QATAR UNIVERSITY

COLLEGE OF ENGINEERING

APPLICATION OF MACHINE LEARNING TECHNIQUES FOR THE PREDICTION OF

DECOMPRESSIVE HEMICRANIECTOMY PROGNOSIS IN ACUTE ISCHEMIC

STROKE

BY

RAHMA SALEH ALI

A Thesis Submitted to

the Faculty of the College of Engineering

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Computing

June 2019

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Rahma Saleh Ali defended on 16/04/2019.

_____
Dr. Uvais Qidwai
Thesis/Dissertation Supervisor

_____
Dr. Sumaya Al-Maadeed
Committee Member

_____
Dr. Hashim Raza Khan
External Examiner

Approved:

_____
Abdel Magid Hamouda , Dean, College of Engineering

ABSTRACT

ALI, RAHMA, S., Masters : June : 2019, Masters of Science in Computing

Title: <u>Application of Machine Learning Techniques for the Prediction of Decompressive Hemicraniectomy Prognosis in Acute Ischemic Stroke</u>

Supervisor of Thesis: Dr. Uvais A Qidwai.

Stroke is one of the leading causes of death in the world with the number of people suffering from it increasing every year. Ischemic strokes, one of the two main types of stroke, occur when blood clots block brain arteries which leads to infarction eventually leading to brain edema. If not addressed quickly enough it may lead to disability and in worst case scenario may even lead to death.

In this thesis we proposed a machine learning based MATLAB tool that aids in speeding up the prognosis of acute ischemic stroke patients. From a set of patient medical data such as patient age, blood pressure reading and infarction volume from first CT scan, we created three prediction models which predict second infarction volume, decision for surgery and treatment time. We also experimented with utilizing the technique of feature reduction and implementing Fuzzy Inference System to consider improving the generated models and combined the best performing models into a MATLAB application.

DEDICATION

To the kind souled pillars of my life that have dedicated their lives to making mine a

smooth sailing one – my beautiful mother and father. I am forever indebted to them

and pray they are granted the highest level of paradise.

Also, to all my loved ones, including my family and friends, for their support and most

importantly for their ability to tolerate my missing-in-action episodes gracefully, well

at least for the most part of it!

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CT	Computed Tomography
ANFIS	Adaptive Neuro-Fuzzy Inference System
MRI	Magnetic Resonance Imaging
TPA	Tissue Plasminogen Activator
DC	Decompressive Craniectomy
DH	Decompressive Hemicraniectomy
ML	Machine Learning
IVGR	Infarction Volume Growth Rate
SVM	Support Vector Machine
ANN	Adaptive Neural Network
PCA	Principal Component Analysis
SVC	Support Vector Classification
SVR	Support Vector Regression
CDSS	Clinical Decision Support Systems
FIS	Fuzzy Inference System
FCM	Fuzzy C-Means
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MSE	Mean Square Error
KPCA	Kernel Principal Component Analysis

Chapter 1: Introduction

1.1. Overview and Motivation

There is a worldwide trend in the increase of the number of stroke cases reported every year. According to the World Health Organization, stroke is one of the leading causes of death in 2016 with approximately 5.8 million deaths reported in 2016 alone. While a stroke could happen to anyone, multiple risk factors such as age, hypertension, obesity and diabetes contribute significantly to its occurrence. Stroke, also known as brain-attack (similar to heart-attack), is of two main types – Ischemic, which occurs 85% of the time and Hemorrhagic, occurring 15% of the time [1]. Ischemic strokes occur when blood clots block brain arteries. This leads to infarction, a condition in which the brain cells end up dying due to the blockage since they are deprived of blood that is necessary for oxygen and nutrients transportation. Consequently, brain edema or swelling of the
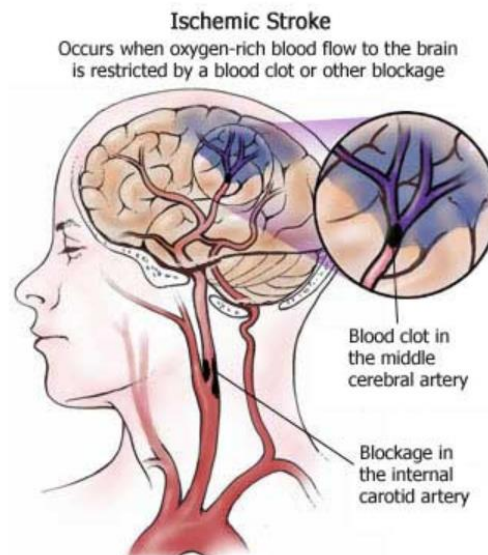
Figure 1. Illustration showing patient with ischemic stroke.

brain starts to occur due to the chemical imbalance caused in that area. If no immediate action is taken to address this, and given that the skull cannot expand, the built-up pressure will then be directed inwards towards the brain stem. This may result in the development of physical disabilities and, in worst case scenarios, may even cause death. Thus, early diagnosis and therefore treatment of stroke is of paramount importance.

Identification of a stroke is relatively easy - a sudden onset of acute neurologic symptoms that peak within a few minutes define a stroke attack in most cases unless proven false. While the initial identification of stroke is straightforward, a more detailed diagnosis, in terms of the type or intensity of the stroke, and the appropriate management depend on the clinical assessment of the patient history and results from a few physical examinations. This is due to the fact that the underlying symptoms vary depending on the region of the brain that is affected [1]. These physical examinations should necessarily include brain and neurovascular imagining, currently done by performing a non-contrast computed tomography (CT) scan of the head. A head CT scan when interpreted by an expert can lead to a major stroke diagnosis in about 70% of the cases in which signs of ischemia are evident. However, minor strokes could consist of a mere 10-12 small volume ischemic changes, changes so small, they are beyond the resolution of a normal CT scan. This makes CT scans highly insensitive to the confirmation of ischemia for the diagnosis of minor strokes. In such cases, Magnetic Resonance Imaging (MRI), an imagining process that is better equipped to detect brain ischemia in transient ischemic attack or in minor ischemic stroke due to its greater spatial resolution is used [2]. While MRI has greater sensitivity for small-volume ischemia observed in minor strokes, it can only be considered in situations that have no urgency for treatment and is typically used as a follow-up imaging.

Ischemic strokes are conventionally treated either by drugs, catheter

embolectomy or surgery depending on the severity of the stroke onset with drugs being the treatment choice for lesser form of stroke severity. Tissue plasminogen activator (TPA) is an example of drug-based treatment. TPA, given via intravenous therapy or IV, works by dissolving the clot and works to stop a stroke by dissolving the clot that is causing it thereby improving blood flow to the part of the brain deprived of blood flow. For TPA to be effective, it should be administered within a minimum of 3 hours and a maximum of up to 4.5 hours in certain patients when the symptoms first appear. Only in cases where drugs are not effective in dissolving blood clot and if the stroke is acute or localized to only one area, then the doctors may perform catheter embolectomy. In this, they use a catheter to thread through the patient's blood vessels to reach the area where the clot is lodged at and then manually remove the clot. A large stroke, however, can lead to swelling of the brain requiring the need for surgical intervention. The surgical response to acute ischemic stroke is Decompressive Craniectomy (DC). DC is a neurosurgical procedure wherein a portion of the skull is temporarily removed so as to relief high intracranial pressure that often occurs in traumatic brain injury patients, in patients that get raised intracranial pressure or in stroke patients. By effectively increasing the volume that the brain can occupy and thereby allowing increased cerebral blood flow and tissue oxygenation, DC helps minimize the damage caused by ischemia. DC is thus considered a lifesaving intervention that effectively decreases mortality and often improves outcome especially when performed in the early period of the disease [3]. When DC is performed in the region of large cerebral hemisphere mass, this procedure is termed Decompressive Hemicraniectomy (DH). If the swelling is not quickly addressed, by not performing DH fast enough, high intracranial pressure building up within the skull may swiftly lead to more serious consequences which could range anywhere from secondary brain damage to permanent neurological damage or

even death [5].

At least one half of major ischemic stroke patients arrive more than 4.5 hours after stroke onset making them illegible for thrombosis [4]. In most cases, this delay is usually caused due to the inability of nearby people to recognize the stroke symptoms of the patient when they undergo stroke. Add to that stroke renders a patient unable to make the emergency call themselves. While identifying a stroke is easy, as mentioned earlier, confirming if a stroke is ischemic or hemorrhagic necessitates the need for imaging which requires time. According to one study, a median time of 60 minutes is required for a stroke patient to be treated after reaching the hospital for treatment [5]. This estimate does not include the delay already caused as a result from the delayed call for help from the patient side. This is alarmingly high given on average for every minute of delay about 1.9 million neurons, 14 billion synapses and 12 km of myelinated fibers are destroyed thus deteriorating an already life-threatening situation. Research is currently undergoing to reduce door to treatment times for stroke patients. Some are suggesting placing of CT scanners in ambulances. This will allow for rapid imaging followed by remote transmission of the imaging data for review by doctors, which will help in the confirmation for thrombolytic process and in its administration before the ambulance reaches the hospital [6]. Remarkably, half or more than half of all ischemic stroke patients are not offered thrombolysis. This is because their symptoms are not considered disabling [1]. These cases must then be managed with more aggressive treatment techniques in order to prevent progression to a major stroke.

In this thesis research we have attempted to address this issue of time criticality by suggesting a machine learning (ML) based approach to determine the appropriate management plan for medical experts. By predicting the different necessary deciding factors for treatment plan, we hope to simulate the ongoing research by suggesting a

fast method that only depends on one CT scan and results from other patient clinical data. Multiple works in the literature address prediction of stroke and its corresponding diagnosis using ML Techniques [7][8][9]. While some address the prediction of infarction volume growth [10], others yet offer a prediction model that determines the outcome of the treatment procedure [11]. To the best of our knowledge, no such tool exists for clinicians' use which includes a prediction model that predicts the infarction volume and determines the adequate treatment plan from the combination input of one CT scan and other clinical data. With the presented thesis, we hope to propose an application that can do so.

