient dataset was used to develop decision/classification tree models to classify patients based on their response classes. When unconstrained, decision trees can over-develop, resulting in over-training, where the generated decisions for classification would be overly personalized and therefore will not generalize to a broader population of patients. To avoid that, the resultant tree was pruned by specifying the minimum number

of samples a leaf node was permitted to have, thereby restricting the maximum size of a tree.

Pruning the tree structure is also a way to prevent overfitting because the generated decisions for classification would not be overly personalized and therefore do not generalize to a broader population of patients. Our experiments started by allowing maximum tree growth (minimum samples per leaf-node of 1) and then gradually increasing the minimum samples per leaf until the performance was no longer acceptable.

The acceptable tree performance was chosen to be an accuracy of no less than 70%. Based on these results and the limited number of patients in the dataset, a pruned tree with minimum leaf-node population size of 5 was selected for further evaluation.

Using the chosen decision tree structure and restriction criteria, we evaluated the performance of the approach by randomly partitioning our data into 80% training patterns and 20% testing patterns. This process was repeated 5 times and averaged, yielding the following results. The classification decision tree produced an average performance of 80% accuracy when predicting the patients' response to the CRT therapy, with average precision being 83% and recall of 80%.

2.4.2 Results for Neural Network Experimentation

The ANN topology for the first classification task consisted of two hidden layers with 16 and 4 neurons in the first and second layers respectively. This ANN was able to identify responders and non-responders based on the previously reported 15% threshold for positive response, with 83 percent accuracy.

Table 2.3 Decision Tree Classification Performance

	Precision	Recall/Sensitivity	F1 Score	Support (N)
Responders	0.92	0.73	0.81	15
Non-Responders	0.69	0.90	0.78	10
Average	0.83	0.80	0.80	25

Although redundant, the repeat of the previous results was necessary to establish equality and continuity between the current and past work [18]. A total of 72 patients were used for this study and 60 of them were accurately classified by the neural network.

Another fully connected feed-forward ANN was developed using the same network topology as the previous experiment (22 16- 4 - 1) to perform the classification experiment on a larger dataset comprising of 122 patients. The criterion for categorization as a responder or a non-responder was the HU elicited doubling (100% increase) of HbF value.

During this experiment, 98 of 122 patients were predicted correctly producing a final accuracy of 80.5%. Specificity (the percentage of responders which were accurately predicted as responders) was 88.5% and sensitivity (the number of responders which were predicted to be non-responders) was 60%.

$$Accuracy = \frac{(TP + TN)}{TP + FP + TN + FN} \qquad Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

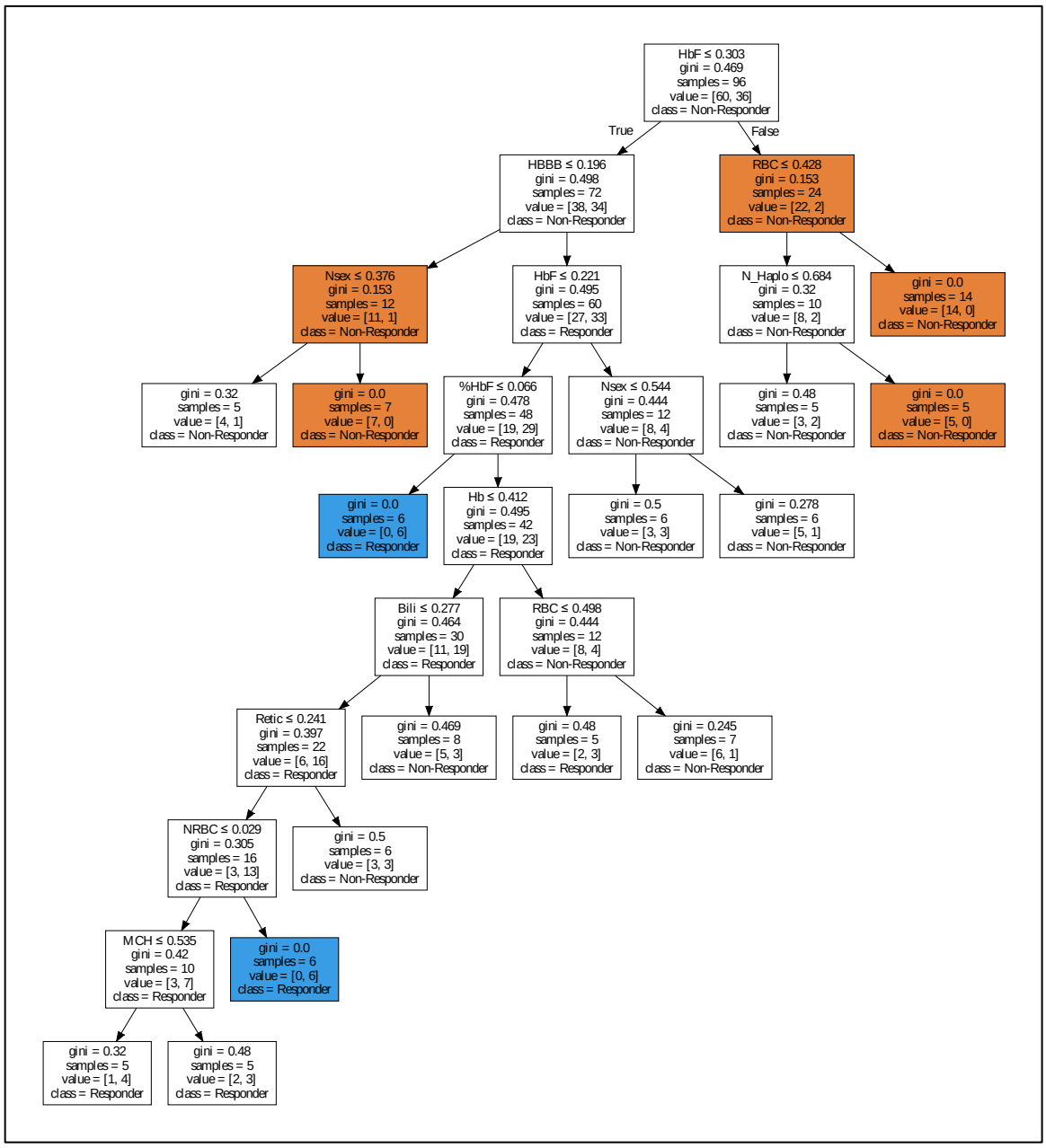


Figure 2.6 Decision Tree for Classifying HFrEF Patients

Table 2.4 Confusion Matrix for the results from the Classification Experiment

	Prediction		
Actual		**1**	**0**
	1	**77**	**10**
	0	**14**	**21**

TP - True Positive = 77, TN - True Negative = 21, FP - False Positive = 14 and FN - False Negative = 10

Predicting the final value of a patient's HbF concentration allows the data analysts to avoid the lack of consensus within the community of physicians and caregivers. In this mode, we can simply predict the final HbF value that a patient will achieve and allow the physician to justify the use of HU. Restatement of the problem turns the problem definition from a classification problem to the regression problem. We experimented with prediction of final fetal hemoglobin using both linear regression and ANN approaches. The models will be predicting a real continuous value and the experiment were performed on the total dataset comprising of 122 patients.

First, we employed a linear regression model for the prediction task which yielded results where the mean and median absolute error values were 0.83 and 0.54 respectively. That amounted to a 19.7% average margin of error, i.e., the average difference between the

actual and predicted values HbF values was +/- 19.7% of the actual HbF value. The distribution of absolute error values can be found in Figure 8.

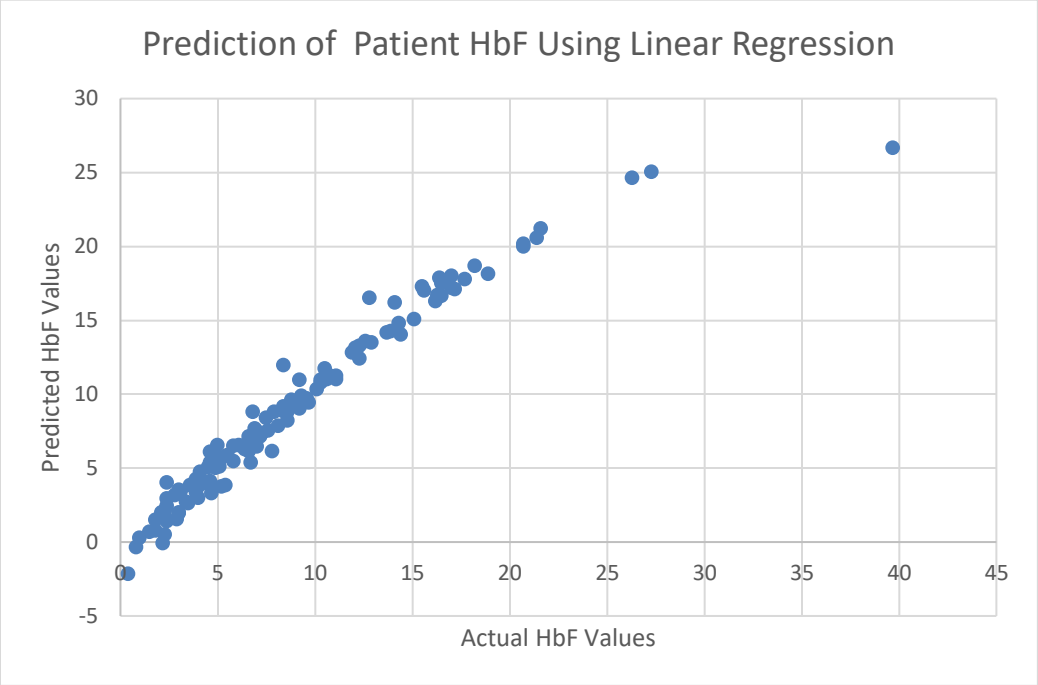


Figure 2.7 Predicted HbF value and their corresponding Actual Values using an ANN

Multiple neural network models were developed and incrementally modified to obtain a network architecture that produced the optimal classification accuracy. We began with a single-layer perceptron and gradually increased the number of neurons and hidden layers while observing the changes in prediction accuracy. During this study, we arrived at and utilized a neural network with 3 hidden layers and 24 (12-8-4) total hidden neurons, which was developed using TensorFlow and Keras.

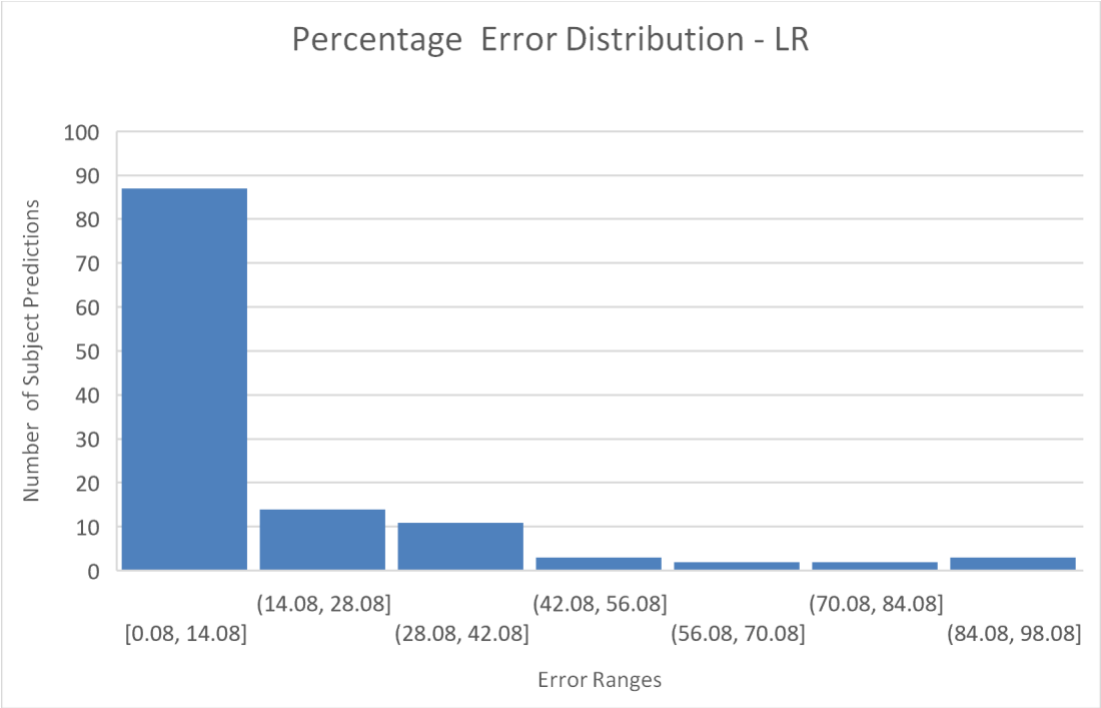


Figure 2.8 Distribution of Percentage Error rates across all subjects final HbF prediction with Linear Regression

Experimentation was carried out using all 122 subjects' data for training and testing. Neural network prediction yielded results where the mean and median absolute error values were 0.75 and 0.34 respectively. That amounted to a 13.37% average margin of error, i.e., the average difference between the actual and predicted HbF values was +/- 13.37% of the actual HbF value. The distribution of absolute error values can be found in Figure 10.

During the evaluation step, we used a percentage error threshold of 30% between the predicted value and the actual value. That is, if the actual value is 10 and the predicted value (P) is greater than or equal to 7 and less than or equal to 13, it is considered correct,

else, the prediction is considered false. With this approach to evaluation, 113 of the 122 predictions were correct giving a 92.6% accuracy level.

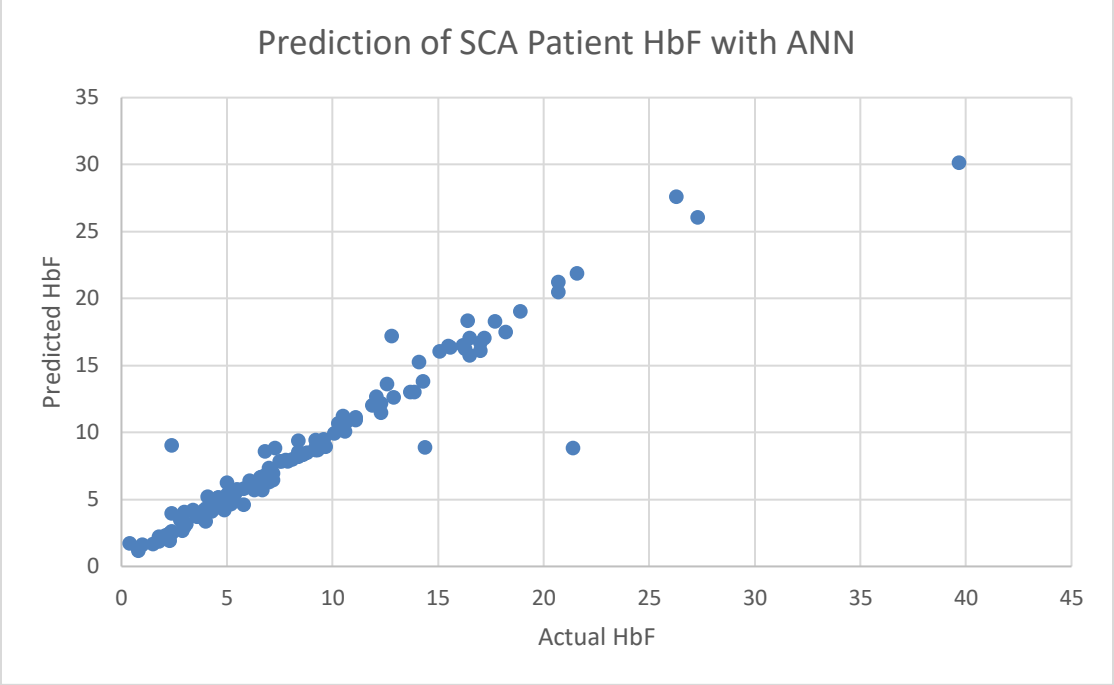


Figure 2.9 Predicted HbF value and their corresponding Actual Values using an ANN

The histogram in Figure 2.10 shows the percentage error regulating from the experiments for all 122 patients. Results show that a substantial amount of patient's predictions had minimal error suggesting that the neural network model was doing a great job of predicting final values close to their actual response values.

Although the reformulation of the problem from classification into the regression domain alleviates the dependency on the accepted definition of a responder, it does not remove other existing logistical limitations. For instance, while the regression approach

may predict correctly the maximum achievable value of %HbF, it fails to provide the conditions under which such outcomes can be achieved.

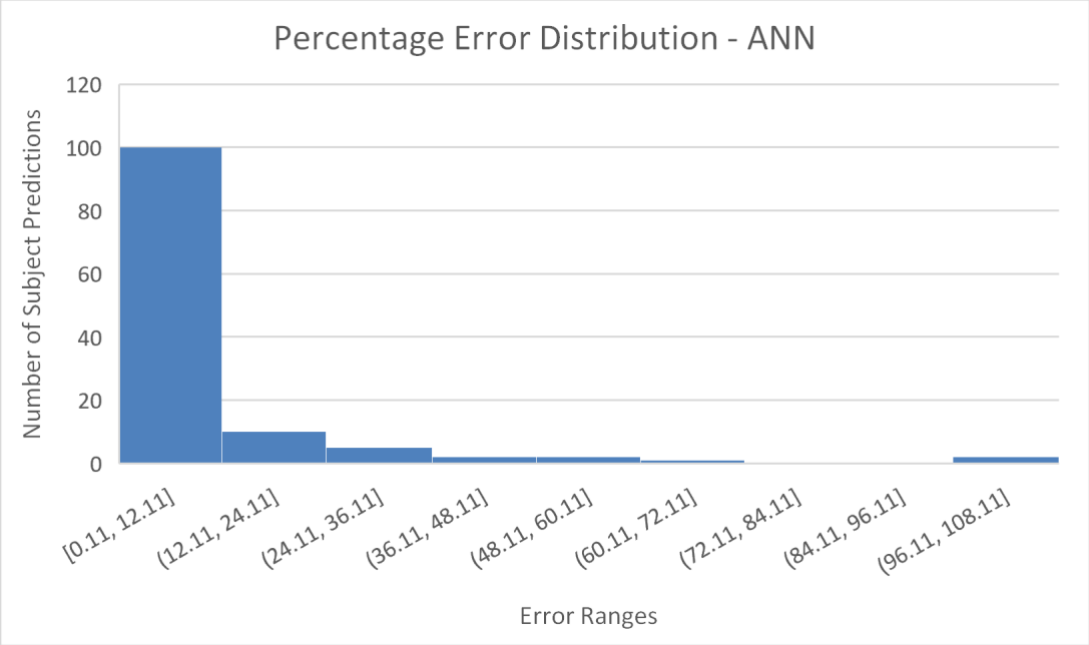


Figure 2.10 Distribution of Percentage Error rates across all subjects final HbF prediction with an ANN

Some of the critical values that may be required in this prediction are the drug dosage and the time needed for the HU therapy to take the maximum effect. Here we present the results of a preliminary approach that allows prediction of %HbF for a given patient 30 days ahead of time. In this exercise, we employed the use of LSTM networks (modeled in TensorFlow and Keras) on a subset of patients. This was performed as a series of regression problems (depending on the number of months the patient underwent therapy) with continuous output values at each step.

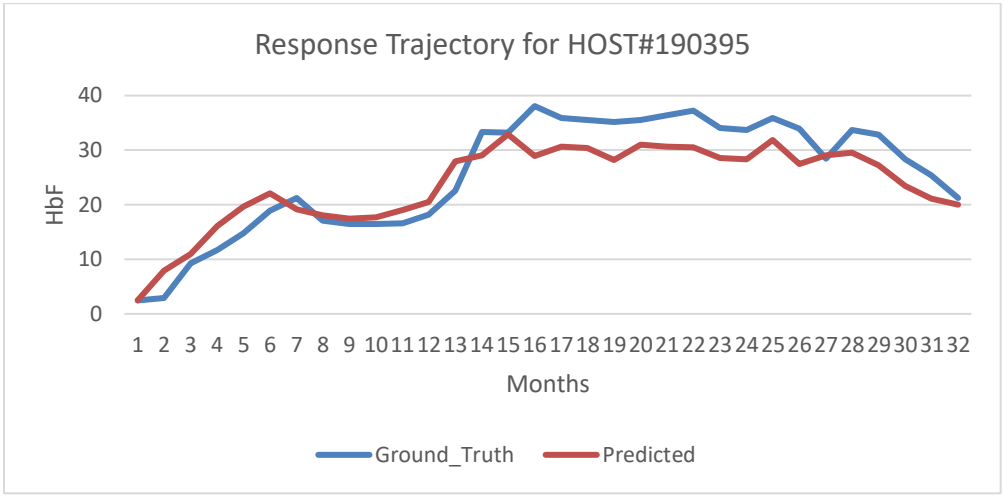
For each patient, we used a series of actual values reflecting the next month's HbF value. The monthly predictions are expected to follow the same sequence of progression or decline to be considered a successful prediction.

The prediction task yielded results where the mean and median cumulative absolute error values were 3.27 and 3.21 respectively. That amounted to a 37.2% average cumulative margin of error, i.e., the average difference between the actual and predicted HbF values was +/- 3.27 of the actual HbF value across all the months of observation.

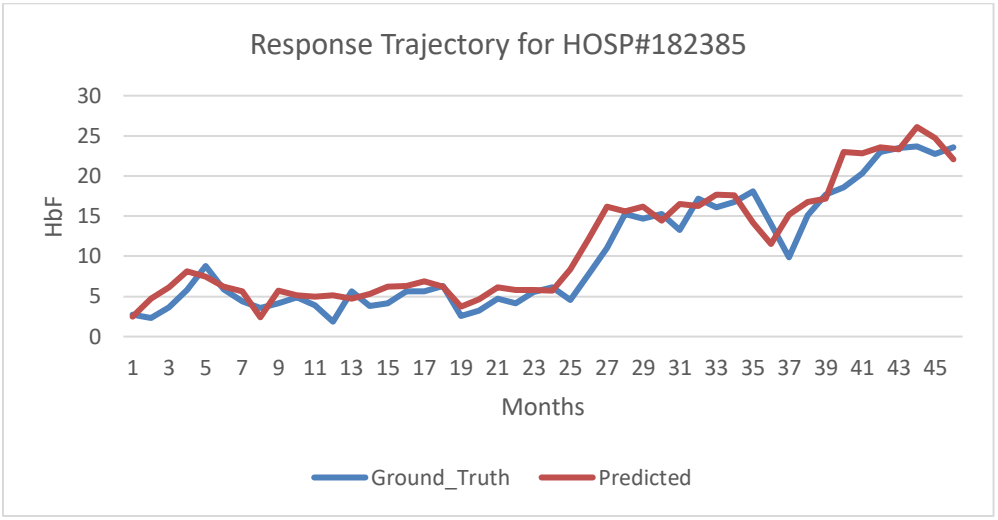
Pearson's correlation coefficient [13] (PCC) was used as a metric to determine the similarity in the monthly progression of the patient's predicted and the expected HbF values. The results corresponding to six of our patient's response trajectory are shown in Figure 11. In this figure, trends illustrated in blue and red correspond to the actual and the predicted response trajectories for the test subjects. Correlation coefficient values range from -1 to +1. Outputs ranging between 0.1 to 0.399 are considered to have poor correlation and values ranging between 0.4 and 1 are considered to have average to high correlation [25]. Considering the above categorization, the ANN exhibited good performance with 66/81 patients' prediction correlation ranging between 0.4 and 1, resulting in an accuracy of about 82%.

