A KNOWLEDGE BASED FRAMEWORK TO PREDICT CORONARY HEART DISEASE

BY

OLUFEKO OLUWATOBILOBA WISDOM

15010301017

BEING A PROJECT SUBMITTED TO THE

DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, COLLEGE OF BASIC AND APPLIED SCIENCES

IN FUFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF BACHELOR OF SCIENCE MOUNTAIN TOP UNIVERSITY, KILOMETRE 12, LAGOS-IBADAN EXPRESSWAY, PRAYER CITY, OGUN STATE, NIGERIA

JULY, 2019

CERTIFICATION

This project titled, **A KNOWLEDGE BASED FRAMEWORK TO PREDICT CORONARY HEART DISEASE**, prepared and submitted by **OLUFEKO OLUWATOBILOBA WISDOM** of matriculation number 15010301017 in fulfilment of the requirements for the degree of **BACHELOR OF SCIENCE (Computer Science)** is hereby accepted.

____________________________ (Signature and Date)

Dr. O.B. Alaba

(Supervisor)

____________________________ (Signature and Date)

Dr. I.O. Akinyemi

(Head of Department)

Accepted as partial fulfilment of the requirements for the degree of BACHELOR OF SCIENCE (Computer Science)

____________________________ **(Signature and Date)**

Prof. A. I. Akinwande

Dean, College of Basic and Applied Sciences

DEDICATION

This project work is dedicated to God Almighty.

ACKNOWLEDGEMENT

I owe much gratitude to God Almighty who gave me the wisdom, knowledge, understanding, strength, divine help and provision from the commencement of this project work to its completion.

I specially appreciate my supervisor Dr. O. B. Alaba who took keen interest in my project work and guided me all along, and taking the pains to ensure the successful completion of this project work.

I will like to acknowledge the Head of Department Computer Science and Mathematics Dr. I.O. Akinyemi, and offer deep gratitude for the efforts, constant encouragement, guidance and support. I also appreciate all the members of staff of the Department of Computer Science: Dr. Alaba O. B., Dr. Oyetunji M.O., Dr. (Mrs.) Kasali F.A., Dr. (Mrs) Taiwo O.O. Mr. Falana O.J., Dr. Idowu P.A., Dr. Ojesanmi O.A., Dr. Adamu O.B., Dr. Okunoye O.B., Dr. (Mrs.) Oladeji F.A., Mr. Ebo I.O and others to mention but a few..

I heartily thank my parents Mr and Mrs Olufeko and my wonderful siblings, thank you all for your moral and financial support. I am grateful for all the investments into my education and future.

I sincerely appreciate my friends and all Mountain Top University colleagues for their help and support during the period of working on this project. I say God bless you all.

ABSTRACT

Coronary heart disease (CHD) is a disease common to both men and women and also use of genetics have been rarely used to predict it.

Coronary heart disease alludes to a narrowing of the coronary veins, the veins that supply oxygen and blood to the heart. It is otherwise called coronay artery disease. It is a note worthy reason for ailment and death. The existing Clinical Decision Support Systems (CDSSs) have not been accurate enough in their prediction and diagnosis of coronary heart disease.

By using genetic information such as Single Nucleotide Polymorphism, Genome Build, Chromosome, Map and the partition coefficient (LogP) gotten from the Duke 2007 dataset and the C4.5 decision tree pattern classification algorithm which was selected amongst other competing classification algorithms including K-Nearest Neighbor, Bayes Classifier and Support Vector Machine after a thorough evaluation on the Waikato Environment for Knowledge Analysis (WEKA version 3.6.7), this study developed a framework for accurate prediction of coronary heart disease. A prediction accuracy of 61.0734% was obtained from training the C4.5 algorithm on the Duke 2007 dataset which gives higher prediction accuracy than the existing CHD Modeling and Execution framework.

An improved framework that enhances the classification/prediction of coronary heart disease which helps to guide patients with Cronary heart disease as to how to best manage their health condition and live a normal life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background to Study

In the course of the most recent couple of years, the advancement of clever basic leadership applications is quick making strides. This idea is known as Artificial Intelligence (AI). Computer based intelligence has diverse sub-fields which incorporate master frameworks, machine vision, AI and normal language handling among others. A choice emotionally supportive network (DSS) is a modernized data framework used to help basic leadership in an association or a business. A DSS gives clients a chance to filter through and examine enormous measures of information, and assemble data that can be utilized to take care of issues and improve decisions (Chen, 2019) .According to the Clinical Decision Support (CDS) Roadmap venture, CDS is "giving clinicians, patients, or people with learning and individual explicit or populace data, astutely sifted or present at proper occasions, to encourage better wellbeing forms, better individual patient consideration, and better populace wellbeing." Clinical choice emotionally supportive networks (CDSS) is utilized to help clinicians in settling on astute antimicrobial treatment choices and to help the antimicrobial stewardship program with ID of patients for potential mediation (Simon et al., 2017). This infers that a CDSS is a choice emotionally supportive network (DSS) that utilizations information the board to accomplish clinical guidance for patient consideration dependent on some number of things of patient information. This facilitates the activity of medicinal services specialists, particularly in regions where the quantity of patients is overpowering.

The writing in resulting areas of this review supports the way that coronary illness is one part of prescription that critically needs CDSS inferable from the way that there are not many authorities here of medication.

Coronary vein can be characterized as the course that provisions blood to the heart. They are in charge of the progression of blood all through the heart muscle.

Coronary heart maladies are supply route based conditions that frequently bring about an assortment of side effects that can influence everyday life activity. Coronary illness alludes to a narrowing of the coronary veins, the veins that supply oxygen and blood to the heart. It is otherwise called coronary corridor ailment. It is a noteworthy reason for ailment and passing (NHLBI, 2016). Significant manifestations of coronary illness incorporate Chest torment (it is frequently portrayed as a crushing, weight, greatness, copying or hurting over the chest (Nordqvist, 2018)), acid reflux, shortcoming, perspiring sickness, cramping, and brevity of breath and so forth. Coronary heart infections are significantly brought about by damage or a sore to the inward layer of the coronary conduit. This harm makes greasy plaques leftovers or store to amass at the site of the damage. These stores includes cholesterol and other cell squander items. The collection of these stores is called atherosclerosis.

Like most different sicknesses, the identification of coronary illness depends on the indications appeared by the person being referred to, anyway this customary strategy isn't exact enough in light of the fact that occasionally an individual may demonstrate manifestations that propose coronary illness then the person in question may experience the ill effects of another coronary illness. Along these lines there is the requirement for choice emotionally supportive networks that can foresee coronary illness dependent on hereditary data.

1.2 Statement of Problem

According to the World Health Organization (WHO), about 3.2 million men and approximately 3.5 million women die every year from coronary heart diseases. Due to this death there is need for CDSSs to predict coronary heart diseases. Existing CDSSs that diagnose coronary heart disease make use of the convention risk methods but little research methods have been carried out on using genetic information to predict coronary heart diseases.

1.3 Aim and Objectives

The point of this undertaking is to build up a structure for a learning based framework to foresee coronary illness. The created framework which utilizes design arrangement calculations will help decide whether an individual has the inclination of creating coronary heart infections.

Consequently, the particular goals of this exploration are to:

1. Complete a relative examination of pertinent grouping calculations so as to choose the most precise calculation that will give insight to the shrewd structure.