The rest of the thesis is structured as follows. For the current chapter, the upcoming sections will offer a detailed description of the thesis objective and goals. This include description of the problem statement in Section 1.2, the objective of the thesis in Section 1.3 and its contribution to the research community will be highlighted in Section 1.4. Chapter 2 details both the relevant preliminary concepts utilized in this thesis and research works present in the literature that address similar focus. Methodology adopted is explained in Chapter 3 while Chapter 4 evaluates the implemented methods. We conclude and offer possible directions for future work in Chapter 5. Published works resulting from this thesis are listed in Chapter 6.

1.2. Problem Statement

The main problem we intend to address in this thesis is prediction related to acute ischemic stroke patients. These predictions are expected to help medical professionals have an idea about the current status of the patient in terms of infarction volume, decide if an ischemic stroke patient needs to undergo DH and if so when to do so. We want to achieve this by determining the possibility of using supervised ML

techniques to accurately predict from a set of clinical data attributes these specific outcomes –

    a. Infarction Volume Growth Rate (IVGR) at a given time.

    b. The decision if DH needs to be performed or if it is possible to balance the chemical buildup in the brain by other less risky options such as medicine.

    c. If DH needs to be done, when it should be done. This is a categorical based output in which output values 0, 1, 2, 3 categorize to 24 hours, 48 hours, 72 hours, 96 hours after the CT scan respectively.

In addition to this, we want to experiment with the utilization of feature reduction technique in this scenario and evaluate the prediction models obtained as a result of utilizing it. In essence, we want to examine and understand if the utilization of feature reduction technique helps in improving the different output predictions. Lastly, we would like to study if there is any difference in performance when one of the predictions is made by integrating a fuzzy inference system as opposed to predicting by ML techniques alone.

1.3. Thesis Objective

The aim of this master thesis is to investigate and evaluate the use of ML techniques to predict the necessary parameters needed to help medical professional make a more informed decision. The main decision we aim to tackle is the decision on whether to proceed with surgery from just one CT scan. The supervised learning techniques that will be considered are Support Vector Machine (SVM), Adaptive Neural Network (ANN) and Adaptive Neural Fuzzy Inference System (ANFIS), a ML technique that combines neural networks with fuzzy inference system. The best

performing of these models for each of the three outcomes will then be integrated in a MATLAB application for medical professionals as an easy to use tool that will offer prediction to necessary insights that relate to prognosis of acute ischemic stroke.

1.4. Thesis Contribution

While certain well-known symptoms that are observed in all patients contribute to the diagnosis of stroke, determining the appropriate treatment plan is usually dependent on the intensity and the elapsed time between stroke onset and patient treatment time. In this thesis we attempt to define a complete system that will aid doctors in determining the most suitable treatment plan. This is accomplished by predicting the infarction volume, the medical treatment procedure to be adopted – which could either be surgery or not (medicinal based option), and if surgery, the time at which the patient needs to be operated upon. By proposing such a system, we contribute to automating the process of determining treatment plans.

Another contribution of this thesis is in its addition in the domain of use of ML techniques in medical settings, in particular, in stroke. Unlike in most other similar works in the literature wherein prediction is made for the diagnosis or detection of a disease, we are for this thesis attempting the prediction of severity of an existing condition, which is argumentatively more complicated than most other diseases addressed in the literature. Complication, in our case, arises from the fact that unlike in most other medical conditions already addressed in the literature that usually have a larger sized offline dataset available to model their proposed solution on, the dataset available for stroke is on the other hand very limited. Also unlike in other medical problems as those discussed in Section 2.2 that deal with diagnosis or detection of a medical condition that can be done over a period of time, prediction of infarction

volume on the other hand, is very time critical. This is because, the faster this rate of growth of the infarction volume is ascertained, the more chances the patient has in reducing health complications such as disabilities or death arising from the delayed treatment. We hope to speed up the process by modelling an application model that requires only one CT scan among other easily obtainable patient medical data inputs.

Chapter 2: Background and Related Work

Before we explain the adopted methodology and the obtained evaluation results, we will first need to review the different relevant preliminary concepts adopted in our work and offer a brief overview of the different similar works in the literature. This chapter is therefore dedicated for that. We have split it into main sections; Section 2.1; in which we offer a brief description of the adopted concepts and Section 2.2; wherein we discuss the different relevant research done in the literature. In the latter section, we will also examine the similarities and differences of our thesis to the discussed works.

2.1. Background

2.1.1. Principal Component Analysis

Presence of imbalanced classes is an issue widely seen in medical datasets [12] [13] [14]. Another widely documented issue is the unfavorable feature-observation ratio. This happens when a small dataset has a large number of features thereby resulting in a model that is unable to learn effectively from this dataset causing a detrimental effect on the performance of the classifier [15]. Principal Component Analysis (PCA) is used as a feature reduction technique that reduces the dimensionality of a dataset with minimal loss of information. It is a statistical technique that is used for finding patterns in highly dimensional data or datasets that usually involve a large number of features as seen in medical and hyperspectral images-based datasets. The underlying assumption of PCA is that most of the information of the classes in a classification task is contained in the directions where the variations are the largest. Feature reduction is usually done by creating standardized linear projections that result in maximizing the variance in the projected space, consequently, reducing the number of variables which helps to process the data more quickly and effectively [16]. Usually,

this variance is represented in percentages. These projections, also known as principal components, are linear combinations of the original variables. Depending on how much of the information is needed to be preserved of the observation vector, the subspace spanned by the appropriate number of principal components is selected [17]. Given medical datasets are usually small in size and contain a large number of features, performing PCA on them then becomes a necessity to improve the training model generated.

PCA is also recommended as the initial stage of the multivariate analysis. It provides an initial outlook at the structure of the data, defines the overall data space spread, and helps in identifying outliers [18] which can then be eliminated. This is especially important in medical datasets that have a large number of unrelated variables and in some case have outliers present in the dataset which further confuse the predictor and thus end up reducing its accuracy as a result.

2.1.2. Support Vector Machine

Support Vector Machine (SVM) being a discriminative classifier, is formally defined by one or more separating hyperplanes. Given a labelled training data, the SVM algorithm outputs the most optimal set of hyperplanes that best categorizes the new samples. For a two-dimensional space, considered a linear based separation, the hyperplane is defined as the line dividing a plane into two parts where each class lie on the either side of the line. There are two main types of SVM: support vector classification (SVC) and support vector regression (SVR).

SVM is popularly used in multiple domains from image recognition to text classification to bioinformatics, and so on. Its strength lies in its ability to offer robust performance in less adequate scenarios such as those involving sparse or noisy data.

This makes SVM the ML technique of choice in many applications [19]. Initially developed to classify training data without errors, SVM was extended to support classifying training data with errors making it as powerful and applicable as neural networks [20]. Add to this, SVM can also handle nonlinear classification effectively. It does this by using a nonlinear kernel function to map samples from low dimensional input space to high dimensional feature space [21]. Even though the training time of the fastest SVMs can be very slow, they are exceptionally accurate and powerful in their ability to model complex non-linear decision boundaries [22].

SVR is another version of SVM that is meant for regression. The performance of this version of SVM crucially depends on the shape of the kernel function and other hyper-parameters that epitomize the noise distribution characteristics in the training data. In this proposed solution, we utilized SVM in both classification and regression prediction as explained in further sections.

2.1.3. Artificial Neural Network

Artificial Neural Network (ANN) is inspired by the functioning of the human brain and replicates the way humans learn. It is a ML technique that comprises of a set of networks called neural networks which consist of input and output layers as well as a hidden layer that transform input to output. While the concept of neural networks, or perceptrons have been around for a very long time now, only recently have their use become a major part of artificial intelligence. This is mainly due to the recent introduction of the concept of backpropagation. Backpropagation allows the networks to readjust the hidden layers of neurons in situations where the output is not as expected thereby strengthening the learning and therefore prediction ability of the training model.

The architecture of a typical ANN can be visualized as weighted directed graphs

with artificial neurons as nodes and directed and weighted edges as connections between the output and input neurons. When considering the connections, ANN can be characterized into two categories – feedforward networks and feedback (recurrent) networks. In feedforward based ANNs, the graphs represented by the connections have no loops. While those in feedback networks have loops due to occurrence of feedback connections. The output from a feedforward network is static since they produce only one set of output values rather than a sequence of output values from a given input with no feedback correction option. This makes feedforward network memoryless given their response to input does not depend on the previously modified network state. Feedback networks on the other hand are dynamic systems since new neuron outputs are generated when a new input pattern is presented and given, they possess feedback paths, the input to each neuron can then be modified resulting a new network state [24]. In this paper, we only considered feedforward ANNs.

2.1.4. Adaptive Neuro Fuzzy Inference System

Fuzzy logic is a computing approach that is based on degree of truthfulness rather than crisp true or false values, commonly known as Boolean logic that modern computers are based on. The use of Fuzzy logic is highly recommended in modelling scenarios that are characterized as inherently imprecise or vague. This is due to their ability to consider vagueness in their modelling. Due to this and given that neural networks have the ability to learn from the environment, self-organize their structure and adapt to it in an interactive manner, as a result, incorporating a fuzzy logic controller into a neural network results in a powerfully flexible learning tool. ANFIS is one such technique that combines these two. By having an adaptive neuro fuzzy system that can self-organize its structure according to the environment, we are able to combine

the advantages of both systems.

As mentioned earlier, ANFIS is a ML technique, the core of which is an ANN based on Takagi-Sugeno Fuzzy Inference System. It is implemented by an underlying framework of adaptive networks. One of the advantage of using ANFIS is its ability to allow fuzzy rules to be extract from numerical data or expert knowledge and adaptively construct a rule base on that [25]. Because it uses a hybrid learning procedure, it is able to construct an input-output mapping based on human knowledge, done using fuzzy if-then rules and input-output data pairs [25], [26]. It is in this if-then rules where the medical knowledge of the medical practitioners can be included. Thus, ANFIS combines neural networks along with injected expert heuristics to build a model.

The architecture and the learning procedure of ANFIS are grounded on an adaptive network which is a feedforward based neural network with supervised learning capability. It consists of nodes and directional links that connect these nodes. Some or all of these nodes could be made adaptive thus making the outputs depend on the parameters belonging to these nodes. The learning rules define how these parameters are changed to reach a pre-defined minimum error. It is therefore due to this customization, learning and relearning by going through multiple passes and its consideration of expert knowledge that makes ANFIS an applicable ML option in the medical domain. It is also due to these reasons one would expect it would produce a more realistic functioning model.