The total duration of observation for all patients as well as available patient data per observation varied substantially, therefore, it was necessary to establish exclusion criteria for certain patients and certain observations in order to preserve the integrity of the results.

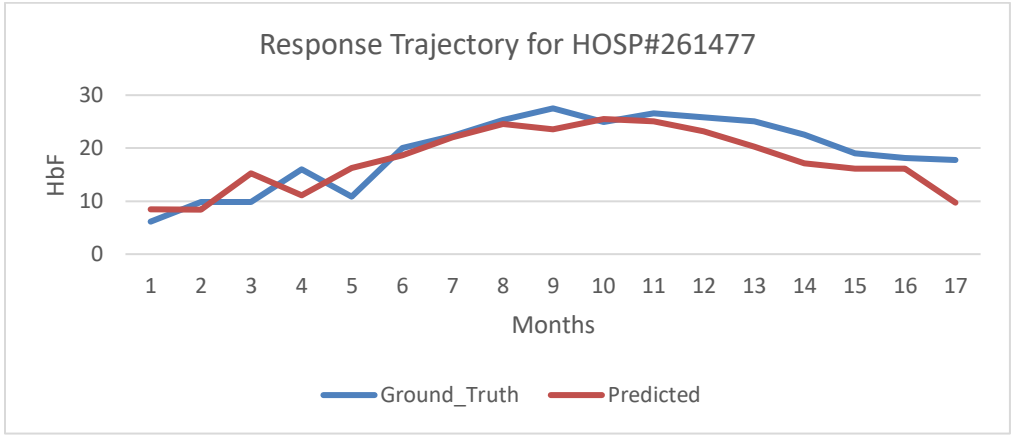
A)


Response Trajectory for HOST#190395
HbF
0
10
20
30
40
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
Months
Ground_Truth
Predicted

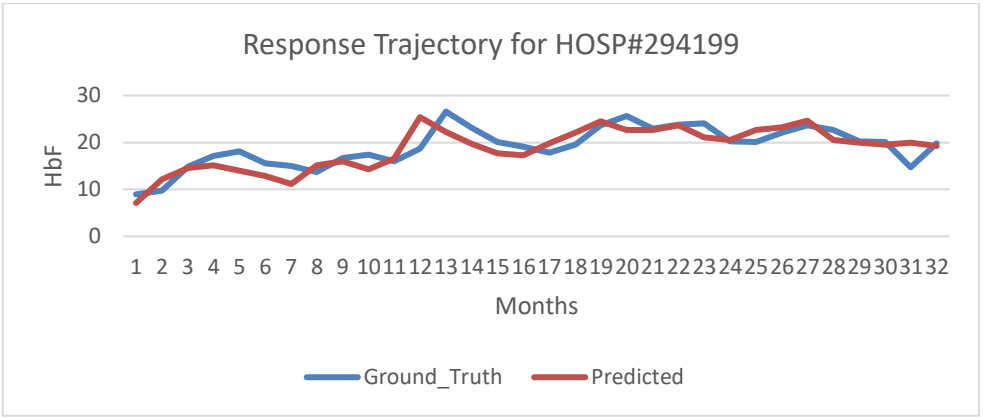
B)


Response Trajectory for HOSP#182385
HbF
0
5
10
15
20
25
30
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45
Months
Ground_Truth
Predicted

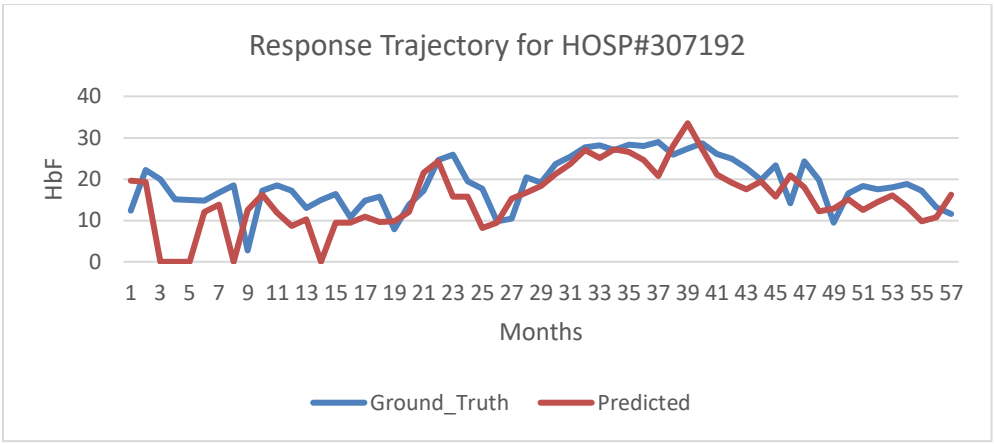
C)


Response Trajectory for HOSP#261477
HbF
0
10
20
30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Months
Ground_Truth
Predicted

D)



E)



F)

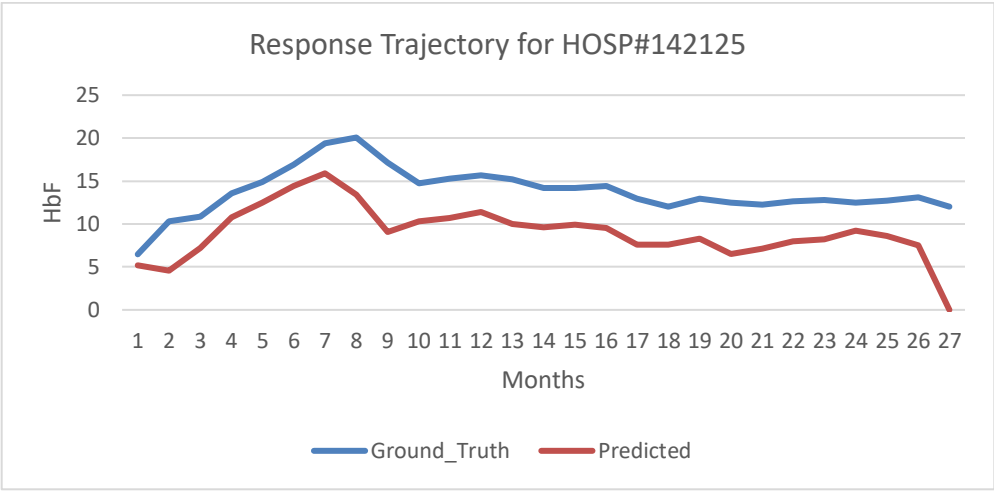


Figure 2.11 (A-F) Some Sample patient's time-series results predicted with the LSTM neural network.

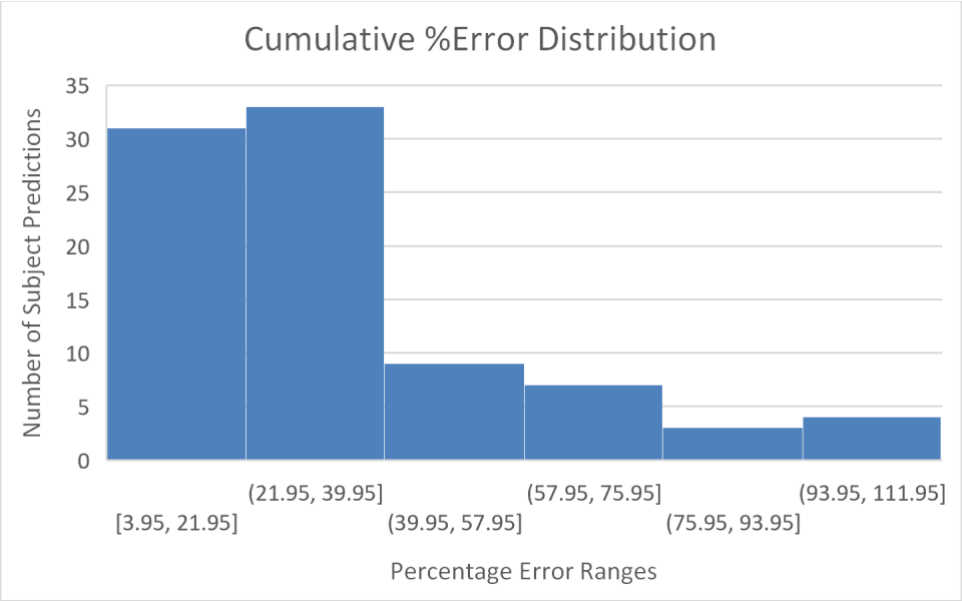


Figure 2.12 Distribution of Percentage Error rates across all subjects' response trajectory predictions.

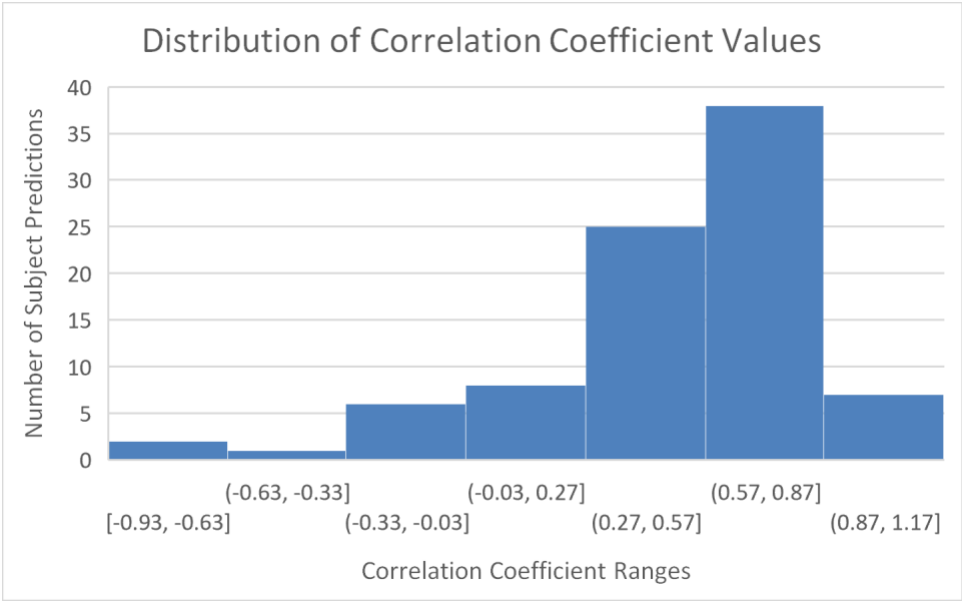


Figure 2.13 Distribution of Correlation Coefficient Values across all subjects' predictions in the dataset.

Patient observations with 3 or more missing values were excluded from the analysis prior to training and testing. Patients having total observations of five (5) or below were also excluded from the analysis. About 18 percent of the total dataset were excluded after the implementation of these exclusion criteria.

We isolated certain patients whom by observation, did not appear to respond positively to medication or patients who experienced a substantial decline in the later stages of taking Hydroxyurea. We altered the dosage administered randomly linearly, mainly during the time steps where the substantial decline was noted. This was done to see the level of impact such changes will have on the predicted response of the patient. This experiment generated mixed results. The periodic HbF predictions of certain patients tentatively responded positively to alterations in the administered dosage while certain patient's responses did not change positively with varying modifications made to the dosage prescribed. This suggests that modification in dosage could not universally sufficient as the only criteria to possibly achieve a positive response to medication.

2.5 CONCLUSION

Neural networks are capable of recognizing patterns and complex relationships between parameters that are not easily discernible by man unaided by a machine. The parameter values used for this experiment were obtained before and after the patients underwent Hydroxyurea therapy.

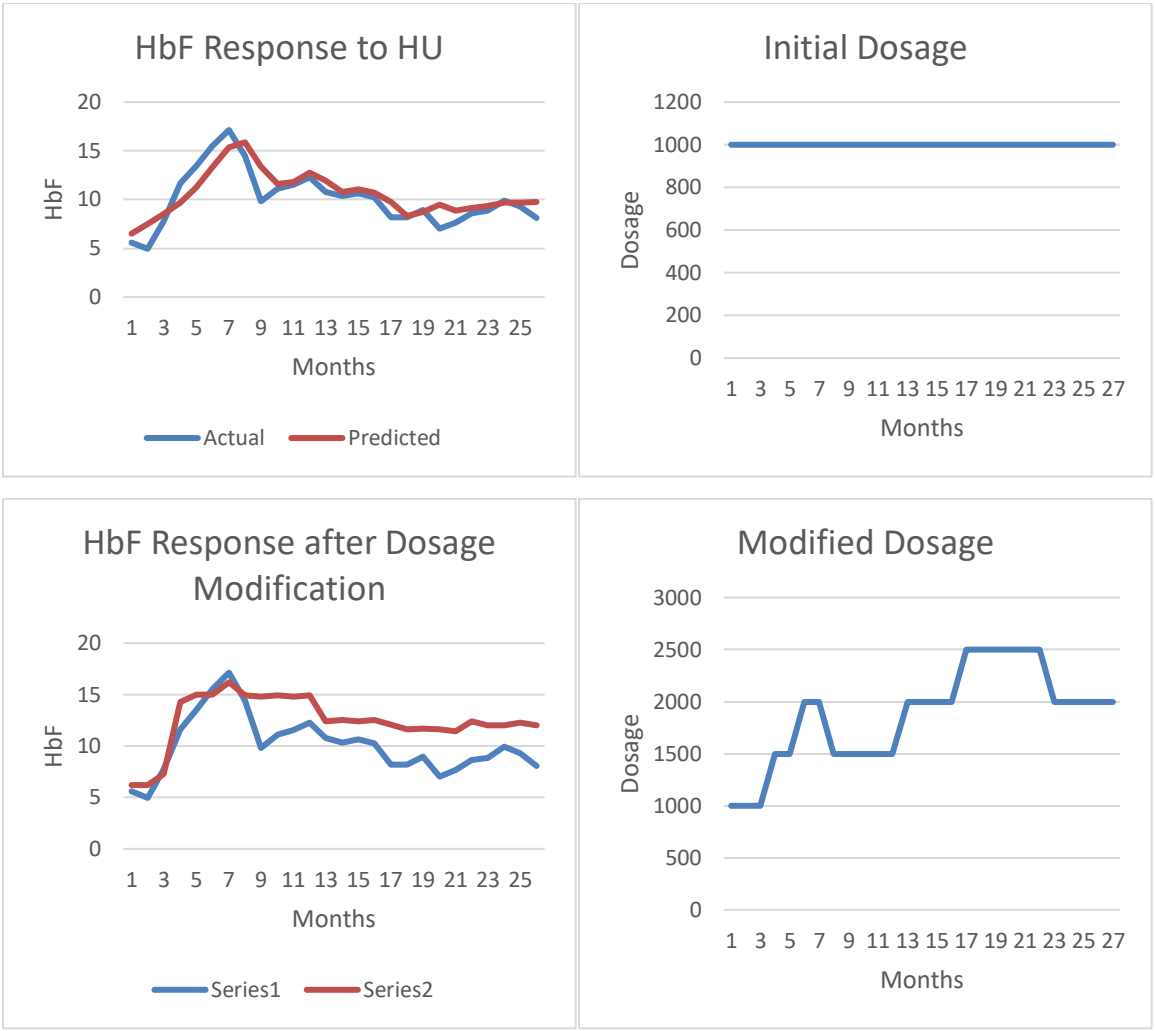


Figure 2.14 Plots showing patient response and their corresponding effect dosage change had on the HbF response of a sample patient.; top-left: Patient response from monthly therapy, top-right: dosage administered to achieve the response in top-left, bottom-left: Simulated response achieved after dosage modification, bottom right: Dosage administration values used to achieve the response in the bottom-left

We have demonstrated that we can (with 83% accuracy) identify patients that will respond to Hydroxyurea (HU) therapy and those that are not likely to respond. We have

been successful at predicting (with 92.6% accuracy) a patient's final fetal hemoglobin value after undergoing HU therapy such that the percentage error between the actual value and the predicted value is less than or equal to 30%.

Regardless of the way the output classes are categorized, training ANNs to accurately distinguish between different classes of patients is of value to the medical community. This experiment shows that ANNs are capable of exploiting correlations between medical data that physicians are not able to identify unaided. This approach and others like it will positively influence medical decision making, administration of intervention procedures and improve the practice of precision medicine.

We have also demonstrated the potential for using a single, well trained general model to predict the response trajectory of Sickle Cell Anemia (SCA) patient to Hydroxyurea (HU) with 81% accuracy.

CHAPTER 3 APPLICATION OF MACHINE LEARNING IN EARLY RECOMMENDATION OF CARDIAC RESYNCHRONIZATION THERAPY TO HEART FAILURE PATIENTS.[2]

3.1 INTRODUCTION

In the era of value-based healthcare, digital innovation, and big data, clinical decision support systems have become vital for organizations seeking to improve care delivery. Clinical decision support (CDS) tools can analyze large volumes of data, flagging potential problems and enhancing team efficiency [12]. Central to many of these improved CDS tools are techniques and procedures that involve techniques firmly rooted in Artificial Intelligence (AI) and Machine Learning (ML) theory.

Artificial intelligence (AI) based predictive tools such as Decision Trees and Artificial Neural Networks (ANNs)[13], [14] have been implemented in recommender systems to enhance customer shopping experiences [15], human activity recognition [16][17], analog circuit design [18], assist with speech recognition, and natural language processing [1]. They have also been incorporated into a plethora of other domains.

---

[2] Brendan E. Odigwe, Alireza Bagheri Rajeoni, Celestine I Odigwe, Frank G. Spinale, Homayoun Valafar. 2022 Proceeding of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. Reprinted here with permission from the publisher. https://authors.acm.org/author-resources/author-rights

Extensive efforts have been made to create new mechanisms of predictive modeling (such as LSTMs, DNNs, and CNNs) [4], [19], as well as to improve already existing methods (such as linear regression and decision trees) [7], [11].

While ANNs have advanced substantially over the last few years, their impact and integration in medical diagnostics have been minimal. The lack of integration of machine learning techniques in Clinical Decision Systems and medical science is due to the challenging nature of analyzing medical data and the absence of well-annotated and relevant data appropriate for machine learning endeavors [20].

Despite these challenges, ML techniques have been applied in the domain of healthcare in applications ranging from medical fraud detection [5] to the prediction of a patient's response to medical treatment [21], thereby paving the way for extensive efforts toward personalized medicine. ANNs have been used in other diagnostic applications that include analysis of medical images[22], cancer detection[23], ophthalmic disease[24], etc. More recently, they have improved patient outcomes by optimizing personalized medical care. In this regard, they hold the potential to predict, and therefore, improve diagnosis and treatment strategies.

Heart failure (HF) is a constellation of clinical signs and symptoms due to underlying defects in left ventricular (LV) function. While pharmacological therapy forms the mainstay for HF treatment, device-driven approaches such as Cardiac Resynchronization Therapy (CRT) have also become a standard of care [25]–[27]. The basis of CRT is to improve LV myocardial conduction patterns in HF patients. Clinical trials have shown that CRT can improve HF symptoms and clinical outcomes in HF patients [28], [29]. However, these studies and retrospective reviews of CRT have

identified that a large proportion of HF patients (30-50%) do not demonstrate an improvement in functional/clinical outcomes [25], [29], [30]. Specific criteria have been established under which certain HF patients qualify for CRT consideration. Such criteria include reduced LV ejection fraction and prolonged myocardial conduction [29]. While HF patients may meet the criteria for CRT, a large portion of these patients derive no benefit. Therefore, developing strategies to identify HF patients with the most significant probability of responding to CRT and conversely identifying a subset of HF patients with a low likelihood of CRT benefit holds significance in improving HF prognosis, treatment, and medical resource utilization.

Considering the amount of HF patients regularly undergoing CRT [31], with relative effective rates averaging 50-70%, it is safe to assume that a robust group of CRT "responders" and "non-responders" exist. As such, we hypothesize that machine learning approaches would be amenable for the prediction of CRT response. For this study, we utilized an existing CRT database, obtained for use in a previously conducted study [31], in which prospective classification of CRT response was established. Demographic, LV function, and biomarker measurements were obtained before CRT placement and subsequent measurements were taken three and six months after CRT placement.

We investigated machine learning methods of classifying candidate patients into responders and non-responders to identify those HF patients with the greatest probability of functional improvement with CRT. Furthermore, we have also explored the prediction of the magnitude of left ventricular function after CRT, observed as the left ventricular end-systolic volume (LVESV).

3.2 BACKGROUND

Heart disease is a leading cause of morbidity, mortality, and significant healthcare costs. It is the leading cause of death in the United States, responsible for about 1 in 4 deaths, and occurs when the heart muscle doesn't pump blood as well as it should [32].

Ejection fraction (EF) is the measurement of how much blood the left ventricle pumps out with each contraction. Preserved EF is a situation where the lower left chamber of the heart is not able to fill properly with blood during the filling phase, so the amount of blood pumped out to the body is less than normal. Reduced EF is a situation where the heart is not able to contract properly, so it doesn't pump out sufficient blood to the rest of the body [32]. All participant patients involved in the study reported in this manuscript were diagnosed with Heart Failure with reduced Ejection Fraction (HFrEF).

The CRT pacemaker coordinates the timing of the upper heart chambers and the lower heart chambers. It is a treatment to restore the regular timing pattern of the heartbeat. It also works on the timing between the left and right sides of the heart by sending electrical signals to the ventricles to pump together, the way they should. This type of electrical stimulation is called biventricular pacing [33].

In medical decision-making (classification, diagnosing, etc.), there are many situations where decisions must be made effectively and reliably. Conceptual simple decision-making models with the possibility of automatic learning are the most appropriate for improving the performance of such tasks.

3.2.1 Previous and Related Work

Classification is a task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain [34]. Binary classification problems aim to observe one of two categories by analyzing a series of attributes. An example is a medical diagnosis for a single medical condition (e.g., disease vs. no disease) based on a battery of tests and measurements. Numerous approaches to classification have been developed [35], but, the choice of technique to apply usually depends on the domain/task to which it is applied as well as the nature and constitution of the data. Classification algorithms have seen increased application in the medical domain in diagnosis[36], prognosis[34] as well as treatment [21].

While several large clinical trials have demonstrated that CRT can be beneficial to a cohort of HF patients, few have examined predictive algorithms. One example is the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT- [31]). This study developed a composite clinical scoring system and related that to CRT functional outcomes. While it was more of a generalized classification scheme and not patient-specific, this was the first approach to develop a relative risk profile for HF patients and relation to CRT response. The SmartDelay Determined AV Optimization: A Comparison to Other AV Delay Methods Used in CRT [30], was a large clinical trial that used a prespecified quantifiable outcome measure at six months - LV volume reduction and collected demographic, functional, and biomarker profiling data at baseline (pre-CRT placement). One outcome of the SMART-AV study was identifying potential biomarkers (using a peripheral blood sample) that could be integrated into a clinical algorithm, such as the MADIT-CRT score[31]. The SMART-AV

study highlighted the complexity of developing modeling algorithms using standard parametric statistical approaches to predict a CRT response and produced the dataset utilized in this paper.