2. Build up a structure for a learning based framework to anticipate coronary heart maladies dependent on hereditary data, for example, Single Nucleotide Polymorphism (SNP), Genome Build, Chromosome, Map and the Partition Coefficient (LogP).

1.4 Methodology

Perform similar examination of example order calculations on suitable datasets utilizing Waikato Environment for Knowledge Analysis (WEKA).

A Distinguishing proof of the considerable number of modules that will make up the structure by broad and comprehensive assessment of fundamentally the same as shrewd systems. Also a distinguishing proof of the considerable number of parts that will make up every module in the wake of deciding the segments that will be required for the particular modules to complete their capacities.

Incorporation of the example characterization calculation picked after assessment of important grouping calculations on WEKA rendition 3.6.7 in its suitable position in the structure.

Proper association of the different segments of the system to one another in order to empower the structure complete the precise forecast of coronary heart infections.

1.5 Scope of Study

The exploration in this proposition makes commitments as far as creating tolerant explicit Clinical Pathways (CPs) for coronary heart ailments that objectives the information needs and clinical duties of General Practitioners (GPs).They will be made to join master restorative learning of cardiologists.

1.6 Significance of Study

This examination helps in the advancement of a structure to foresee coronary illness. a structure for the prescient treatment of different heart maladies. It makes it workable for restorative specialists who have little thought in the part of heart illnesses to have a more prominent information in determination and treatment incorporating of patients with coronary heart disease or different parts of heart sicknesses.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Heart diseases also known as cardiovascular diseases (CVDs) are disorders of the heart and blood vessels and include coronary heart diseases, cerebrovascular disease, rheumatic heart disease and other conditions. According to world health organization (WHO), cardiovascular diseases (CVDs) takes the lives of 17.9million people annually, 31% of all global deaths (WHO, 2011). Over three quarters of CVD occur in low-and middle-income countries.

Heart diseases include coronary heart diseases (CHD), high blood pressure, cardiac arrest, congestive heart failure, arrhythmia, peripheral artery diseases, stroke, and congenital heart disease. Among these diseases, coronary artery disease is the most common and deadliest.

2.2 Coronary Heart Disease

Coronary heart disease (CHD) originally ischemic heart disease (IHD) can also be called coronary artery disease (CAD), is a type of cardiovascular diseases. Coronary heart disease is the most common type of heart diseases. CHD is a severe heart disease that has been recognized as the most debilitating and baffling heart disease ever known, because it deals with the artery supplying blood to the heart.

2.2.1 Symptoms of Coronary heart disease

Chest pain (angina): one might observe a pressure or tightness in one's chest, like something or someone is standing right above the chest. This pain is medically known as angina, which occurs at the middle or towards the left side of the chest. Angina is commonly awoken by physical or mental or emotional stress.

The pain normally fades away moments after quitting the stressful activity. In others, this pain may be sharp feeling in the neck, arm or at the one's back.

Shortness of breath: this occurs when the heart doesn't pump adequate blood to meet the body's requirement. Then it's a sign one might have this disease in question.

Heart attack: a totally blocked coronary artery leads to a heart attack. Some of the archaic signs of heart attack are; pressure in one's chest, and a sharp pain in the arm or shoulder, at times with gasp for breath, and profuse sweating.

2.2.2 Risk Factors That Cause Coronary Heart Disease

The factors responsible for coronary heart disease include:

i. Age: as one gets older, the chances of damaged and narrowed arteries increases
ii. Sex: people generally at risk of coronary heart disease are mostly men. Women mostly have it after menopause.
iii. Family history: heart disease prominent in a the history of a family is associated with a higher risk of heart disease, mostly in cases whereby a close relative or member of the family have developed heart disease at an early stage. If ones father or brother was diagnosed with heart disease before the age 55 and the mother or sister was diagnosed of it before the age 65, one is at greater risk of having coronary heart disease.
iv. Smoking: people who smoke have a higher risk of heart disease. Also those who secondary passers (that is that inhale the smoke from the original smokers)
v. High blood pressure: high blood pressure that isn't addressed or controlled can lead to the hardening of the arteries.
vi. High blood cholesterol level: too much intake of fats results in high cholesterol, and high cholesterol in the blood increases the risk and accumulation of plaques atherosclerosis in the arteries.
vii. Diabetes: this increases the risk of coronary artery disease.
viii. Overweight and obesity: this gives strength to other risk factors
ix. High stress: this may damage the arteries and give strength to other risk factors
x. Unhealthy diet: too much food that has high unsaturated fats increases the chances of coronary heart disease.

2.2.3 Connection between Genetics and Coronary Heart Disease

It is established that coronary heart disease has a high hereditary component. People with one or more close relatives who have or had early coronary heart disease (CAD) are at an increased risk for CAD. For men, early CAD is being diagnosed before age 55, for women, early CAD is being diagnosed before 65.(bill Amos, JD et-al,2017).

For example, third-degree relatives such as first cousins share about 17.5% of their genes, and show a risk of 2% for developing coronary heart disease. Second-degree relatives such as half-siblings share about 25% of their genes and show a risk of 6%. Most first-degree relatives such as siblings share about 50% of their genes and show a risk of about 9%. Monozygotic twins (MZ) share 100% of their genes and show risks near 50%.

Coronary heart disease as a disease has a combination of genetic and environmental components, this explains the reason why the concordance rate for monozygotic twins is not 100% and the concordance rate for dizygotic twins is less than half of the monozygotic rate.

While several forms of molecular genetic analysis are available, linkage analysis gives a particularly versatile procedure that helps explain the familial basis of coronary heart disease. Linkage analysis makes use of events that occur during meiosis, when chromosomes cross over and exchange segments of deoxyribonucleic acid (DNA). Genetic loci that are closer together have a greater probability of being inherited than do loci that are further away. The probability of identifying a disease gene is increased if it co-segregates, this means being linked with an allele whose chromosomal location can be identified among members of a family.

2.2.4 Types of Coronary Heart Disease

Coronary heart disease can be classified into five subtypes which are:

1. Stable Angina: this is also called angina pectoris. This is a sharp pain or pressure to the chest due to inadequate blood supply to the heart muscle.

2. Unstable Angina: this is a type of angina irregular in nature

3. Myocardial infarction: this is generally known as heart attack. This occurs when there is little or no flow of blood to a part of the heart, causing injury to the heart muscle.

4. Sudden Cardiac death: is the sudden loss of blood supply cause from the failure of the heart to pump effectively. Symptoms include loss of breath and abnormal breathing.

2.2.5 Diagnosis of coronary heart disease

A diagnosis of coronary heart disease is made based on full physical examination and order routine blood tests, diagnostic tests also which are as follows;

1). Electrocardiogram: this takes reading of electrical signals as they travel through the heart. An ECG can make known to the examiner the patient previous heart attack or one that is to come.

2). Echocardiogram: this uses sound waves to show images of the heart. During this process, a GP can determine whether or not all parts of the heart are contributing to the heart’s supply of blood.

3). Stress test: if the signs and symptoms of coronary heart disease occur during mostly when one exercises. The doctor may ask the patient to either work on a treadmill or ride a stationary bike during ECG. Another type of stress test is called the nuclear stress test. This measures the blood flow to the heart muscle at rest and during stress.