2.2. Literature Review and Related Work

In order to offer context to the contribution of this thesis and to have a touch of understanding of the different relevant works in the literature, this Section will offer a brief explanation of a selected relevant papers and offer a linkage to how this thesis

learns from them, compliments them and addresses some of the issues and limitations existing in them.

2.2.1. Use of Machine Learning Techniques on Medical Data

Medical diagnosis has traditionally been done entirely manual without the usage of any prediction-based computer programs. This involved having medical doctors look into different medical images and medical readings before deciding the prognosis of the patient. While, to a large extent this is still a common practice, recent developments, such as the revolution of digitization of data in the last few years have provided inexpensive means to collect and store data and have made inclusion of computer incorporated techniques in the diagnosis and treatment of different medical diseases more common. These are often done by computer science-based techniques such as ML techniques as presented in [27], [28]. As a result, modern hospitals are equipped with devices that collect and maintain medical data of their patients. With the advancement of ML techniques, a lot of advancement has been made in their ability for use in analyzing medical data, in particular, in medical diagnosis [29].

There are multiple favorable features that a ML technique needs to possess in order for it to be useful in medical diagnosis tasks. One of which is that it should offer good performance. This is so that significant information is extracted from the available data to the extent that the classification accuracy of the classifier should be at least as good as that of the physicians when offered the same description of the patients. Another feature is the ability of the technique to deal with missing or noisy data given medical data often lack perfectly accurate patient description data of all their patients in their database. A ML technique that is also able to obtain reliable diagnosis in minimal number of tests is another favorable feature. Medical datasets are usually small

sized. This is caused by minimal number of willing participants in the data collection phase. This inherently results in small datasets. It is well known that most ML techniques in principle do not perform well for small datasets. Therefore, a technique that can reliably diagnose with a small amount of data is desirable or necessary.

ML techniques have been used to predict and diagnose multiple diseases in the literature. Of these, SVM and neural networks are the most popular techniques used in medical applications [30]. A similar observation was also noticed in neurological based prediction models. This was confirmed by a study that included 30 citations of publications that evaluated the application ML models in neurological outcome prediction. ANN and SVM were found to be the most frequently used algorithms [31]. SVM is commonly used due to its excellent generalization ability, its ability to perform nonlinear function fitting and due to its ability to perform well with small sample studies [32]. The strength of neural networks lies in their ability to learn from examples. For a neural network to perform well, a set of representative examples is required, making it about quality over quantity [33]. There are also other proposed systems that utilize Adaptive Neuro-Fuzzy Inference System, a technique that is a combination of fuzzy inference system and Adaptive Neural Network, as their underlying system.

Although there is an increasing interest in the automation of prediction of medical diagnosis and treatment procedures in the research world, it is neither appealing to the medical practitioners nor to the general public. One of the reasons is the issue of anxiety caused due to the shift of medical treatment procedures from process performed entirely by humans to the automation of the process by the use computers, robots or Clinical Decision Support Systems (CDSS). CDSS are medical systems that offer decision support to the physicians. Research has shown that many psycho-social reasons contribute to resistance shown by physicians to their use. In one study it was

reported that assumed high implementation cost and staff training requirements were some of the reasons physicians were reluctant to use these systems [34]. There is also the problem of concern of privacy not being maintained that often further clouds the automation of the diagnosis process. The medical opinions and diagnosis of a human being as opposed to that presented by a non-human is often preferred, trusted and accepted by the general public. Add to that the knowledge that no amount of training can help machines make as accurate of a decision for the different situations and previously unseen patterns that may be seen on medical conditions. Although it is well proven that the diagnosis decision made by an expert is more trustworthy, the use of classification techniques would nevertheless help in the decision making [17]. The negative attitude towards the automation of diagnosis can be expected to get better when more people start to see the exceptionally good prediction outcomes of some of these models. There is indication in the literature that this is in fact the case. Studies have shown that some ML models have demonstrated to outperform classical statistical based models, prognostic indices and in certain conditions show similar or better performance when compared to clinical experts [31]. Due to these findings, and the fact that ML based models are comparatively faster in decision making process, makes implementing their use in the diagnosis of stroke even more crucial.

2.2.2. Issues of unbalanced classes in Medical Datasets

While there are many advantages to the utilization of ML techniques for the prediction in medical setting, obtaining a good prediction model can be complicated. These are mainly caused due to the issues pertaining to medical data; in particular their inclinations to being small sized and their tendency to be imbalanced in nature. The former issue is caused due to minimal number of willing participants in the data

collection phase inherently resulting in small datasets. ML techniques in principle do not perform well for small datasets [35]. Also, existing classification methods perform badly in unbalanced datasets [9]. This is because ML techniques optimize their overall performance equally on all classes in the dataset instead of considering relative distribution of each of the classes. This is problematic as many datasets are skewed in nature and do not have an equal distribution of class observations. While skewed datasets are very common in face recognition and gene expression data analysis [8], it is also commonly seen in medical data analysis [36]. The usage of unbalanced datasets often results in commonly known issues of over-fitting and under-fitting in the generated ML models. Understandably, datasets that are small and have an unbalanced set of observations perform even more poorly. Therefore, in order to create a good prediction model, a well-balanced dataset is of prime necessity. Unfortunately, since most medical datasets are usually unbalanced with some class labels represented more than the others, models that are trained on this type of datasets end up having a low accuracy rate mainly due to their tendency to mis-classify the minority class observations to belong to the majority class. This inherently results in increasing the false negative rate. While this could be problematic in any scenario, it could have major implications when the misclassification occurs in a medical prognosis setting. Misclassification of the prognosis of a patient could result in a patient receiving wrong treatment which could in worst case scenarios be lethal. Therefore, extra precaution needs to be taken to seriously consider this issue.

To solve the problem of misclassification caused due to imbalanced classes within a dataset, two categories of techniques have been proposed – sampling-based approaches and algorithm-based approaches. Sampling techniques, the more popular technique, handles data skewness by simply re-balancing the data at the initial stage

specifically at the data preprocessing stage. These include methods that either over sample the minority represented class in dual class dataset to match the size of the majority represented class or methods that under sample the majority class to match the size of the minority class. Due to the simplicity of this technique, it can easily be deployed on top of many of the classification approaches. However, sampling-based techniques also have detrimental effects. Under sampling of the majority sized class may result in the majority class instances loosing potentially useful information against the negative class. Whereas over sampling of a positive class by creating redundant duplicates and training a classifier on that will result in the problem of overfitting. Considering the detrimental results, training models that incorporate over sampling techniques have been found to better perform compared to when under sampling techniques are applied. Algorithm based techniques on the other hand focus on improving the classifier's inherent characteristics instead of tampering with the dataset to eliminate the problem of misclassification of the data. Depending on which ML technique is used – Decision trees, neural network and so on, there is a specific algorithm based technique tailored for each of these techniques [37].

In our case, in order to have a good-sized dataset, we had to collect and compile patient data from a pooled Decompressive Hemicraniectomy database, the constituents of which were obtained from three referral centers in three different countries namely Qatar, UAE and Pakistan. Only patients with three CT scans of the brain and having evidence of acute ischemia were considered resulting in a dataset of 140 records and 150 features for the main dataset. We were also able to collect 203 other records of patient data with 313 features resulting in a respectable sized dataset.

2.2.3. Feature Space Selection Techniques

In addition to their small size, another issue pertaining to medical datasets is the large number of features resulting in small and highly dimensional dataset. Having too many features in a dataset can affect analysis efficiency and in some cases the modelling accuracy [35]. There are two widely known dimensional reduction methods, that can help overcome this issue. In the first one, it is done by selecting a relevant feature subset from the original subset, while in the second method, known as feature extraction, the original features are extracted into a new feature space. In the latter method, the new extracted features are better able to describe the pattern space when compared to the original features. The former method is usually done manually without the use of any known algorithms or techniques. People well aware of the data content and are experts in the represented field of the dataset usually decide which of the features are significant and which are not. However, in the case of the latter, these techniques usually perform the dimension reduction process by projecting the original features into lower feature space to reduce data dimensionality and thereby improve data analytical efficiency. This is normally done in two steps; The first being the extraction of relevant information for classification from the raw data represented by the original feature vector of all its dimensions and the second step is the creation of new feature vector from that with smaller number of dimensions. PCA, independent component analysis and canonical correlation analysis are some of the commonly used methods of feature extraction in the literature, of which PCA is the best known and most widely applied linear feature extraction method [35].

Given our main dataset consisted of 150 features and the second dataset consisted of 313 features, feature reduction was required to make them both more balanced datasets. For each dataset we performed an initial round of dimension

reduction by selecting the most relevant subset of features, deemed relevant by discussing with medical practitioners. This resulted in both datasets having 14 features each. Although our dataset was acceptably balanced after the initial feature reduction process, we experimented with the possibility of applying feature extraction on a dataset to see if this would create a much better prediction model. We decided to use PCA in our experiments given it is widely applied technique as mentioned earlier.

2.2.4. Use of Feature Reduction Techniques in Medical Prediction

While some works in the literature apply ML techniques to their dataset to create a prediction model, others first apply feature selection techniques such as those mentioned in the Subsection 2.2.3 to the dataset to enhance the resulting prediction models. By applying feature reduction techniques like PCA and predicting the output from the resulting reduced dataset they state that their prediction models are able to perform better prediction than most contemporary ML prediction models for the same problem.