While these past studies highlighted the need for prediction algorithms in HF patients being considered for CRT, this remains an unmet medical need. Improper selection of HF treatment impacts both patient health outcomes and the economics of the disease. Studies focused on the economic implications of CRT report [37][38] an estimated value of $53,589/quality-adjusted life-year (QALY) gained over the patient's lifetime for those who respond positively to CRT therapy.

The goal of our study is to expand the scope using multiple variable domains (demographic, functional, biomarker) from the past SMART-AV study [25] using machine learning approaches; particularly cluster analysis, logistic regression, decision trees, and ANNs to classify patients based on their expected outcome to CRT as well as predict the magnitude of left ventricular volume 3 months after CRT.

We conducted an initial study and reported on the use of ML techniques only for patient response classification [39]. In addition to our earlier study, this will be one of the first studies to encompass many variables obtained in HF patients before and following the institution of CRT.

3.2.2 Description of the Data

The dataset was obtained from the SMART-AV clinical trial, for which all methods, approaches, and the way measurements were made have been published [37], [38]. The entire dataset contained 1045 patients with 3 observations (for each patient) recorded at months 0, 3, and 6. Each observation consisted of 80 attributes or parameters.

A prospective 6-month endpoint of a reduction in LV volume was defined as a CRT response. The investigations and results discussed in this paper have been performed using only data observations obtained in month 0 (at baseline). A description of data collection can be seen in Fig 3. 1.

Patients with more than 10 illegitimate parameter values (missing values) were eliminated from our investigation, resulting in 764 remaining patient records. 368 of these patients were classified as responders, while 396 were classified as non-responders at the end of the 6 months of CRT.

Each participant had parameters used for computational experimentation to 53 (52 input parameters and 1 output). Missing or corrupted data points were replaced with the mean across the population cohort to minimize skewing the results of the experiments. The mean is assumed to be within the standard value range for all features and a reasonable estimate for the missing values. Dataset values were normalized to accommodate substantially different unit ranges of various parameters. The ground truth classification of the subjects' outcomes in the dataset were assessed and established by clinicians before the experimentation described in this paper.

3.3 MATERIALS AND METHODS

We explored the use of computational methods, particularly Cluster Analysis, Decision Trees, Linear Regression, and ANNs, for classifying candidate patients as responders and non-responders based on the functional outcomes of HF patients to CRT. We also employed the use of Linear regression and ANNs to predict the value of CRT elicited left ventricular volume that will be observed in patients.

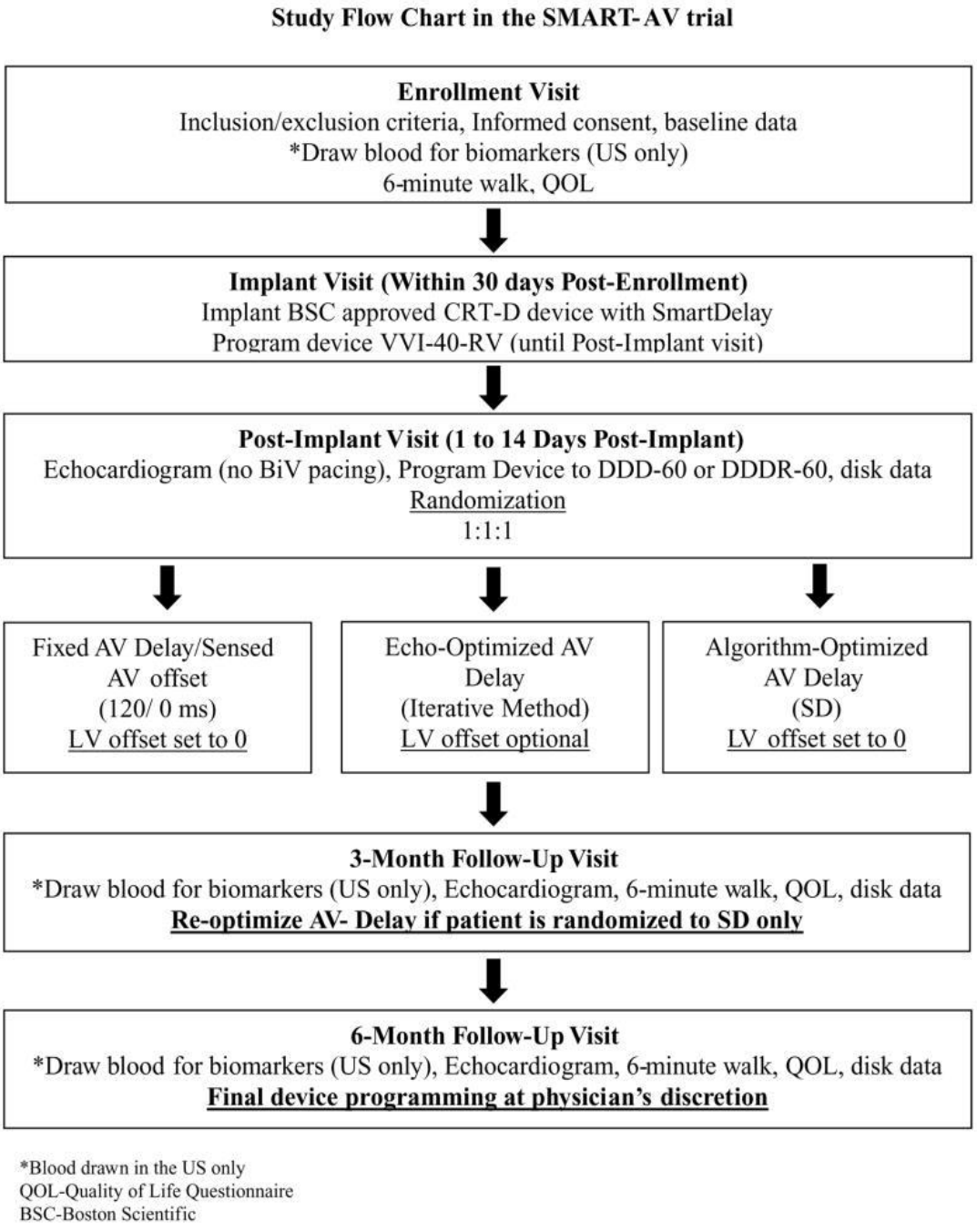


Figure 3.1 Flow Chart Describing Trial Procedure in the SMART AV trial [25]

Our choice of the machine learning techniques is in line with following the minimalist approach of employing the simplest model to satisfy the problem requirements and then extending efforts when more robust techniques are required.

3.3.1 Cluster Analysis

Cluster analysis is a class of techniques used to classify objects or cases into relative groups called clusters without any information from a supervisor. This class of ML is, therefore, one of the most prevalently used unsupervised learning techniques. Cluster analysis is a task of grouping a set of things so that objects in the same group are more like each other than to those in other groups based on some metric of similarity. In cluster analysis, there is no prior information about the group or cluster membership (therefore unsupervised learning) for any of the objects[40]. It is considered one of the more common unsupervised learning techniques.

It involves a series of steps, including formulating a problem, selecting a distance measure, selecting a clustering procedure, deciding the number of clusters, interpreting the profile clusters, and finally, assessing the validity of clustering, all of which were performed during our investigations [40].

We performed K-Means clustering[41] and Hierarchical clustering[42] on the patient data in our dataset. K-Means stores "k" centroids that it uses to define clusters. A point is considered to be in a particular cluster if it is closer (based on some measure of distance) to that cluster's centroid than any other centroid. It finds the best centroids by alternating between assigning data points to clusters based on the current centroids and choosing centroids (points that are the center of a cluster) based on the current assignment of data points to clusters [41]. **Hierarchical clustering** is an algorithm that groups similar

objects into a set of clusters, where each cluster is distinct from the other cluster, and the objects within each cluster are broadly similar to each other [42].

3.3.2 Linear Regression

Linear Regression is a machine learning approach that assumes a linear relationship between the input variables and a single output variable, more specifically, that the output can be calculated from a linear combination of the inputs. It attempts to model the relationship between parameters by fitting a linear equation to observed data. The linear equation assigns one scale factor to each input value, called coefficients. One additional coefficient is also added, giving the line an additional degree of freedom, with which to move up and down on a two-dimensional plot[10].

The least square error is used to derive a function to predict the linear relationship between variables. Suppose, for a dependent variable y and one or more independent variables X, we want to find β that gives us the minimum error following the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

From data, we have observations $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$, and the variables $X_{n\times(p+1)}$ that can be written as follows

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

Therefore, the expectation function is

$$Y' = X\ B$$

$$\text{where } B = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Now the sum of the squared error is

$$SSE = (Y - Y')^2 = (Y - XB)^2 = \sum_{n=1}^{N} \left(y_n - \sum_{p=1}^{P+1} x_{np}\beta_p \right)^2$$

SSE can be minimized by taking its derivitive and solving it for B which will lead to optimum coefficients B [11].

$$\frac{\partial SSE}{\partial \beta_p} = 0 \ => \ B = (X^T X)^{-1} X^T Y$$

Learning a linear regression model means estimating the values of the coefficients used in the representation with the data that we have available. Given the representation of a linear equation, making predictions is as simple as solving the equation for a specific set of inputs.

3.3.3 Decision Trees

Decision trees (DT) are a reliable and effective decision-making technique that provides high classification accuracy with a simple representation of gathered knowledge. They have been used in different areas of medical decision-making [7]. They learn from the provided data to produce an approximation with a set of if-then-else rules. The deeper the tree, the more complex the decision rules, and the fitter the model. DT breaks down a dataset into increasingly smaller subsets while, at the same time, an associated decision tree is incrementally developed. Decision trees use different object attributes to classify

different subsets of objects. Their great advantage is that they do not use a fixed number of predetermined features. In the decision tree approach, the members of a set of objects are classified as either positive or negative instances (in our case, patients that responded positively to CRT and patients that did not). Candidate attributes that may describe the concept are then outlined [8].

A decision tree construction tool uses outlined attributes to formulate the appropriate decision tree that identifies all positive instances of the underlying concept according to subjects with known classification. The first set of objects used for tree generation is usually called the training set. This initial decision tree becomes a basis for:

1. Forecasting whether a new subject (previously unseen) is a positive or negative instance of the concept being modeled.
2. Investigating the hierarchical representation of the most critical attributes. [9]

As the DT is incrementally developed and the dataset is broken down, class entropy is calculated at each decision to determine if maximal information has been gained from the decisions made on our subjects. The subjects, in this case, are the patients. Our primary interest is to predetermine patients' responses and to identify the contributing attributes supporting each decision. A restriction on the minimum number of patients at a decision node was implemented to prevent over-training. Parameters used for this investigation included patient attributes such as left ventricular end-systolic volume, ischemic cardiomyopathy, systolic and diastolic blood pressure, resting heart rate, weight, body mass index, age, etc.

3.3.4 Artificial Neural Networks

Our investigations utilized Artificial Neural Networks (ANNs) as another predictive machine learning approach. ANNs are brain-inspired systems, which are intended to replicate the way that humans learn. Neural networks consist of input and output layers and (in most cases) a hidden layer consisting of units that transform the input into complex abstracted information that the output layer can use. They are excellent tools for finding patterns that are far too complex or numerous for a human analyst to extract. We utilized several ANN architectures with varying complexities, including the legacy shallow ANNs. Our approach involved rapid development and experimentation libraries such as TensorFlow [43] and Keras [44].

3.3.5 Evaluation Process

All patients used in this study were diagnosed with Heart Failure with reduced ejection fraction (HFrEF) and we set out to classify CRT candidate patients into one of two classes: Responders (Positive responses) and non-responders (Negative responses). We repeated the experiments multiple times while exploring different partitioning of the training and testing population. These results were averaged to obtain a fair performance evaluation of the computational approach. For each experiment (except for cluster analysis), we separated the dataset in a 70-30 ratio, i.e., 70% for training and 30% for testing. This distribution amounted to about 534 patients for training and 230 patients for testing. The members of each set were randomly assigned each time; therefore, the class composition changed but still maintained a balanced distribution due to stratification.

For each experiment, we obtained accuracy values for each class separately, and then we used these metrics to evaluate the performance of the model. Some of these metrics involved keeping count of;

- Sensitivity – the rate at which Responders are accurately classified as such.
- Specificity – the rate at which non-Responders are accurately classified as such.
- Accuracy – the total number of correct classifications over the entire data set.

We also set out to predict the magnitude of patient response, measurable by their observed left ventricular end-systolic volume. For each LVESV prediction experiment, we evaluated model performance by obtaining mean error and error percentage values from the predicted values. We also established error margins with which to enumerate the number of patients that achieved good prediction based on our selected threshold.

3.4 RESULTS AND DISCUSSION

Our initial step in understanding the complexity of this task consisted of performing cluster analysis. This exercise aims to separate the patients based on the natural separation of the data in some hyperdimensional space and compare the similarity of patient clusters to that of the expert-based classification. The corroboration between the two approaches could be interpreted as a successful outcome. We performed categorical clustering and K-Means clustering of the patient data. Our utility of K-Means used Euclidian distance as the measure of similarity during neighbor identification and cluster assignments. Unlike K-Means, hierarchical clustering permits the use of different distance measures such as Euclidian [45], Manhattan [46] and Cosine [47]. We performed hierarchical clustering using all three distance measures, which produced accuracies of 52%, 36%, and 52%,

respectively. Based on this outcome, we used Euclidian distance for all subsequent experiments. The comparative results between the Hierarchical Clustering and K-Means Clustering are shown in Table 3.1.

With this approach, we found that patients could not be trivially separated into response classes as investigations yielded maximum classification accuracy of 54% and 52% for K-Means and Hierarchical clustering, respectively while employing Euclidian distance as the distance metric (Table 3.1). The clustering algorithms used are inadequate for classifying the patients. Therefore, more robust classification techniques were explored to address the problem better. Figure 3.3 illustrates the most successful dendrogram that was obtained from Hierarchical Clustering using the Euclidian distance metric.

Based on this information, more robust classification techniques were explored to address the problem.

Table 3.1 Table of Cluster Analysis Performance

Method	Accuracy	Sensitivity	Specificity
Hierarchical Clustering	52%	60%	43%
K-Means Clustering	54%	63%	45%

The HF dataset was used to develop Decision Tree models to classify patients based on their response classes. When unconstrained, decision trees can over-develop, resulting in over-training, where the generated decisions for classification would be overly personalized and therefore will not generalize to a broader population of patients.

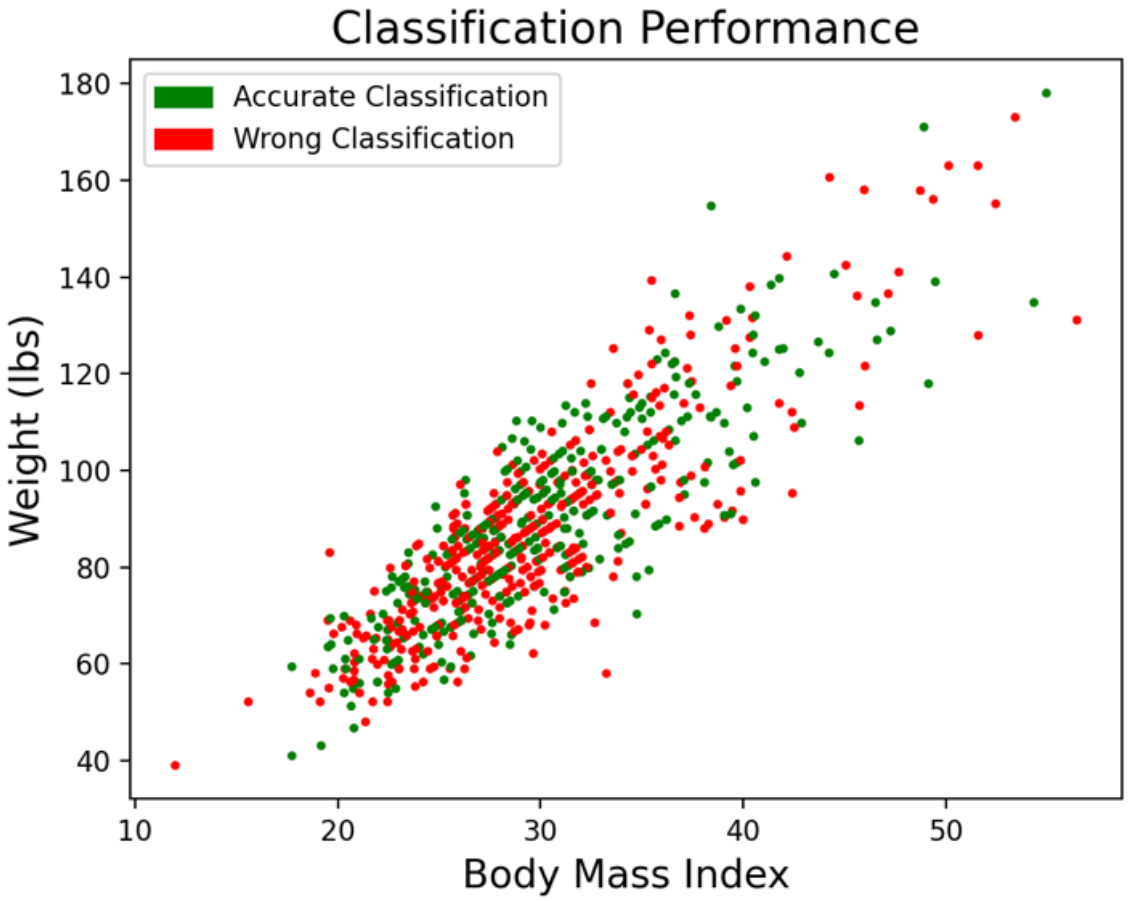


Figure 3.2 K-Means Clustering of HFrEF Patients (Green – Accurate Classification, Red - Misclassification)

To avoid that, the resultant tree was pruned by specifying the minimum number of samples a leaf node was permitted to have, restricting the maximum size of a tree. Pruning the tree structure is also a way to prevent overfitting because the generated decisions for classification would not be overly personalized and therefore do not generalize to a broader population of patients.

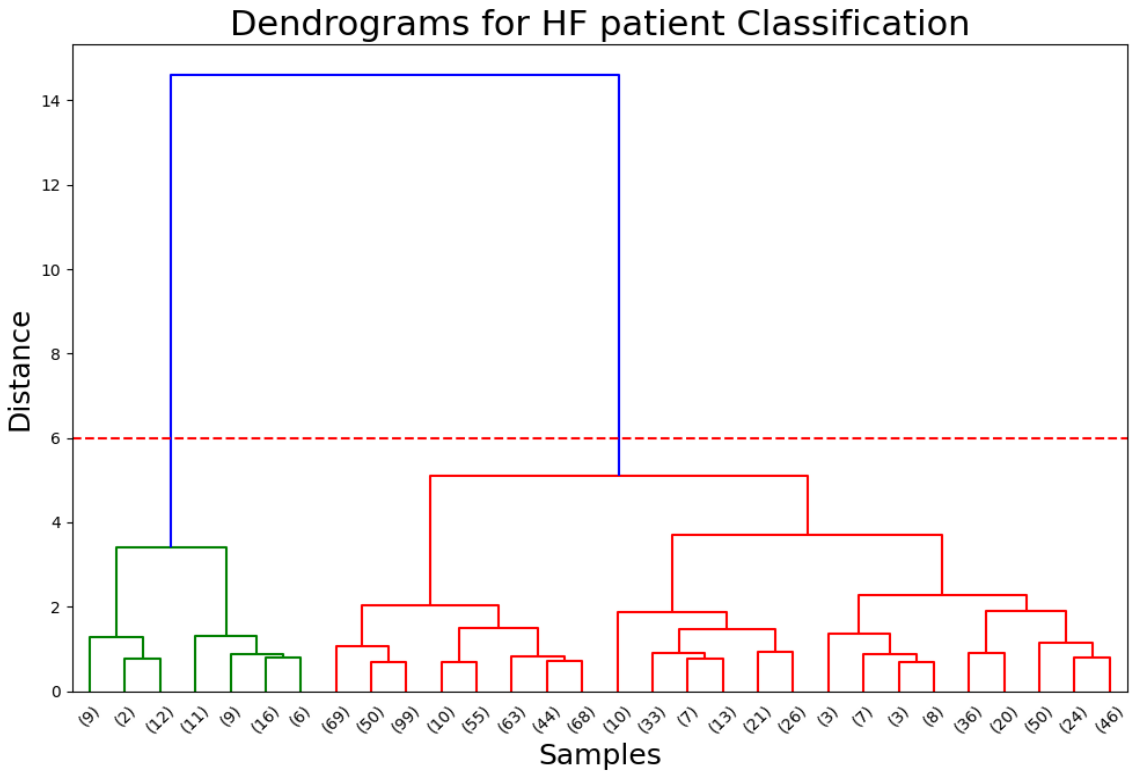


Figure 3.3 Plot for Hierarchical Clustering of HF Patients. Red – Responders, Green – Non-Responders (dendrogram has been truncated before plotting with values on the "samples" axis detailing the number of samples already combined)

Our experiments started by allowing maximum tree growth (minimum samples per leaf-node of 1) and then gradually increasing the minimum samples per leaf until the performance was no longer acceptable. The acceptable tree performance was chosen to be an accuracy of no less than 70%. Figure 3.4 illustrates the performance of the decision trees as a function of the minimum population size of the leaves. Based on these results, a pruned tree with minimum leaf-node population size of 50 was selected for further evaluation.