4). Cardiac catheterization and angiogram: to record the blood flow to the heart, the doctor may inject a special dye to the coronary arteries; this is referred to as angiogram.The dye is injected into the heart arteries through a long, thin, flexible tube that is threaded through an artery, usually in the leg, to the arteries in the heart.

5). Heart scan: technologies like computerized tomography (CT) can help doctor see calcium deposits in the arteries that can block the arteries. If a little amount of calcium is discovered, the person may have coronary heart disease.

2.2.6 Treatment of coronary heart disease

Treatment of coronary heart disease mostly involves lifestyle change, drugs and particular medical procedures (if necesaary).

Lifestyle changes include;

i. Ability to quit smoking
ii. Consumption of healthy foods
iii. Regular exercise

iv. Loss of excess weight
v. Reduction of stress

Different types of drugs that can be used to treat coronary heart disease include;

1. Cholesterol modifying medications: this is by lowering the amount of cholesterol in the blood, particularly low-density lipoprotein cholesterol. These drugs reduce the primary material that accumulates on the coronary arteries. The doctor chooses from a wide range of medication, which includes statins, niacin, fibrates, bile acid sequestrants etc.
2. Aspirin: the doctor may prescribe daily taking of aspirin or blood thinner. This reduces blood clotting tendency, which may help prevent obstruction of the coronary arteries.
3. Beta blockers: these drugs slow down the heart rate and reduce blood pressure. This reduces oxygen demand by the heart.
4. Calcium channel blockers: these drugs may be used with beta blockers, if beta blockers aren't effective alone, or they are used instead of beta blockers, if you are not able to take beta blockers. They help strengthen symptom of chest pain.
5. Ranolazine: this medication may work with chest pain. It may be given with beta blocker, or given instead of beta blocker, if one can't take it.

Medical procedures used includes;

1. Angioplasty and stent placement (percutaneous coronary revasculation): the doctor inserts a long, thin tube into the blocked part of the artery. A wire with a deflated balloon is passed into the blocked area. The balloon is then inflated, compressing the accumulations again the artery walls.

   A stent is mostly left in the artery to help keep the artery open. Most stents slowly reduce the medication to help keep the arteries open.
2. Coronary artery bypass surgery: A surgeon creates a graft to bypass blocked coronary arteries using a vessel from another part of your body. This allows blood to flow around the blocked or narrowed coronary artery. Because this requires open-heart surgery, it's most often reserved for cases of multiple narrowed coronary arteries.

2.3 Overview of CDSSs

The current computer-based patient record systems include CDSS but older CDSSs are either standalone or part of a non-commercial computer-based patient record system (Berner, 2009). Research has showed that as a result of healthcare organizations utilizing CDSSs there is enhanced outcome in the performance of the clinician as well as better patient outcome (Menachemi&Collum, 2011). Presently, about 15-20% of doctor offices use Electronic Medical Records (EMR), and as more doctors use EMRs, the chance for CDSS integration within the healthcare setting will increase (Miller, 2010). EMRs are considered to be the foundation for providing quality healthcare because EMRs are designed to improve both accessibility and legibility of information (Berner, 2009). In order to achieve the full potential of EMR, CDSS is essential.

2.3.1 Target Area of care

- Prevention of such disease through screening and disease management
- Diagnosis done based on the patients signs and symptoms
- Follow-up management which has to do with frequent checkups
- Provide clinicians, staffs and other individuals with knowledge and specific information

2.3.2 System Design

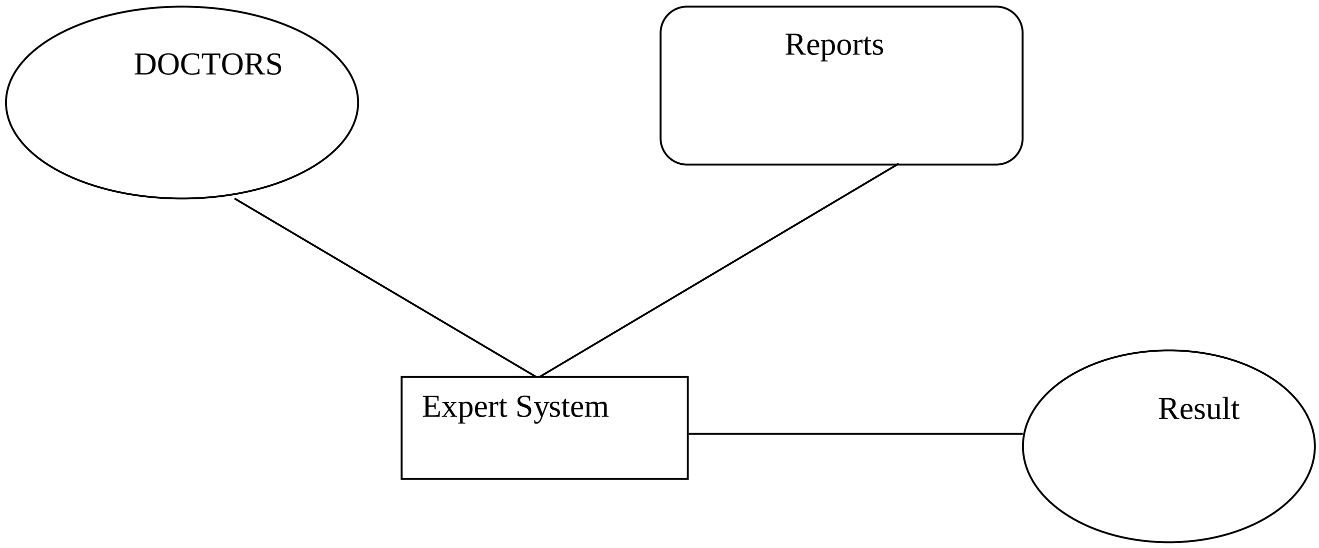

Figure 2.1 Architecture of Proposed System

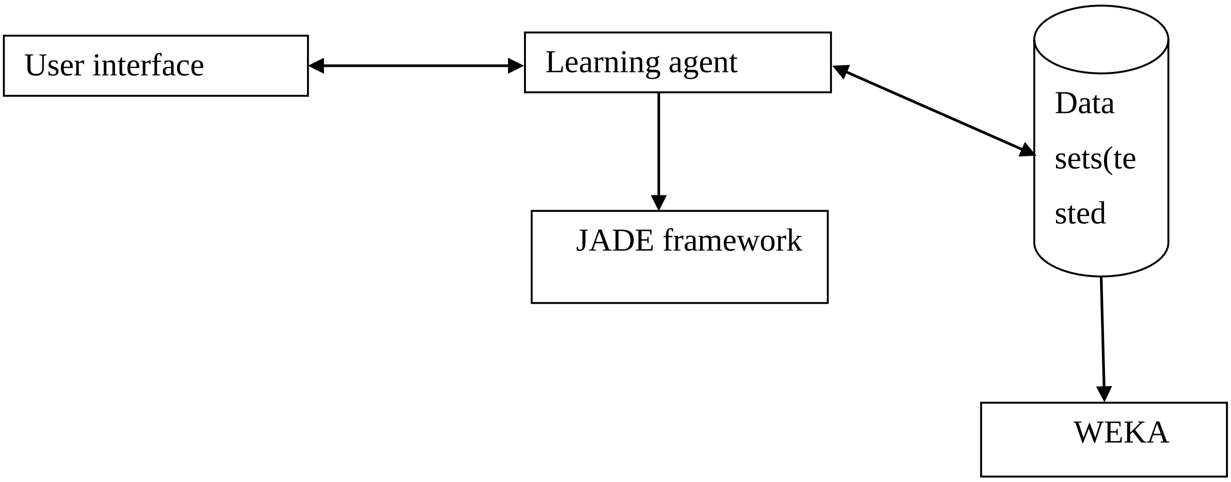

Figure 2.2 Processing view of the system

Advantages of System Design

- Prediction of coronary heart disease is obtained
- Less complexity is obtained

2.3.3 Factors leading to successful CDSS implementation

The following under listed factors lead to the successful implementation of CDSS:

- Simple, user friendly interface
- Automated decision support
- Timely result
- Workflow integration
- Continuous Knowledge-base and update support (Peleg

2.3.4 Agent-based Systems

An example of an agent based system is the expert system. It consists of components which include the knowledge-base (rules), inference engine, working memory (facts) and a user interface. The knowledge-base can be the production rules that are in form of "IF condition THEN action". The condition portion can be called antecedent and it represents a particular fact. If some facts are in the database then the ac tion is performed. The action portion is called consequent.