One such work is that which is presented in [38]. In this paper they present a system that is able to diagnose heart valve diseases. In this work, a Doppler ultrasound first measures doppler signal waveforms at the heart value. The wavelength entropies it obtains as a result are used as features in a combination of feature extraction and classification phase. Their solution has three main phases – a pre-processing phase, feature extraction phase and a classification phase. In the pre-processing phase, two main steps are performed – data acquisition and the pre-processing of doppler signals obtained from the ultrasound. Then in the feature extraction phase, for every signal, 12 wavelength entropy values are calculated to create the feature vector. These extracted 12 features are then reduced to 6 features by the use of PCA. At this stage, the 6

extracted features belonging to the reduced wavelength entropy feature set are then utilized in the next phase of the solution which is the classification phase by inputting them into the ANFIS classifier. Their proposed solution obtained classification accuracies of 96% and 93% for normal and abnormal subjects respectively for a sample size of 215 samples. Similarly in [16], instead of utilizing ANFIS for heart related illness assessment, they utilized SVM. In this, by using PCA they were able to reduce the study dataset consisting of 480 patients with 23 features to 18 features and obtain an optimal SVM model that was better able to diagnose coronary artery disease. The obtained diagnostic accuracy was 79.71% which was an improvement from 76.67% when no feature reduction performed on the dataset. They also noticed that the use of PCA decreased the training error and the total training and testing times.

Yet in another work [17], they use a ANFIS-PCA combination to improve the diagnostic accuracy of diabetes. By using their system, they were able to reduce the number of features in their dataset from 8 to 4 and apply that reduced dataset to the ANFIS classifier. They obtained a classification accuracy of 89.47%. Unlike in this previous work [38] where they perform only one pass in the classifier, in this work, they perform k-fold passes in the classifier. While their accuracy is reported to be better than others reported from different classifiers such as KNN, Bayes and CART, however used without any feature selection procedure, it is nonetheless less in accuracy when compared to the initial discussed work. This is could either be due the low features-observations ratio in this work as compared to that presented in [38].

Another work [39] also use the same combination of ANFIS-PCA in addition to that they pre-process their dataset using fuzzy weighting pre-processing for the automatic detection of diseases related to lymph system data obtained from lymphography test results. Their proposed solution was implemented in three stages –

the feature reduction stage in which 18 features were reduced to 4 features, then a pre-processing phase in which a fuzzy weighting scheme was used and finally that dataset was fed into an ANFIS classifier. The obtained accuracy was 88.83% and was better than other classification results presented by the other works in the literature for the same problem. The input part of their dataset consisted of 148 instances with 18 numerical attributes. While the output could be any a four class which were numeric – 1, 2, 3 and 4 representing normal, metastases, malign lymph and fibrosis respectively. Different to this work as compared to the previous ones was that in this one, each feature was given a new value that is derived from its old value using fuzzy logic. Since fuzzy logic is used, the concept of membership functions is utilized, and two triangular membership functions are defined for this – input and output membership functions for each of the feature in the dataset. These functions are realized by utilizing the mean values of each of the features. The output, that is the transformed features, are then fed into a 5-layer ANFIS having 2 layers of adaptive nodes and the rest fixed nodes. Because ANFIS is based on adaptive system framework that facilitates learning and adaptation, this therefore made it more systematic with a little dependency on expert knowledge.

Given we have an argumentative small training dataset consisting of only 73 records for training, we will apply feature reduction technique of PCA and study if that will have a positive impact in the prediction models generated.

2.2.5. Machine Learning techniques in Stroke Prediction and Diagnosis

Just as the use of ML techniques is not new in medical settings, accordingly, the use of ML techniques in the prediction of various aspects of stroke patients is also not new. Most use either pure images or a combination of images and medical data as

opposed to only medical data for prediction as done in this thesis. Several works such as those discussed in [40][41][42] only use medical images such as CT scans in order to train their models. Yet others, such as those discussed in [43][44], use a combination of medical images along with other clinical data in order to train the model. However, works that use only clinical data such as proposed in our work are very limited in number. This limitation is further confirmed in the survey paper of Jiang et. al [30] in which they discuss how most of the papers in the literature use ML and deep learning based techniques for medical image analysis more than they do for medical data analysis. For instance, one such work that uses ML on only medical data is found in the work proposed by Asadi et. al [45]. They modelled and compared two supervised ML techniques – SVM and ANN for stroke analysis. Their models considered multiple inputs, including risk factors of the patients to predict the expected outcome of endovascular intervention performed on patients with acute anterior circulation ischemic stroke in terms of mRS score. They trained, validated and tested their models on randomly divided data from a dataset of 107 records. They concluded that SVM had a better performance, when compared with the performance of ANN. The overall precision of their proposed system was 87% while their model accuracy was approximately 70%. On the other hand, an example of prediction model that combines medical numerical data along with medical image data is that proposed by Bentley et. al [43]. In this, they trained an SVM model that would take in, as input, raw CT scan images along with other clinical information and as output predict if the stroke patient is at risk of Symptomatic Intracerebral Hemorrhage, a complication that affects patients who are treated by intravenous thrombolysis. To validate their work, they used k-fold cross-validation by splitting their dataset into 106 and 10 training and testing sets respectively for over 1760 repetitions. They compared their work to the outcome of

another work that used radiologists' interpretations of the CT scan images (along with clinical variables) in their model, rather than using raw images and found theirs to outperform the latter in terms of predictive performance. Another version of inputs is that in which they consider symptoms along with numerical inputs as proposed by Mirtskhulava et. al [46]. They modelled a neural network model having two hidden layers which accepted 16 inputs to predict the risk of stroke in the examined patient. These inputs included a combination of patient symptoms and stroke related risk factors. Their proposed system had a binary output, which predicted whether or not the patient would suffer a stroke. Therefore, an output of one meant the patient is at a risk of stroke, while zero meant otherwise.

Predictions of stroke infarction volume using medical data of patients and only one CT scan is not unique to our proposal. There is one such other work that does so in the literature. The work proposed in [10] addresses the prediction of infarction volume growth rate. In this, they experimented with using ANFIS, an algorithm that combines ANN and Fuzzy Inference System (FIS) to investigate infarction growth pattern and use that to predict infarction growth rate and infarction volume at a given instance for stroke patients. They focused on patients that had large vessel occlusion in their anterior circulation. Prediction using this method had not been reported previously and appeared to offer promising results when compared with other models used in the literature. Their proposed solution was able to predict the infarction growth rate and infarction volume of the third CT scan without any statistical difference when compared to the ground truth (P = 0.489). While it is possible to effectively predict the infarction volume of a stroke patient, it can nevertheless be an immensely dispersed scenario. This is due to the fact that it not only depends upon the patients' current conditions but also on their conditions at the time at which the CT scans were performed, which in its self is another

variable with a big variance connected to it. Interestingly, it has been found in [10] that the rate of volumetric growth of an infarction remains fairly consistent for a sizeable time duration, and any volume can be predicted with reasonable accuracy within a predetermined set time span.

Other works such as those presented in [11] developed a binary logistic regression based prediction model that predicted the prognostic outcome in terms of functional dependence and death after 100 days for acute ischemic stroke patients. They trained their model on 1754 records of patients. They used Barthel Index as the clinical scale for measuring disability after stroke. They assumed functional independence as any value equal to or above 95 in the of Barthel Index scale after 100 days. Their internally calibrated and cross-validated models correctly classified over 80% of the patients.

However, no work, as far as we are aware, has combined all of the prediction outputs we proposed in this thesis to create an application that aids in the prediction of the best treatment plan in ischemic stroke patients. We are addressing this gap in the literature by proposing an application that accepts input in the form of medical data of the stroke patient and the infarction volume size obtained from one CT scan to offer a prediction of different parameters that aid in determining the appropriate treatment plan.

Chapter 3: Methodology

This chapter offers a detailed explanation of the steps implemented in this thesis in order to complete the requirements. Figure 2 offers an overview of the implemented methodology. As explained earlier and as shown in the figure, the methodology can principally be divided into 3 main steps. The first step consists of the creation of the prediction model, where in ML models, with and without feature reduction applied to them, are modelled for each of the three outputs. In the second step the use of FIS is experimented with, mainly to see if it is able to offer a better prediction model for the time of surgery output. The input of the FIS, as seen in the figure, is selected from the best performing models from the rest of the two outputs. The last step is the creation of the MATLAB application wherein the best prediction models for each of the outputs are utilized to create a real-world application.

3.1. Data Collection and Preparation

The initial step was the collection of data and preparing it for use. How this was done is further discussed in the succeeding Subsections.

3.1.1. Datasets

For this thesis we utilized two datasets. For both datasets we only considered patients with evidence of acute ischemia and who have undergone at least three CT scans of their brains. The first dataset presented to us by the doctors contained about 140 records having 150 variables and the second dataset was much bigger containing 203 patients having 313 features. While the second dataset contained most of the features present in the first, it also consisted of patient data that had unsure values for some features namely Collateral Score, METS, Clot Burden Score and Modified TICI

Score. While understanding this is a limitation, we nevertheless decided to utilize it as a testing dataset given this was the closest dataset we have to our main dataset.

3.1.2. Preprocessing

Once the dataset was obtained, the first step to the process was to prepare it for use. This was carried out in two main stages. In the first stage feature reduction and null values removal was carried out. The second stage consisted of normalization of the data and the creation of training and testing sets.

In the first stage of preprocessing, the first step was to reduce the number of features in the dataset by using feature reduction techniques. As mentioned earlier this was done by consulting with medical experts and looking into the literature what features would mostly likely aid in stroke related prediction. By doing so, the number of features in both of the resulting datasets were reduced to 14 features. The remaining features included relevant medical details such as the patient age, if they are diabetic patients or not, their HbA1c readings, among others. In addition to these, the initial infarction volume rate was also considered and included in the list of features. A detailed description of the features present in the datasets and their respective median and standard deviation values is as presented in Table 1. The next step was to remove the null values. So, any rows with any null values were removed resulting in 122 records in the first dataset and 18 records in the second dataset. The datasets were at this point ready for the next stage of preprocessing.