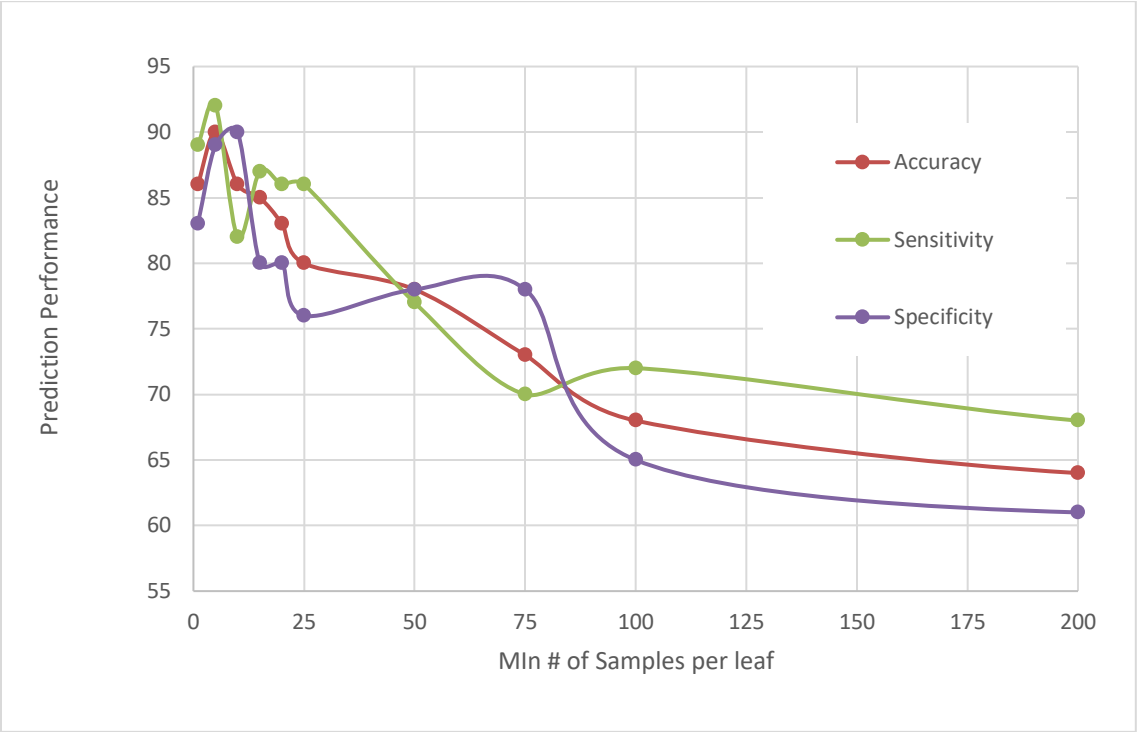


Figure 3.4 Plot of Decision Tree performance with varying Minimum # of samples per node

Using the chosen decision tree structure and restriction criteria, we evaluated the performance of the tree by randomly partitioning our data into 70% training patterns and 30% testing patterns. This process was repeated and averaged, yielding results shown in Table 3.2. The classification decision tree produced an average performance of 72% accuracy when predicting the patients' response to the CRT therapy. Average specificity (the percentage of responders who were classified as responders) was 70%, and average sensitivity (the rate of non-responders classified as non-responders) was 73%.

Table 3.2 Decision Tree Classification Performance

Exp #	Accuracy	Specificity	Sensitivity
1	71	69	73
2	72	65	78
3	71	75	67
4	73	69	75
5	72	73	71
Avg	72	70	73

Multiple neural network models were developed and incrementally modified to obtain a network architecture that produced the optimal classification accuracy. We began with a single-layer perceptron and gradually increased the number of neurons and hidden layers while observing the changes in classification accuracy. Previous research in deep learning [48] suggests that most problems can be solved with 1-2 hidden layers, in which the number of neurons in each hidden layer is likely between the number of neurons in the previous and subsequent layers.

We performed experiments using a single hidden layer starting from 5 neurons and increasing the number of neurons in steps of 5 until the number of neurons was approximately equal to the number of input neurons. With a single layer performance limited to less than 56%, a second hidden layer was added, and experimentation continued.

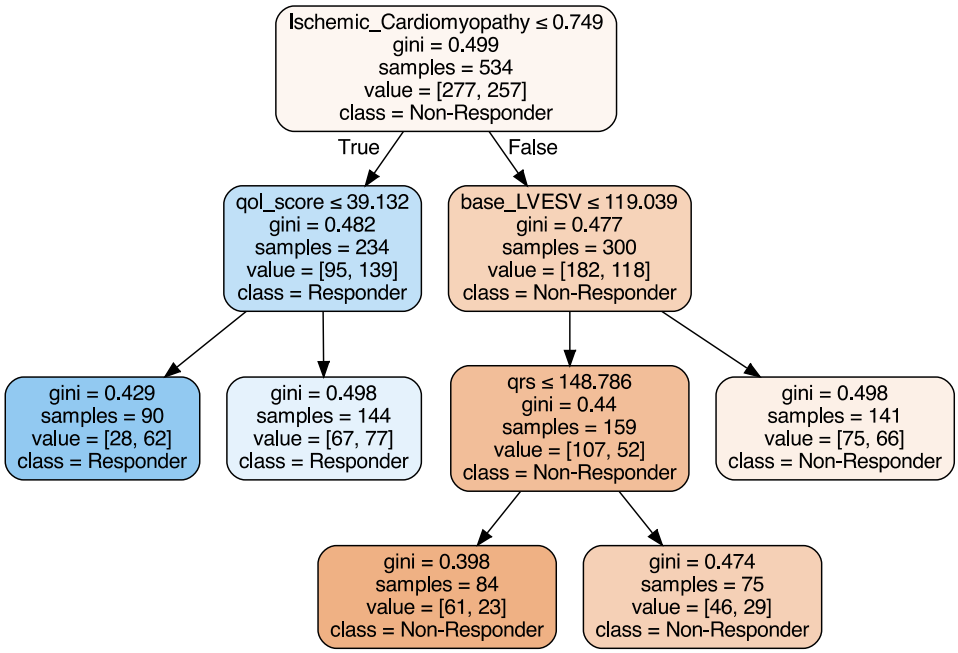


Figure 3.5 Subset of the Decision Tree for Classifying HFrEF Patients

While investigating an appropriate number of neurons for the second hidden layer, we intuitively chose 4 different values for the number of neurons in the first hidden layer (10, 20, 30, and 40). Our exploration of the size of the second hidden layer consisted of starting with two neurons and increasing in steps of 2 until no more improvement was observed in the performance. The impact of growing the second hidden layer on performance was observed to obtain the most efficient network architecture (i.e., the smallest network that produced acceptable results). Based on the results from those experiments, the optimal neural network architecture was; 53 -> 20 -> 14 -> 1 (53 input neurons, 20 neurons in the first hidden layer, 14 neurons in the second hidden layer, and one neuron in the output layer).

ANN experiments were repeated using this network architecture, sharing the dataset into 70% for training and 30% for testing. The classification experiment yielded results showing that 70% of the participating patients were accurately predicted into their response classes. Specificity, and sensitivity values were 68% and 72% (see Table 3.3).

Table 3.3 HFrEF Patient Classification with ANNs

Exp#	Accuracy	Specificity	Sensitivity
1	72	71	73
2	70	69	70
3	68	66	70
4	69	69	69
5	72	68	74
Avg	70	68	72

In general, the Neural network classification had about the same performance as that obtained using the decision tree.

Heart failure with reduced ejection fraction occurs when there is a deficiency in the contraction of the heart muscle during blood pumping. Poor contraction is identified when the left ventricular end-systolic volume (LVESV) is lower than expected. To that end, we experimented with linear regression and ANNs to predict the LVESV a patient will observe, 3 months after CRT implantation.

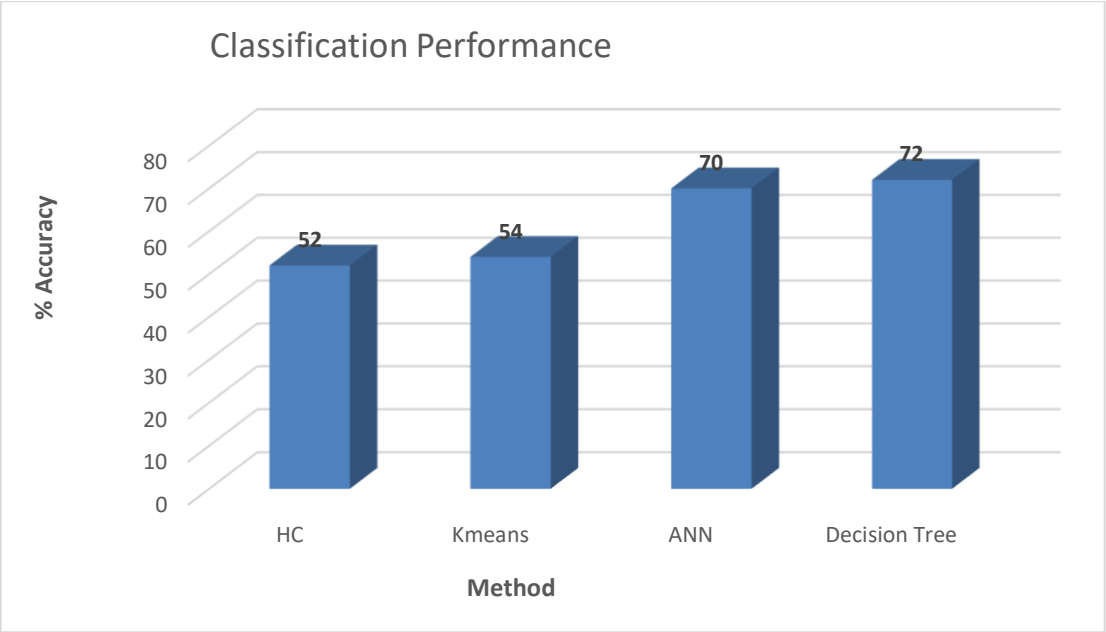


Figure 3.6 Comparison of Classification Performance Between Approaches Employed

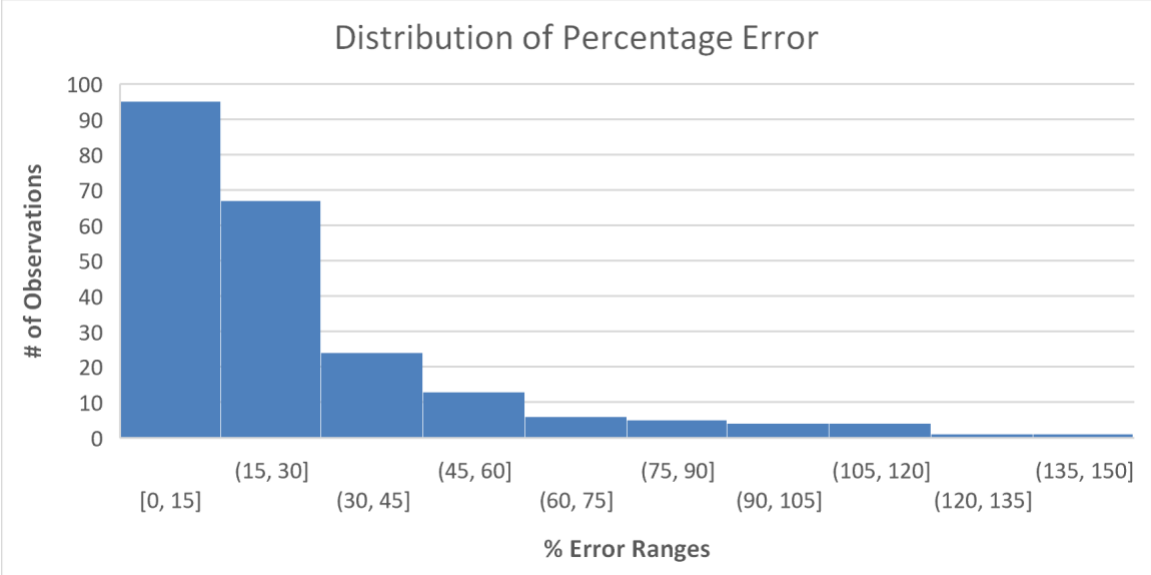


Figure 3.7 Distribution of Percentage Errors Observed using Linear Regression

First, we employed a linear regression model for the prediction task which yielded results where the mean and median absolute error values were 24.3 and 19.6 respectively. That amounted to a 22% average margin of error, i.e., the average difference between the actual and predicted values LVESV values was +/-22% of the actual LVESV value. The distribution of percentage error values can be found in Figure 3.7.

Following the linear regression experiments, we investigated the application of ANNs for the prediction task. ANNs predicted the LVESV with mean and median absolute error values of 23.6 and 17.5 amounting to an average percentage error of 21%. i.e., the average difference between the actual and predicted values LVESV values was +/-21% of the actual LVESV value. The distribution of absolute error values can be found in Figure 3.9.

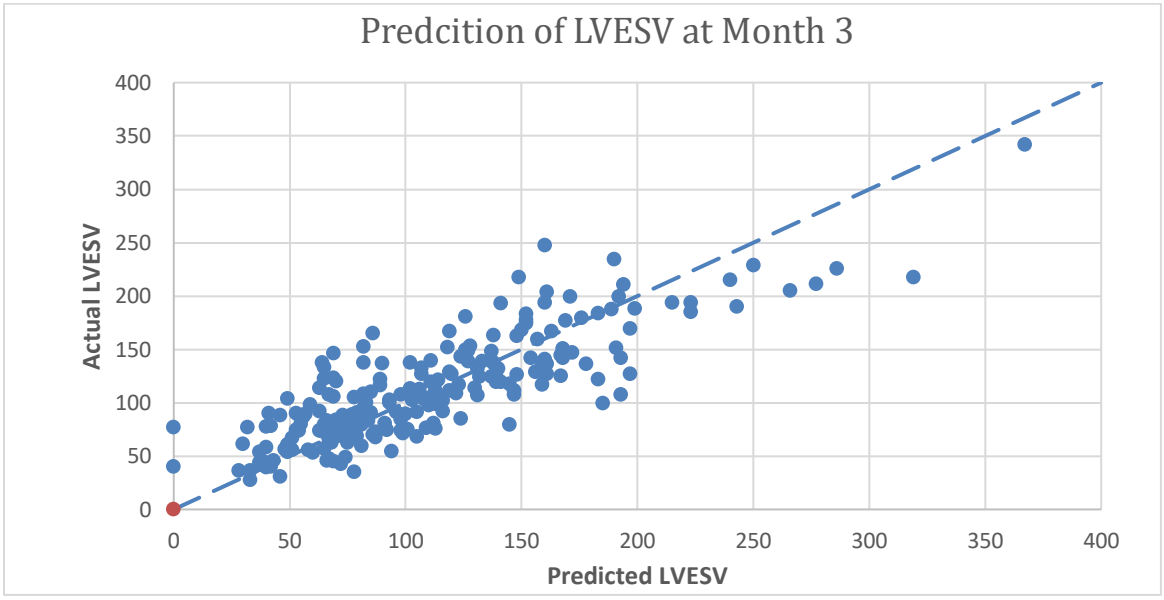


Figure 3.8 Regression Plot showing Actual vs Predicted LVESV values at month 3 after CRT using Linear Regression

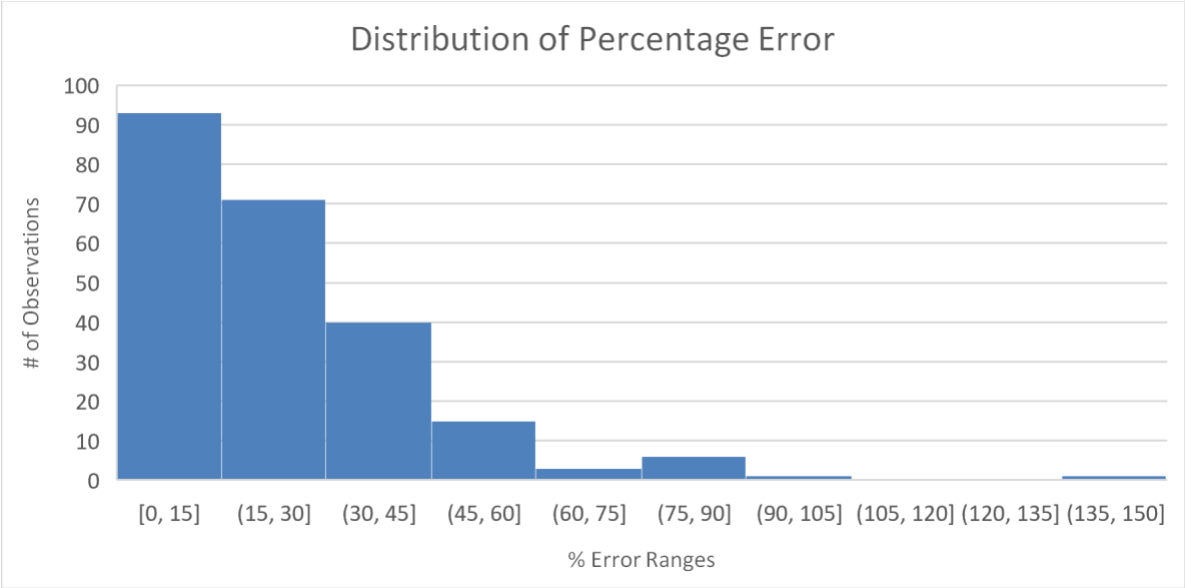


Figure 3.9 Distribution of Percentage Errors Observed Using ANNs

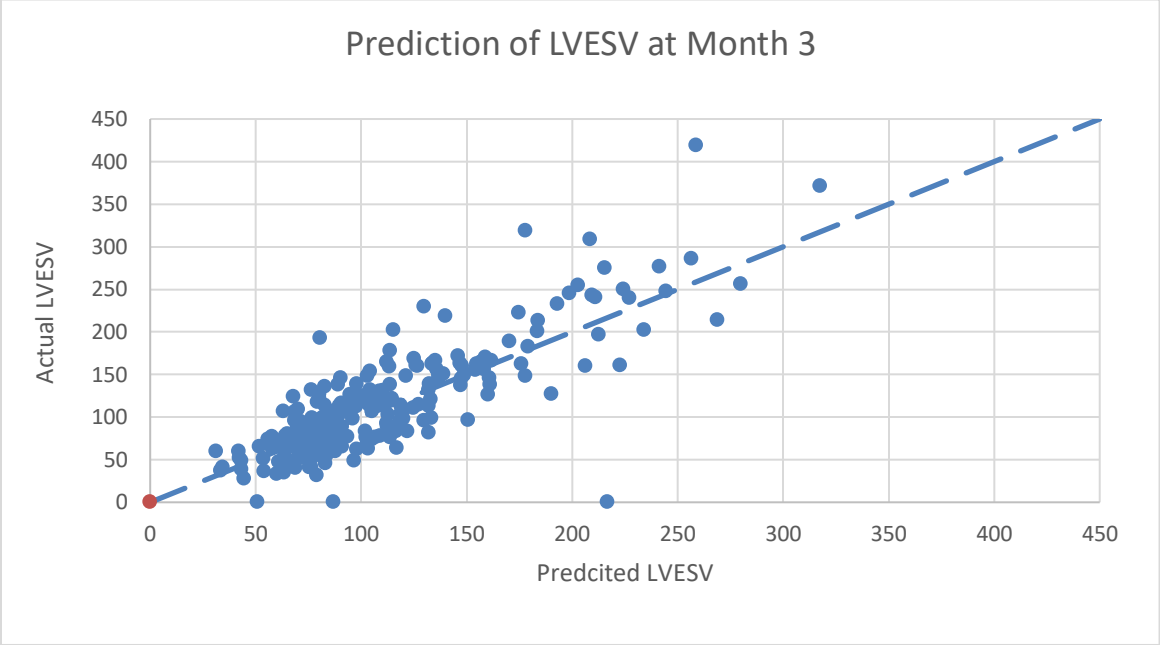


Figure 3.10 Regression Plot Showing Actual vs Predicted LVESV at Month 3 using ANNs

3.5 CONCLUSION

The advancements made in the field of machine learning make ML techniques suitable for use in the development of tools that aid in the medical decision-making process even

though, for various justified reasons, their impact and integration in the field of medical diagnostics have been minimal.

Machine learning models can recognize patterns and complex relationships between parameters that are not easily discernible by man. The parameter values used for this experiment were obtained from patients undergoing CRT.

We have demonstrated that we can (with 72% accuracy) identify HF patients likely to respond positively to CRT with the employment of artificial neural networks. We have also demonstrated that a classification accuracy value of up to 72% is achievable following the sets of rules/questions produced by the decision tree. Regardless of how the output classes are categorized, training ML models to accurately distinguish between different classes of patients is valuable to the medical community.

In addition to the classification of patients based on response outcomes, we have demonstrated that we are able to predict the LVESV at month 3 after implantation with an median percentage error of 17.5%. Our investigations demonstrate that ML models are capable of exploiting relationships with and between HF patient parameters.

This approach and others like it will positively influence medical decision-making, and administration of intervention procedures, and further the practice of precision medicine. The decision tree approach and the rules generated produce a means of prioritizing patient data parameters and present us with the opportunity to extend our

investigation efforts into the identification of the parameters most responsible for patient outcomes.

CHAPTER 4 APPLICATION OF MACHINE LEARNING IN MODELLING THE PROGRESSION OF CHRONIC KIDNEY DISEASE

4.1 BACKGROUND OF CHRONIC KIDNEY DISEASE

Chronic kidney disease (CKD) is a disease where a person's kidneys are damaged and cannot filter blood the way they should [49]. Chronic kidney disease (CKD) is an important medical condition that is prevalent in ~14% of Americans $\geq$ 30 years of age and is predicted to be ~17% by 2030 [50]. Total cost of care for patients with CKD in the United States is approximately 50% higher than other high-cost conditions (stroke or cancer[lung, breast, or colon]) and is more than 10 times higher than individuals without any of these conditions [51]. Much of this is due to two associated significant health consequences: cardiovascular events and the development of end stage kidney disease (ESKD) requiring renal replacement therapy (RRT) (e.g. hemodialysis). These consequences are associated with progressive morbidity (hospitalizations, long term care use) as well as increased risk for mortality (Mathew). Slowing the rate of progression of CKD is an important means of mitigating these risks [52].

The challenge for providers caring for patients with CKD is predicting the exact course of disease in individual patients, as there is marked variability among the rates of progression of CKD among patients with similar stages of CKD (reference). Several risk prediction markers exist to predict the development of ESKD that utilize a surprisingly small number of clinical variables (estimated Glomerular Filtration Rate (eGFR) and Albuminuria). Subsequent analyses have attempted to utilize a larger scope of clinical variables including research assessments that are not readily available in current routine clinical practice. The findings from these analyses suggest a greater fit to outcomes of interest in the CKD population with advanced prediction algorithms using machine learning (ML) with a larger scope of clinical data. However, the use of research assays among the predictor variables limits their current generalizability. Ferguson and colleagues have published the first report of utilizing ML with administrative, readily available clinical data to categorize patients at risk for development of ESKD. This report demonstrates a significant advancement in the science of predicting adverse renal outcomes. Evaluations in multinational cohorts as well as varying ML models and outcomes are needed to maximize the efficiency and generalizability of ML algorithms for CKD progression prediction. Herein we report on the use of feed-forward neural network ML algorithm for prediction of CKD rate of progression utilizing administrative health data from the Veterans Health Administration (VHA).

4.2 METHODS

The Columbia VA Health Care System Institutional Review Board (IRB) and subsequently the Loma Linda VA Health Care System IRB provided local regulatory oversight for this project and approved this project under expedited review. All data were acquired from the VHA Corporate Data Warehouse (CDW) through the Veterans Integrated Networking and Computing Infrastructure. Research-ready data was made available in a Microsoft SQL Server Studio. All data analysis was performed using R statistical software.

4.2.1 Clinical Data

All patients were included who had an eGFR measurement between 2000 and 2020 between 60 and 15 ml/min/1.73m^2 (N = 3,050,675). The population was restricted to those with an available creatinine each year for the subsequent 3 years (N = 662,675). After excluding patients without any urine protein determination (microalbumin, urinalysis with dipstick evaluation for protein, or urine protein) (N = 478,297), and implausible eGFR values (> 120 ml/min/1.73m^2) (N = 1324), a final cohort size of 183,054 was available for analysis (Figure 4.1).