The knowledge-base is responsible for storing factual and heuristic knowledge using one or more knowledge representation schemes to express knowledge about the application domain of interest. The inference engine is the mechanism that is used to manipulate symbolic information and knowledge so as to solve problems through reasoning. The inference engine determines which rule antecedent are satisfied by the fact. The working memory houses the data that is specific to the problem being solved. The individuals who interact with the system include the domain expert (the human expert whose domain knowledge is being sought), knowledge engineer (the one who encodes the expert's knowledge in a declarative form that can be used by the expert system), system engineer (the one who builds the user interface, designs the declarative format of the knowledge-base and implements the inference engine) and the user (the one who will consult the system to get advice as would have been provided by the expert)

2.3.5 Design considerations and goals

- CDSS must be simple
- User friendly
- Must give timely result
- Must cost low
- Must support continuous knowledge

2.4 Pattern Classification Methods

Pattern classification refers to the theory and algorithms of assigning abstract objects into distinct categories, where these categories are typically known in advance. For this research, the pattern classification methods considered are Decision Trees (DTs), K-Nearest Neighbor (KNN), Naïve bayes Classifier and Support Vector Machine (SVM).

2.4.1 Decision Trees

A decision tree is a predictive model which is a mapping from observation about an item to conclusion about its target value (Lin Tan, 2015). In the tree structures, leaves represent classification, non-leaf nodes are features and branches denotes conjunction of features that lead to the classification. The classification of decision trees are carried out in two phases:

- Tree Building or top down: This is computationally intensive and requires the tree to be recursively partitioned until all the data items belong to the same class.
- Tree pruning or bottom top: It is conducted to improve the prediction and classification of the algorithm and minimize the effects of over-fitting which may lead to misclassification of errors (Anyanwu & Shiva, 2009).

Some notable decision tree algorithms include Classification and Regression Trees (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0.

The advantages of decision trees include:

- They are easy to interpret and comprehend
- They can handle both metric and non-metric data as well as missing values which are frequently encountered in clinical studies.
- Little data preparation is required since data does not need to be normalized.
- They can handle data in a short time frame.
- They can be developed using common statistical techniques.

The disadvantages associated with decision trees include:

- They can over fit the data and create complex trees that may not generalize well.
- A small change in the size of a dataset could result in a completely different tree

2.4.2 K-Nearest Neighbour

The idea in K-Nearest Neighbour (k-NN) is to dynamically identify k observation in the training data set that are similar to a new observation. It constructs a classification technique using no assumption about the form of the function, y = f(x1,x2....xp) that relates the independent variable y, to the independent variable x1,x2....xp. It is also an instance based learning for classifying objects based on closest training examples in the feature space. It is a type of lazy learning where the function is only approximated locally and all computations are deferred until classification. The benefit is that higher values of k provide smoothing that reduces the risk of over fitting due to the noise in the training data. The k-nearest neighbour algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours. The simplest case k=1, where we look for the observation that is closest (the nearest neighbour), and set v=y, where y is the class of the nearest neighbour.

2.4.3 Bayes Classifier

The Naive Bayes classifier is a typical and popular example of a suboptimal classifier. The primary assumption is that the components in the feature vector are statistically independent. It greatly simplify learning by assuming features are independent given class. Although independence is generally a poor assumption, in practice Naive Bayes often competes well with more sophisticated classifiers. Bayesian classifiers assigns the most likely class to a given example described by its feature vector. A Bayesian network is a model that encodes probabilistic relationships among variables of interest. This technique is generally used for intrusion detection in combination with statistical schemes, a procedure that yields several advantages, including the capability of encoding interdependencies between variables and of predicting events, as well as the ability to incorporate both prior knowledge and data.

2.4.4 Support Vector Machine

SVMs were introduce by Boser Guyon Vanik in 1992. SVMs are now important and active field of all machine learning research and are regarded as a main example of kernel methods. SVM is a family of machine learning that are used for mathematical and engineering problems including for example handwriting digit recognition, object recognition, speaker identification, face detection in images and target detection. SVM performs classification by building an N-dimensional hyperplane that optimally separates the data into two categories.

2.5 Review of Existing Related CDSSs

2.5.1 A Clinical Decision System For External Beam Radiation Oncology For Prostate Cancer

Here a methodology which can be utilized to select proton or photon radiotherapy in prostate cancer patients. Four state-of-the-art competing treatment modalities were compared (by way of an in silico trial) for a cohort of 25 prostate cancer patients, with and without correction strategies for prostate displacements. Metrics measured from clinical image guidance systems were used. Three correction strategies were investigated; no-correction, extended-no-action-limit, and online-correction. Clinical efficacy was estimated via radiobiological models incorporating robustness (how probable a given treatment plan was delivered) and stability (the consistency between the probable best and worst delivered treatments at the 95% confidence limit). The results obtained at the cohort level enabled the determination of a threshold for likely clinical benefit at the individual level. Depending on the imaging system and correction strategy; 24%, 32% and 44% of patients were identified as suitable candidates for proton therapy. For the constraints of this study: Intensity-modulated proton therapy with online-correction was on average the most effective modality. Irrespective of the imaging system, each treatment modality is similar in terms of robustness, with and without the correction strategies. Conversely, there is substantial variation in stability between the treatment modalities, which is greatly reduced by correction strategies. This study provides a ‘proof-of-concept’ methodology to enable the prospective identification of individual patients that will most likely (above a certain threshold) benefit from proton therapy.

2.5.2 Decision Support System for the Diagnosis of Schizophrenia Spectrum Disorders.

In 2013, Kahn built up a choice emotionally supportive network for the conclusion of schizophrenia range issue. The advancement of this framework is depicted in four-phases: learning procurement, information association, the improvement of a PC helped model, and the assessment of the framework's presentation. The information is extricated from a specialist through open meetings. These meetings went for investigating the master's analytic basic leadership process for the conclusion of schizophrenia. A diagram strategy was utilized to recognize the components associated with the thinking procedure. Learning was first composed and displayed by methods for calculations and after that moved to a computational model made by the covering approach. The presentation appraisal included the examination of the finding of 38 clinical vignettes between a specialist and the choice emotionally supportive network. The outcomes demonstrated a moderately low rate of misclassification (18-34%) and a decent presentation by the choice emotionally supportive network in the determination of schizophrenia, with an exactness of 66-82%.