In the second stage of preprocessing, each column of the dataset was normalized individually. As a result, each column would store its values within the range of 0 to 1. The data was then split into training and testing set each of which consisted of indices

picked at random. This random selection of training and testing set of indices was performed 100 times. The most statistically similar training and testing sets were selected from these 100 sets. Our definition of statistically similar was that the corresponding columns in the training and testing sets had the least difference in terms
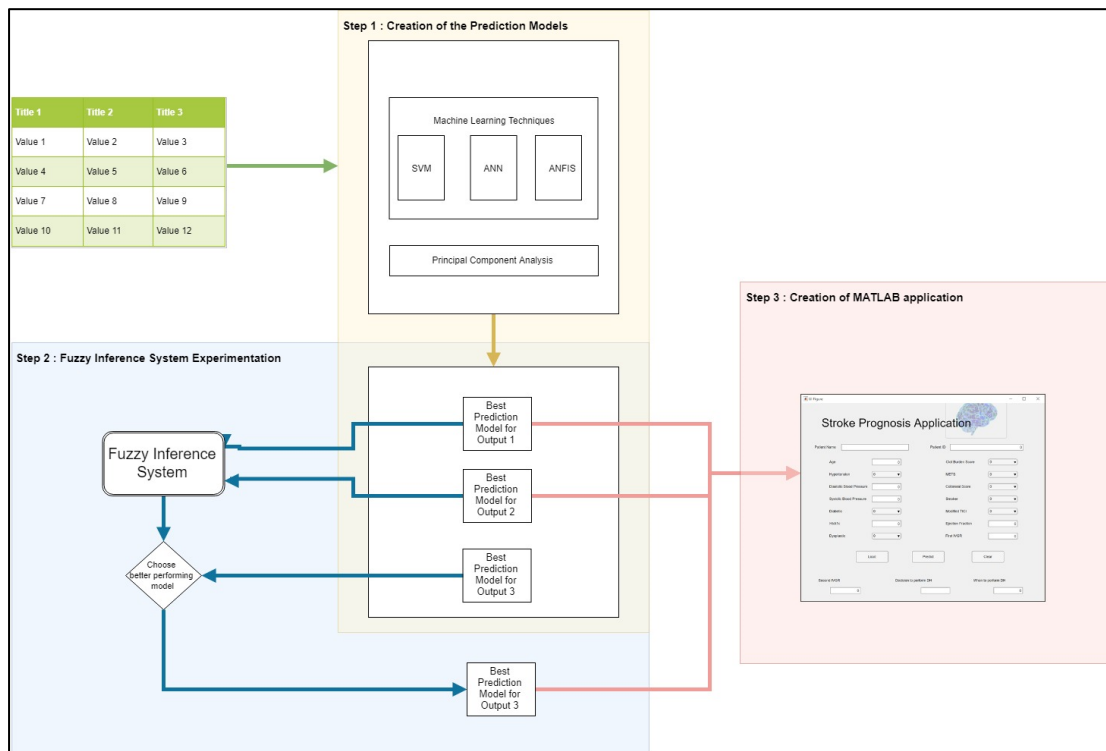
Figure 2. Methodology overview.

of standard deviation and median. By opting for statistically similar testing and training sets, we attempted to create a more representative training and testing sets thereby setting up a more realistic and fair evaluation of the training models generated. This thus resulted in a randomly split yet statistically similar training and testing sets.

Once that was done, 60% instances of the original dataset were used for training, while the rest of the 40% was used for testing. The dataset was now ready for the next step of the process, namely classification.

Table 1. Details of Features Included in the Dataset 1 and Dataset 2

Feature	Description	Value	Median ± Standard Deviation	
			Dataset 1	Dataset 2
Age	-	In years	50 ± 12.614	52 ±14.571
HTN	Hypertension diagnosis	0 – Absent 1 - Present	1 ± 0.502	1 ± 0.514
SBP	Systolic blood pressure reading	-	138 ± 13.441	150 ± 35.968
DBP	Diastolic blood pressure reading	-	77 ± 9.4	89 ± 21.191
DM	Diabetes diagnosis	0 – Absent 1 - Present	0 ± 0.492	0 ± 0.507
HBA1C	HBA1C reading	5.4 to 16.5	2 ± 0.869	9.2 ± 74.84
DYS	Dysplasia diagnosis	0 – Absent 1 - Present	0 ± 0.500	0 ±0.493
Ejection Fraction	Time taken to pull out the blood clot.	In minutes	50 ± 9.982	38 ± 7.617

Feature	Description	Value	Median ± Standard Deviation	
			Dataset 1	Dataset 2
METS	Metabolic Syndrome – a combination of multiple conditions that increase the risk of stroke and that are combined into one score.	0 - All absent 1 - At least one present	1 ± 0.500	0 ± 0.514
Clot Burden Score	This score represents how much pressure the clot will be causing in the brain area where it is blocking the flow. The score depends upon length, density, volume, location, twisted or straight and chemical level.	0 – 10	8 ± 1.512	6 ± 2.781
Collateral Score	Temporary blood vessels that replace the blocked blood vessel.	0 – 3	2 ± 0.823	0 ± 0.485

Feature	Description	Value	Median ± Standard Deviation	
			Dataset 1	Dataset 2
Modified TICI	A score that determines the response of thrombolytic therapy for ischemic stroke.	0 – 3	2 ± 1.742	0 ± 0.470
Smoking	States if the patient is a smoker	0 – Non-smoker 1 - Smoker	0 ± 0.411	0 ± 0.507
IGR1	First infarction growth rate per hour	-	2.778 ± 15.879	69.1 ± 65.663

3.2. Classification

Once we preprocessed the data and prepared the training and testing set, the next step was to model the classification models. We primarily implemented our proposed prediction models in 3 ML models - SVM, ANFIS and ANN and then applied feature reduction technique of PCA to these models which resulted in 6 prediction models for each of the outputs. We started by first attempting to model for IVGR2. For the prediction of IVGR2, our initial experimentation was inspired by the work presented in [10] in which they perform similar stroke-based predictions. We extended their version by applying feature reduction technique of PCA; details of whose

implementation to be explained in Subsection 3.2.1, on the dataset and then the second infarction growth rate was predicted from the reduced dataset using ANFIS. Consequently, we were able to model a better performing model with a lower RMSE. More details on our proposal and methodology that can be read in [47]. We then implemented the rest of ML techniques before progressing to the other outputs. The performance of each of these models is discussed elaborately in Chapter 4.

Given prediction of the second infarction volume growth rate and treatment time are both regression problems, while predicting if a patient needs to undergo DH is a binary classification, we modelled both regression-based prediction models and binary-based classification models.

For SVM, we selected the MATLAB based SVM classifier FITCSVM to train an SVM model for binary classification. Since this classifier is intended for use in low or moderate dimensional predictor dataset, we utilized it in both of our datasets. FITCSVM supports mapping predictor data using kernel functions, we selected the linear kernel function to be underlying function. In the case of regression predictors, we utilized FITRSVM, to fit an SVM regression model.

In the case of ANFIS, because there was only output, we utilized the MATLAB function GENFIS3 to generate an initial FIS for ANFIS training. Using Fuzzy C-Means (FCM) clustering, the function GENFIS3 generates an FIS by extracting a set of rules that model the input data behavior. The number of Epochs was set at 30.

Similarly, for the case of ANN, a simple feed-forward neural network was considered. Given we have 14 features in both datasets, we set the number of hidden layers size to 12. The error tolerance rate was set at 5%.

3.2.1. Feature Reduction

As explained earlier, PCA was used as the feature reduction technique. The extent to which reduction was made was dependent on value set for the percentage of variation, also known as variation percentage. Since one of the objectives of this thesis is to study the application of feature reduction technique of PCA in the creation of the best prediction model, one important step would be to determine the variation percentage at which PCA best performs for a given ML technique. This was done with the assumption that different ML models behave differently to feature reduction techniques. In order to determine the optimal variation percentage for each of the classification techniques, we decided to select the percentage which had the least Root Mean Square Error (RMSE) for regression models or highest accuracy for binary based classifiers. For each of the three ML techniques and for each of the 3 outputs, RMSE or accuracy value was recorded for variation percentage ranges of 95 to 50, with a step size of 10. Because we wanted to train a ML model that had the least RMSE or highest accuracy we needed the variation percentage value to be set accordingly. In order to do so, the variation percentage at the lowest RMSE or highest accuracy for each output was then selected and the mode of these was set as the value for each ML technique. The obtained results are explained in Chapter 4.

This experimentation was done in only the first dataset and the obtained variation percentage were then applied to both the datasets for both training and testing in the first dataset and for testing in the second dataset.

3.3. Fuzzy Inference System

One other study we wanted to look into was to see if the integration of FIS into the prediction model would offer a better prediction performance as opposed to utilizing normal ML based prediction model. Consequently, we formulated a FIS with rules that

would determine the time of surgery based on the predicted outputs of the two ML prediction models – IVGR2 prediction model and prediction model that would predict whether a patient would need to undergo surgery. The best performing models for each of these were selected and their prediction outputs were the intended inputs of the FIS.

In order to achieve this, we needed to implement the FIS by defining the membership functions of the inputs and specifying the rule set for the inference system. In terms of the membership functions, for IVGR2, we selected Gaussian membership functions to represent the predicted infarction volumes and classified it into small, medium and large. Similarly, for the predicted output values of the second prediction model that which represented the decision for surgery (DHCN), we used trapezoidal shaped membership functions. The inputs were classified into Surgery or No Surgery (NoSurgery). For simplicity, the output, time of surgery (TimeOfSurgery), was split into three categories – No Surgery Required (NoSurgery), Surgery Required in 24 hours (Surgery24Hr) and Surgery Required in 48 hours (Surgery48Hr). We utilized triangular-shaped membership function for the output. Once we defined these, we outlined the rules as follows –

1. If (IVGR2 is Small) and (DHCN is NoSurgery) then (TimeOfSurgery is NoSurgery)

2. If (IVGR2 is Medium) and (DHCN is Surgery) then (TimeOfSurgery is Surgery48Hr)

3. If (IVGR2 is Large) and (DHCN is NoSurgery) then (TimeOfSurgery is Surgery48Hr).

4. If (IVGR2 is Large) and (DHCN is Surgery) then (TimeOfSurgery is Surgery24Hr)

3.4. MATLAB Application

To consolidate everything, we finally combined all the selected prediction models for each of the outputs into a MATLAB application. For the application we developed, we designed it to accept all the input values of the different patient clinical data along with first infarction volume.

Figure 3 shows the user-friendly MATLAB application we developed. While the user may manually insert all the patient medical data, he may optionally insert the patient ID and click on button Load for the patient details to get displayed. When the user clicks the Clear button, all the field are cleared of their current contents. With the click of the single button - Predict, the user will be able to predict the second IVGR, the decision to perform DH and when to perform it.

Figure 3. MATLAB application developed for stroke prognosis prediction.

Chapter 4: Evaluation

Evaluation of the applied ML method is usually done by measuring the classification accuracy, the information score or its misclassification cost [29]. In order to evaluate our work, we selected multiple evaluation metrics for both regression-based prediction model and binary based classification model. Given infraction volume prediction is a regression problem, we evaluated our proposed solution using regression appropriate evaluation techniques such as RMSE and Mean Absolute Error (MAE). Likewise, since determining whether a stroke patient requires a surgery or not is a binary classification problem, we utilized evaluation techniques such as Accuracy, Precision and Recall. In this chapter, we will first describe and discuss the different evaluation metrics and then discuss the results obtained for each of them.

4.1. Evaluation Metrics

Before we delve into the obtained evaluation results, we will first offer a brief description of the evaluation metrics we selected to evaluate our work in this section. Accordingly, we will initially discuss evaluation metrics used to evaluate regression-based models by discussing Root Mean Square in Subsection 4.1.1 and Mean Absolute Error in Subsection 4.1.2. We will then discuss evaluation metrics appropriate for evaluating binary classification models in the subsequent Subsection 4.1.3 by defining Accuracy, Precision and Recall.

4.1.1. Root Mean Square Error

Root Mean Square Error (RMSE), which is the square root of the Mean Square Error (MSE), is a popular technique used as an evaluation metric in the literature. The reason we selected this measure was because it empowers large number deviations and

when compared to MSE gives higher weightage and thereby punishes large errors more. This is critical in our solution since we are modelling a prediction model for a medical setting and this involves predicting health related outcomes. Therefore, utmost accuracy must be obtained while avoiding any minute errors as much as possible. The formula RMSE is as shown below.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(actual_i - predicted_i \right)^2}$$

(1)

4.1.2. Mean Absolute Error

Another measure we used to evaluate the regression prediction model was the Mean Absolute Error (MAE). It measures the mean value of error in a set of observations without considering their direction. The score averages over the absolute differences of a set of target and predicted observation, while giving equal weights to all individual differences. The equation of MAE is as shown below.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| actual_i - predicted_i \right|$$

(2)

4.1.3. Accuracy

Accuracy is defined as the fraction of predictions the model being evaluated got correct.

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

(3)

4.1.4. Precision

A simple definition of Precision is the fraction of observations that are predicted positive and are actually positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{4}$$

4.1.5. Recall

Recall calculates how many actual positives the model labels positive.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{5}$$

4.2. Evaluation Results

This section will elaborate on the obtained evaluation results. Subsection 4.2.1 will discuss the results obtained by varying PCA variation percentage for different ML models, while the subsequent subsections, Subsections 4.2.2 until Subsection 4.2.4 will discuss the evaluation results obtained from the different ML models implemented and the reason why they were selected. In addition to discussing the selected ML model, Subsection 4.2.4 will also evaluate the integration of FIS in the prediction of decision for surgery as opposed to only utilizing pure ML prediction models.

4.2.1. PCA variation determination

As mentioned in Chapter 3, we experimented with different PCA variation percentages and selected the variation percentage that resulted with the least RMSE or

highest accuracy per ML technique to obtain the optimal variation percentage for each of the ML techniques implemented.

Table 2. Evaluation Results for SVM at Different Variation Percentages

Variation Percentage (%)	RMSE for IVGR2	RMSE for treatment time	Accuracy for Decision for DC
95	**0.168**	0.226	0.915
90	0.173	0.226	0.915
85	0.198	0.223	0.915
80	0.199	0.222	0.915
75	0.201	0.222	0.915
70	0.201	**0.221**	0.915
65	0.196	**0.221**	0.915
60	0.196	0.221	0.915
55	0.196	0.221	0.915
50	0.196	0.221	0.915

Table 2 offers the obtained values for SVM. We observed no matter what variation percentage we selected; the accuracy of the generated model remained at 0.92% for the SVM model generated to predict the decision for DC. Therefore, since we needed to pick the mode of the best representative values, we selected the variation at 95% since it has the least combined RMSE value.

Table 3. Evaluation Results for ANFIS at Different Variation Percentages

Variation Percentage (%)	RMSE for IVGR2	RMSE for treatment time	Accuracy for Decision for DC
95	0.222	**0.0489**	0.915
90	0.264	0.510	0.893
85	1.093	0.188	0.872
80	0.475	0.266	0.680
75	0.455	0.185	0.680
70	0.324	1.003	0.787
65	0.211	1.028	0.766
60	0.211	0.138	0.915
55	0.211	0.158	0.936
50	**0.199**	0.139	**0.957**

Similarly, Table 3 shows the values obtained for ANFIS. Again, basing our selection reason on the method of mode, that is by selecting variations that had the most number of best evaluation result, the percentage variation at 50% was therefore chosen. Lastly Table 4 shows the values obtained for ANN and variation percentage at 90% was accordingly selected.

Table 4. Evaluation Results for ANN at Different Variation Percentages

Variation Percentage (%)	RMSE for IVGR2	RMSE for treatment time	Accuracy for Decision for DC
95	0.182	0.220	0.787
90	**0.176**	**0.050**	**0.936**
85	0.208	0.207	0.894
80	0.202	0.206	0.894
75	0.202	0.206	0.894
70	0.202	0.205	**0.936**
65	0.178	0.205	**0.936**
60	0.178	0.205	**0.936**
55	0.178	0.205	**0.936**
50	0.415	0.205	0.021

Another insight we wanted to gain from our usage of PCA was to determine the most contributing features for stroke related predictions. As mentioned before, PCA performs feature reduction by creating linear combinations of the inputs. As a case study we studied the contribution of each of the original inputs by considering one of the outputs; second IVGR and studying the generated principal components. Table 5 below shows the obtained RMSE values along the number of principal components returned at each variation percentage.

Table 5. RMSE Values and Principal Components at Each Variation Percentage

Variation Percentage (%)	RMSE for IVGR2	Number of Principal Components
95	**0.168**	**8**
90	0.173	6
85	0.198	5
80	0.199	4
75	0.201	4
70	0.201	3
65	0.196	3
60	0.196	2
55	0.196	2
50	0.196	1

As shown, Principal Component representing a total of 95% of variance of the original dataset had the least RMSE score at 0.168. The dimensionality of the Principal Components or the corresponding number of Principal components associated with that variation percentage was 8. This signifies that 95% of the original information was maintained in this reduced dataset of 8 principal components each consisting a combination of a subset of the original features. Given that each principal component is a linear combination of a subset of the original feature set, the reduced dataset consisting of 8 principal components was then examined to determine the most contributing features. The following features – Age, Ejection Fraction, METS, Clot Burden Score, DBP - in order of decreasing contribution were found to be the most contributing features.
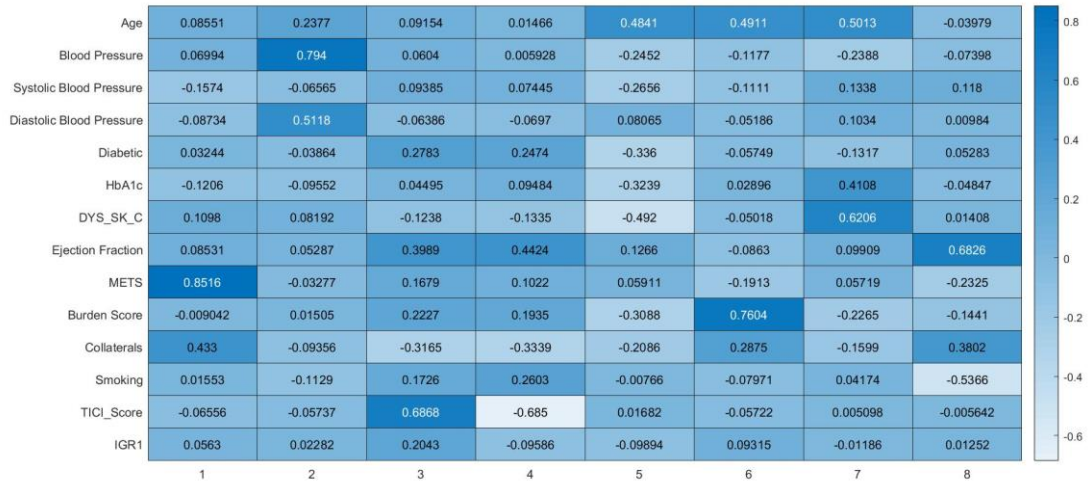
Figure 4. Heatmap showing each feature's contribution on 8 principal components.

4.2.2. Prediction of Second Infarction Volume

As explained earlier, we started with reducing the dataset using PCA and modelling an ANFIS based ML model for that reduced set before we experimented with other ML techniques for both reduced and non-reduced dataset. The obtained results are as shown in Table 6.

Table 6. Evaluation Results of ML Models for IVGR2 Prediction

ML Technique	MSE	RMSE	MAE
SVM	0.012	0.108	0.085
SVM + PCA	**0.008**	**0.092**	**0.068**
ANFIS	0.067	0.259	0.176
ANFIS+PCA	0.021	0.144	0.122
ANN	0.021	0.146	0.088
ANN + PCA	0.011	0.106	0.080

As mentioned earlier and as shown in Table 6 applying feature reduction on the dataset and modelling an ANFIS model on that reduced dataset does in fact result in a better performing model. However, as clearly seen, that is not best performing model. In fact, an SVM model that is modelled on a reduced dataset performs the best in predicting IVGR2. While reduction or the application of PCA always works in the favor of creating a better prediction model, regression based SVM seems to outperform both ANFIS and ANN. The plot is of the output of the best performing model and the original IVGR2 values is shown in Figure 5.
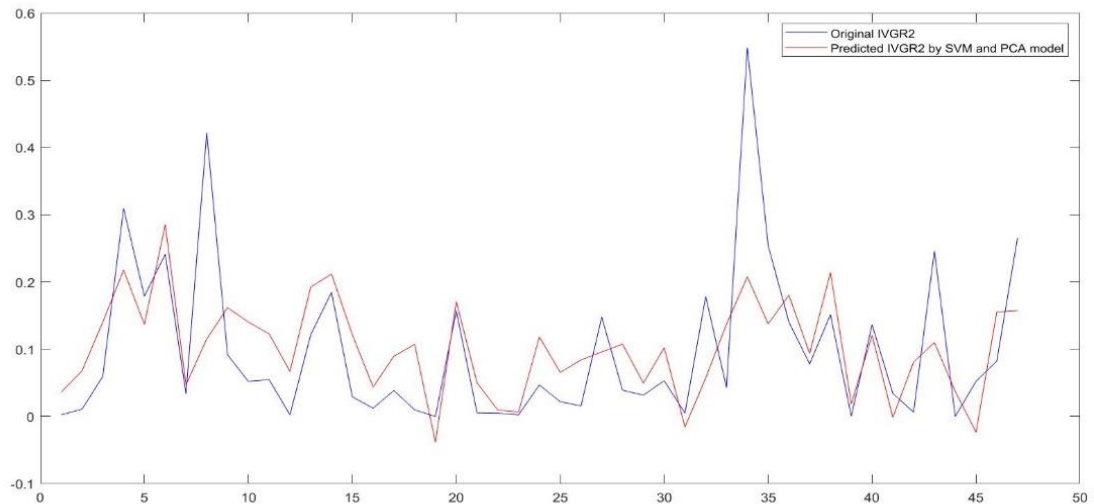

Figure 5. Original and prediction model plots for IVGR2

We tested the obtained best prediction model on the second dataset and the obtained evaluation results are as shown in Table 7. The obtained error is higher than that obtained in the first dataset, however we still obtain an RMSE value way below 50%.

Table 7. Evaluation Results from Selected IVGR2 Prediction Model on Second Dataset

MSE	RMSE	MAE
0.069	0.263	0.200

4.2.3. Prediction of Decision for Surgery

Similarly, for the determination of the best model for decision for surgery (DFS) we implemented and compared between 6 binary classifiers. Their evaluation results are as shown in Table 8. The best performing ML technique was ANFIS that was modelled on a reduced dataset. While SVM performed as good, it however had a lower accuracy.

Table 8. Evaluation Results of ML Models for Decision for Surgery Prediction

ML Technique	Accuracy	Precision	Recall
SVM	0.915	1	0.915
SVM + PCA	0.915	1	0.915
ANFIS	0.872	0.884	0.974
ANFIS+PCA	**0.978**	**1**	**0.915**
ANN	0.830	0.861	0.948
ANN + PCA	0.787	0.814	0.946

When the second dataset was tested using the selected model, the obtained results were however very poor in terms of the returned accuracy as shown in Table 9. This could be due to the fact that some of features such as METS and Clot Burden Score

in the second dataset had unsure values as explained in Chapter 3. Assuming some of these features with unsure values could potentially be the most contributing features to the prediction model, such poor performance can therefore be expected. However, the precision remained at 1, thus proving the reliability of the prediction model.

Table 9. Evaluation Results from Selected DFS Prediction Model on Second Dataset

Accuracy	Precision	Recall
0.428	1	0.5

4.2.4. Prediction of Treatment Time

For the prediction of Treatment Time (TT), we proposed two ways to make this prediction. In the first method we implemented ML models with or without reduction as done for the other two outputs. In the second method, we proposed to utilize the output of the previously obtained best prediction models for the IVGR2 and DFS outputs as input into a FIS and utilize the generated FIS to predict the treatment time.

In the first method, similar to DFS prediction model's performance outcome, ANFIS outperformed the rest of the models, given when a reduced dataset was fed as input. The evaluation results are as shown in Table 10.

Table 10. Evaluation Results of ML Models for Treatment Time Prediction

ML Technique	MSE	RMSE	MAE
SVM	0.037	0.192	0.136
SVM + PCA	0.036	0.189	0.131
ANFIS	0.289	0.538	0.332
ANFIS+PCA	**0.028**	**0.169**	**0.138**
ANN	0.066	0.258	0.204
ANN + PCA	0.042	0.204	0.166

Even though when compared with the rest of the models, ANFIS performs the best with the reduced dataset, it nevertheless has an RMSE error of 16% which is high and reflects when the original treatment time is compared with the predicted treatment time as shown in Figure 6.
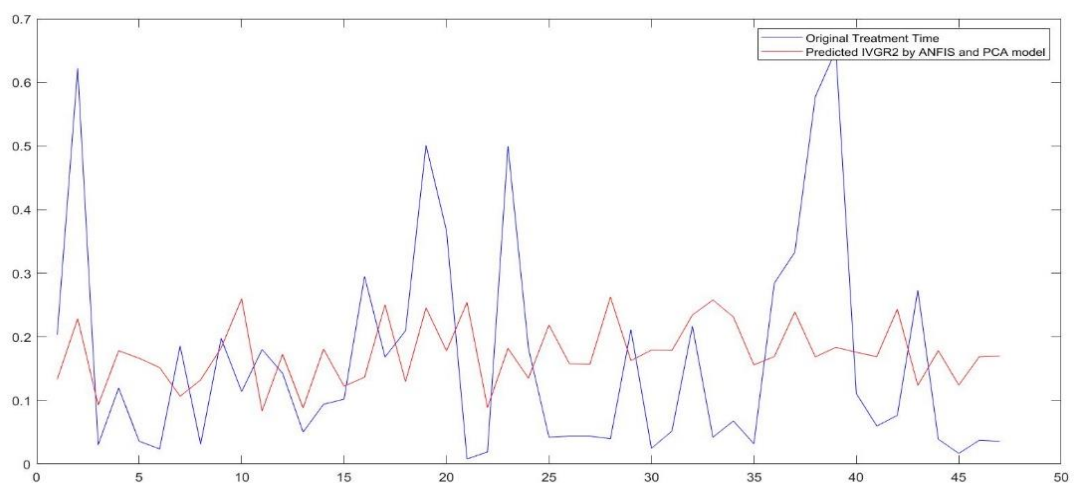

Figure 6. Original and prediction model plots for treatment time.

However, when the selected model was tested using the testing dataset, a similar bad performance was noticed as shown in Table 11.

Table 11. Evaluation Results from Selected TT Prediction Model on Second Dataset

MSE	RMSE	MAE
0.566	0.752	0.606

When considering the usage of FIS, we input the best performing models of both the inputs as explained above. Table 12 below shows the confusion matrix obtained by the prediction results of FIS.

Table 12. Confusion Matrix of FIS

		Predicted Output	
		0	1
Actual Output	0	45	0
	1	2	0

The number of true positive is at 45, resulting at an accuracy of 95.7% which is shown in Table 13.

Table 13. Evaluation Results of Overall FIS System

Accuracy	Precision	Recall
0.957	1	0.957

Using only ML based techniques on the other hand resulted in an accuracy of almost 92% as shown in Table 14.

Table 14. Evaluation Results of Second TT Prediction on First dataset

Accuracy	Precision	Recall
0.915	1	0.915

However as much as it is important for a system to have high accuracy; for our problem, it is more important that our system is always able to make the right decision, which is the essence of precision. Both of them obtained a precision of 100%. Thus, FIS had a better overall performance and was selected as the prediction model. The generated 3D surface of the FIS is as shown in Figure 7.
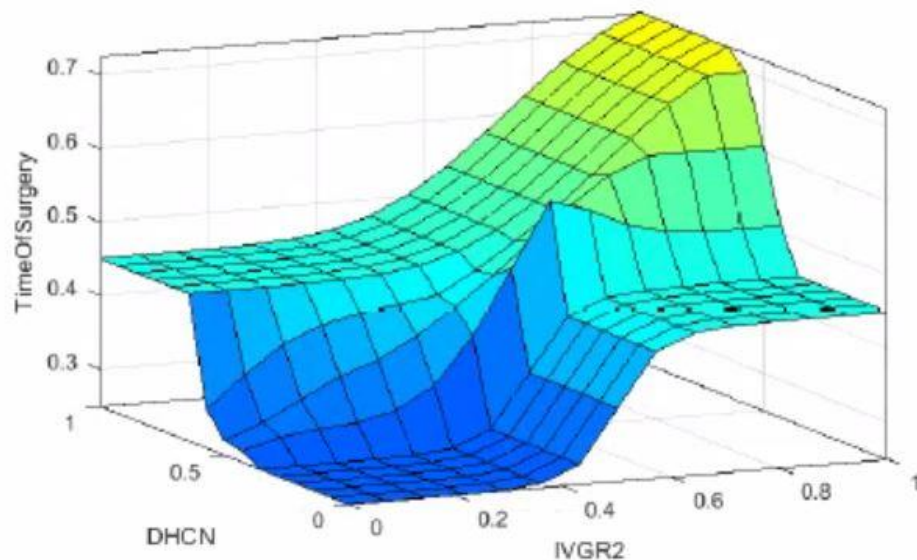

Figure 7. 3D decision surface of generated FIS input and output.

Chapter 5: Conclusion and Future Work

For this thesis, we applied ML techniques for predictions of outputs related to stroke prognosis. We based our prediction model on patient data and the output values from the first CT scan. Because most of the used patient data are readily available and easily accessible for most patients, this made our proposed solution a realistic one. We developed an application with the best performing ML model for each of the outputs. In the case of prediction of second infarction volume, SVM model that is modelled on a reduced dataset performed the best prediction with a root mean square error of 0.092 for the first dataset. Similarly, in the case of decision for surgery prediction, the best performing ML technique was ANFIS that was modelled on a reduced dataset. This model obtained an approximate accuracy of 98% on the first dataset. FIS outperformed the generated ML models for the prediction of the treatment time obtaining an accuracy of 96%. Given these results, we believe our work offers the first step to assisting doctors and other medical professionals make faster decisions in the prognosis of stroke.

The work done in this thesis offers a simple contribution to a very promising large medical domain. While our focus was on stroke prognosis-based prediction, we have experimented within it a number of topics and these can nevertheless offer as starting points to future work.

Given we have used the feature reduction technique of PCA in this thesis, one direction to be considered could be to experiment with the usage of Kernel Principal Component Analysis (KPCA), a nonlinear extension of PCA. KPCA non-linearly maps data into a higher dimensional feature space, instead of lower dimensional feature space, by using kernel methods. While seeming counterintuitive to the expected outcome, multiple studies have shown that attempting this results in better performance than when PCA is used [48], [49].

Another future work that could be considered would be to focus on the utilization of PCA to obtain the most contributing features to the prediction process. As done for the prediction of second IVGR, this could further be expanded to the other outputs to obtain a common and small set of representative features. By doing so, we could narrow down the number of inputs required by doctors to utilize this application without jeopardizing its effectiveness or accuracy. While this will certainly simplify the use of the application for the users and encourage its use, it will also help gear development focus to the improvement of the currently developed prediction models.

Yet another future work could be to tune the ML techniques so as to improve their performance. Admittedly not much has been done in this thesis in that regard. Also, other techniques could be considered and attempted to see if better performing models can be modelled.

Chapter 6: Related Publications

6.1. Conference Paper

[1] R. Ali, U. Qidwai, and S. K. Ilyas, "Use of Combination of PCA and ANFIS in Infarction Volume Growth Rate Prediction in Ischemic Stroke," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2018, pp. 324–329.

6.2. Journal Paper

[1] R. Ali, U. Qidwai, and S. K. Ilyas, "Adaptive Neuro-Fuzzy Inference System for Prediction of Surgery Time for Ischemic Stroke Patients" (Accepted to be published in International Journal of Integrated Engineering (IJIE) (ISSN: 2229-838X) and will appear in upcoming Special Issue on Innovation in Global Healthcare)

REFERENCES

[1] T. D. Musuka, S. B. Wilton, and M. D. H. Mouhieddin Traboulsi, "Diagnosis and management of acute ischemic stroke: speed is critical," *Can. Med. Assoc. J.*, vol. 187, no. 12, pp. 887–893, 2015.

[2] C. S. Kidwell *et al.*, "Comparison of MRI and CT for Detection of Acute Intracerebral Hemorrhage," vol. 292, no. 15, pp. 1823–1830, 2004.

[3] C. M. Schirmer, A. A. A. Jr, and A. M. Malek, "Decompressive Craniectomy," pp. 456–470, 2008.

[4] D. Kleindorfer *et al.*, "Eligibility for Recombinant Tissue Plasminogen Activator in Acute Ischemic Stroke," no. 0525, pp. 1–3, 2004.

[5] D. Tong *et al.*, "Times From Symptom Onset to Hospital Arrival in the Get With The Guidelines – Stroke Program 2002 to 2009 Temporal Trends and Implications," pp. 1912–1917, 2012.

[6] M. Ebinger *et al.*, "Protocols PHANTOM-S : the prehospital acute neurological therapy and optimization of medical care in stroke patients – study," vol. 7, no. June, pp. 348–353, 2012.

[7] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model," *Ann. Appl. Stat.*, vol. 9, no. 3, pp. 1350–1371, Sep. 2015.

[8] T. Lumley, R. A. Kronmal, M. Cushman, T. A. Manolio, and S. Goldstein, "A stroke prediction score in the elderly: validation and Web-based application," 2002.

[9] A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data*

mining - KDD '10, 2010, p. 183.

[10]   S. Kamran *et al.*, "Prediction of infarction volume and infarction growth rate in acute ischemic stroke," *Sci. Rep.*, vol. 7, no. 1, p. 7565, Dec. 2017.

[11]   C. Weimar, A. Ziegler, I. R. König, H.-C. Diener, and on behalf of the G. S. S. Collaborators, "Predicting functional outcome and survival after acute ischemic stroke," *J. Neurol.*, vol. 249, no. 7, pp. 888–895, Jul. 2002.

[12]   M. M. Rahman and D. N. Davis, "Addressing the Class Imbalance Problem in Medical Datasets," *Int. J. Mach. Learn. Comput.*, vol. 3, no. 2, pp. 224–228, 2013.

[13]   L. Mena and J. A. Gonzalez, "Machine Learning for Imbalanced Datasets: Application in Medical Diagnostic."

[14]   P. Yang, L. Xu, B. B. Zhou, Z. Zhang, and A. Y. Zomaya, "A particle swarm based hybrid system for imbalanced medical data sampling," *BMC Genomics*, vol. 10, no. Suppl 3, p. S34, Dec. 2009.

[15]   M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks*, vol. 21, no. 2–3, pp. 427–436, Mar. 2008.

[16]   I. Babaoğlu, O. Fındık, and M. Bayrak, "Effects of principle component analysis on assessment of coronary artery diseases using support vector machine," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2182–2185, Mar. 2010.

[17]   K. Polat and S. Güneş, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," *Digit. Signal Process.*, vol. 17, no. 4, pp. 702–710, Jul. 2007.

[18]   S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom.*

Intell. Lab. Syst., vol. 2, no. 1–3, pp. 37–52, Aug. 1987.

[19] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.

[20] C. Corinna and V. Vladimir, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[21] J. Xie and C. Wang, "Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5809–5815, 2011.

[22] S. Vijayarani, M. S. Dhayanand, and M. P. Research Scholar, "KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS," *Int. J. Comput. Bus. Res.*, vol. 6, no. 2, 2015.

[23] B. Debasish, S. Pal, and D. Chandra Patranabis, "Support Vector Regression," *Neural Inf. Process. – Lett. Rev.*, vol. 11, no. 10, pp. 203–224, 2007.

[24] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer (Long. Beach. Calif).*, vol. 29, no. 3, pp. 31–44, 1996.

[25] F. Chang and Y. Chang, "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir," vol. 29, pp. 1–10, 2006.

[26] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man. Cybern.*, vol. 23, no. 3, pp. 665–685, 1993.

[27] A. J. C. Sharkey, N. E. Sharkey, and S. S. Cross, "Adapting an Ensemble Approach for the Diagnosis of Breast Cancer," in *International Conference on Artificial Neural Networks*, 1998.

[28] Q.-H. Ye *et al.*, "Predicting hepatitis B virus–positive metastatic hepatocellular

carcinomas using gene expression profiling and supervised machine learning," *Nat. Med.*, vol. 9, no. 4, pp. 416–423, Apr. 2003.

[29]   I. Kononenko, "Machine learning for medical diagnosis : history , state of the art and perspective," *Artif. Intell. Med.*, vol. 23, no. 1, pp. 89–109, 2001.

[30]   F. Jiang *et al.*, "Artificial intelligence in healthcare: Past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, pp. 230–243, 2017.

[31]   J. T. Senders *et al.*, "Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review," *World Neurosurg.*, vol. 109, p. 476–486.e1, 2018.

[32]   J. Zhang *et al.*, "Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images," *Biomed Res. Int.*, vol. 2017, pp. 1–9, Jan. 2017.

[33]   J. B. Siddharth Jonathan and K. Shruthi, "A TWO TIER NEURAL INTER-NETWORK BASED APPROACH TO MEDICAL DIAGNOSIS USING K-NEAREST NEIGHBOR CLASSIFICATION FOR DIAGNOSIS PRUNING."

[34]   T. A. Horan, B. Tulu, B. Hilton, and J. Burton, "Use of online systems in clinical medical assessments: an analysis of physician acceptance of online disability evaluation systems," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, 2004, p. 10 pp.

[35]   D.-C. Li, C.-W. Liu, and S. C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets," *Artif. Intell. Med.*, vol. 52, pp. 45–52, 2011.

[36]   N. H. Vo and Y. Won, "Classification of Unbalanced Medical Data with Weighted Regularized Least Squares," in *2007 Frontiers in the Convergence of Bioscience and Information Technologies*, 2007, pp. 347–352.

[37]   Y. Liu, X. Yu, J. X. Huang, and A. An, "Combining integrated sampling with SVM ensembles for learning from imbalanced datasets," *Inf. Process. Manag.*,

vol. 47, no. 4, pp. 617–631, Jul. 2011.

[38]   E. Avci and I. Turkoglu, "An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2873–2878, Mar. 2009.

[39]   K. Polat and S. Güneş, "Automatic determination of diseases related to lymph system from lymphography data using principles component analysis (PCA), fuzzy weighting pre-processing and ANFIS," *Expert Syst. Appl.*, vol. 33, no. 3, pp. 636–641, Oct. 2007.

[40]   A. Lisowska *et al.*, "Context-Aware Convolutional Neural Networks for Stroke Sign Detection in Non-contrast CT Scans," Springer, Cham, 2017, pp. 494–505.

[41]   Y. Wang, A. K. Katsaggelos, X. Wang, and T. B. Parrish, "A deep symmetry convnet for stroke lesion segmentation," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 111–115.

[42]   Y. Choi, Y. Kwon, H. Lee, B. J. Kim, M. C. Paik, and J.-H. Won, "Ensemble of Deep Convolutional Neural Networks for Prognosis of Ischemic Stroke," Springer, Cham, 2016, pp. 231–243.

[43]   P. Bentley *et al.*, "Prediction of stroke thrombolysis outcome using CT brain machine learning," *NeuroImage Clin.*, vol. 4, pp. 635–640, 2014.

[44]   J. Hampton-Till *et al.*, "AUTOMATED QUANTIFICATION OF STROKE DAMAGE ON BRAIN COMPUTED TOMOGRAPHY SCANS : e-ASPECTS," *Eur. Med. J.*, vol. 3, no. 1, pp. 69–74, 2015.

[45]   H. Asadi, R. Dowling, B. Yan, and P. Mitchell, "Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy," *PLoS One*, vol. 9, no. 2, pp. 14–19, 2014.

[46]   L. Mirtskhulava, J. Wong, S. Al-Majeed, and G. Pearce, "Artificial Neural

Network Model in Stroke Diagnosis," *Proc. - UKSim-AMSS 17th Int. Conf. Comput. Model. Simulation, UKSim 2015*, pp. 50–53, 2016.

[47]   R. Ali, U. Qidwai, and S. K. Ilyas, "Use of Combination of PCA and ANFIS in Infarction Volume Growth Rate Prediction in Ischemic Stroke," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2018, pp. 324–329.

[48]   B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[49]   S. Mika, S. Mika, B. Schölkopf, A. M. Smola Klaus-Robert, M. Scholz, and G. Rätsch, "Kernel PCA and De-Noising in Feature Spaces."