Input variables were chosen as to the general availability during routine clinical follow-up. See table 4.1.

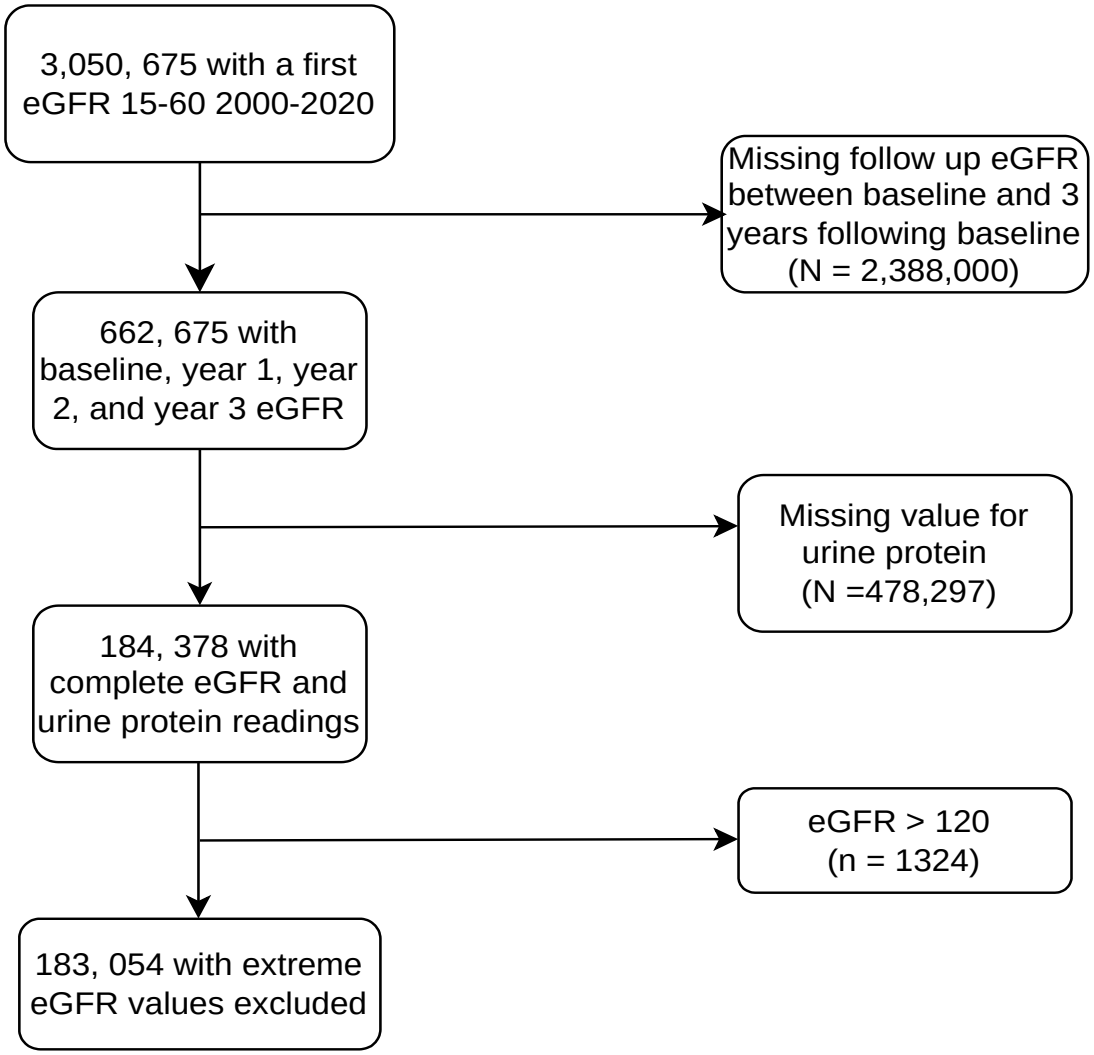


Figure 4.1 Workflow diagram of data filtration steps to aggregate the final usable data set from Veteran Affairs patients from 2018-2020.

Dependent Variables: Serum creatinine and estimated glomerular filtration rate (eGFR) were extracted at baseline and annually for 3 years following the index value. eGFR was not calculated rather extracted from the electronic record; standardized recommendations for estimating equations for eGFR were in place for VA facilities.

Table 4.1 Table of Parameter categories utilized during the study

Variables Utilized	
Demographics	Age, gender, race, ethnicity
Medical History	Hypertension, diabetes, peripheral arterial disease, congestive heart failure, coronary artery disease, stroke, cancer
Vital signs	Blood pressure, heart rate, body mass index
Medications	Diabetes medications, antihypertensives, proton pump inhibitors
Laboratory parameters	comprehensive metabolic panel, complete blood count, urinalysis, parathyroid hormone, urine protein and microalbumin corrected for creatinine.

4.2.2 Analysis

Supplemental table 1 demonstrates distribution of variables utilized for this current analysis. Missingness of variables ranged from <1% to 98%.

Patients were evaluated by baseline CKD stage (staging was determined according to Kidney Disease: Improving Global Outcomes recommendations: CKD G3a, eGFR 45-59 ml/min/1.73m^2; CKD G3b eGFR 30-44 ml/min/1.73m^2; CKD G4 eGFR 15-29 ml/min/1.73m^2. Descriptive statistics were performed across CKD stages.

When there is a discernible relationship between two or more variables in the dataset, those features/variables are said to have covariance. The higher the covariance, the more substantial the correlation between these variables and that implies redundancy in the dataset since there is more information than is actually needed. Correlation coefficient analysis was performed on the dataset used for these experiments. The correlation coefficient analysis results are shown in the heat map (shown in Figure 4.2). As shown in this figure, most of the input parameters of this problem show correlation values within the mid-range of -0.5 to 0.5, indicating that there is no substantial correlation between any two parameters in the dataset to assist with dimensional reduction. The variables with higher correlation, e.g yr1_gfr_delta_rate was excluded due to their correlation with the actual output (yr3_gfr_delta_rate).

In this work, the missing data points were substituted with a constant value, -1, which is out of the range of the parameter values within the input space. This is typically suitable for neural networks and decision trees as the model will make splits for these "out of range" values and minimize its progression towards the results. The dataset also contained parameters with substantially different ranges. For instance, the age of the patient is measured in days (range was in thousands), while other parameters such as hemoglobin A1c (HbA1c) units have single digit values. This disparity in the range of data was resolved by normalization of data prior to the training, validation, and testing processes.

The Normalization approach employed was the use of a min-max scaler which sets all values in the dataset to within a set range (0-1).

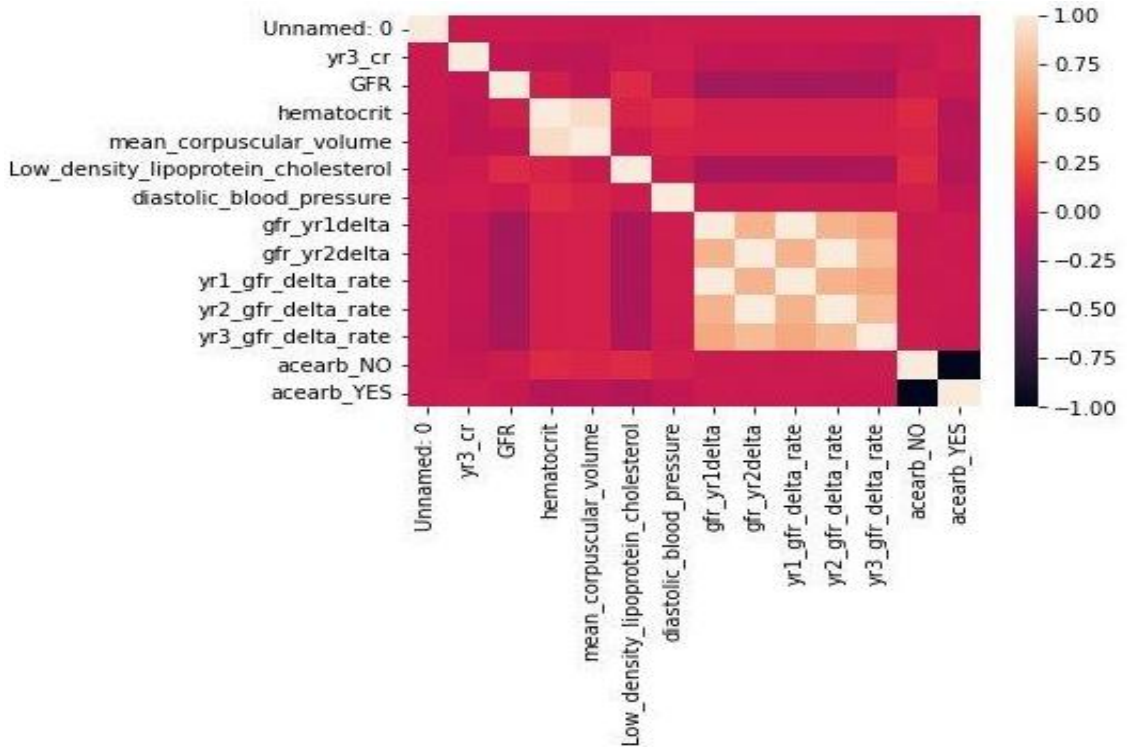


Figure 4.2 Correlation Coefficient Heatmap for the most highly correlated variables in the dataset

Our investigations utilized Artificial Neural Networks (ANNs) as a predictive machine learning approach. ANNs are brain-inspired systems, which are intended to replicate the way that humans learn. Neural networks consist of input and output layers and (in most cases) one or more hidden layers consisting of units that transform the input into complex abstracted information that the output layer can use. They are excellent tools for finding patterns that are far too complex or numerous for a human analyst to extract. They provide several pragmatic advantages over the conventional statistical approaches such as relinquishing the need for statistically well-behaved data and accommodation of high-dimensional data. We utilized several ANN architectures with varying complexities,

including the legacy shallow ANNs. Our approach involved rapid development and experimentation in python employing libraries such as TensorFlow and Keras.

First, we set out to classify CKD candidate patients into one of two classes: Patients experiencing rapid decline (Positive Reponses) and patients not experiencing rapid decline (Negative Responses). We developed a fully connected, feed-forward artificial neural network with 2 hidden layers in python. Due to the binary nature of the expected output, we used the sigmoid activation function, and chose a threshold of 0.5. Patients whose eGFR experienced a decline of greater than or equal to 50% we categorized as experiencing rapid decline. Patients whose GFR exhibited a decline of less than 50% were categorized as not experiencing rapid decline.

As a next step in understanding the varying manifestations of the kidney function progression caused by CKD, we experimented with the classification of patients into one of 5 different categories based on their risk levels, Improved, Stable, Mild, Rapid Decline and Very rapid Decline. In a similar way as the previous experiment, we developed a fully connected, feed-forward artificial neural network with 2 hidden layers for this classification task. Because there are 5 categories in the output, to which a patient can be classified into, our network was configured to output 5 values, one for each category of ES-CKD risk, and the SoftMax function was used to normalize the outputs, converting them from weighted sum values into probabilities that sum to one. Each value in the output of the SoftMax function is interpreted as the probability of membership for each category.

Patients who experience absolute change in their eGFR with values greater than 0.6ml/min/1.73m^2/yr were categorized as **Improved**, those between – 0.9 and 0.6 were

categorized as **Stable**, those between -2.5 and – 0.9 were categorized as **Mild**, those between -5 and -2.5 were categorized as experiencing **Rapid Decline**, those with absolute change of less than -5 were categorized as experiencing **Very Rapid Decline**.

Finally, we experimented with predicting the actual expected percentage change in eGFR between first measurement at baseline and the eGFR, 3 years into observation. Again, we developed a fully connected, feed- forward artificial neural network with 2 hidden layers. This time, we utilized a linear activation function in the output layer of the network because we are attempting to predict the change in value and the change can be positive, negative or 0.

4.3 RESULTS

Baseline parameters at index renal function assessment are displayed in Supplemental Table 4.5. Most variables demonstrated a significant difference across CKD groups. Characteristics that were not different across CKD stages included history of diabetes, history of stroke (no participants had evidence of history of stroke), use of glp1ra, total cholesterol levels, and BMI. Overall, the population was older with average age around 67 years. Consistent with the VA population, the majority of individuals were male (>96%) and Caucasian race (>70%).

Notable differences across CKD groups included greater proportion of Black Americans among those with CKD G4 (12.9%, 14.6%, 23.1% for CKD G3a, G3b and G4, respectively, $P < 0.001$ for trend); higher systolic blood pressure (132, 133, 135 mmHg for CKD G3a, G3b, and G4, respectively; $P < 0.001$ for trend); and greater albumin: creatinine

ratio in a spot urine specimen (74.3, 237.5, 274.8 mg/gm for CKD G3a, G3b, and G4, respectively; $P < 0.001$ for trend).

4.3.1 Progression of Kidney Disease

The majority of individuals (>60% across CKD groups) demonstrated stable or improved eGFR over the course of 3 years. Median [interquartile range – IQR] rate of eGFR change per year at year 3 for CKD G3a, G3b, and G4 were 1.00 [-1.87, 3.97] ml/min/1.73m^2/year, 1.05 [-1.92, 5.86] ml/min/1.73m^2/year, and 1.69 [-2.28, 11.65] ml/min/1.73m^2/year, respectively ($P<0.001$ for trend) (Table 1). In addition, the CKD group with the greatest proportion of individuals demonstrating > 50% reduction in eGFR (a clinically significant reduction in kidney function) at year 3 following index value was CKD G4 with 14.8% (2.8% in CKD G3a, and 5.8% in CKD G3b; $P < 0.001$ for trend).

4.3.2 Predicting the Rate of Change of Kidney Function over 3 years Using Artificial Neural Networks

The parameters utilized for the prediction are displayed in supplemental table 4.2. A total of 42 distinct variables were utilized for the prediction. The result of the feed forward neural network for predicting the 3-year rate of kidney function change is displayed in figure 1. The approach provided an overall average absolute error of 12.5% and median absolute error of 7.9%. Percentage of predictions with <20% error was 97%.

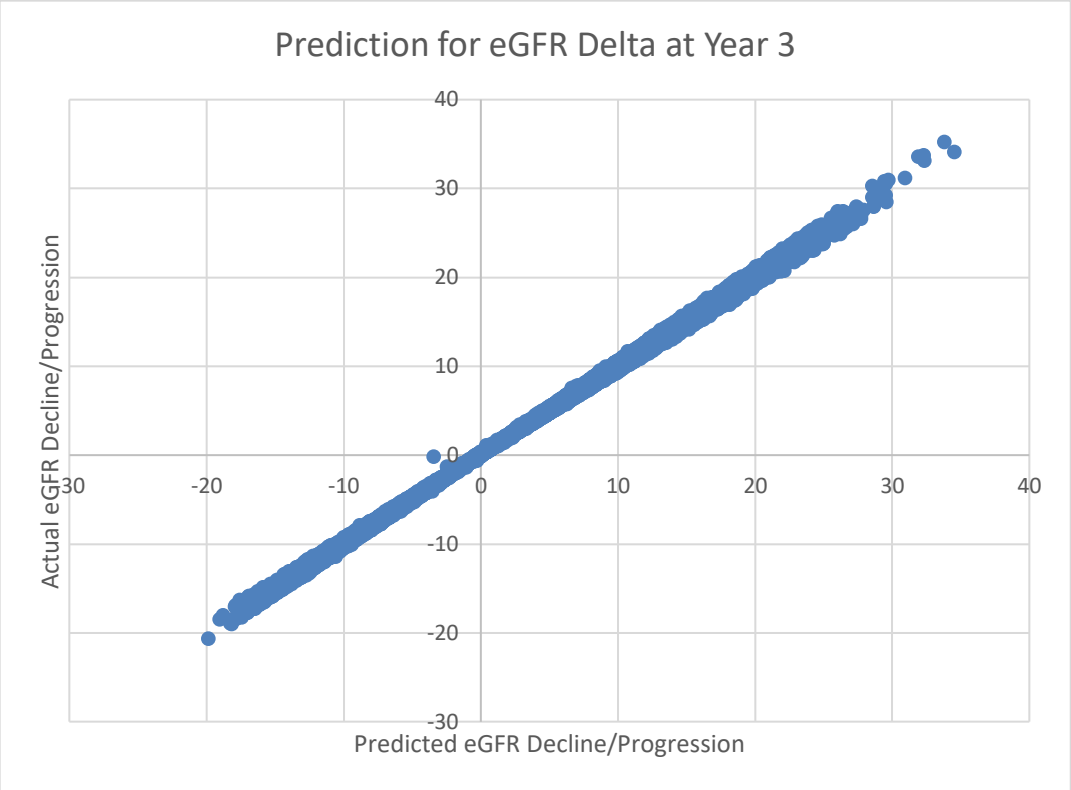


Figure 4.3 Prediction of Actual eGFR Rate of Change vs Predicted Rate of Change

4.3.3 Predicting Categories of Change in Renal Function

To assess the ability to predict categories of change, renal function decline was categorized in 2 separate manners.

First was a multiclass designation here patients who experience absolute change in their eGFR with values greater than 0.6ml/min/1.73m^2/yr were categorized as **Improved**, those between – 0.9 and 0.6 were categorized as **Stable**, those between -2.5 and – 0.9 were categorized as **Mild**, those between -5 and -2.5 were categorized as experiencing **Rapid Decline**, those with absolute change of less than -5 were categorized

as experiencing **Very Rapid Decline**. The second categorization was binary looking at ≥ 50% or <50% decline in 3 years.

Figure 4.4 demonstrates the confusion matrix (comparison of accuracy across categories). As can be seen, the diagonal represents an exact match, and the majority of matches fall within the diagonal. The majority of misclassifications were essentially 1 category difference.

Table 4.2 demonstrates the confusion matrix for the binary classification and similarly the majority of predictions fell along the diagonal. The precision of the algorithm (analogous to the positive predictive value) was 90% for having a 50% or greater decline. Similarly, the recall (analogous to the sensitivity) was 99.6%.

Table 4.2 Confusion Matrix for CKD Binary Classification

	Predicted		
Actual		Not Rapid Decline	Rapid Decline
	Not Rapid Decline	52718	132
	Rapid Decline	209	1858

4.4 DISCUSSION

From a large, nationally representative sample of health data, and utilizing readily available laboratory, pharmacologic and clinical variables, we have demonstrated the ability to achieve >90% precision with predicting the rate of progression of kidney disease among patients with existing CKD.

Table 4.3 Classification Summary for CKD Binary Classification

	Precision	Recall	F1- Score	N
Not Rapid Decline	1.00	1.00	1.00	52850
Rapid Decline	0.93	0.90	0.92	2067

Table 4.4 Confusion Matrix for Multi-Category Classification of CKD Progression

	Predicted Categories					
True Categories		Improved	Stable	Mild Decline	Rapid Decline	Very Rapid Decline
	Improved	1.9e+04	1.5e+02	9	0	0
	Stable	65	4.7e+03	56	0	0
	Mild Decline	0	85	4.2e+03	1e+02	0
	Rapid Decline	0	0	1e+02	4.1e+03	2e+02
	Very Rapid Decline	0	0	0	57	3.5e+03

In addition, we have demonstrated >90% precision in predicting a decline of 50% in eGFR, a clinically important decline in kidney function. These findings are highly relevant

in identifying high-risk individuals for risk reduction measures and possibly for evaluation by novel clinical trials.

Prior to the current era of available large EHR data, and software for rapid machine learning algorithm implementation, models for CKD progression were limited in size and scope. The most widely used algorithm for predicting risk of progression uses 2 variables - the current eGFR and amount of albumin in the urine. (KDIGO CKD heat map) Despite the importance of these variables, the variability in clinical course within each of the categories remains high. Tangri et al enhanced this in a model that included 8 total variables, including other readily available clinical parameters (age, gender, serum calcium, serum albumin, serum phosphorus, and serum bicarbonate) in the Kidney Failure Risk model. [53]. However, widespread adoption of this model has not occurred, and availability of clinical data and computing capability has advanced enough to allow extension of these models.

The only other current report of readily available clinical variables being utilized in machine learning algorithms for risk prediction was provided recently by Ferguson et al [54]. Ferguson et al utilized a random forest method to categorize patients into high-risk (>40% decline in estimated glomerular filtration rate – eGFR, the measure of kidney function; or development of ESKD requiring RRT) or not. The model performed very well with area under the receiver operating curve (AUC) of 0.92 for 1-year outcomes and 0.82 for 5-year outcomes, and similarly Brier scores of 0.02 and 0.07 (closer to 0 suggesting perfect prediction, and the opposite when closer to 1) for the same time points. Our data support and supplement these findings. We have supported their findings by demonstrating

the ability to accomplish such predictions with readily available clinical variables from electronic health record data. Furthermore, we replicated the findings in a national US sample with more representation of Black individuals and some from Pacific Islander populations. In addition, given the availability of such data in the VA EHR, we were able to incorporate pharmacologic and vital sign data, making our report a more clinically useful advancement.

The current data also highlight the high variability in progression of CKD among patients with similar starting eGFR. The majority of patients in the CKD categories studied demonstrated either a stable or improving kidney function over the 3 year follow up period. Less than 20% of the population had a clinically significant 50% decrease in kidney function. This is consistent with prior analyses that demonstrated slow rate of progression across CKD stages (references), further highlighting the need for enhanced prognostication tools that can precisely identify individual risk.

4.5 LIMITATIONS

This is a US Veteran population that is predominantly white (with a considerable Black American constituency), male and older, thus limiting generalizability to populations not fitting these characteristics. The breadth of clinical variables utilized for these algorithms require health records with a comprehensive array of data points per individual. The VA is unique in its large geographic span, and the immensity of the type and duration of data available on a large number of individuals. Nevertheless, such comprehensive health records are becoming more commonplace in smaller systems, thus algorithms developed using the VA system will have applicability. It is possible that some

of the kidney function data points captured were acute changes that do not truly reflect chronic states. The possibility of this was limited by focusing on outpatient creatinine values; also, the number of patients sampled, and the duration of time (3 years) will lessen the likelihood that acute values will predominate in the data sampling. Longer term follow up and identification of need for renal replacement therapy was not available for this analysis.

4.6 CONCLUSION

In conclusion, we have demonstrated the ability to provide highly accurate predictions for the rate of CKD progression among a large nationally representative cohort of individuals from the US with underlying CKD stages 3 and 4 using readily available clinical variables. These predictions were created using machine learning techniques in Python. Such algorithms have the potential for easy incorporation into EHR and decision support systems. Future studies should explore the utility of such machine learning supported decision support systems to improve CKD care.

Table 4.5 Supplemental Table with Baseline Parameters at Index Renal Function Assessment.