2.5.3 A Novel Analysis of Diabetes Mellitus By Using expert system Based on Brain derived neurotropic Factor (BDNF)) Levels.

In 2013, Deverapalli, Apparo, Kumar and Sridhar proposed a novel concept of designing and building intelligent experts systems for the detection and diagnosis of Diabetes Mellitus. The expert system is based on critical diabetic parameters like Brain Derived neurotropic factor levels, and Fasting Blood Glucose (FBG). The proposed rule based expert system constructs large scale knowledgebase based on the behaviour of BDNF related diabetic Data. The system will give an expert decision taking into consideration all the valid ranges of diabetic parameters. The proposed expert system can work effectively even for large sets of patients data. The gap in this work is that the minimum size of patient datasets which can be processed at once by expert systems was not specified.

2.5.4 Comments on the Existing CDSSs

In all the cases cited in the review of related works, one salient feature remained paramount and that is the fact that the ability of the different CDSSs to optimally carry out expected tasks was predicated on the ability of the system developers to represent the knowledge of the expert domain in such a way that the decision reached by the system tended towards what is obtainable in reality.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

Research technique has to do with the methods important to methodically proffer well-ordered depiction of how answers for complete expressed destinations of an exploration work will be done. It is some of the time seen as attempting to see how research is experimentally completed. It traces the different advances that are carefully embraced by the analyst in accomplishing the destinations of the exploration alongside the rationale behind the received strategies.

3.1 Project Design

1. Perform relative investigation of example characterization calculations on fitting datasets utilizing Waikato Environment for Knowledge Analysis (WEKA).
   i. Gain significant datasets to be utilized for characterization tests. In this examination the Duke 2007dataset was utilized.
   ii. Set up the obtained dataset to a configuration distinguished by WEKA. In this examination the Attribute Relation File Format (ARFF) was utilized.
   iii. Prompt the picked dataset with grouping calculations (Decision trees, K-Nearest Neighbour, Bayes Classifier and Support Vector Machine) and record their exhibition.
   iv. A ten times cross approval technique is utilized in assessing the grouping calculations.
2. Recognizable proof of the considerable number of modules that will make up the system by broad and comprehensive assessment of fundamentally the same as savvy structures.

ii. Recognizable proof of the considerable number of segments that will make up every module subsequent to deciding the segments that will be required for the separate modules to do their capacities.

iii. Consideration of the example arrangement calculation picked after assessment in its fitting position in the structure.

iv. Proper association of the different parts of the system to one another so as correctly predict CHD

3.2 Project Tools

3.2.1 Weka (Version 3.6.7)

Our examinations will be finished utilizing Weka 3.6.7. Weka (Waikato Environment for Knowledge Analysis) is a prominent suite of AI programming written in Java, created at the University Of Waikato, New Zealand. Weka bolsters a few standard information mining assignments, all the more explicitly, information pre-handling, grouping, order, and relapse, perception, and highlight choice.

Weka (Waikato Environment for Knowledge Analysis) (Witten and Frank, 2005) is maybe the best-realized open-source AI and information mining condition (Zupan &Demsar, 2008). Propelled clients can get to its parts through Java programming or through an order line interface. For other people, Weka gives a graphical UI in an application called the Weka Knowledge Flow Environment highlighting visual programming, and Weka Explorer giving a less adaptable interface that is maybe simpler to utilize. The two conditions incorporate Weka's amazing cluster of AI and information mining calculations. Weka's people group has additionally built up a lot of expansions covering differing regions, for example, content mining, perception, bioinformatics, and network registering.

3.3 Knowledge Identification

A noteworthy test to learning obtaining is the idea of information in the two its unequivocal and inferred structures. Unequivocal learning which is generally founded on printed data which by enormous isn't organized, arranged or composed in a way that is helpful for the information elicitation process for frameworks philosophy improvement, but instead for the area master. Henceforth, this would require a considerable measure of clarification and explanation from the authority cardiologists and the learning designer would need to complete a broad audit to increase comprehension of the issue territory and explain the phrasing of the issue space to build up the applicable pathways utilized in the framework.

Learning ID identifies with the way toward distinguishing wellsprings of patient administration guides for coronary illness so as to build up a Clinical Pathway (CP) for analytic, prescient and preventive cases to manage coronary illness. A noteworthy asset to encourage information recognizable proof incorporates the Clinical Practice Guidelines (CPGs) for coronary illness, master surveys, talks and perceptions.

3.4 Knowledge Acquisition

Information securing is a significant period of frameworks improvement for distinguishing and assembling master space learning that is joined into the prerequisites and structure of the framework. There are issues guaranteeing that exact and valuable information is caught at first and moved to later improvement endeavours (Ritcher, Miller and Funk, 2004). What's more, information securing is one of the most troublesome and blunder inclined errands that Knowledge architects take part in while building a learning based framework. The expense and execution of the application depends straightforwardly on the nature of the information procured (Rhem, 2002).

In this period of the system, data on the activities/methods utilized by pro cardiologists in the treatment of coronary illness both in its demonstrative, prescient and preventive structures are gathered from Experts just as applicable archives and manuals particularly the CPGs. The customary way to deal with information obtaining is utilized to catch both express and implied learning of the space zone. Information obtaining itself is best portrayed as a relationship of intercession between specialists or metaphysics manufacturers and the specific devices utilized (Hacking, 1983).

3.5 Knowledge Synthesis

Information union abridges all relevant investigation on a particular inquiry, it can improve the comprehension of irregularities in differing proof, and it can spot holes in research proof to characterize future research plans. Exercises of learning combination in medicinal services have generally centred on methodical audits of mediations. Different various learning amalgamation exist in the writing over numerous controls yet finding these, particularly for subjective research, present difficulties.

Information union generally alluded to as 'second era learning', speaks to the total of existing learning, a rundown of the restorative writing that utilizations unequivocal strategies to play out an exhaustive writing search just as a basic examination of individual investigations, and fitting factual procedures to join these substantial examinations (KT-CIHR, 2011).

Information amalgamation is a social just as an individual procedure. Sharing unsaid information expects people to share their own involvement and convictions about a circumstance with others. By then of sharing, legitimization winds up open. Every individual is looked with the huge test of supporting his or her convictions before others – and it is the requirement for avocation, clarification, influence and human association that makes information union a profoundly delicate procedure. To bring individual information into a social setting, inside which individual points of view are enunciated, and clashes are settled in the development of higher-level ideas. In a common information area, the field for communication is regularly given as a self-ruling, self-coordinated work group, made of individuals from various units.

It is a basic issue to choose when and how to build up such a group of connection wherein people can meet and collaborate. This group triggers space learning union primarily through a few stages. In the first place, it encourages the structure of common trust among individuals, and quickens the formation of a certain point of view shared by individuals as inferred learning. The key factor for this progression is sharing knowledge among individuals. Second, the common understood viewpoint is conceptualized through continuous discourse among individuals. The prevailing method of information transformation here is externalization. Implicit field-explicit viewpoints are changed over into express ideas that can be shared past the limit of the group. Exchange legitimately encourages this procedure by initiating externalization at the individual levels. It is a procedure wherein one forms ideas in collaboration with others. It gives the chance to one's speculation or suspicion to be tried. As Markova and Foppa (1990) contended, social intercourse is one of the most dominant media for checking one's own thoughts. Thusly, members in the exchange can take part in the common co-advancement of thoughts. Next comes the progression of defense, which is the procedure of assembly and screening, which decides the degree to which the information made inside the group is genuinely beneficial for the association. Regularly, an individual legitimizes the veracity of his or her convictions dependent on perceptions of the circumstance; these perceptions, thus, rely upon an interesting view3point, individual reasonableness, and individual experience. When somebody makes information, the individual in question bodes well out of another circumstance by holding legitimized convictions and focusing on them. Under this definition, learning is a development of reality instead of something that is valid in any unique or all-inclusive way. The formation of learning isn't just an accumulation of realities yet a one of a kind human procedure that can't be diminished or effectively recreated. It can include emotions and conviction frameworks of which we may not be cognizant. All things considered, legitimization must include the assessment gauges for making a decision about honesty. There might likewise be worth premises that rise above accurate or even minded contemplations. The instigations to start a union of information might be various and subjective instead of straightforward and quantitative measures. At long last, we land at the phase of cross-levelling information. During this stage, the idea that has been made and advocated is incorporated into the information base of the association, which involves an entire system of authoritative learning (Vat, 2003).

3.6 Developing the Results Classification/Prediction Algorithms

Non-linearity, unpredictability and fluffy communications are developing highlights of endless degenerative sicknesses which record for most bleakness and mortality in our contemporary world. Lamentably, even the most dominant and entrenched measurable strategies which were created in the primary portion of the previous century when the situation was overwhelmed by intense infective maladies and the accessible data was a lot more straightforward, or at greatest "confused" instead of "complex" in examination with today (Grossi, 2011). The manner by which results from the investigation of these mind boggling sets of data are made utilize must be easy to utilize however top to bottom. Thus, a versatile framework that can without much of a stretch change the intricate idea of these information must be utilized.

3.6.1 Decision Trees (DTs)

DTs are a basic, yet ground-breaking type of various variable examinations. They give one of a kind capacities to enhance, supplement and substitute for:

- Traditional statistical types of examination, (for example, numerous straight relapse)
- An assortment of information mining apparatuses and procedures, (for example, neural systems)
- Recently created multidimensional types of revealing and examination found in the field of business knowledge. A DT is a choice help device that uses a tree-like chart or model of choices and their potential results, including chance occasion results, asset expenses, and utility. A DT is a sort of calculation where a firmly related impact chart is utilized as a visual

and explanatory choice help device, and the normal qualities (or anticipated utility) of contending options are thusly determined.

A DT comprises of 3 sorts of hubs:

1. Choice hubs – regularly spoken to by squares

2. Chance hubs – spoken to by circles

3. End hubs – spoken to by triangles

DTs are regularly utilized in tasks explore, explicitly in choice investigation, to help distinguish a technique destined to achieve an objective. On the off chance that practically speaking choices must be taken online with no review under fragmented information, a DT ought to be paralleled by a likelihood model as a best decision model or online determination model calculation. Another utilization of DTs is as an unmistakable methods for ascertaining contingent probabilities. A DT comprises of a root hub, branch hubs and leaf hubs. The tree begins with a root hub and is additionally part into branch hubs (every one of the hubs speak to a decision of different choices), and ends with a leaf hub which are un-part hubs that speak to a choice (Peng, 2006). Characterization of DTs are led in two stages, including the tree building (top down) and tree pruning (base up). Tree building is computationally serious, and requires the tree to be recursively parcelled until every one of the information things have a place with a similar class. Tree pruning is led to improve the forecast and order of the calculation and decrease the impacts of over-fitting that can prompt misclassification of blunders (Anyanwu and Shiva, 2009). There are various DT calculations that exist including Classification and Regression Trees (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0. This postulation work utilizes C4.5 based DT calculation which is an improvement over the ID3 technique. In this model, the DT gives the outcome correlation and streamlining. DTs are a non-parametric directed learning technique utilized for arrangement and relapse. The objective is to make a model that predicts the estimation of an objective variable by taking in straightforward choice standards construed from the information highlights. Information mining is tied in with separating designs from an association's put away or warehoused information. These examples can be utilized to pick up knowledge into parts of the association's activities, and to foresee results for future circumstances as a guide to basic leadership. Examples frequently concern the classifications to which circumstances have a place. For instance, is an advance candidate financially sound or not? Will a specific section of the populace overlook a mail out or react to it? Will a procedure give high, medium, or low yield on a bunch of crude materials?

3.6.2 K-Nearest Neighbor

K-Nearest Neighbour (k-NN) is occurrence based learning for arranging items dependent on nearest preparing models in the element space. It is a kind of languid realizing where the capacity is just approximated locally and all calculations are conceded until characterization. The k-closest neighbour calculation is among the least difficult of all AI calculations: an article is characterized by a lion's share vote of its neighbours, with the item being allocated to the class most basic among its k closest neighbours. In the event that k=1, at that point the article is just relegated to the class of its closest neighbour. The k-NN calculation utilizes all named preparing occasions as a model of the objective capacity. During the order stage, k-NN utilizes a comparability based inquiry system to decide a locally ideal speculation work. Test occurrences are contrasted with the put away examples and are allocated a similar class name as the k most comparative put away occasions.

3.6.3 Bayes Classifier

A Bayesian system is a model that encodes probabilistic connections among factors of intrigue. This strategy is commonly utilized for interruption discovery in mix with factual plans, a system that yields a few focal points, including the capacity of encoding interdependencies among factors and of foreseeing occasions, just as the capacity to consolidate both earlier information and information. In any case, a genuine weakness of utilizing Bayesian systems is that their outcomes are like those gotten from limit based frameworks, while extensively higher computational exertion is required.

3.7.4 Support Vector Machine

Bolster Vector Machines have been proposed as a novel system for interruption identification (Mukkamala et al, 2002). A SVM maps input (genuine esteemed) include vectors into a higher-dimensional component space through some nonlinear mapping. SVMs are created on the standard of basic hazard minimization (Valdimir, 1995). Auxiliary hazard minimization tries to discover a speculation (h) for which one can discover most reduced likelihood of mistake though the customary learning methods for example acknowledgment depend on the minimization of the observational hazard, which endeavour to advance the presentation of the learning set. Processing the hyper plane to isolate the information focuses for example preparing a SVM prompts a quadratic streamlining issue. The usage of SVM interruption discovery framework has two stages: preparing and testing. SVMs can get familiar with a bigger arrangement of examples and have the option to scale better, in light of the fact that the order unpredictability does not rely upon the dimensionality of the component space. SVMs additionally can refresh the preparation designs powerfully at whatever point there is another example during grouping.

CHAPTER FOUR

RESULT AND RESULT DISCUSSION

4.0 Analysis of Existing Framework

A comparative analysis of four pattern classification algorithms was carried out. The algorithms were trained on 20% of the Duke 2007 data set using a ten-fold validation test mode in a Weka (Waikato Environment for Knowledge Analysis) environment. The result of the experiment is reported in table 5.1.

Table 4.1: Comparison of different Classification Algorithms on Weka (Source: Adaptive, 2014)

Algorithm	**Method Name**	Correctly Classified Instances in 20% of Full Dataset (%)	**Incorrectly Classified Instances in 20% of Full Dataset (%)**	Time Taken To Build Model (seconds)
SVM (Support Vector Machine)	functions.SMO	51.0314 49.9485 50.0515	51.0314 49.9485 50.0515	59.56
KNN (K-Nearest Neighbor)	lazy.IBk	49.0100	44.5	0.08
NB (Naïve Bayes)	bayes.NaiveBayes	44.0510	48.2123	0.97
DT (Decision Tree)	trees.J48	60.0102	36.333	9.03

4.1.1 National Service Framework for Coronary Heart Disease on Treatment and Outcome Of Patients With Acute Coronary Syndromes

The investigation depended on patients with ACS or myocardial localized necrosis admitted to the coronary consideration units of the Royal London and Newham General clinics. The NSF came into task on 1 April 2000. We analyzed information for patients conceded in the 27 months before the NSF usage with information for patients conceded in the 21 months a short time later. Benchmark clinical information were gathered tentatively and put away electronically as beforehand described. Information recorded included patient statistic information, heart history, hazard factors for CHD, ECG highlights (starting and consequent), crisis treatment, confusions, and further examinations masterminded. Optional anticipation measures, release drugs, and follow up courses of action were additionally archived. A conclusion of diabetes was recorded if the patient required insulin, oral hypoglycaemic specialists, or dietary sugar confinement. The determination of left ventricular disappointment (LVF) was recorded for patients treated with diuretics with radiological proof of interstitial or alveolar pneumonic oedema or indications of dyspnoea related with basal inspiratory crepitations or a third heart sound.

4.1.2 Statistical methods

Patient characteristics and clinical outcomes before and after introduction of the NSF were compared by *t*tests or Mann-Whitney U tests for continuous variables and by χ^2 or Fisher's exact test for categorical variables. Logistic regression models were used to analyse trends in treatment over time. Patients were grouped into nine categories of three month according to their time of admission. This variable was fitted as a linear constant to determine overall trend and trends before and after the introduction of the NSF. An interaction term was fitted to assess any differences in trend before and after introduction of the NSF. Where no trend was observed before introduction of the NSF, rates before and after were directly compared. Where trends were observed, rates after the NSF was introduced were compared with those expected if the pre-NSF trend had continued.

4.1.3 Results

Information from 3371 patients were recorded, 1993 patients in the 27 months before the presentation of the NSF and 1378 patients in the two years a while later. After the presentation of the NSF in-hospital mortality was essentially decreased (95 patients (4.8%) v 43 (3.2%), $p = 0.02$). This was related with a decrease in the advancement of Q wave myocardial localized necrosis (40.6% v 33.3%, $p < 0.0001$) and in the occurrence of left ventricular disappointment (15.9% v 12.3%, $p = 0.003$). The extent of patients getting thrombolysis expanded (69.4% v 84.7%, $p < 0.0001$) with a decline in the time taken to get it (extent thrombolysed inside 20 minutes 12.1% v 26.6%, $p < 0.0001$). The solution of β blockers (51.9% v65.8%, $p < 0.0001$), angiotensin changing over catalyst inhibitors (37% v 66.4%, $p < 0.0001$), and statins (55.2% v 72.7%, $p < 0.0001$) improved and the extent of patients alluded for obtrusive examination expanded (18.3% v 27.0%, $p < 0.0001$). Pattern investigation demonstrated that enhancements in mortality and thrombolysis were legitimately connected with distribution of the NSF, though the upgrades found in medicine of β blockers and statins were the continuation of pre-existing patterns.

4.2 Performance of the Classification Models

4.2.1 Performance of K-Nearest Neighbour Model

Table 4.2: Summary result of K-Nearest Neighbour absolute error analysis

Variable	10-fold
Correctly classified instances	44.0510
Incorrect classified instances	48.2123
Kappa statistics	0.0025
Mean absolute error	0.4988
Root mean squared error	0.7062

The exhibition of K-Nearest Neighbour model utilizing 10-overlay cross approval demonstrates that 50.1235% of the examples in the Duke 2007 dataset were effectively ordered while 49.8765% were inaccurately grouped. The Kappa insights uncover that the arrangement of this model is 0.25% solid with a mean total mistake of 49.88%. The root mean squared blunder (RMSE)- a proportion of the contrasts between qualities anticipated by the model is under 71%.

4.2.2 Performance of Bayes Classifier Model

Table 4.3: Summary result of Bayes Classifier absolute error analysis

Variable	10-fold
Correctly classified instances	49.9485
Incorrect classified instances	48.7955
Kappa statistics	0.0043
Mean absolute error	0.5
Root mean squared error	0.5001

The exhibition of Bayes Classifier model utilizing 10-overlap cross approval demonstrates that 50.2045% of the occurrences in the Duke 2007 dataset were accurately characterized while 49.7955% were erroneously ordered. The Kappa measurements uncover that the grouping of this model is 0.43% dependable with a mean outright mistake of half. The root mean squared blunder (RMSE) - a proportion of the contrasts between qualities anticipated by the model is under 51%.

4.2.3 Performance of Support Vector Machine Model

Table 4.4: Summary result of Support Vector Machine absolute error analysis

Variable	10-fold
Correctly classified instances	51.0314
Incorrect classified instances	49.9485
Kappa statistics	0.0009
Mean absolute error	0.4995
Root mean squared error	0.7067

The performance of Support Vector Machine model using 10-fold cross validation shows that 50.0515% of the instances in the Duke 2007 dataset were correctly classified while 49.9485% were incorrectly classified. The Kappa statistics reveal that the classification of this model is 0.09% reliable with a mean absolute error of 49.95%. The root mean squared error (RMSE)-a measure of the differences between values predicted by the model is less than 71%.

4.2.4 Performance of C4.5 Decision Tree Model

Table 4.5: Summary result of C4.5 Decision Tree absolute error analysis

Variable	10-fold
Correctly classified instances	61.0734
Incorrect classified instances	38.9266
Kappa statistics	0.221
Mean absolute error	0.3955
Root mean squared error	0.4451

The exhibition of C4.5 Decision Tree model utilizing 10-overlay cross approval demonstrates that 61.0734% of the cases in the Duke 2007 dataset were accurately ordered while 38.9266% were inaccurately characterized. The Kappa insights uncover that the characterization of this model is 22.1% solid with a mean total blunder of 39.55%. The root mean squared blunder (RMSE)- a proportion of the contrasts between qualities anticipated by the model is under 45%.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

A viable strategy for decreasing the hazard of CHD is to draw in GPs in the administration of CHD since GPs are the primary purpose of consideration. There are difficulties in the determination of CHD given that a significant number of its clinical highlights are time-explicit and now and then non clear however present. CPGs and CPs have the affinity to decrease this hole. They can help GPs to embrace complex indicative and the executives situations identified with CHD. In perspective on the previous we have planned an improved structure that upgrades the grouping/forecast of schizophrenia that carries with it versatility which aides patients with CHD concerning how to best deal with their wellbeing condition and carry on with a typical life.

5.2 Recommendations and Suggestion for Further Studies

Having constructed the prescient structure, full execution will be completed as pursues to:

I. construct legitimate and helpful information based CPs (rules/system) through the securing of pertinent datasets especially those identified with CHD

ii. Model the demonstrative and treatment ideas of CHD indicating between connections in formal language in manners that control each type of conceivable uncertainty. This is done so as to guarantee that the encoded learning and the basic choice rationale can be executed through mechanized clinical choice emotionally supportive networks to give persistent explicit CPG-based proposals.

5.3 CONTRIBUTIONS TO KNOWLEDGE

This examination offers the accompanying commitments to learning:

I. Cautious layout of a stepwise portrayal of how to assemble substantial and valuable information based CPs through learning recognizable proof, procurement, amalgamation, formalization and arrangement of applicable datasets much the same as CHD.

ii. The structure of a versatile system that is symptomatic, preventive and prescient in degree. This implies the framework when operationalized will foresee with attractive precision, the inclination of a patient to create CHD. Moreover, because of the way that this structure is nonexclusive in its viewpoint, it very well may be adjusted for other perilous illnesses of enthusiasm inside the examination network.

REFERENCES

Berner, E.S. (2009). Clinical Decision Support Systems: State of Art. Rockville: AHRQ Publication.

Frize, M. (2005). Conceptual Framework of Knowledge Management for Ethical Decision Making Support in Neonatal Intensive Care. IEEE Transactions of Information Technology in Biomedicine, 9, 205-215.

Gilchrist, J. (2012). Clinical Decision Support System Using Real-Time Data Analysis for a Neonatal Intensive Care Unit. PhD. diss. Carleton University.

Gordon, C., Johnson, P., Waite, C., & Veleso, M. (1997). Algorithm and Care Pathway: Clinical Guidelines and Healthcare Processes. In E.T Keravnou, C. Garbay, R.H. Baud, and J.C. Wyatt, (Eds). Lecture Notes in Computer Science 1211 (pp.66-69). Springer-Valag, London.

Hacking, I. (1983). Representing and intervening: Introductory Topics in the Philosophy of Natural science. Cambridge Cambridgeshire; New York, Cambridge University Press.

Jha, G.K. (2011). Artificial Neural Networks.http://www.iasri.res.in/ebook/EB_SMAR/e-book_pdf%20files/Manual%20IV/3-ANN.pdf

Matthew S. Simon, David P. Calfee, in Infectious Diseases (Fourth Edition), 2017.

N. Friedman, D. Geiger, and Goldszmidt M. Bayesian network classifiers. Machine Learning, 29:131–163, 1997.

Witten. I.H, Frank E. (2005) ”Data mining: practical machine learning tools and techniques with Java implementations". 2nd edition. San Francisco (CA): Morgan Kaufman.

Vadicherla, D. & Sonawane, S. (2013). Decision support system for heart disease based on sequential minimal optimization in support vector machine. International Journal of Engineering Sciences & Emerging Technologies, Feb. 2013. ISSN: 2231-6604 Volume 4, Issue 2, pp: 19-26.

APPENDIX

package com.schizo.tree;

TreeNode.java

```
public class TreeNode {

	public static final int T_SNP = 1;

	public static final int T_GENOME = 2;

	public static final int T_CHROMOSOME = 3;

	public static final int T_MAP = 4;

	public static final int T_LOGP = 5;

	enum SCHIZOCLASS{

		schizophreniaprone,

		notschizophreniaprone

	}

	private int type;

	private int value;

	private Vector<TreeNode> children;

	private int depth;

	public int getType() {

		return type;

	}

	public void setType(int type) {
```

```java
		this.type = type;
	}

public TreeNode(int value, int type, int d){
		this.value = value;
		this.type = type;
		this.children = new Vector<TreeNode>();
		this.depth = d;
	}
	public TreeNode(TreeNode node){
		this.value = node.value;
		this.type = node.type;
	}
	public int getValue() {
		return value;
	}
	public void setValue(int value) {
		this.value = value;
	}
	public Vector<TreeNode> getChildren() {
		return children;
	}
```

```
public void setChildren(Vector<TreeNode> children) {
        this.children = children;
}
public TreeNode getChildrenAt(int index){
        try{
                return this.children.elementAt(index);
        }catch(NoSuchElementException e){
                e.printStackTrace();
                return null;
        }
}
public TreeNode addNode(int val, int t){
        TreeNode node = new TreeNode(val, t, this.getDepth() + 1);
 children.addElement(node);
        return node;
}
public int getDepth(){
        return this.depth;
}
public SCHIZOCLASS makeDecision(){
        switch(this.type){
        case TreeNode.T_SNP:
```

```
        if((this.value % 2) == 1)
            return SCHIZOCLASS.schizophreniaprone;
        else{
            for(TreeNode node : this.children){
                return node.makeDecision();
            }
        }
        break;
    case TreeNode.T_GENOME:
        for(TreeNode node : this.children){
            return node.makeDecision();
        }
        break;
    case TreeNode.T_CHROMOSOME:
        for(TreeNode node : this.children){
            return node.makeDecision();
        }
        break;
    case TreeNode.T_MAP:
        for(TreeNode node : this.children){
            return node.makeDecision();
        }
```

```
break;
case TreeNode.T_LOGP:
    if((this.value %2) == 0)
        return SCHIZOCLASS.notschizophreniaprone;
    else
        return SCHIZOCLASS.schizophreniaprone;
default:
    return SCHIZOCLASS.notschizophreniaprone;
}
return null;
}


}
package com.schizo.tree;
SchizoTree.java
public class SchizoTree{
    private TreeNode root;
    public SchizoTree(int value, int type, int d){
        setRoot(new TreeNode(value, type, d));
    }
    public TreeNode getRoot() {
        return root;
```

```
        }

        public void setRoot(TreeNode root) {

        this.root = root;

        }

}

package com.schizo.ui;

Schizo.java

public class Schizo {

        public static void main(String[] args) {

                SwingUtilities.invokeLater(new Runnable() {

                        @Override

                        public void run() {

                                try {


        UIManager.setLookAndFeel(UIManager.getSystemLookAndFeelClassName());

                                } catch (Exception e) {

                                        e.printStackTrace();

                                }

                                new Frame();

                        }

                });

        }
```

```java
}

package com.schizo.ui;

Frame.java

public class Frame {

	private String value;

	private SchizoTree schizoTree;

	public Frame(){


		do{

			value = JOptionPane.showInputDialog(null, "Input the SNP value : ");

		}while(value.isEmpty());

		schizoTree = new SchizoTree(Integer.parseInt(value), TreeNode.T_SNP, 1);




		do{

			value = JOptionPane.showInputDialog(null, "Input the Genome Build
Value : ");

		}while(value.isEmpty());

		schizoTree.getRoot().addNode(Integer.parseInt(value),
TreeNode.T_GENOME);
```

```
do{
        value = JOptionPane.showInputDialog(null, "Input Chromosome value
: ");
}while(value.isEmpty());
schizoTree.getRoot().getChildrenAt(0).addNode(Integer.parseInt(value),
TreeNode.T_CHROMOSOME);

do{
        value = JOptionPane.showInputDialog(null, "Input the MAP value : ");
}while(value.isEmpty());

    schizoTree.getRoot().getChildrenAt(0).getChildrenAt(0).addNode(Integer.parseInt(v
alue), TreeNode.T_MAP);

do{
        value = JOptionPane.showInputDialog(null, "Input the LOGP value :
");
}while(value.isEmpty());
char lastChr = value.charAt(value.length() - 1);

    schizoTree.getRoot().getChildrenAt(0).getChildrenAt(0).getChildrenAt(0).addNode((
char)lastChr, TreeNode.T_LOGP);
JOptionPane.showMessageDialog(null,
schizoTree.getRoot().makeDecision());
```

```
        }
    }
```