	CKDG3a	CKDG3b	CKDG4	p
N	147823	27758	7473	
Demographics				
Age (years) (mean (SD))	67.97 (10.05)	69.16 (10.77)	66.07 (11.16)	<0.001
Gender = M (N(%))	142145 (96.2)	26768 (96.4)	7217 (96.6)	0.023
Race N (%)				<0.001
American Indian or Alaska Native	979 (0.7)	198 (0.7)	60 (0.8)	
Asian	704 (0.5)	154 (0.6)	37 (0.5)	
Black American	19133 (12.9)	4055 (14.6)	1728 (23.1)	
Native Hawaiian or Other Pacific Islander	1122 (0.8)	190 (0.7)	46 (0.6)	
Unknown	11095 (7.5)	2351 (8.5)	521 (7.0)	
White	114790 (77.7)	20810 (75.0)	5081 (68.0)	
Comorbidities (N (%))				
Peripheral arterial disease	751 (0.5)	191 (0.7)	53 (0.7)	<0.001
Heart Failure	15288 (10.3)	4269 (15.4)	1409 (18.9)	<0.001
Coronary Artery Disease	30929 (20.9)	6279 (22.6)	1668 (22.3)	<0.001
Diabetes	18136 (12.3)	3720 (13.4)	1201 (16.1)	<0.001
Hypertension	121388 (82.1)	22899 (82.5)	6160 (82.4)	0.273
Stroke	147823 (100.0)	27758 (100.0)	7473 (100.0)	NA

Cancer	4877 (3.3)	739 (2.7)	202 (2.7)	<0.001
Medications (N (%))				
Insulin	18525 (12.5)	4086 (14.7)	1219 (16.3)	<0.001
Metformin	33711 (22.8)	4020 (14.5)	875 (11.7)	<0.001
Sulfonylurea	24889 (16.8)	5074 (18.3)	1080 (14.5)	<0.001
Thiazolidinediones	4569 (3.1)	1222 (4.4)	254 (3.4)	<0.001
Dipeptidyl peptidase inhibitors	793 (0.5)	106 (0.4)	20 (0.3)	<0.001
Sodium-Glucose co-Transporter Type 2 Inhibitors	136 (0.1)	13 (0.0)	0 (0.0)	0.002
Glucagon-like Peptide receptor agonists	193 (0.1)	28 (0.1)	5 (0.1)	0.158
Any Diabetic Medication	54641 (37.0)	10223 (36.8)	2526 (33.8)	<0.001
Cardiovascular				
Angiotensin Converting Enzyme Inhibitors (ACEI)	47338 (32.0)	9310 (33.5)	2299 (30.8)	<0.001
Angiotensin Receptor Type II Blockers (ARB)	12136 (8.2)	2439 (8.8)	713 (9.5)	<0.001
ACEI or ARB	58579 (39.6)	11503 (41.4)	2917 (39.0)	<0.001
Statins	63207 (42.8)	11449 (41.2)	2893 (38.7)	<0.001
Other				
Proton pump inhibitors	30629 (20.7)	5688 (20.5)	1652 (22.1)	0.008

Laboratory Data				
Sodium (meq/L: mean (SD))	138.75 (3.53)	138.57 (3.70)	138.31 (3.97)	<0.001
Potassium (meq/L: mean (SD))	4.34 (0.50)	4.41 (0.54)	4.44 (0.59)	<0.001
Chloride (meq/L: mean (SD))	103.27 (3.88)	103.74 (4.32)	104.48 (4.89)	<0.001
Bicarbonate (meq/L: mean (SD))	26.79 (2.94)	25.97 (3.18)	24.44 (3.52)	<0.001
Blood urea nitrogen (mg/dl: mean (SD))	21.20 (6.78)	28.34 (10.04)	40.06 (16.60)	<0.001
Calcium (mg/dl: mean (SD))	9.33 (0.55)	9.24 (0.58)	9.01 (0.67)	<0.001
Phosphorus (mg/dl: mean (SD))	3.42 (1.84)	3.51 (0.74)	3.91 (1.01)	<0.001
Magnesium (mg/dl: mean (SD))	2.00 (0.28)	2.01 (0.31)	2.02 (0.34)	<0.001
Albumin (g/dl: mean (SD))	3.98 (0.84)	3.98 (11.21)	3.67 (1.35)	<0.001
Hemoglobin (g/dl: mean (SD))	13.81 (2.34)	13.02 (2.46)	12.02 (2.34)	<0.001

Parathyroid hormone (pg/ml: mean (SD))	67.22 (75.10)	86.37 (77.10)	161.51 (169.25)	<0.001
Hemoglobin A1c (%: mean (SD))	7.02 (1.61)	7.17 (1.74)	7.17 (1.86)	<0.001
Total cholesterol (mg/dl: mean (SD))	166.70 (43.08)	166.75 (46.87)	165.25 (51.54)	0.131
Low density lipoprotein cholesterol (mg/dl: mean (SD))	94.08 (34.62)	92.74 (36.11)	91.01 (37.02)	<0.001
High density lipoprotein cholesterol (mg/dl: mean (SD))	40.99 (12.19)	39.37 (12.33)	38.04 (12.54)	<0.001
Urine microalbumin:creatinine ratio (mg/g: mean (SD))	74.26 (359.59)	237.53 (10301.18)	274.82 (856.36)	<0.001
Any Albuminuria noted (including dipstick+) (N (%))	23499 (17.3)	6189 (24.4)	2190 (31.1)	<0.001
Vital Signs				
Systolic Blood Pressure (mmHg: mean (SD))	132.17 (16.38)	132.94 (18.09)	134.68 (19.11)	<0.001
Diastolic Blood Pressure (mmHg: mean (SD))	73.99 (10.52)	72.48 (11.12)	73.53 (11.41)	<0.001

Heart Rate (beats per minute: mean (SD))	74.04 (20.61)	74.27 (13.39)	75.99 (12.88)	<0.001
Weight (lbs: mean (SD))	213.06 (45.52)	211.20 (46.64)	208.52 (48.90)	<0.001
Body Mass Index (kg/m^2: mean (SD))	31.07 (6.10)	30.95 (6.28)	31.35 (72.36)	0.172
Renal Function Changes (eGFR in ml/min/1.73m2)				
Follow-up time from Yr0 to Yr3 eGFR (median [IQR])	1087.06 [1059.02, 1114.87]	1081.91 [1055.83, 1111.01]	1073.01 [1049.80, 1104.92]	<0.001
Year 0 eGFR (median [IQR])	55.70 [52.00, 58.00]	40.20 [36.30, 43.00]	24.40 [20.00, 28.00]	<0.001
Year 1 eGFR (median [IQR])	59.00 [51.60, 67.50]	44.50 [36.00, 58.00]	31.00 [21.00, 59.70]	<0.001
Year 2 eGFR (median [IQR])	58.70 [50.00, 67.20]	44.00 [35.00, 58.00]	31.00 [19.00, 59.80]	<0.001
Year 3 eGFR (median [IQR])	58.00 [48.50, 67.00]	44.00 [33.90, 57.60]	30.22 [17.00, 58.00]	<0.001
Renal Function outcomes				
Change from Year 0 to Year 3 eGFR (median [IQR])	3.00 [-5.60, 12.00]	3.00 [-5.80, 17.17]	5.00 [-6.90, 34.60]	<0.001

Rate of eGFR change year 0 to year 3 (ml/min/1.73m2/year: median [IQR])	1.00 [-1.87, 3.97]	1.05 [-1.92, 5.86]	1.69 [-2.28, 11.65]	<0.001
Categorical change in renal function from year 0 to year 3 (%)				<0.001
Improved	77581 (52.5)	15378 (55.4)	4142 (55.4)	
Stable	20289 (13.7)	3060 (11.0)	720 (9.6)	
Mild Decline	18202 (12.3)	3551 (12.8)	886 (11.9)	
Rapid Decline	17018 (11.5)	3174 (11.4)	1173 (15.7)	
Very Rapid Decline	14733 (10.0)	2595 (9.3)	552 (7.4)	
Decrease in eGFR > 50% at year 3 (Yes: N(%))	4180 (2.8)	1606 (5.8)	1105 (14.8)	<0.001

CHAPTER 5 AUTOMATED ANALYSIS OF FEMORAL ARTERY CALCIFICATION[3]

5.1 INTRODUCTION

Peripheral arterial disease (PAD), which results from atherosclerotic plaque buildup with or without calcification in the large arteries of the extremities, constitutes a growing medical burden to the aging population in the United States. Estimated prevalence of PAD increases dramatically with age, from 0.9% in the population 40-49 years of age to 14.5% in people older than 69.[55] In 2001, estimated cost to the US Medicare program for PAD-related treatment was greater than $4.3 billion.[56] This figure does not take into account lost wages and productivity in PAD patients as a result of decreased mobility. Clinical presentation and outcomes of PAD are highly variable, ranging from asymptomatic disease to intermittent claudication (IC, limb pain during exercise) or rest pain in the affected limb. In the most severe cases, chronic limb threatening ischemia (CLTI) may result in tissue loss or gangrene and the need for amputation. Vascular calcification, which is commonly observed in PAD patients with co-morbid diabetes or

[3] © 2019 IEEE. Reprinted, with permission, from L. Zhao, B. Odigwe, S. Lessner, D. Clair, F. Mussa and H. Valafar, "Automated Analysis of Femoral Artery Calcification Using Machine Learning Techniques," *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2019, pp. 584-589, doi: 10.1109/CSCI49370.2019.00110.

chronic kidney disease (CKD), complicates interventional treatment and correlates with increased morbidity and mortality.[57]–[59] Currently, there is no validated, widely accepted metric to quantify the extent of arterial calcification in the lower extremities,[59] that would be comparable to the widely accepted Agatston score for coronary artery calcification.[60] Thus, there is an urgent need for better methods to predict the progression of PAD and to identify patients at greatest risk for the most severe clinical outcomes, in order to treat these patients more aggressively. To that end we have begun an exploration in the use of Machine Learning techniques to automate the diagnostic and prognostic processes in relation to PAD.

Artificial intelligent (AI) based predictive tools such as Artificial Neural Networks (ANNs)[61]–[63] have been implemented in recommender systems to enhance shopping experience (Netflix and Amazon), used to assist with speech to text recognition[64] on cell phones, and incorporated into automated driving systems[65] (Uber and Google self-driving vehicles). Extensive work has been performed to create new mechanisms of predictive modeling[66] (such as LSTMs, DNNs, and CNNs) as well as to improve already existing methods (such as ANNs, decision trees, etc). In the closer domain of healthcare[63], [67], ANNs have been used in medical fraud detection[68], [69] and in prediction of a patient's response to a drug[21], [70], thereby paving the way to personalized medicine. More specific to our investigation, the key question for clinicians, patients, regulators and insurance providers is: Can Artificial Intelligence use personalized data in patients with PAD to derive a predictive model of individual outcomes following femoral endarterectomy in terms of complications (death, lower extremity amputation, disease progression and need for intervention) at presentation?

While ANNs have advanced substantially over the last few years, their impact and integration in the field of medical diagnostics has been minimal. The lack of integration of Machine Learning techniques in Clinical Decision Systems and medical science is due to the challenging nature of analyzing medical data, and the absence of well annotated and relevant data appropriate for use in Machine Learning endeavors. In this report, we have investigated both of these challenges and present our intermediate progress with the potential of mitigating some of the long-standing challenges. More specifically, we have developed an unsupervised mechanism of tracking the aorta in X-ray computed tomography angiogram (CTA) scans in a small cohort of 5 patients. Our object-tracking mechanism has successfully tracked the aorta and its branches descending as far down as the patellar surface of the femur with more than 85% success. Using this information, we have created a clustering mechanism of identifying arterial wall, lumen, and the total blockage at each slice of the CTA scan. Finally, we discuss a mixed method approach to creating annotated data appropriate for the task of supervised learning while minimizing the time-requirement of the field-experts.

5.2 BACKGROUND AND METHOD

5.2.1 Previous and Related Work

Image segmentation is the process of identifying one or multiple objects in an image. Multiple approaches to image segmentation have been proposed and investigated over the years including thresholding[71], clustering[72], motion[73], edge detection[74], or parametric[75] methods, to name a few. Applications of image segmentation have been explored in different domains including content-based image retrieval, object detection,

face recognition, and medical imaging. More specifically, automated image processing techniques have been used to detect aortic valves using C-arm CTs[76], aortic aneurysms[77], and aortic dissections[78].

5.2.2 Background Anatomy

The primary objective of the current work is to track the aorta as it descends in the human body and branches into other smaller arteries. The aorta is the largest artery in the human body, originating at the outlet of the left ventricle and extending anterior to the vertebral column along the abdominal wall before bifurcating into paired common iliac arteries proximal to the pelvis. Major branch arteries extend from the aorta to provide blood to all parts of the systemic circulation. In the abdominal cavity, these branches begin with the celiac artery, just distal to the diaphragm, followed in order by the superior mesenteric, right renal, left renal, and inferior mesenteric arteries. These abdominal branch arteries are directed perpendicular to the longitudinal orientation of the aorta and will largely be ignored for our purposes.

Our interest in this study lies primarily in the vasculature distal to the iliac bifurcation, which provides blood to the lower extremities. Distal to the bifurcation, the common iliac arteries diverge, with each one supplying blood to one leg. The common iliac arteries further divide, giving rise to the internal and external iliac arteries. After passing through the inguinal ligament, the external iliac artery continues as the femoral artery, which further branches to give rise to the deep femoral and superficial femoral arteries.

To better understand the performance of our automated detection of the arterial system starting with the aorta, it is useful to enumerate a simplified number of main arteries that should be expected based on visual inspection of the images.

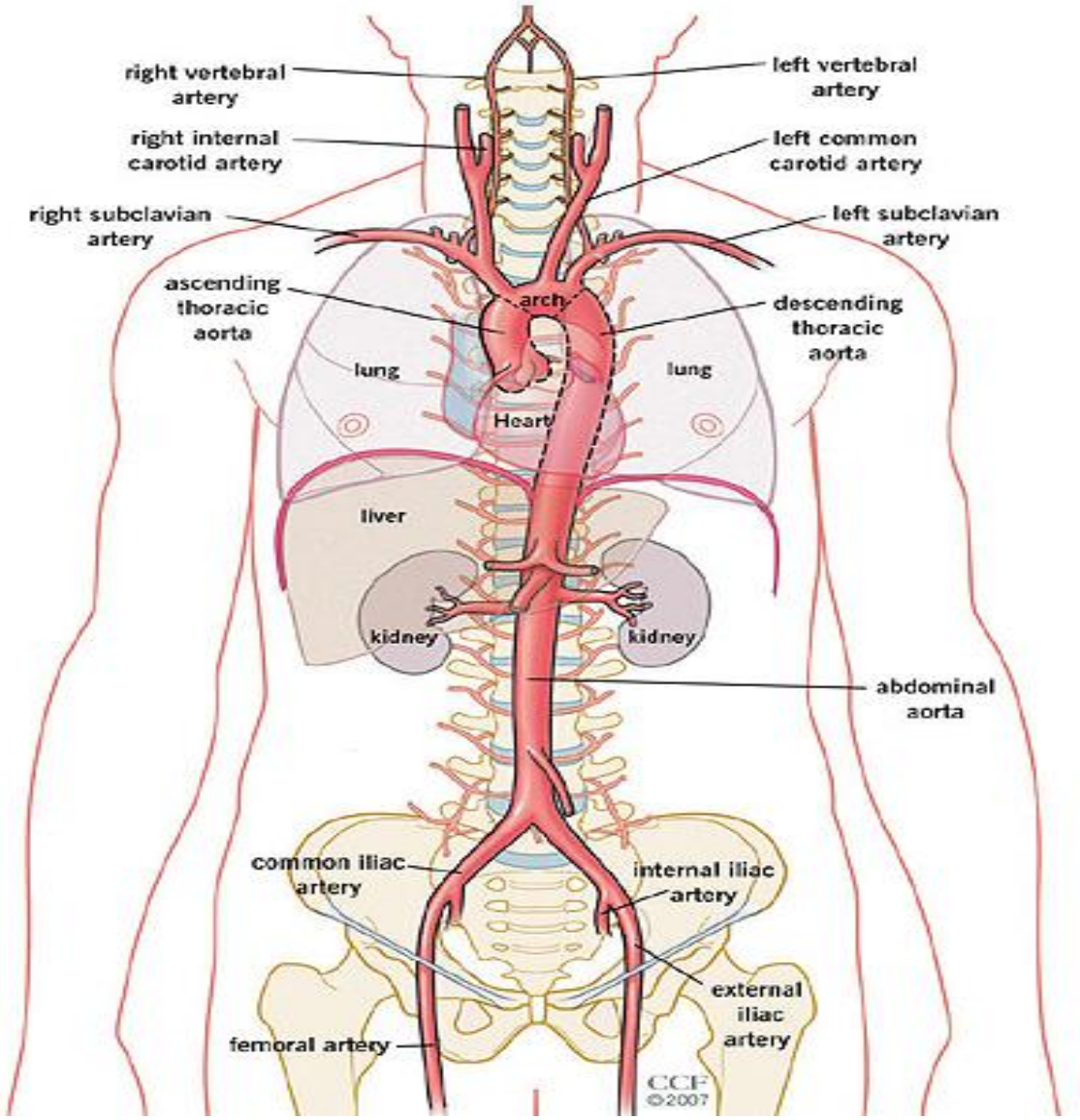


Figure 5.1 An illustration of descending aorta and the major arteries including iliac and femoral arteries.

In this study, visual inspections of the data should identify a single aortic artery in the approximate range of transverse slices of 1-100 a brief range of ~20 slices that include two distinguishable arteries (common iliacs), 40 slices with 4 clearly identifiable arteries (internal and external iliac arteries in each leg), that finally branch to numerous smaller arteries.

5.2.3 Description of the Data

In this study, we analyzed CTA scans with infused contrast agent of five patients who have been diagnosed with peripheral arterial disease. The dataset consisted of both male and female subjects, between 55 and 66 years of age. The CTA scans of these patients were used to perform the automated object tracking approach for identification of the arteries that descend from the aorta. Each image file consisted of a series of approximately 600 transverse images with the resolution of 512x512 that spanned the mid-sternum to the bottom of the subject's feet in the conventional DICOM format. Because of the limited resolution of these images (5mm slices) recognition of arteries below the knee was nearly impossible. Therefore, our region of investigation included the mid-sternum to first appearance of the patella. DICOM files were converted to TIFF files and de-identified prior to analysis.

5.2.4 Automated Object Tracking and Identification of Calcifications

The final objective of our investigation is to develop an unsupervised mechanism of annotating the arterial system that starts with the descending aorta. To accomplish this task, we have developed a multistage image segmentation routine using constrained object-tracking that is suitable for this study.

Stage I – Identification of aorta: During the first stage of the operation, the first slice of the images is subjected to filtration of pixels based on the min/max thresholds that are provided by the user. In this transverse image, the intensities corresponding to all pixels outside of the allowed thresholds will be set to a value of zero while all pixels with intensities corresponding to the acceptable range will be set to the maximum level. Furthermore, all clusters of pixels that do not adhere to an approximate ovular shape receive an intensity of zero (filtered out).

Stage II – Tracking of the arterial network: The primary objective of this stage is to track the descending aorta and the network of arteries that emanate from the aorta. Two primary constraints are used in this stage to identify arteries. The first criterion is based on a cluster of pixels that did not filter during Stage-I and exhibit ovular shape. The second criterion required overlap between any such cluster of pixels and a cluster of pixels identified as an artery from the previous slice.

Stage III – Identification of Calcification: After the completion of Stages I and II, the algorithm proceeds with the identification of the calcifications confined to the interior of the identified arterial lumen. The identification of the arterial lumen is a relatively simple task that is accomplished by the clustering of the pixels based on intensity. Pixels with a relatively higher intensity are associated with calcifications.

Stage-IV: Production of the Final Images: As the final step in our analysis, all the unobstructed interior portions of the arterial system are marked with blue color, while the calcified regions inside of the arterial lumen are marked as red. Other associated data are produced in CSV file format that summarize the number of vessels, the number of lumen pixels, and the number of calcified pixels for each slice of the images. Such data will be

the subject of our future investigations in order to establish the correlative relation between patient outcome and their current CT scans.

5.2.5 Evaluation Technique

To evaluate our automated detection mechanism, we will investigate two components. During the first phase of the evaluation, we record the number of vessels identifiable by a human in each slice for each patient. In the second phase, we record the same results for the automated detection of the vessels. Both manual and automated inspections of the images should generally correspond to the general expectations described in section IIB, while anticipating some variations due to natural anatomical variations among people. We also compare the results of the manual annotation of the arteries to that of the automated identification mechanism. Finally, we utilize the summary information reported for each slice of the images, to generate various visual analytics for each patient.

5.3 RESULTS AND DISCUSSION

5.3.1 Automated Identification of the Arterial Network

The automated detection process starts with the detection of the aorta in the first slice of the images. Figure 5.2a illustrates the first slice for patient 6572. After completing Stage-I of our analysis the masking binary image shown in Figure 5.2b is identified. This mask was then used to track the descending aorta (show in Figure 5.2c) while identifying the internal calcifications (Figure 5.3)

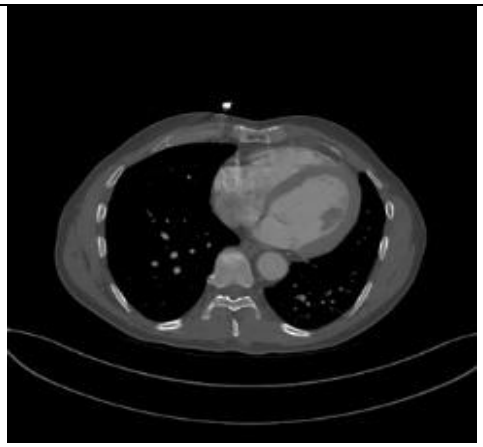
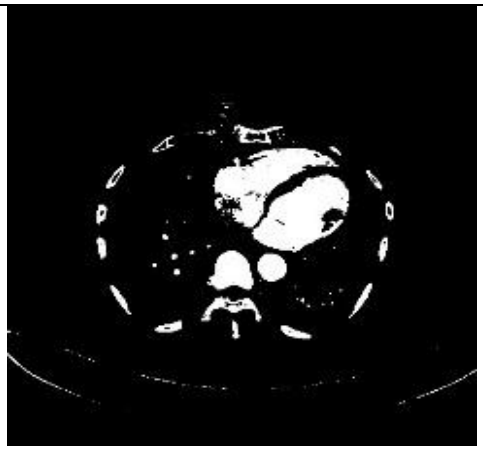
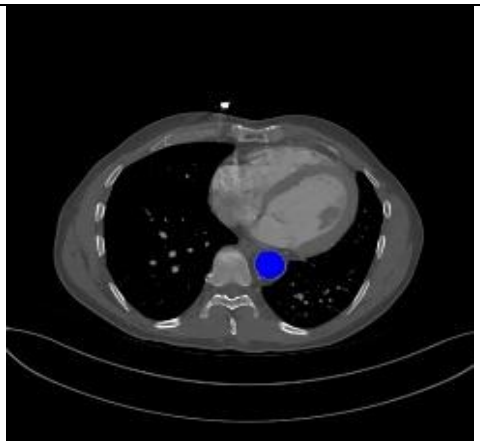
(a)	
(b)	
(c)	

Figure 5.2 The first slice for patient 6572 (a) at the beginning of the process,

(b) binarized based on cutoff thresholds, and (c) after the completion of the Stage-III.

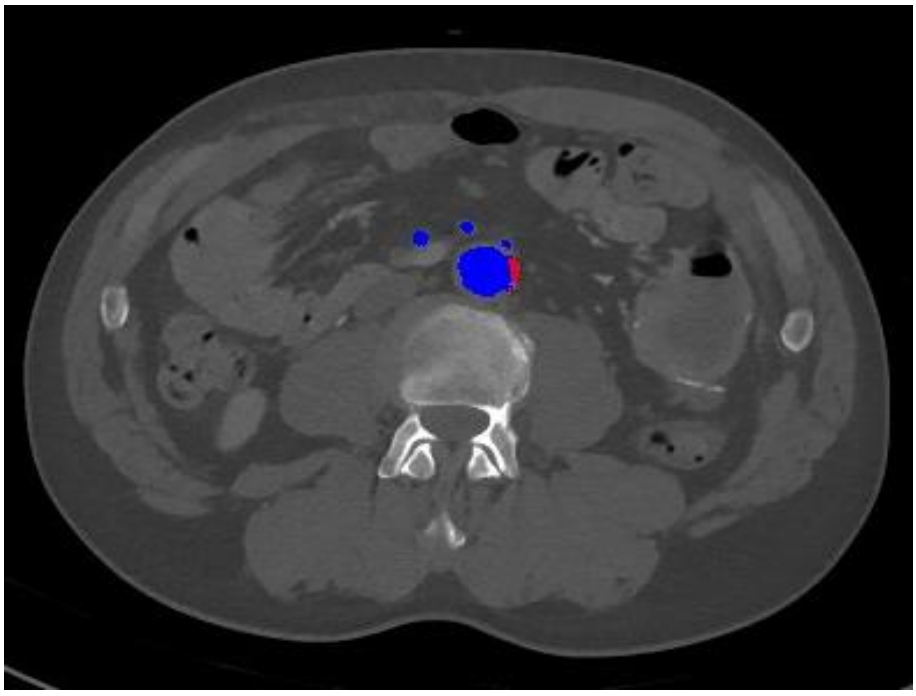


Figure 5.3 Slice #84 for patient 6572 at the end of Stage-IV of the automated analysis. Blue indicates the unobstructed lumen of vessels, while red denotes the identified calcification

5.3.2 Comparison of Automated and Manual Tracking of Vessels

To validate the outcome of our automated tracking of the descending aorta and the emanating network of arteries, we compare the results to manual inspection and identification of arteries. This comparison can be performed at various levels of granularity spanning from coarse total number of arteries detected for each person, to the fine-grained pixel level. In this investigation, we focus on an intermediate level that compares the number of arteries in each slice detected by the automated system versus manually by a trained human. Figure 5.4 illustrates the results for manual and automated detection of vessels in slices 1-385. Note the degree of similarity between the two approaches.

Furthermore, it is noteworthy that not only in some instances the automated approach has missed detection of vessels (usually slices 250-300), however in slices 80-120, the automated method has over-detected the number of vessels.

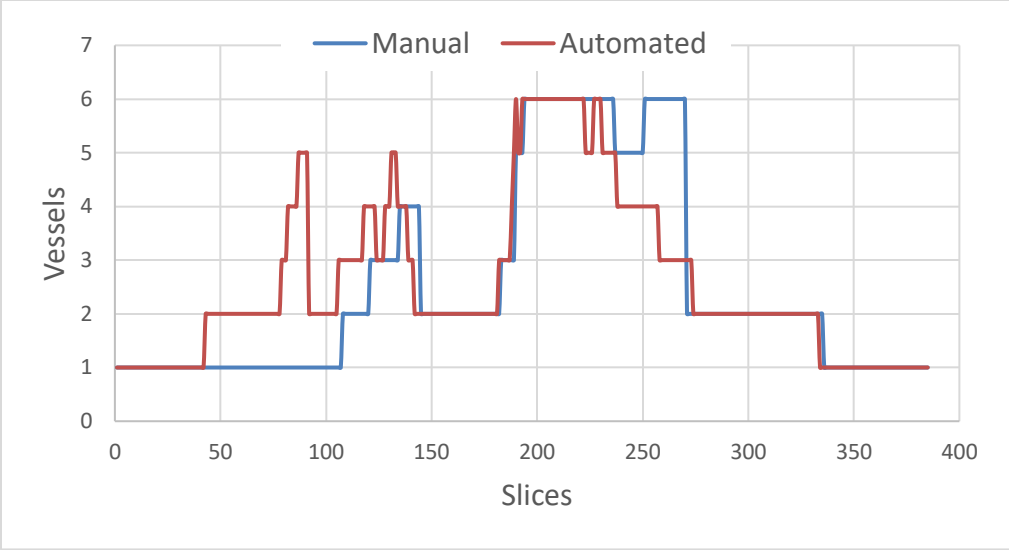


Figure 5.4 Comparison of the number of vessels detected manually (blue) to automated (red) for slices 1-385 in patient 6572.

5.3.3 Stenosis Profile

Subsequent to the detection of the arterial system, our automated analysis can discriminate between the lumen, or a calcified region of a vessel based on the intensity of the pixels. Using this method, we can quantify several relevant parameters including the volume of the cross-section of a vessel, relative unobstructed area, and the relative area of the obstruction. Using this information, a profile of vessel stenosis for a given patient can be constructed.

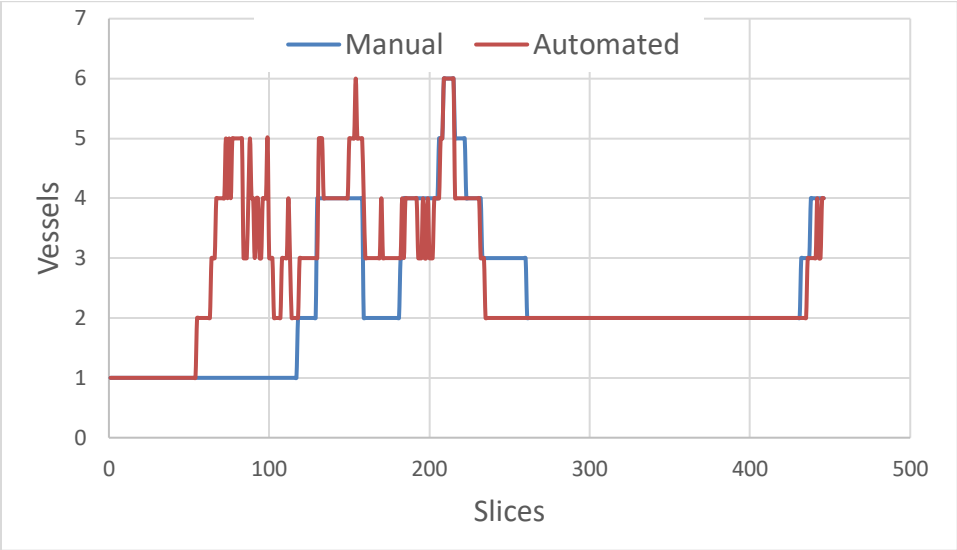


Figure 5.5 Comparison of the number of vessels detected manually (blue) to automated (red) for slices 1-385 in patient 6573.

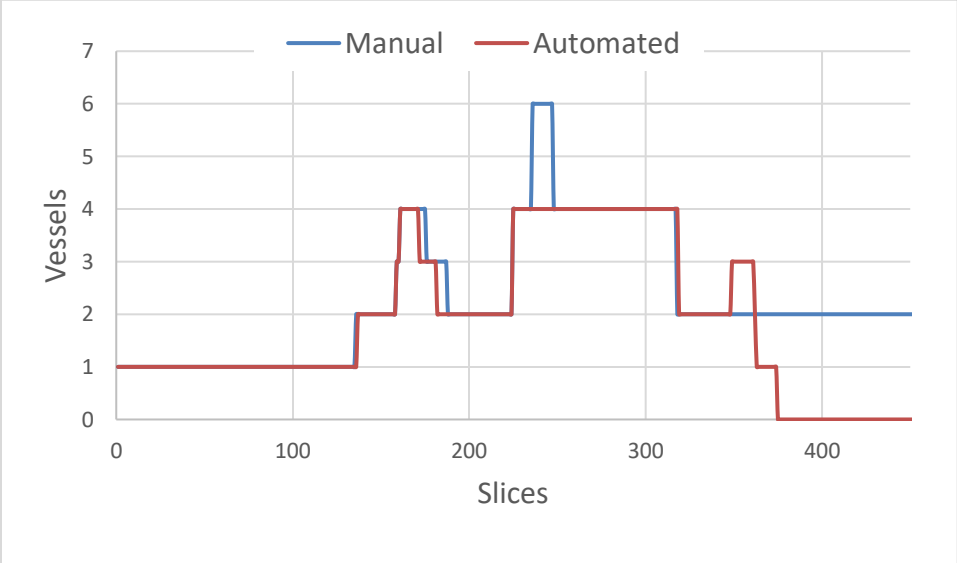


Figure 5.6 Comparison of the number of vessels detected manually (blue) to automated (red) for slices 1-385 in patient 6574.

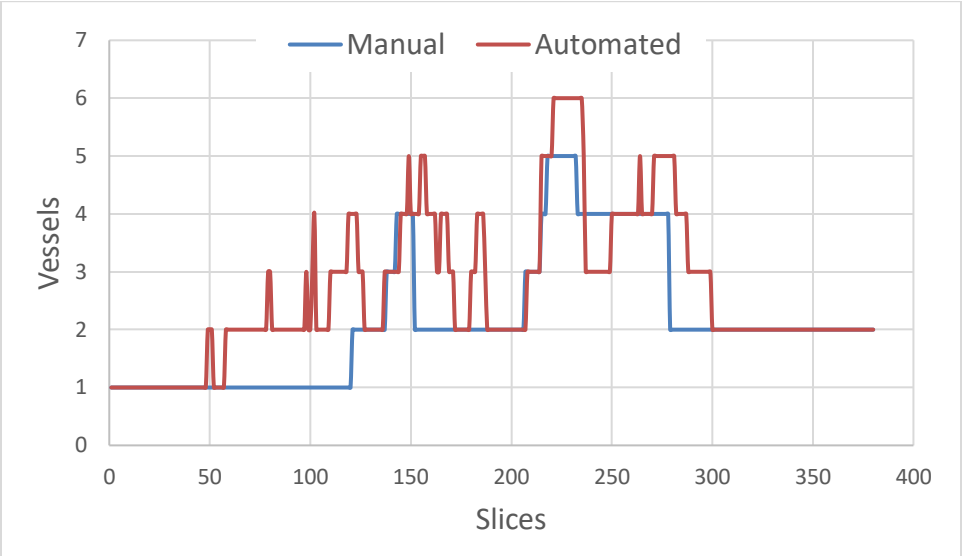


Figure 5.7 Comparison of the number of vessels detected manually (blue) to automated (red) for slices 1-385 in patient 6575.

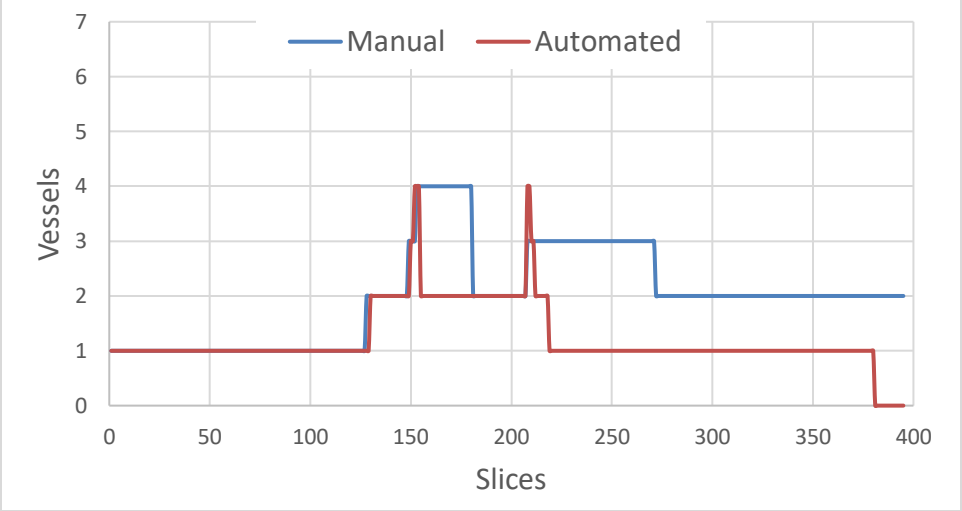


Figure 5.8 Comparison of the number of vessels detected manually (blue) to automated (red) for slices 1-385 in patient 6576.

Figure 5.9 illustrates the automatically generated stenosis profile for patient 6572 that indicates the highest level of blockage (20%) at slice range of 96-98. In contrast to this patient, the stenosis profile for patient 6573 exhibits a more serious vascular disease with nearly 100% blockage at slice range of 400-408.

In addition to the volumetric information that is purely based on the number of pixels, our automated system can provide a measure of blockage that includes a degree of calcification. Although not discussed here, this score is closely correlated to the Agatston score[60] and is based on the integral of intensities of pixels that are identified as calcifications.

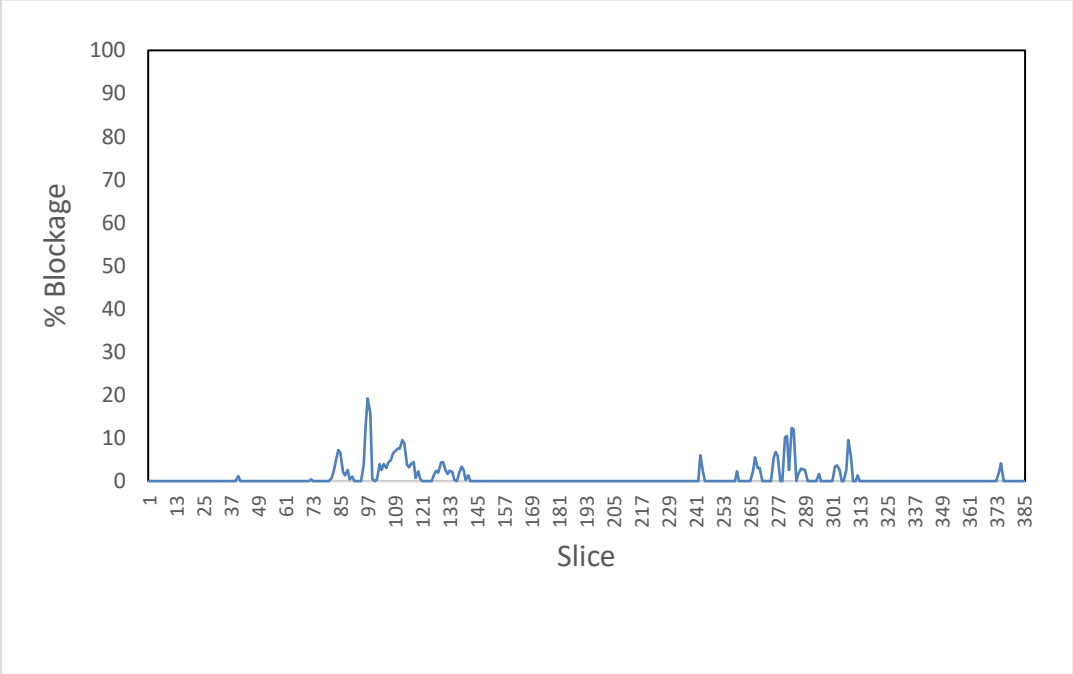


Figure 5.9 Automated calculation of the stenosis profile for patient 6572.

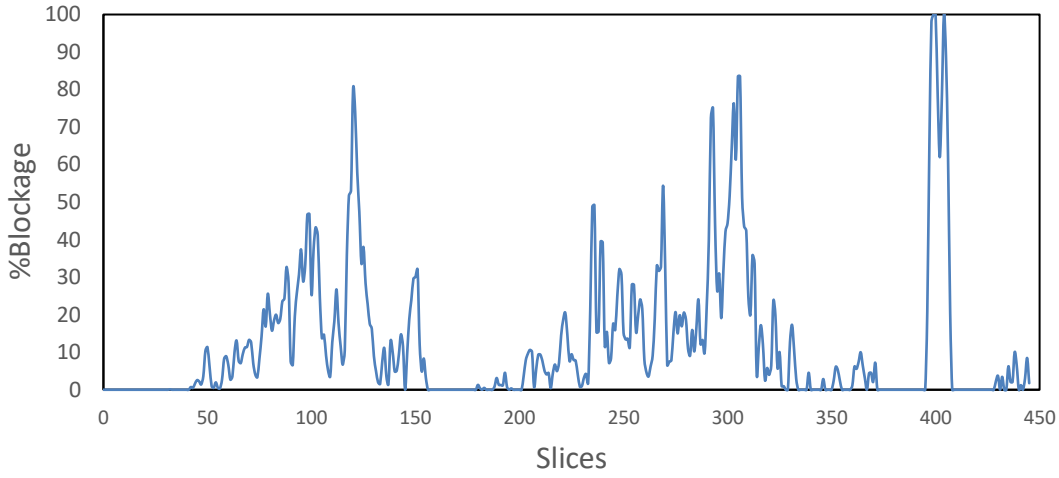


Figure 5.10 Automated calculation of the stenosis profile for patient 6575.

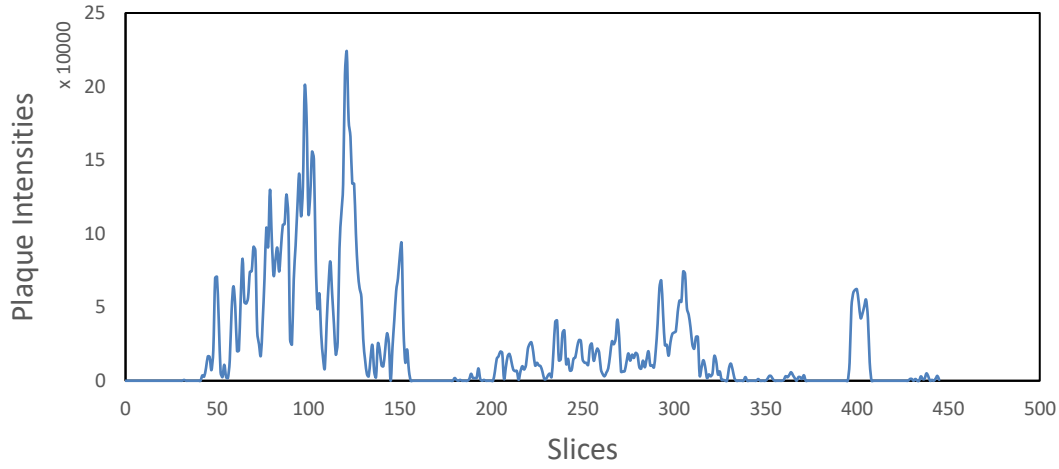


Figure 5.11 Sum of the calcification intensities in each slice for patient 6575

Our ultimate goal in this work is to automate identification and quantitation of vessel calcification in CT angiograms of the extremities in patients diagnosed with peripheral arterial disease as a prognostic tool for risk stratification. To this end, we have initiated clinical studies to track patient outcomes and relate adverse outcomes (disease

progression, need for additional intervention, amputation, or death) to the extent and distribution of calcification at initial CTA evaluation.

5.4 CONCLUSION

Detailed annotation of medical images at the pixel level is one pragmatic limitation in initiating multidisciplinary research between experts in the fields of Machine Learning and medical sciences. Here we have presented a mixed approach that combines thresholding, motion detection, and other geometrical constraints in order to facilitate an effective means of tracking the descending aorta, as it expands into a network of arteries in the lower abdominal and femoral regions. Our object tracking approach has exhibited more than 85% agreement with human annotation of images. Such an approach can be used to provide a more efficient means of image annotation by experts and physicians who cannot dedicate as much time to this kind of activity.

In addition to annotation of the arteries, we have also demonstrated the feasibility of detection and quantification of the calcifications in the major arteries extending into the femoral regions. Tools such as the stenosis profile and Agatston scoring profile can be useful in diagnosis and to correlate with patient outcomes for the eventual development of prognostic tools.

CHAPTER 6 CONCLUSION

The advancements made in the field of machine learning make ML techniques suitable for use in the development of tools that aid in the medical decision-making process even though, for various justified reasons, their impact and integration in the field of medical diagnostics have been minimal.

Machine learning models can recognize patterns and complex relationships between parameters that are not easily discernible by man.

The approach / machine learning technique used for each experiment in our studies was chosen out of necessity with primary considering given to the study objective.

The datasets used in our work were varied in size as well as the hypothesis guiding our experiments, and as such, our approach to predictive analysis also varied. When number of subjects are incredibly limited, as they were in our research into sickle cell anemia response to hydroxyurea, a greater degree of caution is employed when considering dataset splits for training and testing.

In our sickle cell anemia patient's response study, as with all our studies, improved patient selection was the goal, but primary to sickle cell anemia and Hydroxyurea was the lack of consensus on acceptable classification criterion. This necessitated the employment of different categorization criteria and ultimately a reformulation of the task as a regression problem, to get around the lack of consensus. Our approach to data splitting was also one

made necessary by the unique challenges with the dataset size. we decided to use an N-fold cross validation approach, primarily due to the dataset constraints.

In our Heart Failure study, the primary goal was patient classification and patient response quantification, but the improvement to clinical practice lied with identifying parameters most relevant to observed response patterns. The primary approach needed to be explainable, which is why decision/classification trees, and more specifically pruned trees were the main approach. This was compatible with the size of the dataset, comprising less than 1000 subjects and the demand for more explainable predictive models, as classification trees exposes the sets of rules used in arriving at its decisions.

In Our Chronic Kidney Disease Study, having the largest and most diverse (particularly due to the nationwide nature of data sources) dataset used in all our studies, of primary importance was the modeling and prediction of the rate of change for a patient parameter. With the size of the dataset came the propensity for it to be less “well behaved”. As is often seen with medical datasets of that size, it had a lot of missing values and the predictive modeling approach had to be one that is robust to such occurrences. Therefore, use of ANNs was the primary predictive approach used in this study. Also, explainability of the decision making and/or the identification the most relevant parameters was not the primary objective, so there was no deterrent to the use of a more “black-boxed” approach such as neural networks.

These approaches and others like it will positively influence medical decision-making, and administration of intervention procedures, and further the practice of precision medicine. The approaches and the rules generated produce a means of prioritizing patient

data parameters and present us with the opportunity to extend medical practice and ultimately improve patient outcomes.

REFERENCES

[1] "Understanding Convolutional Neural Networks for NLP." https://www.kdnuggets.com/2015/11/understanding-convolutional-neural-networks-nlp.html/3 (accessed Nov. 11, 2020).

[2] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[3] S. Hadush, Y. Girmay, A. Sinamo, and G. Hagos, "Breast Cancer Detection Using Convolutional Neural Networks," Mar. 2020, Accessed: Nov. 11, 2020. [Online]. Available: http://arxiv.org/abs/2003.07911.

[4] E. Choi, M. Taha Bahadori, A. Schuetz, and W. F. Stewart, "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks."

[5] J. M. Johnson and T. M. Khoshgoftaar, "Medicare fraud detection using neural networks," *J. Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0225-0.

[6] R. K. Price, E. L. Spitznagel, T. J. Downey, D. J. Meyer, N. K. Risk, and O. G. El-Ghazzawy, "Applying artificial neural network models to clinical decision making.," *Psychol. Assess.*, vol. 12, no. 1, pp. 40–51, 2000, doi: 10.1037/1040-3590.12.1.40.

[7] J.-M. Bae, "The clinical decision analysis using decision tree," *Epidemiol. Health*,

p. 36, doi: 10.4178/epih/e2014025.

[8] G. Karalis, “Decision Trees and Applications.,” *Adv. Exp. Med. Biol.*, vol. 1194, pp. 239–242, 2020, doi: 10.1007/978-3-030-32622-7_21.

[9] “ROSE: decision trees, automatic learning and their applications in cardiac medicine - PubMed.” https://pubmed.ncbi.nlm.nih.gov/8591548/ (accessed Sep. 23, 2020).

[10] M. Alizamir, S. Kim, O. Kisi, and M. Zounemat-Kermani, “A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions,” *Energy*, vol. 197, p. 117239, Apr. 2020, doi: 10.1016/J.ENERGY.2020.117239.

[11] D. M. Bates and D. G. Watts, “Nonlinear regression analysis and its applications,” p. 365, 1988.

[12] “How Machine Learning is Transforming Clinical Decision Support Tools,” 2020. https://healthitanalytics.com/features/how-machine-learning-is-transforming-clinical-decision-support-tools (accessed Nov. 06, 2020).

[13] J. Dalton and A. Deshmane, “An Approach To Increasing Machine Intelligence,” *IEEE Potentials*, vol. 10, no. 2, pp. 33–36, 1991, doi: 10.1109/45.84097.

[14] “Artificial neural networks - IEEE Journals & Magazine.” https://ieeexplore.ieee.org/document/84097 (accessed Nov. 11, 2020).

[15] S. Wu, W. Zhang, F. Sun, and B. Cui, “Graph Neural Networks in Recommender Systems: A Survey,” Nov. 2020, Accessed: Nov. 11, 2020. [Online]. Available: http://arxiv.org/abs/2011.02260.

[16] C. Odhiambo, P. Wright, C. Corbett, and H. Valafar, “MedSensor: Medication

Adherence Monitoring Using Neural Networks on Smartwatch Accelerometer Sensor Data," May 2021, doi: 10.48550/arxiv.2105.08907.

[17] C. O. Odhiambo, C. A. Cole, A. Torkjazi, and H. Valafar, "State transition modeling of the smoking behavior using lstm recurrent neural networks," *Proc. - 6th Annu. Conf. Comput. Sci. Comput. Intell. CSCI 2019*, pp. 898–904, Dec. 2019, doi: 10.1109/CSCI49370.2019.00171.

[18] A. B. Rajeoni, "ANALOG CIRCUIT SIZING USING MACHINE LEARNING BASED TRANSISTOR," 2021.

[19] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10253 LNCS, pp. 477–492, doi: 10.1007/978-3-319-59536-8_30.

[20] C. H. Lee and H. J. Yoon, "Medical big data: Promise and challenges," *Kidney Res. Clin. Pract.*, vol. 36, no. 1, pp. 3–11, Mar. 2017, doi: 10.23876/j.krcp.2017.36.1.3.

[21] B. E. Odigwe, J. S. Eyitayo, C. I. Odigwe, and H. Valafar, "Modelling of Sickle Cell Anemia Patients Response to Hydroxyurea using Artificial Neural Networks," Nov. 2019, Accessed: May 26, 2020. [Online]. Available: http://arxiv.org/abs/1911.10978.

[22] L. Zhao, B. Odigwe, S. Lessner, D. G. Clair, F. Mussa, and H. Valafar, "Automated Analysis of Femoral Artery Calcification Using Machine Learning Techniques," *Proc. - 6th Annu. Conf. Comput. Sci. Comput. Intell. CSCI 2019*, pp.

584–589, Dec. 2019, Accessed: Nov. 11, 2020. [Online]. Available: http://arxiv.org/abs/1912.06010.

[23] L. Álvarez Menéndez, F. J. de Cos Juez, F. Sánchez Lasheras, and J. A. Álvarez Riesgo, "Artificial neural networks applied to cancer detection in a breast screening programme," *Math. Comput. Model.*, vol. 52, no. 7–8, pp. 983–991, Oct. 2010, doi: 10.1016/j.mcm.2010.03.019.

[24] W. Lu, Y. Tong, Y. Yu, Y. Xing, C. Chen, and Y. Shen, "Applications of Artificial Intelligence in Ophthalmology: General Overview," *Journal of Ophthalmology*, vol. 2018. Hindawi Limited, 2018, doi: 10.1155/2018/5278196.

[25] M. R. Bristow, "Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION): Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure," *N Engl J Med*, vol. 350, pp. 2140–2150, 2004.

[26] J. G. F. Cleland *et al.*, "The effect of cardiac resynchronization on morbidity and mortality in heart failure," *N. Engl. J. Med.*, vol. 352, no. 15, pp. 1539–1549, 2005.

[27] S. A. Hunt *et al.*, "2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed," *J. Am. Coll. Cardiol.*, vol. 53, no. 15, pp. e1--e90, 2009.

[28] K. A. Ellenbogen *et al.*, "Primary results from the SmartDelay determined AV optimization: a comparison to other AV delay methods used in cardiac resynchronization therapy (SMART-AV) trial: a randomized trial comparing

empirical, echocardiography-guided, and algorithmic atrioventri," *Circulation*, vol. 122, no. 25, pp. 2660–2668, 2010.

[29] I. Goldenberg *et al.*, "Predictors of response to cardiac resynchronization therapy in the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT)," *Circulation*, vol. 124, no. 14, pp. 1527–1536, 2011.

[30] L. A. Pires *et al.*, "Clinical predictors and timing of New York Heart Association class improvement with cardiac resynchronization therapy in patients with advanced chronic heart failure: results from the Multicenter InSync Randomized Clinical Evaluation (MIRACLE) and Multice," *Am. Heart J.*, vol. 151, no. 4, pp. 837–843, 2006.

[31] F. G. Spinale *et al.*, "Development of a biomarker panel to predict cardiac resynchronization therapy response: Results from the SMART-AV trial," *Hear. Rhythm*, vol. 16, no. 5, pp. 743–753, May 2019, doi: 10.1016/j.hrthm.2018.11.026.

[32] "Heart Failure: Investigation of an Epidemic," 2013, doi: 10.1161/CIRCRESAHA.113.300268.

[33] K. M. Stein *et al.*, "SmartDelay determined AV optimization: A comparison of AV delay methods used in cardiac resynchronization therapy (SMART-AV): Rationale and design," *PACE - Pacing Clin. Electrophysiol.*, vol. 33, no. 1, pp. 54–63, Jan. 2010, doi: 10.1111/j.1540-8159.2009.02581.x.

[34] E. Papaemmanuil *et al.*, "Genomic Classification and Prognosis in Acute Myeloid Leukemia," *N. Engl. J. Med.*, vol. 374, no. 23, pp. 2209–2221, Jun. 2016, doi: 10.1056/nejmoa1516192.

[35] S. Kumar Srivastava, R. Kumari MTech Scholar, and S. Kr Srivastava Sr Asst Professor, “Machine Learning: A Review on Binary Classification Development of Efficient Algorithms for Predicting the Incidence of Malaria Cases in India using Spiking Neuron Models View project Cardiovascular Risk Assessment and Prediction View project Machine Learning: A Review on Binary Classification,” *Artic. Int. J. Comput. Appl.*, vol. 160, no. 7, pp. 975–8887, 2017, doi: 10.5120/ijca2017913083.

[36] R. Cuocolo, T. Perillo, E. De Rosa, L. Ugga, and M. Petretta, “Current applications of big data and machine learning in cardiology,” *J. Geriatr. Cardiol.*, vol. 16, pp. 601–607, 2019, doi: 10.11909/j.issn.1671-5411.2019.08.002.

[37] M. R. Gold, A. Padhiar, S. Mealing, M. K. Sidhu, S. I. Tsintzos, and W. T. Abraham, “Economic Value and Cost-Effectiveness of Cardiac Resynchronization Therapy Among Patients With Mild Heart Failure Projections From the REVERSE Long-Term Follow-Up,” 2017.

[38] G. Boriani *et al.*, “Cardiac resynchronization therapy: a cost or an investment?,” doi: 10.1093/europace/eur079.

[39] B. E. Odigwe, F. G. Spinale, and H. Valafar, “Application of Machine Learning in Early Recommendation of Cardiac Resynchronization Therapy,” Sep. 2021, doi: 10.48550/arxiv.2109.06139.

[40] B. S. Everitt, “Commentary: Classification and cluster analysis,” *BMJ*, vol. 311, no. 7004, p. 535, Aug. 1995, doi: 10.1136/bmj.311.7004.535.

[41] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, “Application of k-Means Clustering algorithm for prediction of Students’ Academic Performance,” 2010.

[Online]. Available: http://sites.google.com/site/ijcsis/.

[42] M. J. Guess and S. B. Wilson, “Introduction to hierarchical clustering.,” *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, vol. 19, no. 2, pp. 144–151, Apr. 2002, doi: 10.1097/00004691-200203000-00005.

[43] N. Ketkar and N. Ketkar, “Introduction to Tensorflow,” in *Deep Learning with Python*, Apress, 2017, pp. 159–194.

[44] N. Ketkar and N. Ketkar, “Introduction to Keras,” in *Deep Learning with Python*, Apress, 2017, pp. 97–111.

[45] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, “Euclidean Distance Geometry and Applications,” *SIAM Rev.*, vol. 56, no. 1, pp. 3–69, 2014, doi: 10.1137/120875909.

[46] D. Sinwar and R. Kaushik, “Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering,” no. May, 2014.

[47] B. Li and L. Han, “Distance Weighted Cosine Similarity Measure for Text Classification,” in *Intelligent Data Engineering and Automated Learning -- IDEAL 2013*, 2013, pp. 611–618.

[48] D. Stathakis, “How many hidden layers and nodes?,” *Int. J. Remote Sens.*, vol. 30, no. 8, pp. 2133–2147, Apr. 2009, doi: 10.1080/01431160802549278.

[49] P. F. Orlandi *et al.*, “CLINICAL EPIDEMIOLOGY Slope of Kidney Function and Its Association with Longitudinal Mortality and Cardiovascular Disease among Individuals with CKD,” 2017, doi: 10.1681/ASN.2020040476.

[50] T. J. Hoerger *et al.*, “The future burden of CKD in the United States: a simulation model for the CDC CKD Initiative,” *Am. J. Kidney Dis.*, vol. 65, no. 3, pp. 403–

411, Mar. 2015, doi: 10.1053/J.AJKD.2014.09.023.

[51] C. Small *et al.*, “Non-dialysis dependent chronic kidney disease is associated with high total and out-of-pocket healthcare expenditures,” *BMC Nephrol.*, vol. 18, no. 1, Jan. 2017, doi: 10.1186/S12882-016-0432-2.

[52] C. Wanner *et al.*, “Empagliflozin and Clinical Outcomes in Patients With Type 2 Diabetes Mellitus, Established Cardiovascular Disease, and Chronic Kidney Disease,” *Circulation*, vol. 137, no. 2, pp. 119–129, 2018, doi: 10.1161/CIRCULATIONAHA.117.028268.

[53] N. Tangri *et al.*, “A predictive model for progression of chronic kidney disease to kidney failure,” *JAMA*, vol. 305, no. 15, pp. 1553–1559, Apr. 2011, doi: 10.1001/JAMA.2011.451.

[54] T. Ferguson *et al.*, “Development and External Validation of a Machine Learning Model for Progression of CKD,” *Kidney Int. reports*, vol. 7, no. 8, pp. 1772–1781, Aug. 2022, doi: 10.1016/J.EKIR.2022.05.004.

[55] E. Selvin and T. P. Erlinger, “Prevalence of and Risk Factors for Peripheral Arterial Disease in the United States,” *Circulation*, vol. 110, no. 6, pp. 738–743, Aug. 2004, doi: 10.1161/01.CIR.0000137913.26087.F0.

[56] A. T. Hirsch, L. Hartman, R. J. Town, and B. A. Virnig, “National health care costs of peripheral arterial disease in the Medicare population,” *Vasc. Med.*, vol. 13, no. 3, pp. 209–215, Aug. 2008, doi: 10.1177/1358863X08089277.

[57] M. A. Allison *et al.*, “Calcified atherosclerosis in different vascular beds and the risk of mortality.,” *Arterioscler. Thromb. Vasc. Biol.*, vol. 32, no. 1, pp. 140–6, Jan. 2012, doi: 10.1161/ATVBAHA.111.235234.

[58] M. M. Chowdhury *et al.*, “Lower limb arterial calcification (LLAC) scores in patients with symptomatic peripheral arterial disease are associated with increased cardiac mortality and morbidity.,” *PLoS One*, vol. 12, no. 9, p. e0182952, Sep. 2017, doi: 10.1371/journal.pone.0182952.

[59] K. J. Rocha-Singh, T. Zeller, and M. R. Jaff, “Peripheral arterial calcification: prevalence, mechanism, detection, and clinical implications.,” *Catheter. Cardiovasc. Interv.*, vol. 83, no. 6, pp. E212-20, May 2014, doi: 10.1002/ccd.25387.

[60] A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano, “Quantification of coronary artery calcium using ultrafast computed tomography.,” *J. Am. Coll. Cardiol.*, vol. 15, no. 4, pp. 827–32, Mar. 1990, doi: 10.1016/0735-1097(90)90282-t.

[61] K. J. Holyoak, “Parallel Distributed-Processing - Explorations in the Microstructure of Cognition - Rumelhart,De, Mcclelland,Jl,” *Science (80-.).*, vol. 236, no. 4804, pp. 992–996, 1987.

[62] D. E. Rumelhart and J. L. Mcclelland, *Parallel distributed processing: explorations in the microstructure of Cognition. Volume 1. Foundations*, vol. 327, no. 6122. MIT Press, Cambridge, MA, 1986.

[63] N. Boltin, D. Vu, B. Janos, A. Shofner, J. Culley, and H. Valafar, “An AI model for Rapid and Accurate Identification of Chemical Agents in Mass Casualty Incidents.,” *HIMS 2016 Proc. 2016 Int. Conf. Heal. Informatics Med. Syst. HIMS (2016 Las Vegas, Nev.)*, vol. 2016, pp. 169–175, Jul. 2016.

[64] A. Graves, A. Mohamed, and G. Hinton, “Speech recognition with deep recurrent

neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649, doi: 10.1109/ICASSP.2013.6638947.

[65] M. Bojarski *et al.*, "End to End Learning for Self-Driving Cars," Apr. 2016.

[66] S. Haykin, *Neural networks: a comprehensive foundation*. 1994.

[67] N. Boltin, D. Valdes, J. M. Culley, and H. Valafar, "Mobile Decision Support Tool for Emergency Departments and Mass Casualty Incidents (EDIT): Initial Study.," *JMIR mHealth uHealth*, vol. 6, no. 6, p. e10727, Jun. 2018, doi: 10.2196/10727.

[68] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," Sep. 2010, doi: 10.1016/j.chb.2012.01.002.

[69] S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neural-network," in *Proceedings of the Hawaii International Conference on System Sciences*, 1994, vol. 3, pp. 621–630, doi: 10.1109/hicss.1994.323314.

[70] H. Valafar *et al.*, "Predicting the effectiveness of hydroxyurea in individual sickle cell anemia patients.," *Artif. Intell. Med.*, vol. 18, no. 2, pp. 133–48, Feb. 2000.

[71] K. J. Batenburg and J. Sijbers, "Adaptive thresholding of tomograms by projection distance minimization," *Pattern Recognit.*, vol. 42, no. 10, pp. 2297–2305, Oct. 2009, doi: 10.1016/j.patcog.2008.11.027.

[72] L. Barghout and J. Sheynin, "Real-world scene perception and perceptual organization: Lessons from Computer Vision," *J. Vis.*, vol. 13, no. 9, pp. 709–709, Jul. 2013, doi: 10.1167/13.9.709.

[73] D. Wang, "Unsupervised video segmentation based on watersheds and temporal

tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 539–546, 1998, doi: 10.1109/76.718501.

[74] L. Barghout, "Visual Taxometric Approach to Image Segmentation Using Fuzzy-Spatial Taxon Cut Yields Contextually Relevant Regions," in *Communications in Computer and Information Science*, 2014, vol. 443 CCIS, no. PART 2, pp. 163–173, doi: 10.1007/978-3-319-08855-6_17.

[75] D. Cremers and S. Soatto, "Motion competition: A variational approach to piecewise parametric motion segmentation," *International Journal of Computer Vision*, vol. 62, no. 3. pp. 249–265, May 2005, doi: 10.1007/s11263-005-4882-4.

[76] Y. Zheng *et al.*, "Automatic aorta segmentation and valve landmark detection in C-arm CT: application to aortic valve implantation.," *Med. Image Comput. Comput. Assist. Interv.*, vol. 13, no. Pt 1, pp. 476–83, 2010, doi: 10.1007/978-3-642-15705-9_58.

[77] F. Lareyre, C. Adam, M. Carrier, C. Dommerc, C. Mialhe, and J. Raffort, "A fully automated pipeline for mining abdominal aortic aneurysm using image segmentation," *Sci. Rep.*, vol. 9, no. 1, p. 13750, Dec. 2019, doi: 10.1038/s41598-019-50251-8.

[78] E. Dehghan, H. Wang, and T. Syeda-Mahmood, "Automatic detection of aortic dissection in contrast-enhanced CT," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Apr. 2017, pp. 557–560, doi: 10.1109/ISBI.2017.7950582.

Appendix A: Permissions

Brendan Odigwe November 9, 2022 at 11:59 AM
Permission to use manuscript in Dissertation
To: LIANG ZHAO

Hello Liang, I hope you have been having a great week.

I am graduating this semester and would like to use our paper in my dissertation.

Because you are listed as the first author on the paper, I need your written permission.

So, I am officially requesting your permission to include the contents of our paper in my dissertation.

The paper - L. Zhao, B. Odigwe, S. Lessner, D. Clair, F. Mussa and H. Valafar, "Automated Analysis of Femoral Artery Calcification Using Machine Learning Techniques," *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2019, pp. 584-589, doi: 10.1109/CSCI49370.2019.00110.

I look forward to hearing from you soon.

Thank you and Kind Regards

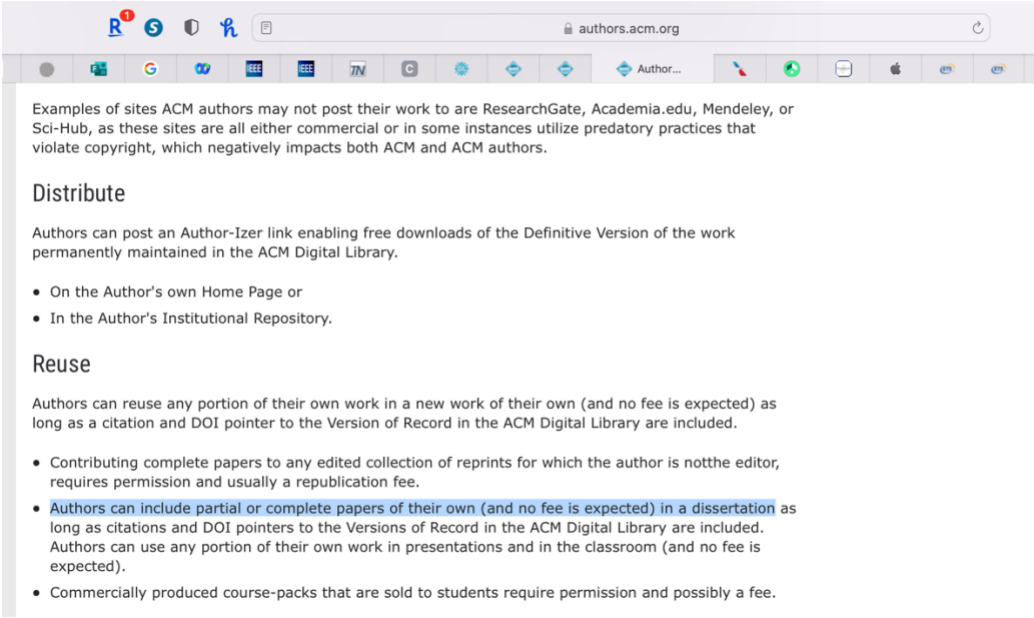
Brendan Odigwe
PhD Candidate, Computer Science
UofSC

Zhao, Liang November 9, 2022 at 1:49 PM
Re: Permission to use manuscript in Dissertation
To: Odigwe, Brendan

Hello Brendan,

Congratulation! I am totally glad that you would like to include our paper and it is derived from Dr. Valafa's lab. So this is my written permission that you can feel free to use the paper. Wish you everything good!

Best,
Liang



Examples of sites ACM authors may not post their work to are ResearchGate, Academia.edu, Mendeley, or Sci-Hub, as these sites are all either commercial or in some instances utilize predatory practices that violate copyright, which negatively impacts both ACM and ACM authors.

Distribute

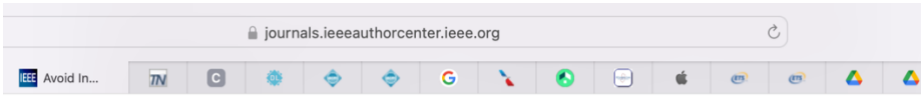
Authors can post an Author-Izer link enabling free downloads of the Definitive Version of the work permanently maintained in the ACM Digital Library.

- On the Author's own Home Page or
- In the Author's Institutional Repository.

Reuse

Authors can reuse any portion of their own work in a new work of their own (and no fee is expected) as long as a citation and DOI pointer to the Version of Record in the ACM Digital Library are included.

- Contributing complete papers to any edited collection of reprints for which the author is notthe editor, requires permission and usually a republication fee.
- Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included. Authors can use any portion of their own work in presentations and in the classroom (and no fee is expected).
- Commercially produced course-packs that are sold to students require permission and possibly a fee.



Can I Reuse My Published Article in My Thesis?

You may reuse your published article in your thesis or dissertation without requesting permission, provided that you fulfill the following requirements depending on which aspects of the article you wish to reuse.

- **Text excerpts:** Provide the full citation of the original published article followed by the IEEE copyright line: © 20XX IEEE. If you are reusing a substantial portion of your article and you are not the senior author, obtain the senior author's approval before reusing the text.
- **Graphics and tables:** The IEEE copyright line (© 20XX IEEE) should appear with each reprinted graphic and table.
- **Full text article:** Include the following copyright notice in the references: "© 20XX IEEE. Reprinted, with permission, from [full citation of original published article]."

When posting your thesis on your university website, include the following message:

> "In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [name of university or educational entity]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation."