PREDICTIVE MODEL FOR DETECTING THE OVERLAPPED SYMPTOMS OF CARDIOVASCULAR DISEASES

BY

NAJWA FADHIL ABBAS

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy in Information Technology

Kulliyyah of Information & Communication Technology
International Islamic University Malaysia

NOVEMBER 2023

ABSTRACT

Cardiovascular diseases (CVD) have a significant impact on increasing the mortality rate in the Middle East and The United Arab Emirates (UAE) has one of the highest age-standardized death rates for cardiovascular disease (CVD). Recently, based on the Assessment Risk Tools for Cardiovascular Diseases (CVD), World Health Organization (WHO) reported that 40% of all fatalities in the UAE are attributed to CVD, which has been linked to the main Risk Factors (RF) advances as obesity, hypertension, tobacco, and high cholesterol. In most cases, angiography is a reliable method for the diagnosis and treatment of cardiovascular diseases. However, it is a costly approach associated with various complications. The significant increase in the prevalence of cardiovascular diseases and the subsequent complications and treatment costs have urged researchers to plan for the better examination, prevention, early detection, and effective treatment of these conditions The present study aimed to detect the patterns for the overlapped symptoms of cardiovascular diseases using integrated Deep Learning classification techniques for analyzing the data of internal medicine patients who are at the risk of heart failure with 2621 samples and 40 characteristics. Selecting the characteristics and evaluating the influential factors are essential to the development of classifiers and increasing their accuracy. The proposed work suggested a model based on Gini-Entropy-Regression Model (GERM). The objective is to predict future risk with a certain probability and compare its performance with Deep Learning MLP Model. Statistical analysis and methods were used in this research to detect the symptoms of CVD that overlapped and to accurately identify a specific heart condition. The dataset utilized to train the computer consists of medical records from more than 14 hospitals in UAE which were collected based on four main categories such as basic information, symptoms, inducement and history, and physical sign and assistant examination. The suggested model consisted of four levels, level 1: Preprocessing data, Level 2: Feature Extraction, Level 3: Feature Selection, Level 4: Feature Detection. The results of the suggested model were as follows: the result was 84.4% when the symptoms of (CVD) is overlapping DSYP and CHEP. When Accuracy measured with combination DSYP, CHEP, and CYAN it has been increased up to 88.9%. DSYP, CHEP, CYAN, showing values of 89.8%. in 5th Neural Network (NN) the combinations were DSYP, CHEP, CYAN, DBPH, WFAT, EMPT showing ideal value of accuracy measured up to 90.6% and with Fever this combination of Neural Network has been showing accuracy = 91%. From the findings the previous seven predictors (Risk Factors) give the best overlapping and diagnosis for CVD.

ملخص البحث

تؤثرأمراض القلب والأوعية الدموية بشكل كبير على زيادة معدل الوفيات في دول الشرق الأوسط بما في ذلك دولة الإمارات العربية المتحدة التي لديها واحدة من أعلى معدلات الوفيات القياسية حسب العمر تشير تقارير الصحة العالمية استنادا إلى أدوات تقييم المخاطر الخاصة بإحصاءات منظمة الصحة العالمية (WHO) ، وتظهر التقارير أن 40٪ من جميع الوفيات في دولة الإمارات العربية المتحدة يعزى السبب إلى أمراض القلب والأوعية الدموية ، ومن المتوقع أن يرتفع عدد الوفيات من 17.7 مليون في عام 2015 إلى 23.6 مليون في عام 2030 ، والتي تم ربطها بعوامل الخطر الرئيسية وتقدمها مثل السمنة وارتفاع ضغط الدم والتبغ وارتفاع الكوليسترول في الدم ... إلخ . قد تختلف أعراض الأمراض القلبية الوعائية من شخص لآخر. على سبيل المثال، بعض الناس أكثر عرضة للإصابة بألم في الصدر. من المرجح أن يعاني الأشخاص الآخرون من علامات وأعراض أخرى إلى جانب عدم الراحة في الصدر، مثل ضيق التنفس والغثيان والتعب الشديد الذي يمثل العملية المعقدة لجمع البيانات الطبية وتقييمها وتحليلها وتفسيرها لصياغة واحد أو سلسلة من القرارات. لذلك كانت الحاجة إلى تطوير نموذج تنبؤي لاتخاذ القرار الصحيح بناء على الأعراض المتداخلة للأمراض القلبية الوعائية مهمة جدا في هذه الدراسة. تم استخدام نموذج التنبؤ المقترح القائم على تقنيات التعلم العميق (DL) وتقسيمه إلى أربعة أجزاء، الجزء الأول هو جزء المعالجة (يتضمن تنظيف وتصفية مجموعة البيانات التي تم الحصول عليها من مرضى الأمراض القلبية الوعائية وتقسيم البيانات إلى مجموعات التدريب والاختبار. الجزء الثاني هو اختيار الميزات (يتضمن استخراج الخصائص الديموغرافية والخصائص الأكثر فعالية للمريض من خلال العثور على نوع الارتباط بين عوامل الخطر (RF)) لمرضى الأمراض القلبية الوعائية. الجزء الثالث هو اكتشاف الميزة (يتضمن اختيار أعلى المتنبئات (الأعراض المتداخلة) التي لها أعلى تأثير على المرضى. الجزء الرابع اتخاذ القرار (الذي يتضمن اتخاذ

القرار الصحيح المبكر بناء على الأعراض المتداخلة). تم استخدام هذه الدراسة القائمة على الطرق الإحصائية وشبكة الإدراك الحسي متعدد الطبقات للكشف عن الأعراض الأكثر تداخلا والتي لها أعلى تأثير على تحديد حالة قلبية معينة بدقة سواء كانت امراض أوعية قلبية أم لا. تتكون مجموعة البيانات المستخدمة من سجلات طبية من أكثر من 14 مستشفى في دولة الإمارات العربية المتحدة، ويتكون حجم العينة من 2621 مستشفى، وتم جمع معلومات العينة ومسحها في أبوظبي عاصمة دولة الإمارات العربية المتحدة وتم نشرها مؤخرا بناء على أربع فئات رئيسية مثل المعلومات الأساسية، الأعراض والإغراءات والتاريخ والعلامة البدنية والفحص المساعد. نتيجة للنتائج، تم اكتشاف أن مستوى دقة الأمراض قد زاد عندما تمت محاكاتها في أزواج من مرض واحد مع مرض آخر بسبب تداخل الأعراض. كانت نتائج دقة التنبؤ المبكر (CVD) بناء على 8 مستويات (تجارب) على النحو التالي: المستوى الأول كانت الدقة 84.4٪ في حالة تداخل DYSP و CHEP. المستوى الثاني كانت الدقة 86.7٪ في حالة تداخل DSYP و Uتشب و CYAN. المستوى الثالث كانت الدقة 88.9٪ في حالة تداخل DSYP و CHEP و CIAN و WFAT. المستوى الرابع كانت الدقة 89.8٪ في حالة تداخل DYSP و CHEP و CYAN و WFAT مما يفسر 88.9٪. المستوى الخامس كانت الدقة 90.6 في حالة تداخل DYSP و CHEP و CIAN و WFAT و EMPT و DPCH. المستوى السادس، كانت النتيجة 91٪ في حالة تداخل DYSP، CHEP، CYAN، WFAT، EMPT، DPCH و FEVE

APPROVAL PAGE

The dissertation of Adam Hakim Khairuddin has been approved by the following:

_____

Akram M.Z. Khedher
Supervisor

_____

Asadullah Shah
Co-supervisor

_____

Noor-Aziza Bt. Mohamadali
Co-supervisor

_____

Amelia Ritahani Bt. Ismail
Internal Examiner

_____

Rahmita Wirza O.K Rahmat
External Examiner

_____

Mohamed Elwathig Saeed Mighani
Chairman

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted for any other degrees at IIUM or other institutions.

Najwa Fadhil Abbas

Signature.....Najwa..... Date....24 Nov. 2023.

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

EARLY PREDICTION DEEP LEARNING MODEL FOR OVERLAPPING SYMPTOMS OF CARDIOVASCULAR DISEASES

I declare that the copyright holder of this thesis/dissertation are jointly owned by the student and IIUM.

Copyright © 2014 Student Name and International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.

3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Najwa Fadhil Abbas Al Hadeethy

…… …… ……..

Signature Date

ACKNOWLEDGEMENTS

Bismillahirrahmannirrahim. Praise and thanks to The Almighty Allah S.W.T for blessing me with strength, grit, and determination to finish the research for my PhD thesis. Undertaking this PhD has been a truly life-challenging experience for me, and it would not have been possible to complete it without the support and guidance that I received from many people.

Firstly, it is my utmost pleasure to dedicate this work to Allah, to my dear mother, my brother and daughter, the soul of my father, and who have granted me the gift of their unwavering belief in my ability to accomplish this goal: thank you for your support and patience.

I would like to express my countless gratitude to my supervisor, Prof. Dr. Akram M.Z. Khedher and co-supervisors, Prof. Dr Asadullah Shah and Dr. Noor Azizah for their advice, beneficial and constructive comments, patience, and kindness through this whole time guiding me without fail.

I shall thank all the lecturers and staffs from Kulliyyah of Information and Communication Technology (ICT), for helping me with pre-processing procedures,
Lastly, my special appreciation would be to my both parents, my late father, and my daughter, all members of my family for always believe and being thoughtful for everything I have done to complete this journey. Millions of thanks for their untold commitment, continuous support and understanding throughout my PhD journey. May each single person who is involved in this journey, directly or indirectly, is granted with His blessings for this life and hereafter. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRIVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
AS	Atherosclerosis
BMI	Body Mass Index
CAD	Computer Aided Design
CAD	Coronary Artery Disease
CART	Classification And Regression Tree
CD	Chronic Disease
CHD	Coronary Heart Disease
CHEP	Chest Pain
CIFARIO	Canadian Institute for Advanced Research
CMRIC	Cardiovascular Magnetic Resonance Imaging
CNN	Convolutional Neural Network
CVD	Cardiovascular Diseases
CYAN	Cyanosis
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
DBPH	Diastolic Blood Pressure with High Rate
DPCH	Discomfort Pressure in Chest
DT	Decision Tree
DYSP	Dyspnea
ECG	Electrocardiography
EKG	Electrocardiograph
EMPT	Emptysis
EMRs	Electronic Medical Records
FEVE	Fever
GDP	Gross Domestic Product
GERD	Gastroesophageal Reflux Disease
HEADACHE	Headache:
HER	Electronic Health Record
IOT	Internet of Things
IT	Information Technology
MDSS	Medical Decision System Structure
MI	Myocardial Infraction
ML	Machine Learning
MLP	Multilayer Perceptron Neural Network
MMRE	Mean of Mean Relative Error
MNIST	Modified National Institute of Standards Technology
MRE	Mean Relative Error
MRI	Magnetic Resonance Imaging
MSA	Multiple System Atrophy
NN	Neural Network
NCD	Non-Communicable Disease
PCA	Patient Controlled Analgesia

RF	Risk Factors
RBM	Restricted Boltzmann Machine
SAS	Sleep Apnea Syndrome
SVT	Supra Ventricular Tachycardia
UAE	United Arab Emirates
WFAT	Weakness or Fatigue:
AI	Artificial Intelligence
ANN	Artificial Neural Network
AS	Atherosclerosis
BMI	Body Mass Index
CAD	Computer Aided Design
CAD	Coronary Artery Disease
CART	Classification And Regression Tree
CD	Chronic Disease
CHD	Coronary Heart Disease
CHEP	Chest Pain
CIFARIO	Canadian Institute for Advanced Research
CMRIC	Cardiovascular Magnetic Resonance Imaging
CNN	Convolutional Neural Network
CVD	Cardiovascular Diseases
CYAN	Cyanosis
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
DBPH	Diastolic Blood Pressure with High Rate
DPCH	Discomfort Pressure in Chest
DT	Decision Tree
DYSP	Dyspnea
ECG	Electrocardiography
EKG	Electrocardiograph
EMPT	Emptysis
EMRs	Electronic Medical Records
FEVE	Fever
GDP	Gross Domestic Product
GERD	Gastroesophageal Reflux Disease
HEADACHE	Headache:
HER	Electronic Health Record
IOT	Internet of Things
IT	Information Technology
MDSS	Medical Decision System Structure
MI	Myocardial Infraction
ML	Machine Learning

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF STUDY

In the 21st century, information and communication technology has changed the way people live and do business. It would not be incorrect to state that in the current age, information technology has changed almost every element of human lifestyles. Rouse (2016) stated that "health information technology is a study of health information management resources and techniques, this field of study promotes information technology for health, medical practice, medical research and information technology. The use of information technology in the health sector has gained popularity quickly and the advantages of this new paradigm are experienced across the world.

Due to the necessity for secure and effective administration of medical data, the significance of health information, as is often called, has grown considerably in recent years. The health information technology field is engaged in the effective gathering of medical data, safe archiving, and prompt recovery, which enhances the diagnosis and treatment of patients. Health IT also enables appropriate administration, analysis, and use of health-related data to provide customers and patients with more efficient health care delivery and service. The underlying concept of health information is to provide healthcare professionals and patients with more control. Additionally, it emphasizes the necessity and sensitivity of the obligations of healthcare workers who handle and manage data (University of Toronto, 2013). Also prevalent are other words such as clinical computer science and health information management, which concentrate on how to integrate the power of technology in contemporary health practices and medical data administration. Although the idea has various names, the idea underlying in all is basically the same. It is a process in which data is examined and used to create information that is effectively used with a view to address clinical issues and facilitating fast delivery of medical treatment in a time-sensitive way (Dalrymple, 2011). Chronic Prediction is a deep learning technique

that uses user experience to predict the impact of risk factors of non-communicable diseases (Pittoli et al., 2018).

A heart or blood vessel disease is referred to as a cardiovascular disease (arteries and veins). Heart disease has overtaken infectious illnesses as the leading cause of mortality and disability in the globe for those over 65 years old and is now considered a ‘second epidemic' in many gulf nations due to the alarmingly high and steadily growing prevalence of the condition (Gale Nutrition Encyclopedia, 2011). Early identification of cardiovascular illness may help decrease the death rate. The complex connection between data, information, knowledge, and decision to which all the are subscribed is shown in Figure 1.1 below as an example of Informatics pyramid for CVD case which had been developed on the parametric variables containing the numeric information, knowledge, and data, which can be implemented using deep learning models.

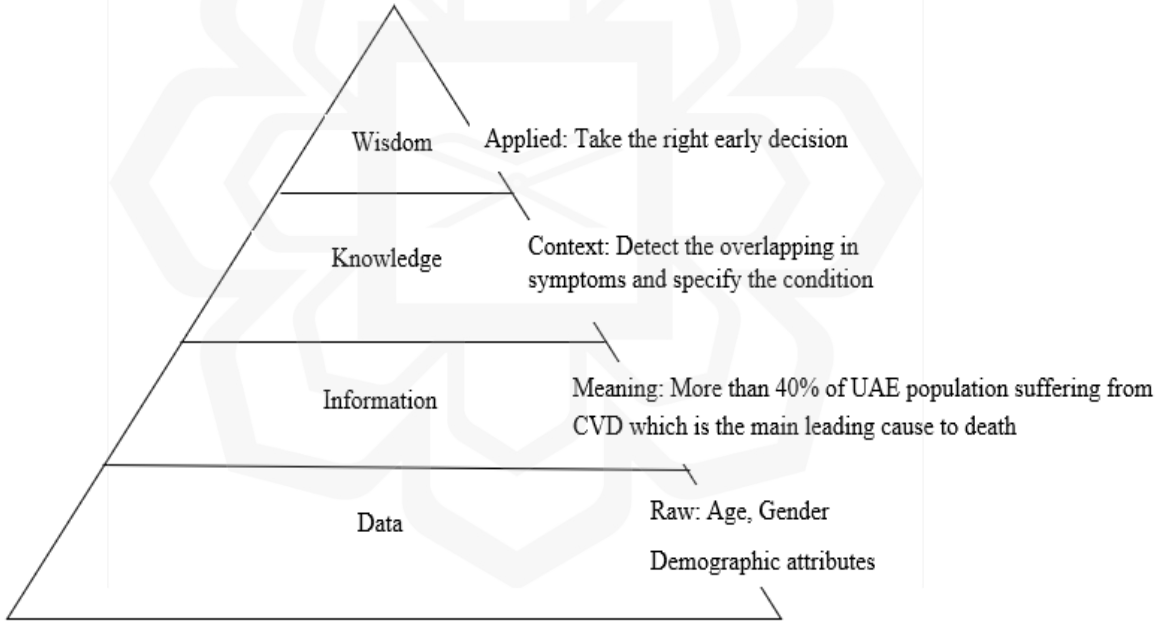


Figure1.1 Medical Example based on a traditional Data-Information-knowledge-Wisdom (DIKW) pyramid (Muhson, 2018)

Data mining techniques have been used to find useful patterns in the data, to support medical decision making such as diagnosis process, choosing treatment option, and prognosis prediction. When applied properly, it could help the healthcare provider to

improve patient's care. However, medical data are typically complex and diverse, and discovering patterns in such kind of data using traditional method is difficult. The Electronic Health Record (EHR) is an electronic data gathering of individual patients and populations. It may be shared across health care professionals in a certain state or nation (Gunter & Terry, 2005). Health records may contain general health history, patient reviews, patient treatments, medical records, allergies, vaccination status, test findings, radiological pictures, and some helpful information for inspection. This large data collection may enable researchers to examine and diagnose illnesses using computer methods. Using EHRs, legacy systems may save costs, improve care quality, and increase record mobility or sharing. Deep learning and machine learning are techniques that use computational models with multiple processing layers to learn from representative data (LeCun, Bengio, & Hinton, 2015).

This information may be presented in a variety of formats, including visual objects (pictures), voice (speech recognition), or sequential data presented in text form. Based on the representation of each data layer, deep learning models create output parameters using backend propagation algorithms (Goodfellow, Bengio, Courville, & Bengio, 2016). Artificial intelligence has proved its potential uses across a wide range of disciplines, as well as in the realms of medical research.

The combination of large datasets from hospitals and deep learning methods such as convolutional neural networks may be utilized to develop prediction models for risk management in disease control (Cheung *et al.*, 2020; Usama *et al.*, 2018; Zhu *et al.*, 2020). The gap between normal human physiological states to diseased state can be potentially filled by examining and monitoring clinical, biomedical, and healthcare wellness data from individuals.

In this study, treatment of such diseases require a deep learning model (GERM) model that can predict and detect the overlapping in CVD symptoms in early stage, with precise accuracy and reliability. Intensive research is carried out by various researchers using diverse machine learning algorithms to forecast heart disease taking different datasets which consists of different attributes that result in heart attack. This study will discuss the

different deep learning approaches and build a predictive model to detect overlapping symptoms in cardiovascular diseases. The deep neuraCONTENTSl networks, which in the case can act as alerts and warnings in biomedical depending on the electronic medical record (EMRs).

1.2 STATEMENT OF THE PROBLEM

The incidents and mortality rates due to cardiovascular diseases are increasing in recent years, where how to diagnose and prevent such diseases has become a challenging situation. Acute forms of cardiovascular catastrophe such as thrombosis and atherosclerosis plaque formation involve 30-70 years. Such long periods of silent symptoms has made it difficult to diagnose early screening for health care physicians (Escárcega *et al.*, 2018). Future Medical Multi-Level studies should be underpinned by multiple theoretical perspectives (Frynas & Yamahaki*, et al*., 2016). Most cases, such as ischemic cardiomyopathy and acute coronary syndrome, reached progressive phases of the condition at the first appointment. And though AS plaques are discovered, it is a struggle to address whether or not lesions need involvement or weakness in clinical practice (Cao *et al.*, 2019). Meanwhile, some of the CVDs may share similar characteristic symptoms like chest pain, irregular heartbeat, shortness in breath, Pain in the neck, jaw, throat, upper abdomen or back, and numbness, which increases the chances of physicians diagnosing misleading. Based on the reports of WHO, Number of deaths will be increased at the rate of 1.7% during the next 10-15 years in the Middle Eastern countries as explained in Table 1.1 which explains the demographics and characteristics of risk assessment tools.

Table1.1 Demographic and Characteristics of Cardiovascular Risk Assessment tools **(**Abderrahim Oulhaj et al, 2020)

Tool Name	Region /Area	Ages	Outcomes
WHO-MENA	Middle Eastern Countries	40–70 years	10-year risk of fatal or non-fatal MI or stroke
SCORE-H	12 European cohorts	40–65 years	10-year risk of CHD death or stroke death

PCRE-W PCRE-AA	CARDIA, Framingham, ARIC, CHS, USA	40–79 years	10-year risk of fatal or non-fatal CHD or fatal or non-fatal stroke
FRAM-ATP III	Framingham Heart Study	20–79 years	10-year risk for myocardial infarction and coronary death
FRAM-LAB	Framingham, MA, USA	30–74 years	10-year risk of coronary death, MI, coronary insufficiency, angina, ischemic stroke, haemorrhagic stroke, transient ischemic attack, peripheral artery disease, or heart failure

Such post-determined symptoms and silenced indications for cardiovascular diseases can be estimated and assessed using the deep learning neural networks. Deep Belief Networks, Medical Image Segmentation, and Bayesian Networks can use the large clinical dataset to predict risk factors' impacts. In this research the researcher is going to develop a predictive model based on machine learning algorithms to assist and predict precisely early symptoms of heart diseases in UAE.

1.3 RESEARCH OBJECTIVES

This study aims to build a deep learning model for early predictions of related symptoms of cardiovascular diseases and focusing on using UAE clinical dataset as a predictor of risk factors and the impact of overlapping symptoms in cardiovascular diseases and contribution in the literature to improve the diagnostic measures for CVDs, predictive models in cardiovascular disease prediction are essential tools that enable early detection, personalized risk assessment, efficient resource allocation, informed treatment planning, and contribute to ongoing research and prevention efforts. They have the potential to improve patient outcomes and reduce the societal impact of cardiovascular diseases.and the best prediction accuracy is achieved in the following ways:

1. To compute and deduct the highest influential indexes of the Risk Factors (RF) for the Cardiovascular Disease (CVDs) and understand their effects in the prediction model.

2. To develop a profound learning model based on (GINI –Entropy) learning Mode for the predictive analysis of overlapping CVD symptoms.
3. To detect the most effective overlapped predictors (symptoms) of CVD in the proposed model (GINI –Entropy) learning Model.

1.4 RESEARCH QUESTIONS

The problem of this study can be expressed by the following main questions that must be answered to achieve the objectives of study:

RQ1 Which Risk Factors (RF) are the most significant causes of CVD in Deep Learning Model?

RQ2 Why it is important to detect the positive and negative correlations for the Risk Factors of CVD in the proposed Model?

RQ3 How is the profound learning model detect the overlapping in cardiovascular diseases predictors (symptoms)?

RQ4 How many predictors (overlapped symptoms) be effective in identifying the specific and correct decision?

RQ5 How the proposed deep learning model will achieve more accurate and reliable results in the early Prediction of cardiovascular as compared to the previous models?

Table 1.2 presents the research paradigms methods for satisfying the Research questions with Research objectives as follows:

Table 1.2 Research Paradigm Methods

Research Questions	Research Objectives	Research Methods	Chapter
RQ1 Which Risk Factors (RF) are the most significant causes of CVD in Deep Learning Model?	1. To compute and deduct the highest influential indexes of the Risk Factors (RF) for	Statistical Analysis Methods CART Algorithm Gini-Entropy Index Algorithm	Chapter Three Chapter Four

RQ2 Why it is important to detect the positive and negative correlations for the Risk Factors of CVD in the proposed Model?	the Cardiovascular Disease (CVDs) and understand their effects in the prediction model.	Correlation Plot Matrix Visual plot Analysis Dimensional Reduction	
RQ3 How is the profound learning model detect the overlapping in cardiovascular diseases predictors (symptoms)? RQ4 How many predictors (overlapped symptoms) be effective in identifying the specific and correct decision?	2. To develop a profound learning model for the predictive analysis of overlapping CVD symptoms. 3.To detect the effectiveness overlapped predictors (symptoms) which have a high impact on CVDs prediction	Neural networks as Multilayer Perceptron Regression Analysis Correlation test	Chapter Three Chapter Four
RQ5 How the proposed model will achieve more accurate and reliable results in the early Prediction of cardiovascular as compared to the previous models?	To detect the effectiveness overlapped predictors (symptoms) which have a high impact on CVDs prediction	Descriptive Analysis K-fold Cross Validation methods	Chapter Three Chapter Five

1.5 SIGNIFICANCE OF THE STUDY

In recent years, with the advent of deep learning, cardiac imaging analysis has shown great development prospects. Deep learning can help to analyze coronary angiography, echocardiography and electrocardiogram (ECG). Cardiac intervention has been the main treatment for cardiovascular disease in recent decades, including CHD and acute coronary syndrome (ACS). In the current age, information technology has practically changed every area of human existence. The use of IT in the health sector is gaining popularity quickly and the advantages of this new paradigm are being recognized across the world. This development generated enormous amounts of patient data that can be used by computer technology and machine learning methods and converted into valuable information and knowledge. This data may be utilized to build expert systems to diagnose life-threatening illnesses such as cardiovascular illnesses with lower costs, time, and enhanced diagnostic accuracy. Although contemporary medicine produces a lot of data every day, nothing has been done to utilize this accessible information to address difficulties in diagnosing cardiac illnesses successfully. Identifying the need to investigate the use of robust data mining methods in the detection of heart disease and other weakening illness by health professionals.

A fundamentally new way of doing simulation that focuses on limited human insight is the enormous machine learning area (Bishop, 2006). A change is going on in the world of machine learning. Over the last few decades, it has changed from a specialized field to a significant economic development driver as innovation revolutionizes web browsing, speech-to-text, and many other economic significance fields (He *et al.*, 2015; Silver *et al.*, 2016). Computational models are beginning to be used by biomedical researchers to interpret knowledge and define the fundamental concepts (Jonas & Kording, 2017).The need for predictive model is to assess the forecasting of overlapping in symptoms based on Deep learning techniques. The need for predictive model is to assess the forecasting of overlapping in symptoms based on Deep learning techniques.

Deep learning techniques for overlapped symptomatic cardiovascular diseases and risk heterogeneity must not only be reliable to be useful, but they will need to offer visibility about where they are likely to deliver false outcomes and be explainable in the way that

health care professionals can appreciate when the model obtains a given outcome. These proposed methods assist in guaranteeing that the model can be extended diligently to the population about which it is most precise (Schlesinger & Stultz, 2020). Deep learning can derive actionable insights from patterns in troves of data. Computer Aided Diagnosis (CAD) has been used to help radiologists interpret medical images in clinical practice for decades. Deep learning algorithms outperform classical algorithms on benchmark image recognition tasks and are only recently being implemented into medical imaging research.

The use of machine learning and deep learning inside the health industry has quickly acquired worldwide significance. In recent years the significance of health informatics has increased considerably since medical data needs to be managed securely and efficiently. Health information technology also makes it possible to properly store, analyze and utilize health related data for the more Productive delivery of healthcare. It is also crucial to assist doctors discover successful therapies and patients get better and better healthcare services that are more inexpensive (Zarb, 2016).

1.6 TERMS DEFENITIONS

This section defines the basic terms that has been used in this study as follows:

- Artificial Intelligence: Machines that are trained to think and behave like humans are referred to as artificial intelligence (AI) systems (Dustin Harris, 2022)
- Artificial Neural Network: It is a calculation method that builds several processes based on interconnected connections, consisting of an arbitrary number of cells or nodes or units or neurons that connect the input set to the output. (Ingre & Yadav, 2015; Landahl et al., 1943; McCulloch & Pitts, 1943).
- Cardiovascular Diseases: It is a group of diseases that affect the heart and blood vessels (Organization, 2013).
- Computer Aided Diagnosis: In the realm of medical imaging, the computer-based system that aids clinicians in making judgments quickly is known as computer-aided diagnosis (CAD) (Halalli & Makandar, 2016).

- Convolutional Neural Networks: is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other (Prasoon *et al.*, 2013)
- Correlation Test: Test that shows if there is a relation between two or more variables (Richard A, 2010).
- Decision Tree: Decisive trees are models of choices and their probable consequences, such as chance event outcomes, resource costs and utility values (Ben-Haim & Tom-Tov, 2010; Brijain *et al.*, 2014).
- Electrocardiography: Electrocardiography is the method of creating an ECG. Electrodes are inserted on the skin to record the electrical activity of the heart.
- Machine Learning: Automated algorithmic improvement via experience and the utilization of data is the focus of machine learning (Nature, 2015)
- Matrix Plot: It is an array that is used to assess the relationship among several pairs of variables at once (Will Koehrsen, 2018).
- Multilayer Perceptron: In artificial neural networks, a multilayer perceptron (MLP) belongs to the feed forward category (ANN). An MLP may refer to any feed forward ANN or a network with many layers of perceptron's, depending on the context in which it is employed (Bishop, 1995).
- Naïve Bayesian: Naive Bayes classifiers are a series of basic "probabilistic classifiers" based on applying Bayes' theorem with strong independence assumptions between the features in the dataset (Bellazzi & Zupan, 2008; Frank et al., 2000; Webb, 2010).
- Non-Communicable Diseases: A non-communicable disease (NCD) is one that cannot be spread from one person to another by direct contact (Pittoli et al., 2018).
- P- Value: Determine the appropriateness of rejecting the null hypothesis, p-value ranges from -1 to 1 (Rebecca Bevans, 2020).
- Performance Chest Pain: "performance linked to the number of resources applied under the Constituted conditions".

- Portability: "degree of consciousness and Chest Pain of a system, product or Constituent that will be transferred from one hospital to another".
- R^2 (adjusted): Percentage of response variable variation that is explained by its relationship with one or more predictor variables, adjusted for the number of predictors in the model (Rebecca Bevans, 2020).
- Recurrent Neural Networks: It is a kind of artificial neural network where the connections between nodes create a directed graph in a temporal sequence. As a result, it's capable of displaying temporal dynamics.
- Regression Test: Test that shows if the variation in one variable causes variation in another variable (Rebecca Bevans, 2020).
- GERM: Gini Entropy regression Model is a Deep learning model used in this study to detect the overlapping in CVD symptoms based on GINI –Entropy statistical methods and regression classification techniques.

1.7 THESIS ORGANIZATION

This section gives a brief outline of the proposed thesis chapter as follows:

- Chapter One presents the background, significance of the study, statement of the problem, research objectives and research Questions, and terms.
- Chapter Two discusses the currently available research on deep learning algorithms, how they relate to medical diagnostics and how they are employed in early cardiovascular disease prediction. This section contains literature on prior research of the issues mentioned, a theoretical framework connected to data analysis, philosophy, obstacles, errors, best practices, and highlights knowledge gaps.
- Chapter Three: this chapter provides an introduction about research strategy and design, and approach, data sampling involves description of data, the source of data collection, characteristics, and demographics (RF) and describe the dependent/independent variables and the main methods for finding the correlation type between them and their effects in detecting the overlapped symptoms, as well

as presents some of the deep learning models which have been used in disease prediction.

- Chapter Four presents the design and implementation of the developed system in details the methodology used for detecting the overlapping and find the best accuracy for the proposed model with 8 experiments (levels).
- Chapter Five This chapter "provides for evaluating model values by using K fold cross validation with PRED(X) and MMRE, as well as summarized k-fold models with R-square values.
- Chapter Six illustrates the main conclusions and future works.

1.8 SCOPE OF THE STUDY

This study is becoming more popular as technology and it is being used in a variety of fields Clinical Implementation: Ultimately, the scope extends to the clinical implementation of CVD prediction models. This involves studying how these models can be integrated into healthcare systems and used by healthcare providers to improve patient care and outcomes, Public Health and Policy: Some studies consider the broader public health implications of CVD prediction models. This includes assessing how the widespread adoption of these models can impact healthcare policies and guidelines.

Personalization: Personalized medicine is an emerging area of study within CVD prediction models. Researchers aim to tailor predictions and interventions to individual patient profiles, considering their unique risk factors and genetic makeup.

Model Development: Developing novel predictive models is a significant aspect. Researchers explore different machine learning and statistical modeling approaches, including logistic regression, decision trees, random forests, support vector machines, deep learning, and ensemble techniques.

1.9 CHAPTER SUMMARY

This chapter presented the background of the study and discussed the problem statement, as well as explained the primary aim of this thesis. Following the research objectives that were specifically addressed, the significance of the study highlighted how this study could fill the gaps between previous studies and benefit the future. Finally, a brief introduction into the methodology used in this study was mentioned, followed by an outline of this thesis in brief.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter discusses the currently available research on deep learning algorithms, how they relate to medical diagnostics and how they are employed in early cardiovascular disease prediction. This section contains literature on prior research of the issues mentioned, a theoretical framework connected to data analysis, philosophy, obstacles, errors, best practices, and highlights knowledge gaps.

2.2 DEEP LEARNING

When it comes to the early detection and diagnosis of cardiovascular diseases, as well as the prediction of outcomes and the evaluation of prognoses, AI technologies such as machine and deep learning may play a critical role. Electronic medical records have created enormous amounts of data (quantitative, qualitative, and transactional) that must be interpreted by algorithms powered by AI (EHRs). Using therapeutically relevant information from the massive amount of data collected by AI systems, clinicians may be able to detect organ malfunctions in the early stages, improving patient care and quality of life in the process. Also important in the fight against (CVD) is telemedicine and deep learning. Health care might undergo a drastic transformation thanks to the Internet of Things (IoT), which could allow distant clinicians to constantly monitor the physical condition of patients. The use of cardiac imaging in the diagnosis of cardiovascular illness is very important (CVD). It has so far been limited to the visual and quantitative examination of cardiac structure and function. Big Data and machine learning have opened new opportunities for building artificial intelligence systems that may assist doctors in quickly diagnosing cardiovascular diseases (CVDs).

As a subfield of machine learning, deep learning (DL) may be used to categorize and phenotype new illnesses, as well as make challenging judgments in medicine. In most cases, neural networks with several layers are used to do this. Recent advances in computer hardware and algorithms have boosted its application in e-commerce, banking, and photo recognition. Deep Learning (DL) mirrors the complexity of a human brain in which data that have several abstract levels may develop complicated hierarchical representations. The programmer inserts known data into the system so that algorithms can appropriately react, even in the face of completely new input (Seetharam *et al.*, 2019). Neural network learning via experience, reading data, building hierarchical structures, and offering higher degrees of input and output. It can record intricate nonlinear links between variables of input-output. Average results error and their forecasts may be reduced by estimating input weights and outcomes data. Doctors diagnose their expertise, experience, and cultural background. Deep learning might at this moment be highly productive, extending and increasing medical knowledge, especially for non-expert doctors. DL may use more hidden layers to investigate more intricate data than standard neural networks. This is why DL has lately been popular in medical research as a result of the increasing number and complexity of data, notably in the field of imaging analysis (Romiti *et al.*, 2020). Chronic Prediction is a thorough learning methodology that utilizes experience to forecast the influence of non-Communicable disease risk factors (Pittoli *et al.*, 2018).

2.3 DARTIFICIAL NEURAL NETWORK

ANN is ML model that McCulloch and Pitts first proposed in 1943. (Ingre & Yadav, 2015; Landahl *et al.*, 1943; McCulloch & Pitts, 1943). This is influenced by neuronal processes in the brain and can conduct multiple functions, such as Regression and grouping. It is made up of linked artificial neurons that first acknowledge input data and second measure the output value on the basis of the input values given (W. G. J. A. o. i. m. Baxt, 1991). A major benefit of ANN over traditional mathematical approaches is the potential of ANN to model dynamic non-linear interactions. When modeling non-trivial functions, this gives ANN the strategic advantage; helping it to produce successful success when applied to

different complex problems in science and engineering. However, ANN has many drawbacks: (1) it is extremely responsive to the importance of its parameters; (2) the design and sophistication of the designed network plays a major role in its performance; (3) it has a large computational training cost; and (4) it may be difficult for humans to understand the resulting induction models (Bellazzi & Zupan, 2008).

2.4 DECISION TREES

Decision Tress (DT) is a relevant for decision process that combines a guided acyclic graph to carry out classification, built from training data (Ben-Haim & Tom-Tov, 2010; Brijain *et al.*, 2014). Each non-leaf node is liable for checking a function within the tree layout, which Quinlan created in 1986, is one of the leading DTs (Quinlan, 1986). C4.5 (successor of ID3), See5 (successor of C4.5) (Quinlan, 1992) and CART (Classification And Regression Tree) are some of the most common DT algorithms (Y. Wang & Witten, 1996). One major benefit of DT is its low computational complexity, while the main downside is that since the evaluated dataset includes several attributes, the built tree will become very complex (Y. Wang & Witten, 1996).

2.5 NAÏVE BAYSIAN

The Bayesian Naïve is an efficient probabilistic classifier based on the Bayesian theorem. It predicts multiple probabilities of the input data which implies that the input features vary from one another, i.e., the existence of one function is not related to the absence of another. Since it is a reasonably simple classifier that assumes unreasonable freedom compared with other more sophisticated algorithms, it can work equally well (Bellazzi & Zupan, 2008), And one of the most common classifications used for drug evaluation is (Rish, 2001; Rish *et al.*, 2001). However, the performance of the Naïve Bayesian classifier can be surpassed by advanced algorithms such as ANN and SVM, with non-linear biomarkers (Bellazzi & Zupan, 2008; Frank *et al.*, 2000; Webb, 2010).

2.6 MULTILAYER PERCEPTRON NEURAL NETWORK (MLP)

A multilayer perceptron (MLP) is a form of deep neural network that proposed to fix complex problems with additional perceptron's, stacked in many layers (Bishop, 1995). MLP is one of the most popular neural network architectures used in support structures for medical decision making. And it refers to the subset of neural networks which are supervised. H. Yan *et al.* (2006) explored decision making multilayer perceptron for heart diseases diagnosis. A fluid and fuzzy cognitive method is the medical diagnosis by design, and soft computational strategies. In the advancement of medical decision support structures, such as neural networks, immense capacity has been shown to be applied (MDSS). The researchers have used a multiplayer-based policy support system for the detection of cardiovascular diseases. In the process input layer, forty input variables are used, grouped into four classes, and then coded using the suggested coding schemes. For the cascading learning process, the number of nodes in the hidden layer is determined. Each of the five nodes in the output layer correlates to a severe heart condition. The missing patient data is handled using the substitution medium type. In addition, the computer is trained using an improved back propagation algorithm. Figure 2.1 shows a multilayer perceptron model used for early prediction of CVDs using multiple symptoms. Symptoms of the disease are used as input layer that in the hidden layer are analyzed and given as a given in the output layer of the model.

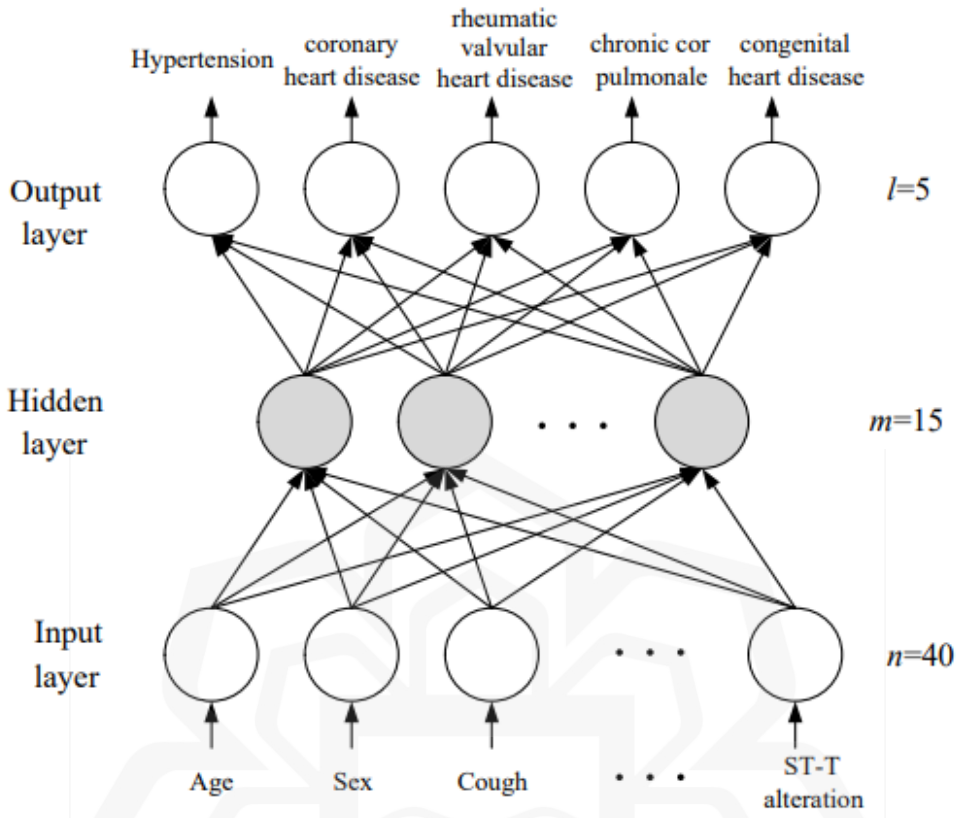


Figure 2.1 Example of Multilayer Perceptron Network (MLP) Source: H. Yan *et al.* (2006)

For the preparation and development of the program, a total of 168 medical reports obtained from patients with five cardiac disorders is used. In specific, to test the program generalizability, three measurement techniques, cross validation, holdout, and bootstrapping, are implemented. The findings indicate that the proposed predictive analytics framework based on MLP will achieve very high diagnostic precision (> 90 percent) and comparably low intervals (< 5 percent), showing its utility in assisting the cardiac disease clinical decision-making method. The following measurements have been used for assessing the model:

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases.

Sensitivity: The sensitivity of a test is its ability to determine the patient cases correctly. To estimate it, we should calculate the proportion of true positive in-patient cases.
Specificity: The specificity of a test is its ability to determine the healthy cases correctly. To estimate it, we should calculate the proportion of true negative in healthy cases.

2.7 CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks took sentiments into pooling layers for analysis. The emotions are being mapped as sentence length and converted into corresponding scattered vectors of fixed size. The technique is more popular in visual representation of emotions, opinions, and sentiments. CNN is expressed using Caffe and Pythons on Linux based devices such as android smartphones. Google Net proposed a 22 layered deep CNN which then best express sentiments, emotions and notions based on its varying size and depth. Word2vec is another seven layered convolutional neural networks developed by Google itself which uses words to express them as vectors of specific shape and size.

CNN has proved to be very Productive in solving image recognition challenges. Analysis work based on CNN substantially increased the best performance for many image datasets, including the MNIST, CIFAR10, and the NORB database. It's also very useful for learning from picture data regarding local and global systems. General picture artifacts such as handwritten numbers or human faces have apparent local and global forms, but it is possible to incorporate simple local characteristics such as edges and curves to become more complicated characteristics such as corners and forms and ultimately objects. CNN has also increasingly been introduced into diagnostic imaging research, such as the segmentation of knee cartilage (Prasoon *et al.*, 2013).

2.8 DEEP RECURENT NEURAT NETWORK

Recurrent Neural Networks are based not on fixed sized context rather than based on hierarchical bidirectional recurrent neural network which is being used evaluate patients reviews about healthcare in the history. The huge data set in the form of text mining;

challenge was gathered to evaluate the longs short term memory of patients' reviews about hospitality. The model is focused on the temporal actions and opinions of user experiences as well as a permanent emotional product. Models give approaches on the product review of users in context of long- and short-term memory.

2.9 DEEP BELIEF NETWORK

RBM-based deep belief networks (DBNs) have a few shrouded layers. Include portrayal has been shown to be Productive using DBN. It makes use of unlabeled data to fill in the gaps left by marked examination problems. The models used in this methodology are Poorly Shared Deep Neural Networks, which support natural language by exchanging sentimental labels. This method, which employs back propagation, allows for the planned work to be shortened. The suggested method is more accurate and efficient than the previous studies in terms of cross-lingual emotion classification, according to the experimental results (Ruangkanokmas *et al.*, 2016).

2.10 CARDIOVASCULAR DISEASES (CVD)

The category of illnesses connected with the heart and/or blood vessels is cardiovascular disease CVD. This involves diseases that involve Blood vessel restriction that brings blood to the heart, such as in coronary heart disease, cardiovascular disorder in the head and peripheral arterial disease in the limbs. Second, cardiac and heart valve tissue damage induced by rheumatic fever such as rheumatic heart attack and heart muscle failure in order to pump more blood into the blood vessels. (Organization, 2013).

Acute myocardial infarction (MI) - generally referred to as a cardiac attack - is an acute occurrence of special significance. This is because people across the globe are facing a deleterious health crisis that induces significant mortality. (Go *et al.*, 2013; A. G. Wilson *et al.*, 2016; A. G. Wilson & Izmailov, 2020). Atherosclerosis - the buildup of plaque in the walls of the arteries, diminished them and raised the blood challenge. MI cases typically arise when myocardial ischemia persists over a prolonged amount of time, overwhelming

the structures of myocardial cellular recovery intended to support the regular functioning process and Cardiovascular function homeostasis. If this imbalanced requirement and blood supply reaches a critical threshold and persists for an extended time, it will result in permanent loss or necrosis of myocardial cells. Such an occurrence is sometimes induced in a coronary artery by plaque breakup with the forming of a thrombus, and in the most tragic situation, it can contribute to the death of an individual.

Several epidemiological experiments have been carried out to further recognize and classify the disorder, considering the adverse influence of MI on culture. Such study has also established essential risk factors for CVD, including age, race, saturated fat, high blood pressure, overweight, asthma, tobacco, drink, psychosocial causes, and poor and morbidly obese eating patterns (Anand *et al.*, 2008; Hubert *et al.*, 1983; Hung *et al.*, 2015; Levin & Stokes, 1989; Psaty *et al.*, 2001; Rosengren *et al.*, 2004). However, when evaluating risk factors, causation is important because the extent of their Predicts on individual health will vary over one generation.

The mortality rate may also be decreased by diligent tracking, updating, and regulating certain risk factors. Includes, emerging approaches, socio-economic cost control criteria and escalating requirements for customized treatment (Olier & Vellido, 2008), new projections for the technologies able to deliver diagnosis, estimates and suggestions for the patient. This was part of a predictive, preventive and tailored Medicine strategy (Snyderman & Williams, 2003). These are some of the goals of the idea of "3P" in medicine was to provide medical practitioners with reliable and productive risk management approaches as well as early, precise, and tailored diagnoses for patients who might delay the onset of MI.

2.11 CORONARY ARTERY DISEASE (CAD)

The cost of CAD to the government and other health stakeholders is enormous (Kuulaasmaa, et al., 2000). CAD is one of the most common causes of mortality in the world, and not only because of the financial hardships it brings (WHO, 2017; Gendeirs, et al., 2012). The buildup of plaque in the arteries supplying blood to the heart (coronary

arteries) results in coronary artery disease (CAD) (BBC, 2013). A plaque forms on the arterial wall when cholesterol and other substances congregate (National Heart, 2016). Every year, CAD kills 7 million people all over the globe.

CAD is becoming more common, and treating it is expensive for the government and other healthcare providers (Kuulasmaa, et al., 2000). It's not only that CAD is financially draining, but that it's also a deadly disease that kills more people than any other in the world (WHO, 2016; Genders, et al., 2012). In the long run, coronary artery disease (CAD) may lead to heart failure, a condition in which the heart is unable to pump blood effectively. Arrhythmia is the initial sign of this condition, which causes an irregular heartbeat/cardiac rhythm (2016). Blood pressure, cholesterol, sugar (overweight/obesity), physical inactivity, a poor diet and smoking are only a few of the indicators used in the diagnosis of coronary artery disease (CAD) (National Heart, 2016). Factors associated with an increased risk of coronary artery disease include things like age, gender, and one's family history (National Heart, 2016). There are several tests that may be done by a doctor to diagnose and recommend the best therapy for a patient who has symptoms or is at high risk of heart disease, such as an EKG, an ECG, and a stress test (NHS, 2015). This method is both time-consuming and resource-intensive, making it exceedingly expensive to diagnose and treat. "Coronary artery disease starts throughout infancy and by the time of teenage years, there is evidence that plaques are created in most individuals that will remain with us for life," said Fisher, past editor of the American Heart Association journal. "Early-established preventive interventions are expected to provide higher lifetime benefits. Healthy living can postpone CAD development and it is hoped that CAD would be rectified before it may cause CHD."A balanced lifestyle with excellent diet, weight control and lots of physical exercise may play a major part in preventing CAD. Figure 2.2 below illustrates the difference between a normal coronary artery and a CAD artery in terms of size and shape.

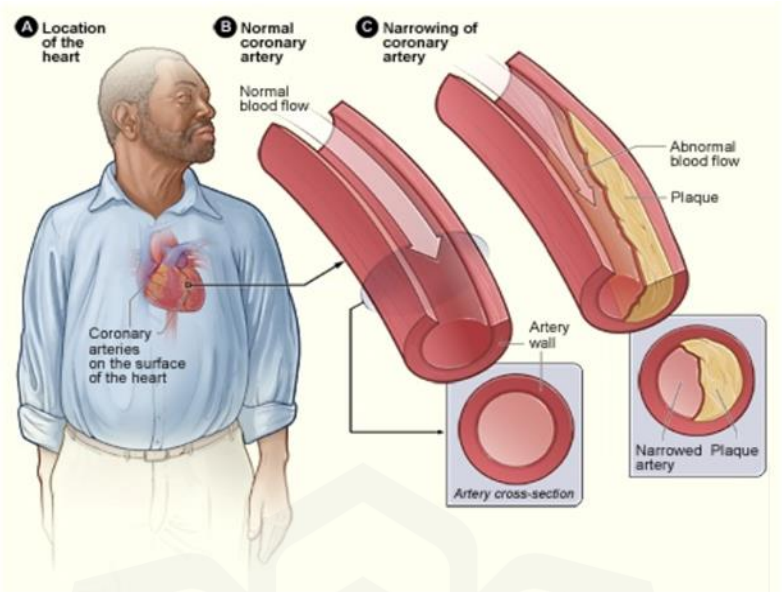


Figure 2.2 The difference between a normal coronary artery and a CAD artery in terms of size and shape Source (National Heart, Lung, and Blood Institute. License: Public Domain)

2.12 SYMPTOMS OF CAD

When it comes to Coronary Artery Disease (CAD), individuals may be completely unaware they have it since they have no symptoms at all at first. Plaque formation in your arteries may take anywhere from several years to many decades. Even minor symptoms may suggest your heart is working harder to carry oxygen-rich blood to your body when your arteries tighten. Chest discomfort and shortness of breath are the most frequent symptoms, particularly after mild physical activity such as going up stairs, although they may occur even while your lying-in bed at night.

Coronary artery disease may be undetected until a heart attack occurs. Chest pain (angina) symptoms include heaviness, tightness, pressure, hurting, burning, numbness, heaviness, squeezing or a dull ache. Angina may also have difficulty breathing. Discomfort may also be felt in your left shoulder, arms, neck, back, or jaw, or it can extend throughout your whole body.

- I'm worn out.

- Feeling dizzy or lightheaded.
- Nausea.
- Weakness.
- Women may have distinct symptoms of a heart attack, such as discomfort or pain in the shoulders, neck, abdominal (belly), and/or back.
- Constipation or a feeling of unease in the stomach.
- Anxiety doesn't seem to have a cause.
- Sweat that feels like ice.

2.13 DIAGNOSIS OF CAD

If you aren't in immediate danger, the cardiologist will talk to you about your symptoms and medical history before doing a physical exam (such as a heart attack or stroke). The following types of diagnostic tests may be used:

These examinations capture the electrical activity of the heart via the use of an electrocardiograph (EKG). Tests for heart attacks, ischemia, and irregular heartbeats. By placing your heart under the greatest amount of strain, this treadmill test might reveal whether or not you have a healthy heart. This test has the potential to identify angina and coronary artery obstructions.

Pharmacologic stress test: Medication is administered to raise your heart rate and imitate exercise instead of utilizing exercise to test your heart while it's working its hardest. Angina and coronary artery blockages may be detected with this test. Atherosclerosis (hardening of the arteries) may be detected by a test that detects calcium deposits in the coronary arteries. It utilizes sound waves to examine how well your heart's structures operate and how well your heart is functioning. There are a variety of blood tests performed to check for conditions that may damage the arteries, such as high triglycerides, high lipoprotein (bad) cholesterol, high C-reactive protein (CRP), high glucose, high hemoglobin, and other conditions. To check for coronary artery disease, small tubes are inserted into the heart's blood vessels and monitored. Cathode-ray angiography. Other types of imaging examinations for diagnosing health issues include A radioactive tracer is

administered prior to the nuclear imaging test, and pictures of the heart are produced. Computerized tomography angiography: CT and contrast dye are used to examine 3D images of the heart in motion and identify blockages in the coronary arteries.

2.14 RISK FACTORS (RF)

According to the American Heart Association's study, the following are likely causes of coronary artery disease:

Age: Most heart disease deaths occur in those 65 and older. Women are more likely than men to die within a few weeks after a heart attack at the age of 65.

Gender: Heart attacks are more common in males than in women, and they occur earlier in life. When it comes to heart disease deaths, women's mortality rate is lower than that of men.

Inheritance (Including Race): If children have parents with heart disease, they will acquire the illness more often. Most individuals with a significant family history of heart disease are likely to produce one or more additional risk factors.

Smoke of tobacco: The chance of a smoker developing CHD for non-smokers is considerably greater. Cigarette smoking is a strong independent risk factor for sudden heart mortality in CHD patients. Cigarette smoking also works to significantly raise the risk of CHD with other risk factors. Even for non-smokers, exposure to smoke of other individuals raises the risk of heart disease.

High cholesterol in the blood: The risk of CHD increases as blood cholesterol increases. This risk rises further if additional risk factors (such as high blood pressure and cigarette smoking) are present. Cholesterol levels in a person are also influenced by age, sex, inheritance, and nutrition.

Blood pressure higher: High blood pressure increases the effort in the heart, thickening and strengthening the heart muscle. This cardiac muscle strengthening is not natural and inhibits the heart from functioning correctly. The risk of stroke, heart attack, renal failure and heart failure is also increased.

Inactivity of the physical: Inactive lifestyles are a risk factor for heart disease. Regular

physical exercise, moderate to vigorous, helps to decrease the risk of heart or blood vessel disease. Even modest activity is beneficial, if done on a regular and long-term basis. Physical exercise may help manage blood cholesterol, obesity and diabetes, and decrease blood pressure in certain individuals.

Obesity and excess weight: People who have extra body fat — particularly in their tail — are more prone to have heart and stroke even if there are no other risk factors. Overweight and obese people with risk factors for cardiovascular disorders, e.g. hypertension, high cholesterol, or elevated blood sugar, may alter their lifestyles and reduce triglycerides, the blood glucose, HbA1c, and the risk of Type 2 diabetes clinically. Although many individuals may have difficulties reducing weight, a sustained decrease in weight of 3% to 5% may lead to clinically significant reductions in some risk factors. Weight reduction may decrease over 5 percent heart rate, cholesterol, and blood glucose.

Mellitus diabetes: Diabetes significantly raises the risk of cardiovascular disease. Much if glucose levels are managed, diabetes raises the risk of heart disease and stroke, but if blood sugar is not properly monitored the dangers are even higher. At least 68% of individuals over 65 years old with diabetes die from heart disease and 16% die from stroke.

2.15 PROMOTING CLINICAL JUDGEMENT

Clinical decision support systems (CDSS) apply to any electronic device built and established to aid with critical clinical decision-making. An area that has been of particular importance that has been in recent years the capacity to evaluate the attributes of different individuals to create patient-specific reviews or guidelines, which are subsequently provided to physicians for Consideration (Bright *et al.*, 2012). Due to its ability of drawing on the abundance of daily clinical knowledge, which often is not exploited, the CDSS has become one of the most critical elements in future health care, including data-driven clinical procedures like diagnosis and prognosis as well as the exploration of new medical insights (e.g., the underlying mechanisms of a disease). Furthermore, the functionality of different programmed may be mixed and given, making it a dynamic tool whose usefulness and Efficacy in the clinical sense cannot be underestimated. (Bright *et al.*, 2012). In professional

decisions, the CDSS not only supports the need to bridge the gaps between practitioners, having them adhere to the same level of practice as their best practice. In spite of this benefit, CDSS has been highly appealing in the healthcare market, with many interesting works from the Artificial Intelligence Research community (W. G. J. A. o. i. m. Baxt, 1991; Brennan *et al.*, 2013; Chawla & Davis, 2013; Green *et al.*, 1995; Grotzinger *et al.*, 2012) - postulating methods that incorporate both knowledge-driven and data-driven principles for medical through the continuing implementation of CDSS.

2.16 ARRYHTHMIA DISEASE

Arrhythmia is an issue with the heartbeat rate or rhythm. It implies that perhaps the heart beats too fast, too slowly, or irregularly. Whenever the heart beats quicker than usual, tachycardia is termed. If the heart beats too slowly, bradycardia is termed. Atrium fibrillation is the most frequent form of arrhythmia, which produces irregular and rapid heartbeat.Many things may alter the rhythm of the heart, such as heart attack, smoking, cardiac abnormalities, and stress. Some chemicals or medications may also induce rhythmic conditions. Arrhythmic symptoms include quick or sluggish heartbeat, skip beats, light-headedness or dizziness, tachycardia, breathlessness, and perspiration (Medicine, 2016).

There are many kinds of rhythms, and each kind relates to a pattern and may thus be identified and classified. The rhythms may be categorized into two main groups. The first group comprises of arrhythmias made of a single irregular heartbeat known as morphological arrhythmias. The other group is made up of rhythms produced by the so-called rhythmic arrhythmic a series of irregular heartbeats. Both kinds of arrhythmias lead to morphological changes in heartbeat wave frequency, which may be detected via the ECG test. The most frequent kinds of rhythms are shown in Figure 2.3 below:

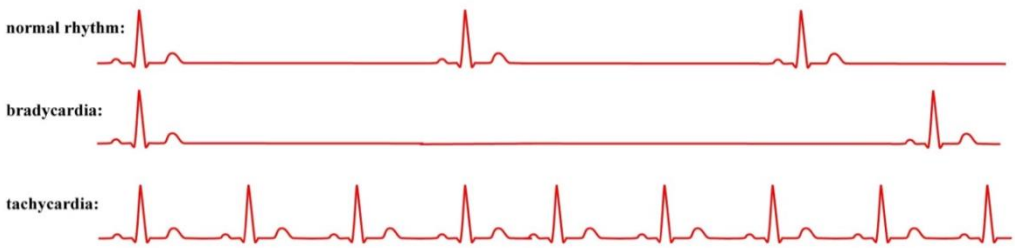


Figure 2.3 The most frequent kinds of rhythms (source: Lemmer, B. 2006)

The procedure of detecting and categorizing arrhythmias may be extremely problematic for a human person, because it is often required to analyze every heartbeat of the ECG data collected over many hours, or even days by a patient wearing a Holter monitor. Furthermore, due to tiredness, there is a potential of human mistake by the individual doing ECG data analysis.

2.17 ARRHYTHMIAS CAUSES & TYPES

Arrhythmia is caused by a cardiac electrical system issue. Some arrhythmia causes include: Irritable cells of the heart: Heart cells may begin to fail and give out aberrant electrical impulses. Signals from these dysfunctional cardiac cells interfere with the correct signals from the heart's normal pacemaker. This 'defuses' the heart that causes an uneven pulse.

Signals blocked: The electronic impulses telling the heart to beat may be 'blocked.' This slowly causes the heart to beat.

Abnormal trajectory: Electrical impulses may begin in the correct location but are stopped and diverted in order not to follow the correct route through the heart and create an arrhythmia.

Drugs and stimulants: in certain instances, arrhythmia may be caused by drugs and other substances, such as coffee, nicotine, and alcohol.

The most known types are:

Bradycardia symptoms (slow heartbeat): This word is used when the heart beats too slowly, usually fewer than 60 beats per minute. When the heart beats so slowly, it's

significant that it can't pump enough blood to fulfil the body's requirements. Untreated bradycardia may lead to extreme fatigue, dizziness, lightheadedness or fainting, as insufficient blood reaches the brain. A sluggish heartbeat may be normal and may be linked to better fitness.

Syndrome of the sick sinus: This phrase explains when the natural cardiac pacemaker works too slowly, and "fires" instruct the heart to beat slowly. It may be caused by age or the fatty tissue of the heart's blood vessels.

The heart block is termed the heart block when the signal from the receiving chambers (atria) to the chambers (ventricles) of the heart is delayed or blocked. It's unusual, but it may be severe. The symptoms may be minor or severe, depending on where and severity of the obstruction.

Tachycardia (a fast heartbeat): Tachycardia is usually above 100 beats per minute when the heart beats excessively quickly. Some types of tachycardia are easy to treat and not severe, while others may endanger life.

Tachycardia super ventricular: "Supraventricular tachycardia" is a fast pulse starting in the collecting chambers of the heart, atria, or electric track in the atria (SVT). Common SVT forms include atrial and atrial fibrillation.

Arterial Flutter: "atrial flutter" is an additional or early electrical signal that travels around the atria in a circle rather than following the usual signal route. This 'over-stimulation' causes the auricular to contract rapidly or 'flutter' considerably faster than usual. Most of this fluttering is prevented from beating too quickly by an electrical route from the atria to inhibit the pumping of the heart, the ventricles. Atrial flutter does not typically endanger life, but may nevertheless induce chest discomfort, faintness or more severe cardiac issues.

Fibrillation of the atrium: "Atrial fibrillation" is the most frequent type of SVT. This occurs when 'waves' of unchecked electrical impulses flow through the atria from the sinus node, instead of the usual controlled signals. The unregulated impulses force muscle fiber in the atria to contract out of time, such that the atria 'quiver.' Some of this electrical abnormality reaches the ventricles and causes a fast and erratic heart rate. If the heart is fibrous, it doesn't pump consistently or function as it should. Atrial fibrillation may induce heartbeat, irregular pulsation, tightness or chest discomfort, faintness, or vertigo. Atrial fibrillation may also raise the risk of stroke since blood in the atria may coagulate. These

coagulations may break free from the heart, enter the circulation, and go into the brain causing a stroke.

Supraventricular paroxysmal tachycardia: "Paroxysmal supraventricular tachycardia" is a "short circuit created by an additional electric connection or a heart-pathway which makes the heart susceptible to events of sudden frequent fast heartbeats which last minutes or even hours (PSVT). While these episodes may be scary, they are seldom harmful and may be handled quite successfully.

White-Wolff-Parkinson syndrome: The "Wolff-Parkinson-White Syndrome" is termed an additional or aberrant electric route which links the atria with the ventricles producing SVT episodes.

Ventricular Tachycardia: It may be extremely hazardous if the ventricles beat excessively quickly, termed "ventricular tachycardia." Tachycardia ventricular which is so severe that the ventricles cannot pump properly may lead to ventricular fibrillation. Ventricular fibrillation occurs when the electrical impulse that should cause heartbeat divides around the ventricles into uncontrolled 'waves.' This life-threatening condition has to be promptly addressed.

2.18 SLEEP APNEA

Tagluk et al. (2010) have developed a novel classification technique for Sleep Apnea Syndrome (SAS) utilizing wavelets and artificial neural network transformation (ANN). The network was pre - trained for various impulses. By multi resolution wavelet transformations, the abdominal respiration signals were split into spectral components. These spectral components were used for the artificial neural network inputs. The neural network was then designed to give three outputs to categories of the patient's SAS condition. The suggested approach for this study is shown in Figure below. A coefficient of a discrete wavelet transform applied to the raw specimens of the hypoxia in the abdominal stress signal constituted the basis of the neural network. The experimental findings produced utilizing 360 apneas from 21 distinct patients showed the validity of the technique suggested. The highest overall accuracy was 78.85 per cent, which in comparison to manual

grading was excellent enough. The more prevalent sleep disorder is obstructive sleep apnea. Obstructive sleep apnea is characterized by repeated episodes of partial or full obstruction of the upper airway while sleeping. Apneic episodes need more effort from your chest muscles and diaphragm since the pressure within your lungs is increasing. A sharp gasp or jolt of the body signals that breathing has resumed. These episodes may make it difficult to get a good night's sleep, decrease the amount of oxygen reaching critical organs, and even induce irregular heartbeat.

Because of unsteadiness in the breathing control center of patients with central sleep apnea, their airways are not closed while they sleep. Central apnea is linked to the nervous system's overall health.

2.19 SYMPTOMS OF SLEEP APNEA

OSA is often detected for the first time by the patient's bed companion. There are a lot of people who are impacted by this who don't have any sleep issues. OSA is most often characterized by the following signs and symptoms:

- Snoring.
- Tiredness or drowsiness throughout the day.
- Frequent nocturnal awakenings due to restlessness during sleeping.
- Wakes up suddenly feeling like you're gasping or choked.
- In the morning, a painful throat or a dry mouth.
- Inability to concentrate, forgetfulness, or impatience are symptoms of cognitive impairment.
- Disturbances of mood (depression or anxiety).
- Sweating throughout the night.
- Urinating often in the middle of the night.
- Headaches.

Those who suffer from central sleep apnea are more likely to have frequent awakenings or sleeplessness, as well as a choking or gasping feeling when they first wake up in the morning Children's symptoms may be more subtle and include:

- Inadequate grades at school.
- In the classroom, sluggishness or drowsiness is frequently misconstrued as laziness.
- The inability to properly breathe via the mouth and swallow food throughout the day.
- Breathing causes ribcage inward movement, or diaphragmatic breathing.
- Inconvenient sleeping arrangements, such as sleeping on your stomach or with your neck outstretched.
- Nighttime perspiration is excessive.
- Diseases of cognition and behavior (hyperactivity, attention deficits).
- Bedwetting.

Sleep apnea, if ignored, may lead to a variety of health issues, such as hypertension, stroke, arrhythmias, cardiomyopathy, heart failure, diabetes, obesity, and heart attacks. Because people with sleep apnea have increased blood pressure, sleep apnea is prone to induce arrhythmias and heart failure. About half of those with heart problems or atrial fibrillation have sleep apnea. The reason for this is that sleep apnea may lead to the following symptoms:

- Oxygen depletion that occurs often (hypoxia).
- Changes in the atmospheric concentration of carbon dioxide.
- Pressure changes in the chest have an immediate impact on the heart.
- Inflammation-related biomarkers are rising in the body.

Sleep apnea is common in cardiac arrhythmias including heart failure (it's a coin flip the weather patient has it or not). Sleep apnea is a potentially serious sleep disorder in which breathing repeatedly stops and starts. If you snore loudly and feel tired even after a full night's sleep, you might have sleep apnea.

The main types of sleep apnea are:

Obstructive sleep apnea, the more common form that occurs when throat muscles relax.

Central sleep apnea, which occurs when your brain doesn't send proper signals to the muscles that control breathing.

Complex sleep apnea syndrome, also known as treatment-emergent central sleep apnea, which occurs when someone has both obstructive sleep apnea and central sleep apnea. Figure 2.4 explains the methodology proposed for classifying Apnea types as follows:

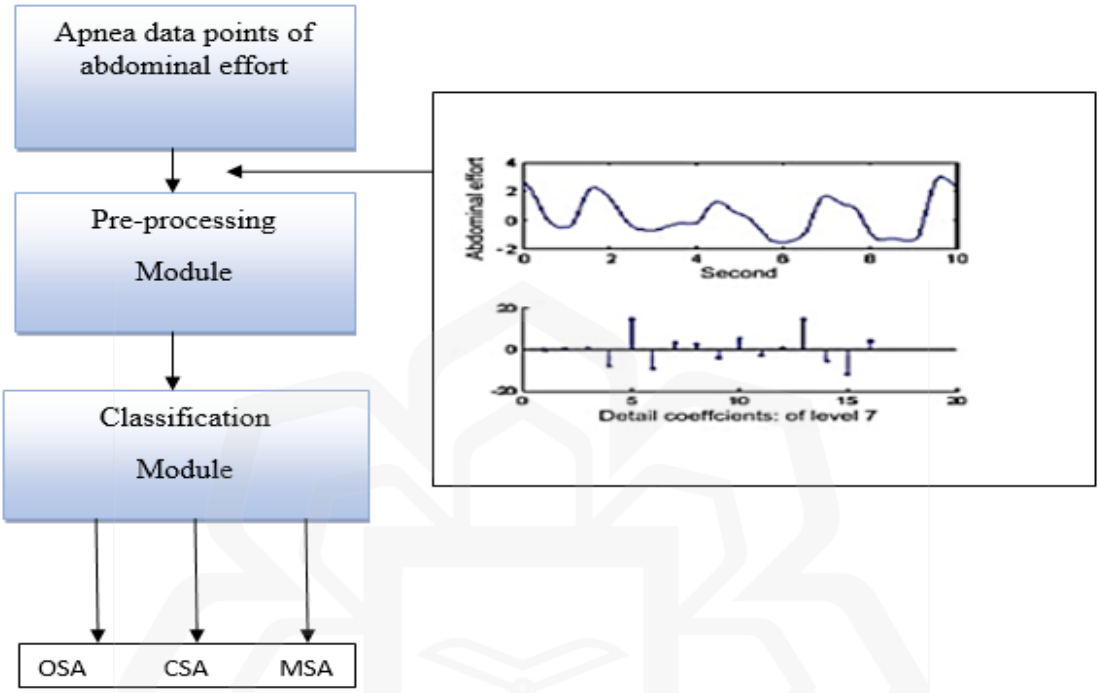


Figure 2.4 Methodology Proposed for classifying Apnea Types (Source: Tagluk et al., 2010

2.20 OVERLAPPING SYMPTOMS

Depression diagnosis also relies on somatic symptoms that coincide with the overlapping of symptoms of certain medical conditions. (Ellis *et al.*, 2006); Milberger *et al.* (1995). Ellis *et al.* (2006) studied Disease signs correlate with depressive symptoms. This study reviewed reported interviews with 46 of 61 qualifying populations residing among older adults with advanced disorders and a significant number of somatic depressive symptoms (AbuKhousa & Campbell, 2012; Alhadeethy *et al.*, 2021). Participants replied to an interactive query regarding emotions and formal queries about depressive symptoms. In the studies by Hsu *et al.* (2017) The overlapping general populace is very frequent with gastro esophageal reflux (GERD) and dyspepsia, although the relationship of each is little known.

The management of signs and risk conditions is the priority of primary care clinicians and cardiologists, allowing less opportunity to cope with thoughts and emotions (Chaddha *et al.*, 2016).Figure 2.5 shows a simple diagram for explaining the overlapping of Heart – CVD symptoms their distribution and redundancy after assigning the attributes with a specific value.

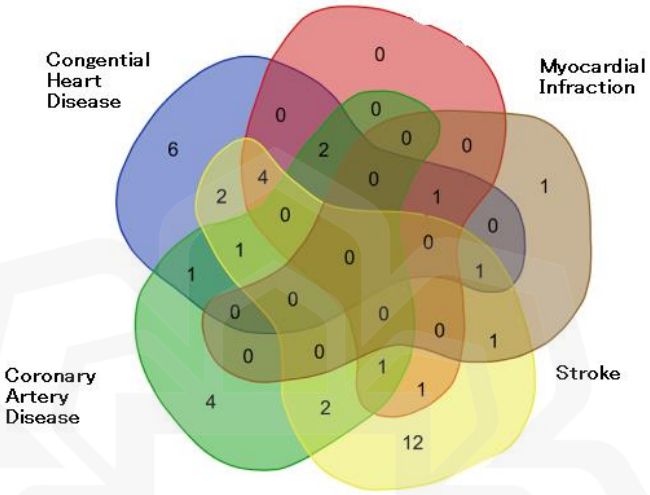


Figure 2.5 Overlapping of Heart – CVD symptoms Diag

General concern over everyday activities, though requirements for the formal diagnosis of generalized anxiety disorder are not fulfilled, has been found to encourage and accelerate coronary disease in psychological pain, including anguish and tension. The impact can also be progressive, indicating that further episode of anxiety, cold and fatigue, despite the usual coronary arteries or palpitations or heart beating, may be correlated with a greater incidence of cardiovascular disorders, as well as with the combination of symptoms or indicators of cardiovascular disease, such as chest pain act as overlapping symptoms of CVDs (Mensah & Collins, 2015; A. J. Thomas *et al.*, 2004).

Heart attack symptoms include the following:

- Discomfort or pressure in the left arm or chest
- Feelings of satiety, bloating, or choking (may feel like heartburn)
- Back, neck, throat, or arm pains are all possible symptoms.

- Dizziness or sweating are symptoms of a more serious condition.
- Shortness of breath, dizziness, or trembling Extreme fatigue

Heart failure signs and symptoms include:

- a feeling of being out of breath (often causes a hacking cough)
- Unhealthy weight gain occurs quickly (a weight gain of 2 or 3 pounds in a day is possible)
- Ankle, thigh, and abdominal swelling
- Dizziness
- Weakness and exhaustion
- An erratic or rapid pulse
- Nausea, heart palpitations, a racing heart, difficulty breathing when you wake up at night, and a shift in sleep patterns are all possible symptoms.
- Weight loss and a lack of appetite are two late signs.

The severity of your heart problems symptoms may or may not be linked to your heart's condition. Despite your symptoms, your heart function may only be slightly impaired. The treatment for heart failure varies depending on the underlying reason, however it often involves medications to manage the symptoms, such as:

- To remove excess fluid from the body, use diuretics or water tablets.
- Beta-blockers are used to stop the effects of adrenaline.
- help regulate sodium levels with ACE inhibitors.

Symptoms of an arrhythmia include:

- a racing pulse (experiencing "flip-flops" or "skip-beats," or feeling like your heart is "running away" is an example of this)
- your heart is pounding.
- a sensation of faintness, lightheadedness, or dizziness
- a feeling of being out of breath
- pain in the sternum
- a feeling of weakness or exhaustion
- Depending on the kind of irregular heartbeat, treatment options include:
- heart-slowing medication

- Drugs that can help you get your heartbeat back to normal.
- Antithrombotic drugs
- Cardioversion is a therapy that includes shocking your heart with a powerful electrical current to restore its normal rhythm.

Heart valve dysfunction may cause a variety of symptoms, including the following:

- Breathlessness or difficulty breathing after exertion.
- fatigue, or a feeling of unsteadiness
- Fainting
- Chest discomfort when participating in a sport.
- A fast pulse, irregular pulse, skipped pulses, or a flip-flop sensation in your chest may be felt as palpitations.
- Symptoms of heart failure caused by valve dysfunction include:
- Ankle, foot, or abdominal swelling
- gaining a lot of weight quickly

Adults with congenital cardiac disease may have the following signs and symptoms:

- a feeling of being out of breath
- Exercise is difficult because of a lack of time or resources.
- Heart failure and valve disease symptoms

Some signs of newborn congenital heart disease include:

- The skin, nails, and lips all have a pale blue tinge to them.
- breathing that is too rapid, as well as inadequate nutrition.
- Fat growth that isn't healthy
- a propensity towards bronchitis
- Absence of desire or capacity to work out.
- Many individuals with cardiac muscle illness have no or mild symptoms and go about their daily lives as usual. Some people's symptoms increase over time as their heart function deteriorates.

Cardiomyopathy symptoms may occur at any age and include the following:

- a tightness or discomfort in the chest (typically after physical exertion, although it may also happen after rest or a meal)

- the signs and symptoms of cardiac arrest
- Reduced blood flow to the legs
- Fatigue\Fainting
- a racing pulse (chest flutters because of abnormal heart beats)
- Cardiomyopathy may cause sudden death in a tiny percentage of patients.

Pericarditis is characterized by the following signs and symptoms:

- A tightness in the chest. Unlike angina, this isn't a heart attack. In the middle of the chest, it may be a stab wound. In certain cases, the discomfort may extend to the neck, limbs, or back. Coughing, swallowing, or taking a big breath makes it worse; sitting forward makes it better.
- Fever of a low grade
- Heart rate has risen.
- Pericarditis usually goes away on its own, but in certain cases, doctors may prescribe medication to help manage it.
- Aspirin and other anti-inflammatory medications

2.21 THEORITICAL REVIEWS

Megha B. et al., (2023) used different practices for the analysis of various heart diseases using Machine Learning (ML) and Deep Learning (DL) algorithms such as Convolutional Neural Networks (CNNs), recurrent neural networks, deep belief networks, long short-term memory, and others investigated by different researchers over the time span. The articles, for this study, were considered from 2018 to 2022 and after the screening, 63 articles were used for primary study. The most common issue was the unavailability of larger and discrete datasets followed by the improvement of the pre-existing models.

Maria T. et al., (2023) proposed a model using deep learning methods, combined with feature augmentation techniques for evaluating whether patients are at risk of suffering cardiovascular disease. The results of the proposed methods outperform other state of the art methods by 4.4%, leading to a precision of a 90%, which presents a significant improvement, even more so when it comes to an affliction that affects a large population.

Krittanawong et al., (2020) evaluated machine learning algorithms' overall predictive ability of predicting cardiovascular disease. The ability to predict diseases such as coronary artery disease, cardiac arrhythmias, heart failure, and stroke was observed. They used the area under the curve metric was d in the prediction analysis. However, because of the heterogeneity of machine learning algorithms, identifying an optimal algorithm for cardiovascular disease remains a challenge.

Cao et al., (2019) studied deep learning models for cardiovascular disease to improve the medical diagnosis and screening of diseases. It was challenging for them to scan and diagnose the CVDs in the early stage. They used the medical image recognition type of deep learning to screen and diagnose various diseases in the present Research. For a refresher, picture detection, distinction, and pattern recognition are important, and it is going to be popular with loads of research scholars. As a component, picture segmentation, grouping, and other factors are used to think about in-depth cardiovascular disease. CVDS are becoming the main Constituent of improving mortality rates in recent years, owing to the lack of information regarding the early treatment of the disease. The machine learning techniques helped them to work better in certain areas, which the author(s) define as computational biomedical. Also, the study is looking forward to creating a patient portal integrated with deep learning technologies to simplify the care of the patient and to include remote evaluation and tracking. Saravana et al., (2023) used in this study, eight ML classifiers which were utilized to identify crucial features that enhance the accuracy of heart disease prediction. Various combinations of features and well-known classification algorithms were employed to develop the prediction model. Neural network models, such as Naïve Bayes and Radial Basis Functions, were implemented, achieving accuracies of 94.78% and 90.78% respectively in heart disease prediction. Among the state-of-the-art methods for cardiovascular problem prediction, Learning Vector Quantization exhibited the highest accuracy rate of 98.7%. The motivation behind predicting Cardiovascular Heart Disease lies in its potential to save lives, improves health outcomes, and allocates healthcare resources efficiently.

Hartmut Hoehle et al., (2016) introduced his research paper on cardiovascular disease symptoms and policies that guide toward their difficulty in problem. This study conceptualizes deep learning model creates and approves the treatment to validate their

measurement that to which level it is measured to be the best. The researcher explains that predictive deep learning model has been gaining a lot of attention from individuals across the world when we Consider patient's interactions with PC.

To conceptualize deep learning model, Hartmut researched Microsoft's deep learning rules that they defined in the latest square's model development in the hospitals and emphasize better enhancement from the perspective of treatment. After that researchers have gone through the survey of multiple hospitals about the cardiovascular diseases and their treatments used the SPSS technique in putting all content gathered from multiple hospitals by this survey. When doing statistical analysis this researcher exploratory factor investigation technique on the German patients that what symptoms influences them and make the changes in their functions or daily tasks. The results showed that square's model is predicting the behavior of German patients this paper can be utilized to direct future research in patients to PC interaction and will help in the feasible structure of deep learning model.

D. Lee et al., (2015) were working on finding out the relationship between the symptoms with patient's irregularities. In this paper, research was conducted that how patients interacting with treatments and how to get recovery easily. Consequences say that for deep learning model on Discomfort Pressure in Chest, trust and brand consciousness are researched to keep up a long-term patient's relationship. 310 patients from Korea participated in this research. The technique that was used to conduct research was utilizing the Partial Least Squares (PSL) technique as it is more reasonable than the covariance-based methodology.

(Lee, Antecedents, and consequences of mobile phone usability: Linking simplicity and interactivity to SATS, trust, and brand loyalty. Information & Management, 2015) Simplicity factor and interaction factor were taken to be two variables defined in this research that need to be analyzed in the research. The reliability symptoms for both variables were higher than 0.80 and the Average security was higher than 0.50, which predicts the strong reliability and validity of the data as well as in variables.

Karima Momani et al., (2016) processed an experimental analysis dependent on a lot of measures to assess the deep learning procedures running on various hospitals that which one is the better to reduce the symptoms. In this analysis usability of deep learning

applications is assessed on multiple Operating systems; IOS, Android, Symbian. A model is created on software quality standards of the ISO 9126 which is utilized to assess that which one of the Operating systems is better. ISO 25062 and ISO 9241 models have been utilized in this research and consist of information collected from 32 clients, which 7.0 questionnaires. Two applications were investigated; google applications and maps. Multiple difficulties were recognized while utilizing applications by the users and the flaws have been identified one of them was screen size, the resolution of the screen and the memory of the deep learning phone, approve the discoveries of our system with some software issues not compatible with the user.

Ali Balapour et al., (2017) published a research paper Meta-regression study on the usability of applications. Fundamentally absence of usability is a significant Common issue in PC sites and deep learning applications. This article was going to investigate the correlation between usability perception and a factor influencing usability.

Bakator *et al.* (2018) did the implementation of deep learning for medical diagnosis in their study. A detailed review of numerous research papers in the use of deep neural networks in the medical field has been carried out. More than 300 research papers were investigated by the authors, and 46 articles were presented in more depth after many selection measures. The findings show that when it comes to deep learning and medical picture processing, convolutional neural networks (CNN) are the most frequently portrayed. Furthermore, MRI also was commonly used as training details. Segmentation is the most represented when it comes to the particular use. It is necessary to remember that the study and interpretation of the paper are weighted against more current papers and articles in the title that contained "deep learning." It can be shown that the form of data that is used to train and implement deep neural networks has a wide range of data. For expert-level diagnostics, CT scan images, MRIs, fundus photography, and other forms of details may be used. The descriptive structure of this analysis can add to the current body of literature on a moderate basis. The goal was to include an article that was objective, clear, and succinct. The individual study findings provide ample knowledge and insight into deep learning applications for identifying, recognizing, segmentation, and diagnosing different diseases and anomalies in particular anatomical regions of concern (ROI). Without even a question, the use of deep learning in the medical sector can continue to grow since it has

already produced promising results in medical image processing and, more specifically, in the identification and diagnosis of image-based cancer. In the long term, this will improve healthcare reliability and Cyanosis, thus reducing the likelihood of chronic illnesses being detected late. Though, as stated earlier, there will still be a fair journey to go until neural networks of ideas associated can be commercially important. Finally, artificial intelligence is supposed to "rise" by the combination of learning regarding interpretation and abstract thinking.

Brunese *et al.* (2020) explored the deep learning approaches for heart diseases through cardiac sounds. Machine learning models were selected to recognize cardiac sound patterns to evaluate the symptoms and risk predictions. Many factors of mortality are linked with cardiovascular disease. In practice, the heart rhythm is disrupted by many irregularities, such as a heart murmur or artifact. They interpreted this function vector as an input for a deep neural network by accumulating a collection of features obtainable directly from cardiac sounds to distinguish whether a cardiac sound belongs to a healthy individual or to a patient with a cardiac condition. One person loses his or her life every 37 seconds because of cardiovascular diseases according to the Centre for Disease Control and Prevention (Ulbricht & Southgate, 1991).

The authors chose two staged evaluations in their study design. Descriptive feature of the samples were explored using Orange frameworks, a collection of machine learning applications that is largely used for empirical analysis in data mining (Demšar *et al.*, 2013). Secondly, the model was created using Keras, a high-level API for neural networks, written in Python, a Google research and development open-source data visualization library. The experiment that this research undertook shows the feasibility of the method suggested in the real world.

Karimi-Bidhendi *et al.* (2020) explored the completely automatic segmentation model for deep learning for cardiovascular magnetic resonance. Improving clinical workflow, diagnostic precision and analysis Efficacy are known to be unmet clinical needs for the increasing patient FEVE unity of congenital heart disease (CHD). Cardiovascular visualization includes non-invasive and non-ionizing tests for CHD patients. While CMR data allows for precise heart function analysis and anatomy, the clinical workflow relies mostly on time-Consuming CMR images manual analysis. Therefore, as the present work

aims to create, an integrated and reliable segmentation tool specifically dedicated to pediatric CMR images will greatly enhance the clinical workflow. Large, annotated datasets are needed to train artificial intelligence algorithms for CMR research, which are not readily accessible for pediatric content, especially in patients with CHD. The authors built a novel approach to mitigate this issue; a Genetic Adversarial Network (GAN) is used to synthesize the training data collection. In addition, the authors trained and validated a large, completely Coevolutionary network (FCN), which the authors made accessible to the public. The findings of the chambers' segmentation from our completely automated system demonstrated good segmentation, and two separate statistical tests did not notice a substantial statistical discrepancy.

Peili *et al.* (2018) estimated the deep learning models on the risk factors of coronary artery disease. One of the prevalent diseases that endanger the health and existence of individuals is heart disease (CHD). The standard form of diagnosis depends largely on the technical expertise of the physicians and the experience of the clinic. In recent years, precision medical therapies focused on big data and deep learning have been a science hotspot.

(Litjens *et al.*, 2017; Shen *et al.*, 2017). In comparison to undertaking algorithms, Machine Learning is Part of a broader class of artificial intelligence techniques that are based on data representation learning (Nature, 2015). This is a neural artificial network with many unknown layers between the input layer and the output layer known as the Deep Neural Network this same machine learning algorithm (Goodfellow *et al.*, 2016; LeCun *et al.*, 1998; Schmidhuber, 2015).The literature, however, focuses mainly on how the CHD early warning models can be designed and customized, while missing the lifecycle management simulation Consequences of data-model-experimental outcomes training. The algorithm of Tracking-Ancestor and The Find-Specified-Ancestor algorithms are programmed to navigate the family of the deep neural model. The results of the Research suggest that would include a Productive automated data processing system for CHD early warning researchers. Srivastava *et al.* (2020) has introduced a new method of evaluating disease using a dataset of Cleveland Heart Disease, integrating computing capacity with separate ML and DM algorithms and has concluded that K-Nearest Neighbors provides the maximum Accuracy of all algorithms by 87 percent. In addition, and therefore not enough

in the medical sector, a web app is built using a python flask that enables to insert attributes and forecast heart failure.

David (2018) used a predictive method to evaluate and forecast the potential of cardiovascular disease in utilizing three data mining classification algorithms such as Random Forecast, Decision Tree and Naïve Bayes. The main purpose of this essential work was to find the best classification algorithm to ensure optimum Cyanosis in classification of ordinary and abnormal individuals, which takes longer to settle on a better classification and achieve the results needed. Prevention of loss of life is also necessary at an earlier point. The experimental setup was developed to test the Chest Pain of algorithms with the help of the UCI machine learning repository's dataset for heart disease. The Random Forest Algorithm is found to be most Productive with 81 percent relative to other algorithms for Prediction of cardiac disease, and this finding in the medical sector is known to be incomplete as it needs stronger and productive performance.

Kwon *et al.* (2019) predicted in-hospital mortality rates using deep network models among the patients of heart diseases through echocardiographic techniques. The leading cause of global death is heart disease (HD); HD has many mortalities forecast models for recognizing chronically sick patients and directing decision-making. However, current templates cannot be included in the process of initial therapy or screening. The purpose of this Research was to extract and test an HD echocardiography-based mortality prediction. Usage of derivative data from a clinic, the authors built a prediction model focused on deep learning. And researchers also carried out external confirmation using hospital echocardiography study. The authors carried out a subgroup study of hospital patients with coronary heart disease and heart failure and associated deep learning with the statistical models commonly used, such as GRACE score, TIMI, and GWTG-HF ratings. Study conclusions have explained that the DL models focused on echocardiography forecast hospital mortality.

Bernard *et al.* (2018) explored MRI prediction using the deep neural networks, and deep learning techniques to diagnose (members *et al.*, 2014; Norris *et al.*, 1992; Silberberg *et al.*, 1989). Over the past decades, the automation of the related functions has therefore been the focus of extensive study. The dataset includes data from 150 CMRI multi-equipment recordings with comparison and description measurements from two medical

professionals. The aim of this paper is to determine how much state-of-the-art techniques of deep learning will go in evaluating CMRI, such as segmenting the myocardium and the two ventricles. The results so far suggest that fully automated CMRI analysis is about to solve the problem. This will reduce raw data processing time, allowing the patient to see test results before visiting the radiology facility. Clinicians can dictate physiological and technological criteria with pre-filled radiological reports and advanced automatic speech recognition technology in new systems. This approach can also use automated CMRI research tools (Isin & Ozdalili, 2017). However, further research is needed until accrediting authorities approve such software for MRI. Isin and Ozdalili (2017) suggested deep learning arrhythmia detection methods. In cardiac arrhythmia diagnosis, the ECG is crucial. This study uses a deep learning algorithm trained on a broad picture data set to automatically diagnose ECG arrhythmia by classifying patient ECGs into cardiac diseases. The transplanted deep Convolutional Neural Network extracts functions and feeds the collected characteristics into a basic back propagation neural network for classification. The key summarized studies are in Table 2.1:

Table 2.1 Comparatives Studies for samples of CVD Previous Models.

Author / Year	Outcomes	Methods /Algorithms	Limitations
Maria et al., (2023)	The results of the proposed methods outperform other state of the art methods by 4.4%, leading to a precision of a 90%, which presents a significant improvement	Sparse auto encoder CNN	time – cost consuming
Alireza et al., -2022	investigating the factors affecting heart attacks Compare the results of the proposed Group Model of Data Handling (GMDH) neural network with other four neural network.	Neural Networks. long short-term memory (lstm), probabilistic neural network,	time consuming in comparison with performances of four neural networks

		redialbasis function	
Ying Huang et al.,-2021	Developing a 10-year cardiovascular disease (CVD) risk prediction models for the contemporary Chinese populations based on the Guangzhou Biobank Cohort Study (GBCS)	Deep Learning Neural Networks	small number of commonly available factors
Enriko, I. K. A-(2019)	diagnosing of cardiovascular diseases based on a Tehran hospital dataset	CART tree, and CHAID tree algorithms, and reinforcement	lack of Risk Factors & time Consuming
Brunese et al. (2020)	Recognizing cardiac sound patterns to evaluate the symptoms and risk predictions	orange frameworks, high-level API for neural networks, written in Python, and development open-source data visualization library	lack in Demographics Information for Patients
Srivastava et al. (2020)	building a web app that enables inserting attributes and forecast heart failure. With accuracy 87%	ML DM algorithms K-Nearest Neighbors	result is not enough in the medical sector
David -2018	finding the best classification algorithm to ensure optimum Cyanosis in classification of ordinary and abnormal individuals, which takes longer to settle on a better classification and achieve the	Random Forecast, Decision Tree and Naïve Bayes.	High error values and hard to evaluate the interpretation.

	results needed with accuracy of 81%		
Stuti Nathaniel. (2018)	the proposed technique predicted an accurateness of about 97, 94% respectively.	Two class Bayes Point and Logistic Regression (LR)	Working on unstructured data which is not enough to give the specific diagnosis
Xiaoqing Zhang et al., -2019	Developing a novel Deep Neural Network Model for Multi Label Chronic Disease Prediction with accuracy of 81.31%	Convolution Neural Network, Binary relevance, and Label Power	Accuracy is less than 90%
Rishabh Wadhawan,-2018	Predicting a Coronary Heart Disease Using A Priori Algorithm with Data Mining Classification	Mafia algorithm K-Nearest Neighbors algorithm	Difficulties in extracting hidden knowledge related to heart disease
Cheng Hsiung et al., -2015	Cardiovascular prediction using Solo & Ensemble classifiers	Single-Neural Network Classifier and Multiple Neural Network Classifiers.	The two classifiers are not working with the same efficiency
Sudhir et al.,-2017	Early Identification of Diseases Based on Responsible Attribute by using Data Mining	Decision Tree, K-Nearest Neighbor, Support Vector Machines, Naive Bayesian Classifiers, Neural Networks	Lack of attributes
Min Chen., et al -2017	Mining Sequential Risk Patterns from Large-Scale Clinical Databases	Classification Modeling techniques.	Complexity in mining
Robert Moskovitch., -2016	Prognosis of Clinical Outcomes with Temporal Patterns and Experiences with One Class Feature Selection	Maitreya Framework	One class Feature selection is difficult to find. vout, implies the

			results have only a small improvement compared to another one
Matthew B Carson., et al (2015)	prediction based on knowledge mining of disease genes	decision tree (AD Tree) classifier	Complexity in features extraction
Alizadeh Sani et., al - 2012	diagnosing cardiovascular diseases using cost-sensitive algorithms and10-fold cross-validation	SVM Classification methods	Unbalanced data Cost sensitive

In summary, Deep learning models can be continuously updated and improved with new data. This adaptability ensures that CAD systems remain up-to-date with the latest medical knowledge and technological advancements. The adoption of deep learning in CAD is paramount for achieving more accurate, efficient, and accessible medical diagnostics. Its ability to handle complex data, extract relevant information, and adapt to new challenges positions deep learning as a transformative technology in healthcare.

2.22 RESEARCH GAP

In summary, from the literature review, while significant advancements have been made in CVD prediction, there are research gaps in the development of accurate models, customization of algorithms, understanding the impact of lifestyle changes, cost-effective interventions, integration of data, and addressing ethical considerations. Closing these gaps can lead to more effective CVD prevention and management strategies has been indicated that the proposed solution could also refer to other classification problems involving data sets of the same type used in this analysis. Yet study into clustering, noise reduction and ambiguous rules-based approaches for the detection of diseases remains a lot to be done to explore both their ability and usefulness. In future, more focus needs to be given to the data sets for the detection and prediction of diseases utilizing gradual deep learning approaches. This approach also needs to be tested on additional data sets and in particular on large data

sets, in order to illustrate the usefulness of the calculation time method on large data. Furthermore, it explores how the approach suggested can be applied to include other forms of medical data sets. Yet there is very little literature regarding the overlapping of symptoms and prediction of overlapping which creates a research gap to be filled. Our research is aimed on finding the early prediction of overlapping symptoms in the CVDs.

Deep learning offers better integration of medical data sources, addresses patient disease type heterogeneity, bridges the gap between omics and bedside phenotypes and finally enables individualized medication. Scientific advancement may require that current limits be overcome, including limited theoretical design and testing, limited interpretability, and over fit strategies. Cardiovascular medicine is good at smart analysis of continuous and massive data streams in this new era of wearable sensors and the integration of traditional data. This integration would form the foundation for a medical web of things between medical devices and analytical systems if data protection and security concerns were met. Seamless integration of various data sources could allow continuous monitoring of diseases, risk stratification, and early warnings on potential decompensation. The literature research showed that the suggested technique might also cover additional classification issues using data sets of the same kind employed in this investigation.

However, there is still more to be done to examine their potential and use in studying clusters, noise reduction, and confusing rules-based techniques for detecting illnesses. In the future, the data sets for the identification and prediction of illnesses using deep learning algorithms must be more focused. Additional data sets and in particular huge data sets must be examined to demonstrate the Empties of the time technique of computation for huge data.

2.23 CHAPTER SUMMARY

This chapter explained some of deep learning methods and algorithms that have been used in deep learning models for CVD prediction and explained the symptoms for each heart-CVD diseases and list all the attributes and risk factors for each disease and how the overlapping occurred with a simple diagram and showed all the related and previous works

with a comparative study for all Heart- CVD predictive models based on Deep learning techniques to find the research gap and future directions.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 INTRODUCTION

This chapter provides an introduction about research strategy and design, and approach, data. Sampling involves description of data, the source of data collection, characteristics, and demographics of the Risk Factors (RF) and describe the dependent/independent variables and analysis of the influential factors by applying the main methods for finding the correlation type between them and their effects (weight factor and redundancy) in detecting the overlapped symptoms, as well as presents some of the deep learning models which have been used in disease prediction, their advantages and limitations.

3.2 RESEARCH STRATEGIES

There are several techniques for recognizing the environment in which testing is performed by researchers Scientists follow two study strategies in particular: quantitative strategies focused on empirical measurements and numerical data analysis, and qualitative strategies which emphasize words rather than quantity in data collection and analysis (Bell *et al.*, 2018). The nature of the sample may be qualitative, quantitative or a combination of the two based on the research issue (Yin, 2000). Due to the complexity of the research and the need to include as many participants as possible a quantitative approach is used in this situation. Selecting characteristics and evaluating the influential factors are essential to the development of classifiers and increasing their accuracy. Therefore, we investigated the influential factors of the Gini index which is rarely used in previous studies and working only with quantitative research.

3.3 RESEARCH DESIGHN

Statistical analysis is an essential component of this study architecture. To the contrary, however, this method of analyzing empirical data harmonizes this report's earlier study problems and eventually with its findings and conclusions (Yin, 2000). The main levels for this study are, data interpretation, data clearing and pre-processing, modeling, evaluation, and implementation each level the process will be as following:

Step 1: Collecting the datasets from the online sources of UAE hospitals.

Step 2: Preprocessing the data by filtering the undesired, noisy, redundant data, and ignore the missing values using statistical methods- Filtering equations – Mean methods.

Step 3: Encoding the numeric and discrete datasets by assigning it to the values belong to a specific interval and dividing the data into subsets for training and testing, and validation as subset 1, subset 2, and subset 3, as explained later.

Step 4: Visualize the data – scatter plot by using the Matrix to find the relation between the Risk factors and known patterns and trends between the Risk factors (RF) using Correlation Test - Matrix Scatter Plots (Feature Extraction)

Step 5: finding the strength of the correlation between the Risk Factors and their effects on the patients using statistical methods as Gini -Entropy Index – Multi Linear regression Model (Feature Recognition)

Step 6: Predicate the overlapping between the most effective Risk Factors (RF) and find the accuracy of the model using Regression Model Analysis – Correlation Analysis (Prediction)

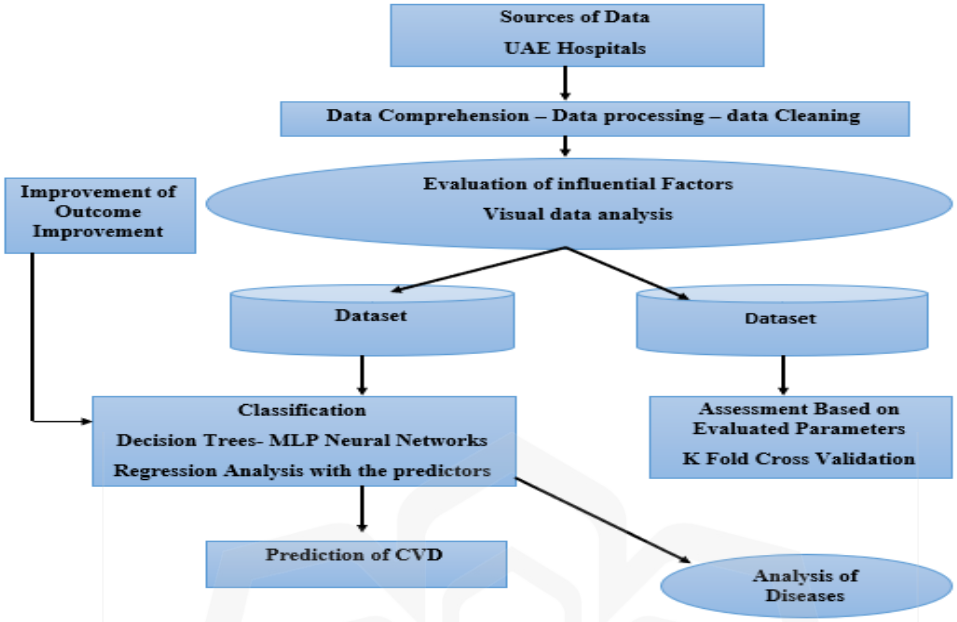


Figure 3.1 General Research Model

3.4 RESEARCH APPROACH

The method of investigation as a result, the focus of the study dictates this. It is critical to choose the greatest study method since it reveals how to answer or at least discuss research problems. For each sample, the deductive, inductive, or combined methods are chosen. The deductive method was introduced because of the nature of this research. The deductive method begins with the acceptance and refinement of the subject matter hypothesis into testable hypotheses. Researchers will be able to verify their initial study assumptions because of this process. Saunders, Lewis, and Thorn hill (2019) discuss the importance of leveraging an existing theory when commencing work from a deductive perspective in research. This approach involves starting with a deductive perspective to test an existing theoretical perspective using qualitative or quantitative methods.

3.5 RESEARCH STUDY POPULATION

In this research study, the research study population or participants selected from the UAE Hospitals with 2621 sample. The medical record from the hospitals were used as dataset from the UAE hospitals collected and surveyed in Abu Dhabi – Capital of UAE, based on SEHA Information Centre, most of CVD data using a convenience sampling technique which is a form of unlikely sampling procedure, where a community of people can be contacted or reached with the sample easily.

3.6 RESEARCH DATASET

The present dataset was collected from the following webpage http://www.figshare.com provided by the (Oulhaj *et al.*, 2020). The links of the dataset are given in the section of citation and is online available. Collecting online datasets comes with ethical considerations to ensure responsible and respectful data practices. Here are some key ethics applied for collecting online datasets:

Data Minimization: Researchers collected only the data necessary for research objectives and avoided collecting excessive or irrelevant information [4].

Respect for Rights: Respect human dignity and rights when collecting data, ensuring that it doesn't infringe on individuals' freedoms or harm them in any way [4].

Transparency: Be transparent about data collection practices, including data sources, methods, and the intended use of the data.

Data Security: Implement strong data security measures to protect collected data from unauthorized access or breaches.

Data Quality: Ensure data accuracy and reliability by employing rigorous data collection and validation processes.

Respect for Cultural Differences: Be sensitive to cultural norms and differences when collecting data from diverse populations.

Honesty and Integrity: Conduct data collection ethically, without misrepresentation, falsification, or manipulation of data.

Accountability: Take responsibility for data collection actions and their consequences, including addressing any ethical violations or breaches.

Compliance with Laws and Regulations: Adhere to relevant laws and regulations governing data collection, such as data protection and privacy laws.

Continuous Monitoring: Regularly review and assess data collection practices to ensure they remain ethical and comply with evolving ethical standards.

These ethics are essential to maintain trust, protect individuals' rights, and ensure the responsible and ethical use of online datasets in research and other applications. Description of the included hospitals is given above. In this study, important features and patterns will be extracted from forty input variables Risk Factors (RF) which are essential to the diagnosis of CVD. Table 3.1 lists information for some of UAE hospitals where the data is collected, the source shown that most of datasets are collected and surveyed in Abu Dhabi which is the Capital of UAE related to heart diseases patients, WHO used this data for implementing the Heart Risk Assessment Tools analysis (Abderrahim Oulhaj et al, 2020).

Table 3.1 Hospitals Information from where CVD Dataset was collected.

Hospital Name	Organization Source	Location	Phone
Ahalia Hospital	Private	Abu Dhabi	02-6262666
Al Corniche Hospital	SEHA	Abu Dhabi	02-6724900
Al Gharbia Hospitals	SEHA	Abu Dhabi	02-8844444
Al Mafraq Hospital	SEHA	Abu Dhabi	02-5011111
Al Noor Hospital	Private	Abu Dhabi	02-6265265
Al Raha Hospital	Private	Abu Dhabi	02-6330440

Al Rahba Hospital	SEHA	Abu Dhabi	02-5064444
Al Salama Hospital	Private	Abu Dhabi	02-6966777
American European Medical Center	Private	Abu Dhabi	02 4455477
The British Clinic	Private	Abu Dhabi	02 6771252
Burjeel Hospital -	Private	Abu Dhabi	02-5085555
Cleveland Clinic	Private	Abu Dhabi	02-6590200
Dar Al Shifaa	Private	Abu Dhabi	02-6416999
Exeter Medical Center	Private	Abu Dhabi	02-6354321
Franco-Emirien Hospital	Private	Abu Dhabi	02-6265722
Gulf Diagnostic Center	Private	Abu Dhabi	02-4177222
Ibn Nafees Medical Center	Private	Abu Dhabi	02-6324200
LLH Hospital	Private	Abu Dhabi	02-6335522
Seha Emirates Hospital	SEHA	Abu Dhabi	02-4438 999
Shaikh Khalifa Medical City	SEHA	Abu Dhabi	02-8194000
Universal Hospital	Private	Abu Dhabi	02-6435555

StudyID	Gender	Age at baseli	CP	Emp	Head	Dsyp	Total cholesterol	HDL-C	LDL-C	Dm	Sm	HbA1C	eGFR	SBP	DBP	BMI
1	0	64	0	0	1	1	4.80	1.22	3.16	0.92	3.93	5.90	93.02	144.00	87.00	40.16
2	0	52	0	0	1	0	6.40	0.95	4.61	1.84	6.74	6.00	105.49	148.00	91.00	44.54
3	0	56	0	0	1	1	6.40	1.70	3.90	1.75	3.76	5.90	99.52	149.00	86.00	40.54
4	0	58	0	0	0	1	5.10	0.99	3.55	1.22	5.15	5.50	90.03	116.00	68.00	32.03
5	0	63	1	0	1	1	5.00	1.30	3.19	1.11	3.85	7.46	79.47	132.00	63.00	30.76
6	0	51	0	0	1	1	4.90	1.05	3.42	0.94	4.67	6.20	103.08	124.00	84.00	36.80
7	0	71	0	0	1	0	5.60	1.24	1.52	6.19	4.52	6.50	88.56	125.00	72.00	35.84
8	0	44	0	0	0	0	6.60	0.72	4.09	3.90	9.17	5.80	92.41	120.00	77.00	44.12
9	0	54	0	0	0	1	5.00	1.12	3.62	0.57	4.46	6.00	96.16	123.00	79.00	39.73
10	0	44	0	0	1	1	5.70	1.55	3.91	0.52	3.68	6.90	110.22	157.00	87.00	35.57
11	0	49	0	0	1	1	6.80	1.01	4.39	3.06	6.73	6.20	109.87	112.00	64.00	35.83
12	1	56	0	1	0	0	4.60	1.20	2.25	2.51	3.83	5.10	82.02	121.00	78.00	25.34
13	0	48	0	0	1	0	4.30	0.76	3.11	0.93	5.66	5.70	98.40	129.00	72.00	29.92
14	1	71	0	1	1	1	5.80	0.83	4.31	1.44	6.99	6.40	69.14	132.00	57.00	27.33
15	0	69	0	0	1	1	5.50	1.61	3.22	1.46	3.42	5.70	93.03	148.00	70.00	23.40
16	1	68	0	0	1	0	6.40	1.03	4.14	2.68	6.21	6.60	88.18	137.00	76.00	22.87
17	0	51	1	0	1	0	3.20	0.58	1.74	1.92	5.52	6.60	100.80	140.00	80.00	39.80
18	1	53	0	0	1	0	7.70	1.71	5.39	1.30	4.50	6.90	98.50	143.00	89.00	28.24
19	0	79	1	0	0	1	3.40	1.38	1.52	1.08	2.46	7.70	90.86	131.00	80.00	19.47
21	0	65	0	0	0	1	4.10	0.98	2.61	1.11	4.18	6.30	92.37	137.00	76.00	32.13
22	0	79	1	0	1	1	5.30	1.91	3.04	0.76	2.77	7.10	84.67	146.00	74.00	33.28
23	0	60	1	0	1	1	4.70	1.13	2.96	1.23	4.16	0.60	77.53	135.00	68.00	33.67

Table 3.2 Samples of some medical dataset which was collected from the mentioned hospitals7.

3.7 DATASET DESCRIBTION AND ILLUSTRATION

the patient has overweighed Table 3.2 lists all the input variables with the description and values. In this study, forty Risk Factors independent variables were essential to the diagnosis of the CVD of interest were extracted from the UAE database. These variables can be divided into four sections: Section One: the basic information of a patient, including the age and the sex, Section Two: the symptoms (14 Risk Factors in total), Section Three: the inducement and the history (5 Risk factors in total), and Section Four: the physical examinations and lab results (19 Risk factors in total). There are 3 types of input features: Objective as patient's information, Examination as results of medical examination. Subjective as information given by the patient. The dataset contained one numeric which is an age feature, and all others are discrete features. However, one common problem in medical databases is the lack of information in the form of missing data values. For instance, some medical examinations may not be appropriate for certain patients, or the

doctor failed to record some symptoms. Different methods have been suggested to handle this missing data problem (Heckerling et al., 1991; Doyle et al., 1995). The simplest method is to ignore the cases with missing data. In addition, notably, none of the variables had missing or repetitive values since the data is cleaned and reprocessed by removing the redundancy data and all the data are coding. For numeric parameters as age, the age will be assigned to value between the interval (0,1) for the range of ages between 0-100 years if the patient age is 50 years, its numeric value will be assigned as 0.5. Discrete parameters with two independent attributes as Gender will be assigned to binary values, for instance, 1 represents male and 0 female. Discrete parameters with three independent attributes as Electrocardiogram will be 1 as positive case for presence of attribute, 0 for normal case (intermediate level), and -1 for negative case, absence of attribute. For weight parameter BMI (Body Mass Index), the value will be -1 if the BMI is less than 18 it means the patient has underweight, the value will be 0 for BMI between 19-15, it means that the patient has normal weight, and the value will be 1 for BMI more than 25 it means.

Table 3.3 Input / Independent Variables for Cardiovascular Diseases in the Selected Dataset

Group name		**Variable description**	**Designation**
Information	**Gender**	0 Female, 1 male	[0, 1]
	Age	numeric values will be assigned to interval (0,1)	[0, 1]
	Weight/BMI	Body Mass Index	[-1, 0.5,1]
	Cgh	Cough	[-1, 0.5,1]
Symptoms	Emp	Emptysis/Expectoration	[-1, 0.5, 1]
	Dys	Dyspnea	[-1, 0.5, 1]
	Gas	Gasping/Panting	[-1, 0.5, 1]
	Nau	Nausea/vomiting	[-1, 0.5, 1]

	Dizz	Dizziness	[-1, 0.5, 1]
	Head	Headache	[-1, 0.5, 1]
	Fev	Fever	[-1, 0.5, 1]
	Cyan	Cyanosis	[-1, 0.5, 1]
	Pal	Palpitations	[-1, 0.5, 1]
	Weak	Weakness or fatigue Discomfort, pressure, heaviness in the chest	[-1, 0.5, 1]
	Cp	Chest pain (precordial region, posterior sternal)	[-1, 0.5, 1]
	Ell	Edema of lower limb	[-1, 0.5, 1]
	Uri	Upper respiratory infection	[-1, 0.5, 1]
	cpd_his	Chronic pulmonary diseased history	[-1, 0.5, 1]
	hyp_his	Hypertention family history	[-1, 0.5, 1]
	CHD_his	Coronary heart diseases family history	[-1, 0.5, 1]
Physical sign and assistant examination	DBPH	Diastolic Blood pressure with High rate	[0,1]
	Chol	Total Cholesterol	[-1, 0.5, 1]
	HDL – C	High Density in Blood	[-1, 0.5, 1]
	LDL -C	Low Density in Blood	[-1, 0.5, 1]
	Cve	Cervical venous engorgement	[-1, 0.5, 1]
	bar_che	Barrel chest	[-1, 0.5, 1]
	sound_b	Sound of breath in pulmonary	[-1, 0.5, 1]
	Moi	Moist rales	[-1, 0.5, 1]
	Sonant	Boundary of heart sonant	[-1, 0.5, 1]
	Sm	Systolic murmur	[-1, 0.5, 1]
	Dm	Diastolic murmur	[-1, 0.5, 1]
	s_thri	Systolic thrill	[-1, 0.5, 1]
	SHS	Second heart sound	[-1, 0.5, 1]

	LK_tend	Liver and kidney tenderness	[-1, 0.5, 1]
	Hyp_note	Hyper resonant note	[-1, 0.5, 1]
	EPR	Eminence in precordial region	[-1, 0.5, 1]
	CRO	Cardiac rhythm in good order or not	[-1, 0.5, 1]
	P_wave	Pulmonic P-wave	[-1, 0.5, 1]
	Neck_vr	Neck venous return	[0,1]
	Cenlar	Cardiac enlargement	[-1, 0.5, 1]
	ECG	Electrocardiogram	[-1, 0.5, 1]
	STT_alter	ST-T alteration	[-1, 0.5, 1]

In this study dataset after preprocessing, cleaning, and encoding will be divided into 30% for testing and 70% for training which will be divided into subsets for training the model, mean value assigned for missing values.

3.1 ANALYSIS OF THE CORRELATED RISK FACTORS (RF)

In this study feature selection is an important process to get the correlation strength between the RF and how they are correlated together, it is often performed as pre-processing step, which aims to reduce dimensionality of data into a subset of relevant or optimal features. The success in performing feature selection depends on, how to carefully choose the proper method to achieve the goal and understanding of the domain knowledge. The following main steps were performed by using Minitab 17 statistical software Data Visualization: Minitab allows users to create various types of graphs and charts to visualize data, which is essential for data exploration and presentation.

Statistical Output: Minitab generates clear and concise statistical output, making it easier for users to interpret and communicate their findings.

Statistical Education: Minitab has been a popular choice in educational settings for teaching statistics and data analysis. Its ease of useas follows:

1. Use of visual plots for clear understanding of correlation between RFs

2. Correlation test (Is there a correlation between Risk Factors and CVD?)
3. Multiple Linear Regression analysis (Which factors are the most significant in impacting the rate of CVD?)

In this study the first step before correlation and regression analysis we need to visualize the data by creating the scatter plots to see the correlation between the two variables if it is negative or positive or normal, statistical software offers two methods of **Correlation:** the Pearson product moment correlation and the spearman rank order correlation. The Pearson correlation (also known as r) which is the most method measures the relationship between two variables. In this study The Pearson correlation was used to measure the linearity relationship between the most effective Risk Factors. For this type of pf correlation p- value which is the amount of risk you are willing to take, Pearson's Correlation Coefficient value must be calculated within the following boundaries:

0< PCC < 0.4 = Weak positive correlation

0.4 <PCC<0.6 = moderately positive correlation

0.6 PCC < 1 = strong positive correlation

- 0.4 < PCC < 0 = weak negative correlation

P-value ₋0.05 must be ignored since there is no correlation between the variables.

The following Figure 3.2 explains how to visualize the Data using Matrix scatter Plot in Minitab statistical Software, it shows examples of visualize the data by creating the scatter plot between the risk factors in Minitab statistical method, in this figure we find that there is a positive correlation as the sloping line goes up between obesity and high Blood pressure, from this plot we noticed that most of correlation sloping line results for obesity/IBM between 25 – 35 which means there is a strong positive correlation.

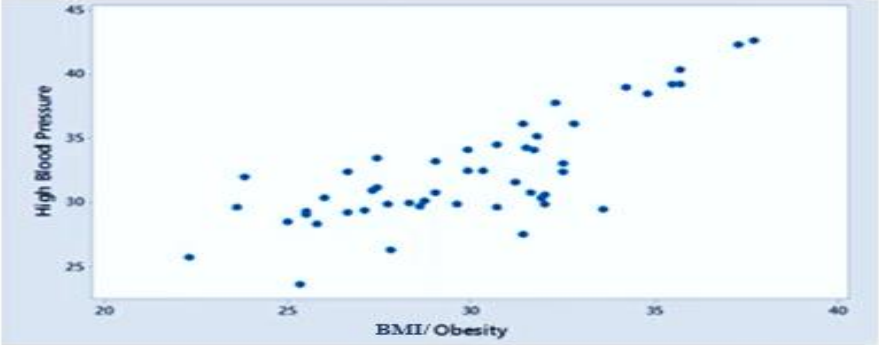


Figure 3.2 Scatter Plot of High Blood Pressure Vias Obesity/BMI

Figure 3.3 shows the scatter plot of the correlation between the high blood pressure via weekly exercise, from this figure we noticed that there is a negative correlation as the sloping line goes down.

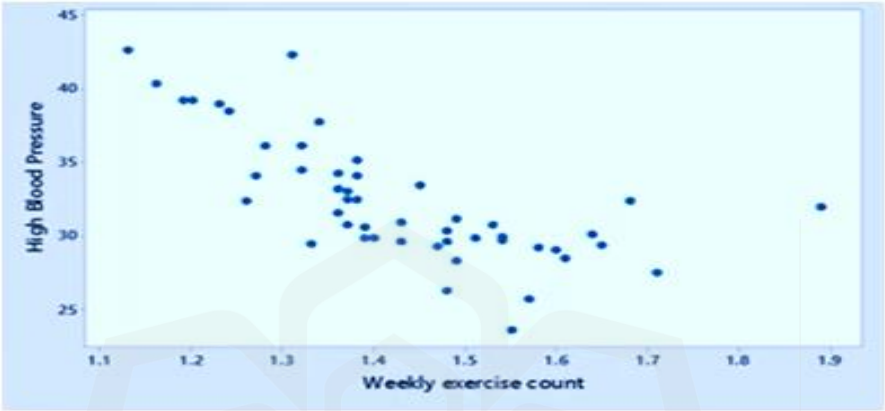


Figure 3.3 Scatter Plot of High Blood Pressure (BP) vias Weekly Exercise Count (WEXC)

Figure 3.4 explains some examples for correlations results to show the strength of correlation between the Risk Factors (RF) based on PCC coefficient calculation in Minitab, in the following figure the blue boxes represent the strong positive correlations, the red boxes represent the strong negative correlation, and the yellow boxes represent the moderate correlation. For example, Body Mass Index (BMI) has a strong positive correlation with the Blood Pressure (BP) since PCC > 0.6 so there is impurity variable, while Weekly Exercise Count WEXC has a strong negative correlation with Blood Pressure (BP) so there is no impurity variable, since PCC < -0.6, in the following example , in the following example we noticed that blocks 1,5,7, and 16 have strong positive correlation, blocks 2,11,20,4,27,22, and 23 have strong negative correlation, blocks 3,12,28,6,15,29, and 30 have normal mediate correlation.

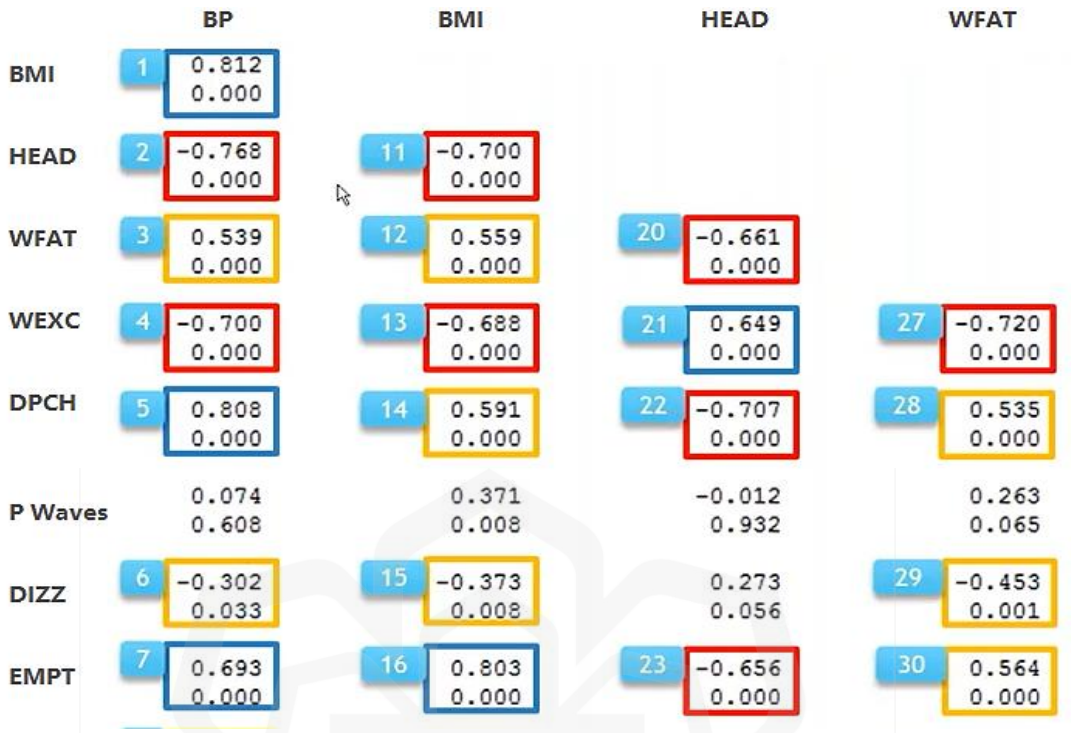


Figure 3.4 Examples for Correlation Results between the RFs

3.2 FINDINDING THE INFLUENTIAL RISK FACTORS USING GINI_ENTROPY INDEX

To evaluate the effectiveness of each Risk Factor and its impact on the patients, this study used Gini Index for calculating the wight factor for each Risk Factor. It is a method used to calculate the impurities of a variable. Gini impurity is a function that determines how well a decision tree was split. Gini method is rarely used in the previous studies of medical research, it helps us to determine which splitter is best so that we can build a pure decision tree. Gini impurity ranges values from 0 to 0.5.A variable with the lowest level of impurity is an unbalanced variable while the variable with the highest level is a balanced variable and has the highest effectiveness on the patients.

The following equation (3.1) is used to calculate the weight index as follows:

$$\text{Gini(D)} = -1 - \sum_{i=1}^{m} pi^2 \qquad \ldots. (3.1)$$

Where p*i* is the relative frequency of class j in Dataset D

If Dataset D is split on A into two subsets D1 & D2, the Gini index Gini(D) is defined as:

Gini(D) = $\frac{|D1|}{|D|}$ Gini (D1) + = $\frac{|D2|}{|D|}$ Gini(D2) …. (3.2)

While Entropy Index as follows:

E(D) = $\sum_{i-}^{c} pi$ - $\log_2 pi$ …. (3.3)

The internal working of Gini impurity is also somewhat like the working of entropy, entropy index is same as Gini Index but the interval for Gini is between 0, 0.5 while the interval for Entropy is 0,1 in the Decision Tree. In the Decision Tree algorithm, both are used for building the tree by splitting as per the appropriate features but there is quite a difference in the computation of both the methods. The Gini and Entropy index are used in the classic CART (Classification & Regression Tree) algorithm and is very easy to calculate. CART creates a binary tree and implements the decision which can be represented by weights which are often to be within the range of 0-1 and the result which is closer to 1 means the variable has the highest effectiveness while the result which is closer to 0 means the variable has the lowest effectiveness on the patients. Figure 3. Explains the difference between Gini and Entropy Index for calculating the weights of the Risk Factors as follows:

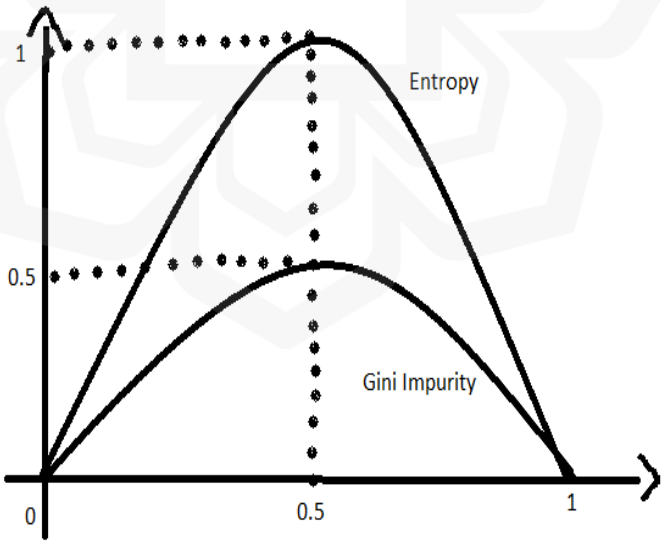


Figure 3. 5 Differences between Entropy and Gini Index

3.3 DEEP LEARNING MODEL BASED ON MLP

The deep learning model Consists of three layers, input to which data is loaded, as given in figure 3.4 below, the hidden layer of the model, which is processing unit of the model. It involves the processing of the given data and present the results in the end in the output layer. The output layer shows the best matches in the given data and solves the mystery. Input in the model in the present study is CVD disease, their demographics, and the symptoms, which are being processed in the hidden layers, and in the end the model gives the best matched overlapped symptoms of the CVD disease. After the training process, the model is used to predict the overlapping in the symptoms to predict that specific disease.

Artificial neural network involves backpropagation algorithms. A backpropagation is used with multilayer Perceptron's (MLP) consisting of an input layer, hidden layers and an output layer. BP networks is refers to the method where error computed between the result of predication output and actual output, it is used to adjustment weight to reach the minimum error (Syafria, Buono, & Silalahi, 2014).

For the input factor weight, a backpropagation algorithm is used. Each input in the input layer is multiplied by identical initial weight after the product sum is gained; activation function is used to output the result; Figure 3.11 shows the MLP Network

$$\text{net}j = \sum_{i=1}^{I} w_{ij} \ . \text{oi}, \qquad (3.4)$$

Where net$j$ denotes the input of hidden layer, i and j are Indicators of different neurons included at the network. The input vector size is represented by I, O indicates to the input element and the weight is w. Every neuron of hidden layer uses their input (net$j$) after get it as an argument for a function and produces an output Oj equation by:

$$Oj = (\text{net}j). \qquad (3.5)$$

The f function is a nonlinear function in a form of sigmoid. This function used with a sum of inputs weighted as a previous step before propagating the signal to the following

layer. This sigmoid function has one trait, that its, can express its derivative by the function itself as shown in (equation 3.6):

$$\tilde{f}(\text{net}j) = (\text{net})\,(1 - f\,(\text{ne}jt)). \quad (3.6)$$

The variance between the output of network, O_k, and value of the target output, d_k, as shown in equation (3.7), can be a definition for the training error to pattern of input.

$$e_k = (d_k - o_k)\,. \quad (3.7)$$

Compute the sum of squared error can be as shown in (equation 3.8).

$$E = \frac{1}{2} \sum_{k=1}^{k} e_{k,}^2 \quad (3.8)$$

Where K is the neurons number in the layer of output. The error is minimized by weight adjustment between layers when it is propagated. the expresses a weight adjustment as following:

$$\Delta wij\,(n+1) = (\delta j \cdot oi) + \alpha\,\Delta wij\,(n)\,, \quad (3.9)$$

Where $\Delta(n+1)$ and $\Delta wij(n)$ are the weight changes in epochs $(n+1)$ and (n), respectively, η is the learning rate parameter, δ is an index of the rate of error change , and α is the momentum coefficient. This operation of feeding forward signals and returns is frequently till the network error is decreases as a whole or arrives an passable value (Alkhasawneh et al., 2013).

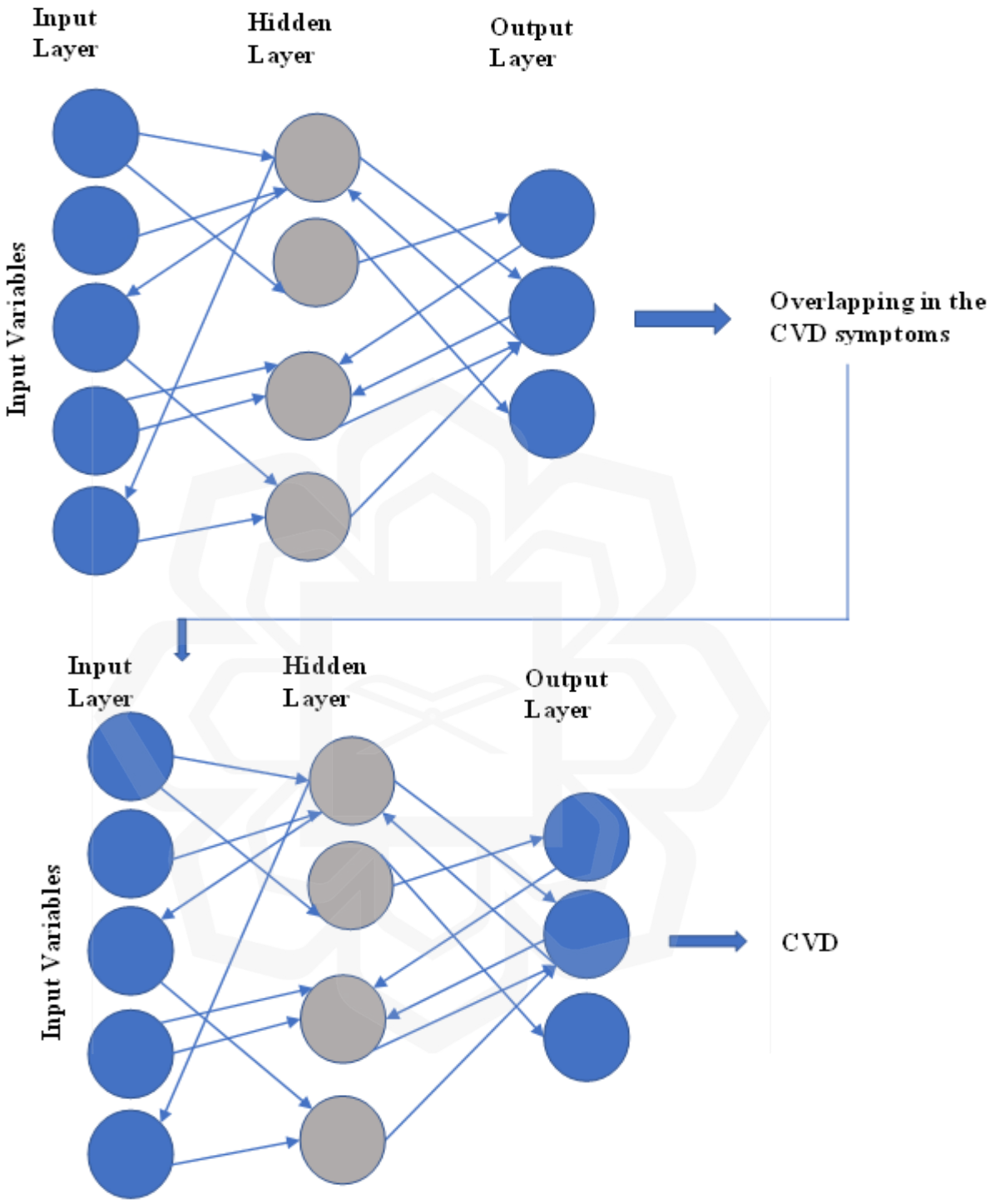


Figure 3.6 MLP Architecture in the Proposed Model

In neural networks, "tansig" and "logsig" are activation functions that are used for specific purposes:

1. Tansig (Hyperbolic Tangent Sigmoid): The "tansig" activation function is a hyperbolic tangent function. It produces output values between -1 and 1, making it suitable for networks where values need to be centered on zero. It is often used in hidden layers of neural networks.

2. Logsig (Logistic Sigmoid): The "logsig" activation function is the logistic sigmoid function. It maps input values to a range between 0 and 1. It is commonly used in the output layer of binary classification networks, as it models the probability of a binary outcome.

The sigmoid function, in general, is preferred because it is smooth, bounded, and has a simple derivative. These properties make it suitable for gradient-based optimization algorithms like back propagation, which is commonly used in training neural networks.

The structure of MLP Neural Network will be two parts each part as follows:

Part 1

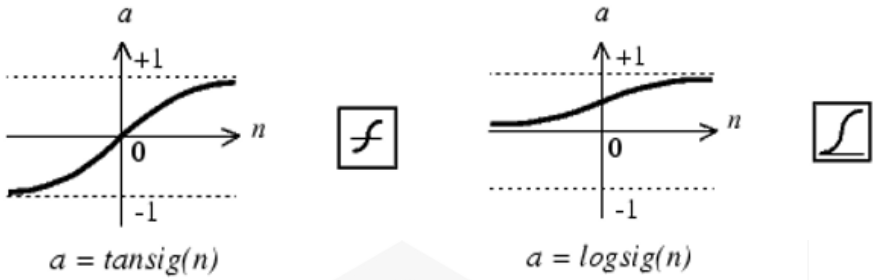
- Input layer: which consists of 40 nodes; the activation function is logsig.
- Hidden layers consist of 50 nodes respectively; the activation function of the hidden layers is tansig, logsig respectively.
- Output layer consisting of 5; the activation function is logsig.

Part 2

- Input layer which consists of 10 nodes; the activation function is logsig.
- One hidden layer consists of 10 nodes; the activation function is tansig and the output layer consists of 3 nodes with a logsig activation function.

As shown in Figure 3.6, each node output (of hidden and output nodes) is calculated by obtaining the sum of the product between weights and values from the previous layer. The result of the summation usually approximately is the weighted sum; it is transformed to working output through an algorithmic process known as the transfer function at which its derivative affects the weight adaptation. Where in this work, two activation functions are used: tansig and logsig activation function. Sigmoid function is the most commonly used activation function; it is preferred because it is smooth and bounded, and it has a simple

derivative. Figure 4.10 shows the tansig and logsig activation function. (Neelima Rajput, & Verma, 2014)



a) Tansing activation function b) logigmoid activation finction

Figure 0.1 Tansing and Logsig Activation Function

During the training phase, the training data are fed into the input layer. The data are propagated to the hidden layers and then to the output layer. This approach is called the forward pass of the BP algorithmic.

In forward pass (implemented using Equations 3.5 and 3.6), each node in hidden layers obtains input from all the nodes in the input layer. They are multiplied with appropriate weights and then summed and passed to the next hidden layers. The output of the hidden node is the nonlinear transformation of this resulting sum. Similarly, the node in output layer obtains input from all the nodes of the last hidden layers, which are multiplied with appropriate weights and then summed. The output of the nodes is a nonlinear transformation of the resulting sum. The output values of the output layer are compared with the target output values; the difference between them (which represents the error) is back propagated to the hidden layer to adapt the weights (i.e. the target difference between actual and desired output values is used during the network learning process). This approach is called backward pass of the BP algorithmic (implemented using equations 3.7 - 3.9). Back-propagated errors of the hidden layer to the input layer are used to update connection strength between nodes, that is, the weight between the input-hidden layers. The learning process proceeds by iteratively performing forward and backward passes until the

network is learned. After the learning process, the resulting weighted vectors are used in the testing phase.

In the testing phase, each test vector is fed into the input layer. Only the forward pass is used with the weight vectors resulting from the training phase. If the output node value equal to one (1), that is mean the speaker is recognized. Initially, the *weight vectors* are initiated randomly. The *stopping condition* used is min square err <= 0.0001, or the number of epochs reaches 200. The network is stopped when it reaches this value. The *constant learning rate* α=0.5 is selected by trial and error. The constant rate goal is presented to the network, the weights leading to an output node are modified slightly during the learning in the direction required to produce a smaller error the next time the same pattern is presented. The amount of weight modification is the learning rate times the error. For example, if the learning rate is .5, the weight change is one half the error. When the learning rate is large, the weight changes are large, and the learning proceeds rapidly. Oscillation or no convergence can occur if the learning rate is too large. (James Martens, 2016)

Figure 4.10 shows the tansig and logsig activation function. During the training phase, the training data are fed into the input layer. The data are propagated to the hidden layers and then to the output layer. This approach is called the forward pass of the BP algorithmic. In forward pass (implemented using Equations 3.5 and 3.6), each node in hidden layers obtains input from all the nodes in the input layer. They are multiplied with appropriate weights and then summed and passed to the next hidden layers. The output of the hidden node is the nonlinear transformation of this resulting sum. Similarly, the node in output layer obtains input from all the nodes of the last hidden layers, which are multiplied with appropriate weights and then summed. The output of the nodes is a nonlinear transformation of the resulting sum.

The output values of the output layer are compared with the target output values; the difference between them (which represents the error) is back propagated to the hidden layer to adapt the weights (i.e., the target difference between actual and desired output values is used during the network learning process). This approach is called backward pass of the BP algorithmic (implemented using equations 3.7 - 3.9). Back-propagated errors of the hidden layer to the input layer are used to update connection strength between nodes, that is, the weight between the input-hidden layers. The learning process proceeds by iteratively

performing forward and backward passes until the network is learned. After the learning process, the resulting weighted vectors are used in the testing phase. During the training phase, the training data are fed into the input layer. The data are propagated to the hidden layers and then to the output layer. This approach is called the forward pass of the BP algorithmic. In forward pass (implemented using Equations 3.5 and 3.6), each node in hidden layers obtains input from all the nodes in the input layer. They are multiplied with appropriate weights and then summed and passed to the next hidden layers. The output of the hidden node is the nonlinear transformation of this resulting sum. Similarly, the node in output layer obtains input from all the nodes of the last hidden layers, which are multiplied with appropriate weights and then summed. The output of the nodes is a nonlinear transformation of the resulting sum.

The output values of the output layer are compared with the target output values; the difference between them (which represents the error) is back propagated to the hidden layer to adapt the weights (i.e., the target difference between actual and desired output values is used during the network learning process). This approach is called backward pass of the BP algorithmic (implemented using equations 3.7 - 3.9). Back-propagated errors of the hidden layer to the input layer are used to update connection strength between nodes, that is, the weight between the input-hidden layers. The learning process proceeds by iteratively performing forward and backward passes until the network is learned. After the learning process, the resulting weighted vectors are used in the testing phase.

In the testing phase, each test vector is fed into the input layer. Only the forward pass is used with the weight vectors resulting from the training phase. If the output node value equal to one (1), that is mean the disease is recognized. Initially, the weight vectors are initiated randomly. The stopping condition used is min square err <= 0.0001, or the number of epochs reaches 500. The network is stopped when it reaches this value. The constant learning rate α=0.4 is selected by trial and error. The constant rate goal is presented to the network, the weights leading to an output node are modified slightly during the learning in the direction required to produce a smaller error the next time the same pattern is presented. The amount of weight modification is the learning rate times the error. For example, if the learning rate is .5, the weight change is one half the error. When the learning

rate is large, the weight changes are large, and the learning proceeds rapidly. Oscillation or no convergence can occur if the learning rate is too large.

3.4 DIFFERENT SOFTWARE QUALITY MODELS

A. MCCALL'S MODEL

A model developed by Jim McCall for the identification of cardiovascular disease signs via deep learning has been hailed as one of the most renowned antecedents of current quality models that prevent patients from serious heart cases. There are several approaches for assessing the quality of software. Of these, a few have been codified as industry best practices. In 1977, McCall created the UU model for the United States Army. Factors that make up McCall's quality model include the following:

- Correction: the functionality conforms to the specific diseases.
- Chest Pain: use of detecting resources such as hospitals and different medications.
- Integrity: protect a system of unauthorized access.
- Reliability: the ability of the system to perform well.
- Usability: the ease of applying the software.

According to McCall, usability comprises of two symptoms namely (i) discomfort in chest (ii) shortness of breath.

B. FUYS SOFTWARE QUALITY MODEL (Suman. & Wadhwa, 2014)

Robert Grady and Hewlett Packard presented the FUYS model. This model is based on quality to build a secure platform for users to maintain their confidentiality concerns about different diseases. They built the FUYS model for Rational Software Company. FUYS model comprises the following non-functional requirements.

C. QUALITY MODEL PROTECTS USERS:

Human considerations, aesthetics in general, Cyanosis, and recordkeeping are all part of the process. Throughout the development process, human considerations are taken into account. One of the paradoxes of general aesthetics is that people find more aesthetically pleasing designs to be more appealing. According to cyanosis, things that are linked should be represented in a similar way, while those that are unrelated should be shown in a different way. Insights into the real user experience may be found in the documentation of products. Below were the qualities of accuracy level of early predictive deep learning model.

Reliability: "frequency and severity of failure, resilience, predictability, accuracy and average time between failures (MTBF).

Performance: "Constitutes conditions on functional necessities such as speed, Chest Pain, availability, accuracy, performance, response time, recovery time and use of resources".

Compatibility: "Consist of test capacity, extensibility, and adaptability, relief of maintenance, compatibility, configurability, service capacity, installation capacity and location".

D. NIELSEN MODEL

(Nielsen, 1994)) As indicated by ISO-9241 that is the extended version of ISO 9126 (UAE Hospital requirements for treatment of heart patients with visual showcase terminals of symptoms) (ISO, 1998) standard, we have the accompanying definition: Usability of early predictive deep learning system and its ability to function Effectively and Efficiently while giving emotional fulfilment to its patients. The usability of an interface is typically connected with five parameters associated with associated with deep learning. According to Nielsen, there are mainly five symptoms that must be considered in each deep learning application to satisfy patient’s requirements. These are learnability, Chest Pain, memorability, mistakes, and Discomfort Pressure in Chest.

3.5 USER DEEP LEARNABILITY TOWARD HEART ATTACK:

The application must be easy to learn for the heart attack using deep learning so that the patients can easily perform their activities to enhance functionality level of heart.

Chest Pain of early prediction: The application must be easy to use and there must be an understanding of the objective associated with deep learning and the Common cardiovascular diseases. Chest Pain refers to the speed with which the patient can use the medication to complete the treatment once the patient has cured the disease.

Memorability: The improvement of the patients from the symptoms to take the treatment of the disease, when the patient does not take the treatment for a while and then return to it and how easily the patient becomes healthy. Measuring the time to complete the treatment after a while using deep learning technique. Also, when considering the number of clicks, steps and deep learning analysis that is required for the patient's treatment.

3.6 BERTOA MODEL

Model of Quality ISO 9126 Model is Bertoa's go-to model of choice. For the Productive of Evaluation Commercial Off-the-Shelf (COTS) it defines several quality attributes (Commercial-off-the-shelf). To protect user data stored in systems and social media platforms from link exposure, software development businesses rely on COTS. Usability, in Bertoa's view, is made up of three components: learnability, which refers to how well a system can figure out what kind of user you are and how you intend to use it, and understandability, which refers to how well users can figure out how to pick the right software for their needs and how to put it to work on specific tasks. The ability of the software product to be controlled by the user is called Weakness and Fatigue. (Spiekermann, 2010)

3.7 G. KUROSU-3 MODEL

Some of the ISO/IEC 25010 principles drew on Kurosu's current model. According to Kurosu-3, usability is comprised of the following elements: the ability of a product or system to empower the customer to figure out how to use it with Emptysis, Chest Pain in criticism, the recognizability of patients toward treatment that can be used by specific patients to achieve specific treatments to secure their sensitive attributes It is important for software products to be able to detect and treat patients' symptoms of weakness and fatigue so that they can be used by people with a wide range of skills and abilities, and accessibility is the degree to which patients can use the provided treatment with ease, even if they have difficulty remembering how to do so. (Kurosu, 2014).

3.8 CHAPTER SUMMARY

This chapter presented the steps of research methodology, research design, and research approach, and illustrated the characteristics of dataset and its encoding and the ways of finding the correlation between the Risk Factors and how they are correlated together, and how to find the most effective Risk Factors on the patients and presented some models of Disease prediction.

CHAPTER FOUR

PREDICTION OF OVERLAPPED SYMPTOMS

4.1 INTRODUCTION

This chapter illustrates the methodology steps and processes used in predicting the CVD by analysis the predictors with 8 levels of symptoms to find the best accuracy and the optimal model for UAE hospitals and finally compare the suggested hybrid model Gini-Entropy-Regression Model with MLP Model and compare the results. This chapter divided into two parts of classification, first part involves analysis of influential RF to find the most effective symptoms on patients using Gini-Entropy-Regression analysis Model (GERM), and second part involves CVD prediction based on MLP model.

4.2 IMPLEMENTATION METHODOLOGY

In the two classification stages of this study, the results were analyzed. To validate and divide the data into training, and testing categories, this study used cross-validation instead of simple validation. After the pre-processing of the dataset, the decision tree by Gini Index, MLP neural network, and deep learning classifications to the learning data in the SPSS, Python and Minitab software. Each method had specific parameters to adjust the models as explained in table 4.1 In addition, the optimal values of the parameters were selected meticulously after multiple stages of testing and simulation based on the accuracy of the models.

Table 4.1 Adjustment Parameters for the Proposed System

Model	Description of Parameters	Software
Gini – Entropy Regression Model (GERM)	Gini – Index Branches Max. depth: 20 Max.leaf size: 3 Confidence level of pruning: 0.25 Min. records to terminate 4	Python SPSS 25 Minitab

Multilayer Perceptron Neural Network (MLP)	Momentum coefficient: 0.7 Learning rate: 0.4 Epochs 10 Batch size 200	Python SPSS 25

4.3 PARAMETERS ASSESMENT BASED ON WEIGHTS OF INFLUENTIAL R F

The proposed work used a Gini- Entropy Index method which is rarely used in previous studies, it is used to calculate the impurities of all Risk Factors (RF). An unbalanced variable is a variable with the lowest level of impurity which means it has the lowest effects on the patients, while a balanced knot has the highest level of impurity which means it has the highest effects on the patients, thereby providing the most beneficial data for the proposed model. This study initially reported the weight of the indexes. Weights are often within the range of 0-1, and the closer they are to one, they have a more significant effect on the incidence or non-incidence of heart failure. Figure 4.1 shows the weights of the highest indexes (variables) based on the entropy index.

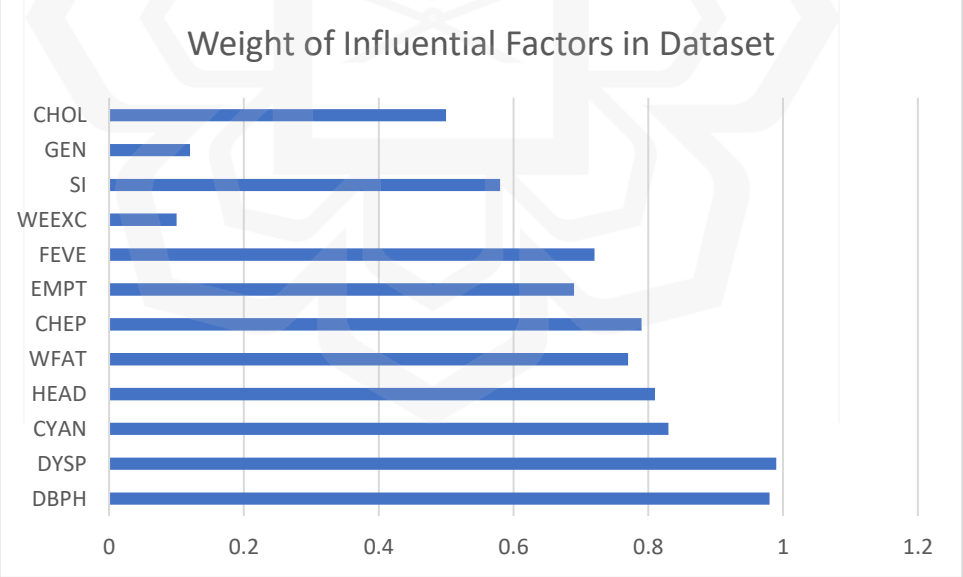


Figure 4.1 The highest effective Risk factors based on Gini- Entropy index.

As a results, DBPH, DSYP, CYAN, HEAD, WFAT, CHEP, EMPT, and FEVE had the most significant influential factor with highest impurity since the weights converges to 1

CHOL, and SI had a moderate effects since they converge to 0.5 while GEN, and WEEXC had no effects since the weights converges to 0.

4.4 DETECTING THE OVERLAPPING IN SYMPTOMS

This study needs to find the relationship between the variables to detect the overlapping in Symptoms, since each variable is related to another, one of the most important analyses is regression methods as Multiple Linear Regression (MLR) was used to predicate the overlapping of symptoms and find the accuracy of disease. SPSS was used to regress all these ML-Combinations on all the data points that the model was able to predict (eight different usability Risk factors). The fundamental reason this technique for stepping regression is the reduction in the distance between predictor responses. This study will detect the overlapping in symptoms therefore may combine many independent variables, such as two factors Independent Variables (IV), three factors IV, and four factors IV, for example. A better model has a higher R2 value. Figure 4.2 explains the General Regression Predictive Model for detecting the overlapping in symptoms

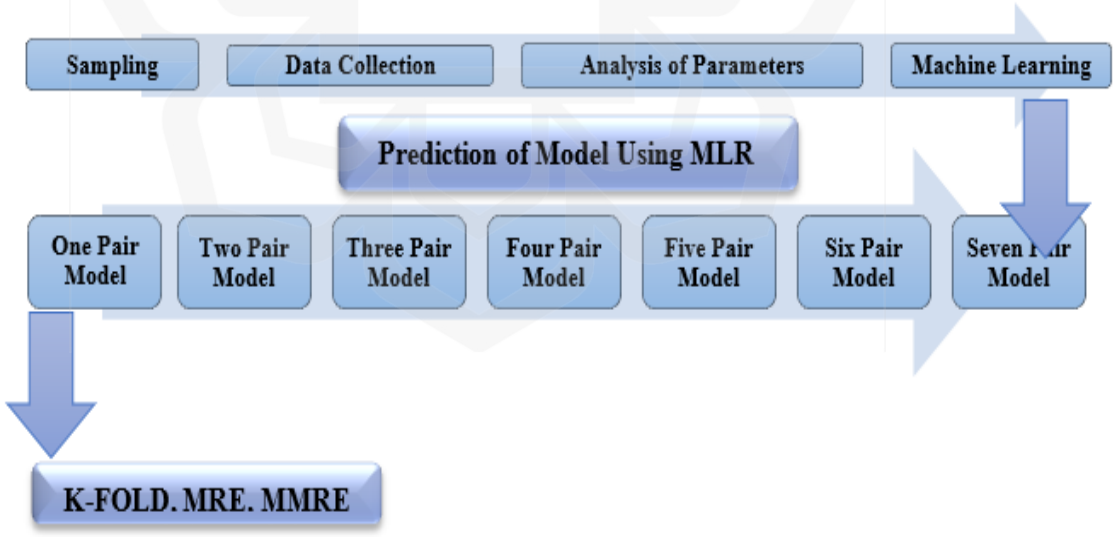


Figure 4.2 Gini-Entropy- Regression Model (GERM)

Figure 4.2 shows the proposed Gini- Entropy Regression Model (GERM) which consists of the following levels:

First Stage: Random Sampling: Random sampling is a common method where data points are selected randomly from the UAE hospitals dataset which stored in SEHA & Figshare site. This helped to reduce bias in the selection process.

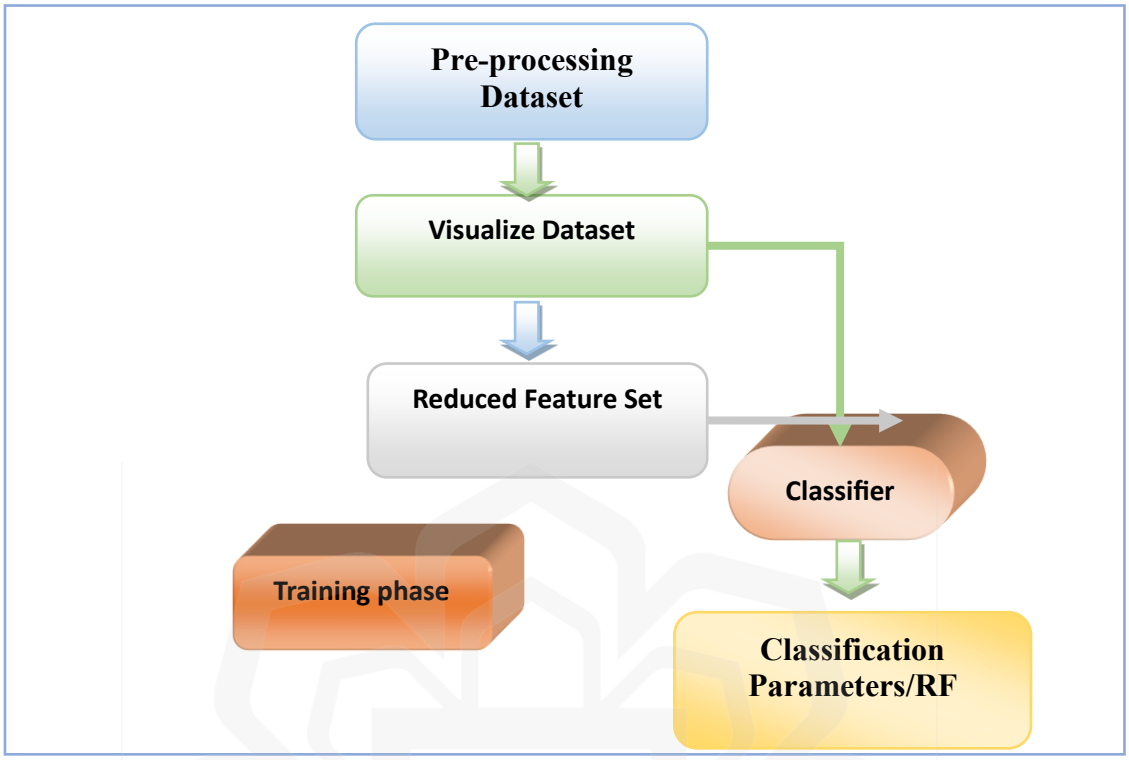
Second Stage: Data collection which included the online Patient Information: Basic demographic data, including age, gender, and ethnicity, are collected to understand the patient population. Medical History: Details of a patient's medical history, such as previous illnesses, surgeries, medications, and family medical history, are documented. Electronic Health Records (EHR): Electronic health records store patient information, diagnoses, treatments, medications, and outcomes, making data accessible for healthcare providers and researchers.

Third stage: Parameters are analyzed by using GINI-Entropy algorithm and visually plotted using Mini Tab Matrix to find the relationship between the parameters and find the most overlapped symptoms which affected in CVD.

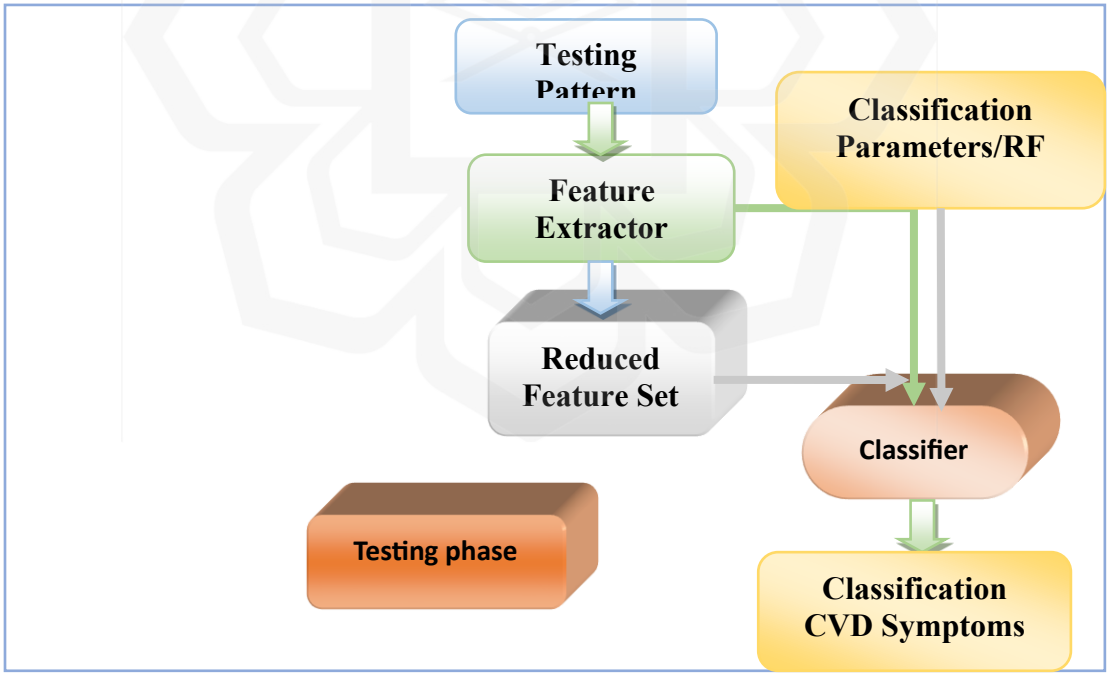
Fourth Stage: Machine learning techniques are used to find the degree of overlapping between the symptoms by check two overlapping symptoms till to 8 overlapped symptoms and find the best detection for the overlapping.

Fifth Stage: Check the model validity & reliability using kth fold validation algorithm.

Figure 4.3 shows the classification operation, where it consists of two phases mainly, (training and testing) phase. In training phase, the dataset will be processed by filtering the noisy and redundant data, then data will be visualized by matrix plots to find the correlation strength between the RF, features will be reduced by using statistical method as Gini-Entropy Index and submitted to the classifier to find the patterns for the training dataset.



a.) Training Phase



b) Testing Phase

Figure 4.3 Classification Operation in the Proposed Model

4.5 ANALYSIS OF PREDICTORES

First, each of the eight potential predictors that cause data leaks should be seen as a simple linear reverse (SLR). After that, multiple linear regressions were performed by the researcher to obtain usability variables predictions. This model will be used to evaluate and identify the usability criteria (Kellar, 2013). In both regression analyses, the SPSS software was used (version 25.0). Now as per findings we have evaluated data in below terms that Empties, Fever, Cyanosis, Weakness or Fatigue, Discomfort Pressure in Chest and Chest Pain that how user rated them in terms of Heart Issues. These are abbreviated CHEP as Chest Pain, dyspepsia as DYSP, Fever is defined as FEVE, Headache as HEADACHE, Cyanosis defined like CYAN, Weakness and Fatigue as WFAT, fulfilment as DPCH and the definition of success is CHEP. Table 4.1 shows the results of SLR.

Table 4.2 Single Linear Regression Model Results for one Independent Variable

Predictor	R-Square	Standardized Beta	Significance
EMPT	0.643	0.802	0.000
DYSP	0.699	0.836	0.000
FEVE	0.149	-0.386	0.000
HEADACHE	0.029	0.171	0.000
CYAN	0.406	-0.637	0.000
WFAT	0.054	0.232	0.002
DPCH	0.299	0.547	0.000
CHEP	0.715	0.846	0.000

The SLR equation of the strongest predictor is provided as:

$$\bar{Y} = 0.347 + (0.846 \times \text{CHEP}) \quad (4.1)$$

4.6 MULTIPLE LINEAR REGRESSION ALGORITHM

Multiple linear regression refers to a statistical technique that uses two or more independent variables to predict the outcome of a dependent variable. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance. This method incorporates the dependent variable with a variety of independent variables, such as two factors Independent Variables (IV), three factors IV, four factors IV.

The following steps describe the process of algorithm.

1. Visualize the data
- Create Matrix Plot
2. Check model and assumptions
- Variance Inflation Factor (VIF)? <5 Exclude correlating factors (OFAT)
- P -Value < 0.05? Exclude non- significant factors (OFAT
3. For a Regression Model to be valid, the assumptions are:
- RF are normally distributed
- RF are independent of the fitted values
- RF don't show a pattern over time
- Calculating the regression Equation
4. Check quality of the model
- R-Sq large?
- S (Unexpected Variation) Small?

The formula for a multiple linear regression is:

Y=BOB1+…… BnXn+S …. (4.1)

Y = the predicted value of the dependent variable

B0 = the y-intercept (value of y when all other parameters are set to 0)

B1X1 = the regression coefficient (B1) of the first independent variable (X1) (a.k.a. the effect that increasing the value of the independent variable has on the predicted y value)

BnXn = the regression coefficient of the last independent variable

S= model error (a.k.a. how much variation there is in our estimate of Y)

4.7 MLR WITH TWO PREDICTORS

In Table 4.3 this study tested two RF to detect the overlapping between them and the strength of their correlation, as a result 29 tested samples out of the total are listed in this table, the samples were collected based on the highest R-Square and smallest significance error (S).

Table 4.3 MLR Results with Two Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
1	FEVE	0.643	-0.002	0.965
	EMPT		0.801	0.000
2	FEVE	0.751	-0.232	0.000
	DYSP		0.791	0.000
3	FEVE	0.157	-0.366	0.000
	HEADACHE		0.095	0.195
4	FEVE	0.407	-0.040	0.000
	CYAN		-0.615	0.000
5	FEVE	0.298	-0.525	0.000
	WFAT		-0.410	0.000
6	FEVE	0.393	-0.310	0.000
	DPCH		0.500	0.000
7	FEVE	0.722	-0.086	0.000
	CHEP		0.814	0.000
8	CYAN	0.672	-0.220	0.000

	EMPT		0.663	0.000
	CYAN	0.774	-0.311	0.000
9	DYSP		0.689	0.000
	CYAN	0.413	-0.625	0.000
10	CHEP		0.084	0.164
	CYAN	0.532	-0.703	0.000
11	WFAT		0.361	0.000
	CYAN	0.541	-0.518	0.000
12	DPCH		0.387	0.000
	CYAN	0.413	-0.625	0.000
13	HEADACHE		0.084	0.164
	CYAN	0.759	-0.250	0.000
14	CHEP		0.709	0.000
	CYAN	0.672	-0.220	0.000
15	EMPT	0.763	0.356	0.000
	CHEP		0.565	0.000
16	EMPT	0.834	0.464	0.000
	DYSP		0.552	0.000
17	EMPT	0.677	0.791	0.000
	WFAT		0.185	0.000
18	EMPT	0.705	0.694	0.000
	DPCH		0.271	0.000
19	EMPT	0.643	0.802	0.000
	HEADACHE		0.000	0.996
20	CHEP	0.844	0.517	0.000
	DYSP		0.487	0.000

21	CHEP	0.753	0.837	0.000
	WFAT		0.195	0.000
22	CHEP	0.747	0.753	0.000
	DPCH		0.202	0.000
23	CHEP	0.719	0.838	0.000
	HEADACHE		0.061	0.147
24	CHEP	0.844	0.517	0.000
	DYSP		0.487	0.000
25	DYSP	0.700	0.829	0.000
	WFAT		0.030	0.492
26	DYSP	0.724	0.747	0.000
	DPCH		0.182	0.000
27	WFAT	0.089	0.246	0.001
	HEADACHE		0.189	0.012
28	WFAT	0.343	0.212	0.001
	DPCH		0.539	0.000
29	HEADACHE	0.312	0.535	0.000
	DPCH		0.117	0.074

The following equation gives the highest accuracy in case of overlapping between the Dyspnea (DYSP) and Chest Pain (DYSP) which is = 0.844:

$Y = -0.248 + (0.487 \times \text{DYSP}) + (0.517 \times \text{CHEP})$ (4.2)

4.8 MLR WITH THREE PREDICTORS

This creates the model by combining three independent variables of usability factors with a dependent variable of user rating. The R Square value is calculated using all possible

combinations of predictors, with most of them statistically relevant at 0.000. Only statistically relevant combinations are included in MLR findings in Table 4.4 There was a total of four variations, four of which yielded no statistically significant results. The MLR results with Three predictors are presented in" Table 4.4

Table 4.4 MLR Results with Three Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
1	EMPT	0.838	0.449	.000
	DYSP		0.655	.001
	FEVE		-0.069	.000
2	EMPT	0.835	0.487	.000
	DYSP		0.644	.000
	HEADACHE		0.022	.408
3	EMPT	0.846	0.413	.000
	DYSP		0.618	.000
	CYAN		-0.113	.003
4	EMPT	0.839	.506	.000
	DYSP		.613	.000
	WFAT		.079	.022
5	EMPT	0.846	.474	.000
	DYSP		.585	.000
	DPCH		.139	.001
6	EMPT	0.846	.280	.000
	DYSP		.522	.000
	CHEP		.377	.000
7	EMPT	0.643	.853	.000
	FEVE		-.002	.964

	HEADACHE		.000	.991
8	EMPT	0.677	.725	.000
	FEVE		.079	.122
	CYAN		-.205	.000
9	EMPT	0.685	.736	.000
	FEVE		-.110	.047
	WFAT		.226	.000
10	EMPT	0.705	.686	.000
	FEVE		-.016	.736
	DPCH		.271	.000
11	EMPT	0.763	.351	.000
	FEVE		-.010	.818
	CHEP		.565	.000
12	EMPT	0.672	0.664	.000
	HEADACHE		-0.001	.975
	CYAN		-0.220	.000
13	EMPT	0.677	.839	.000
	HEADACHE		.013	.723
	WFAT		.195	.000
14	EMPT	0.705	.741	.000
	HEADACHE		-.004	.904
	DPCH		.304	.000
15	EMPT	0.763	.350	.000
	HEADACHE		.022	.579
	CHEP		.567	.000
16	EMPT	0.731	.590	.000

	CYAN		-.313	.000
	WFAT		.254	.000
17	EMPT	0.727	.577	.000
	CYAN		-.195	.000
	DPCH		.257	.000
18	EMPT	0.781	.264	.000
	CYAN		-.175	.000
	CHEP		.541	.000
19	EMPT	0.737	.685	.000
	WFAT		.181	.000
	DPCH		.268	.000
20	EMPT	0.797	.685	.000
	WFAT		.181	.000
	CHEP		.268	.000
21	EMPT	0.79	.338	.000
	DPCH		.186	.000
	CHEP		.493	.000
22	DYSP	0.754	.788	.000
	FEVE		-.220	.000
	HEADACHE		.059	.137
23	DYSP	0.783	.699	.000
	FEVE		-.114	.011
	CYAN		-.242	.000
24	DYSP	0.769	.743	.000
	FEVE		-.293	.000
	WFAT		.151	.000

25	DYSP	0.772	.712	.000
	FEVE		-.223	.000
	DPCH		.165	.000
26	DYSP	0.856	.499	.000
	FEVE		-.118	.000
	CHEP		.466	.000
27	DYSP	0.779	.687	.000
	HEADACHE		.072	.055
	CYAN		-.302	.000
28	DYSP	0.711	.817	.000
	HEADACHE		.106	.013
	WFAT		.041	.352
29	DYSP	0.732	.743	.000
	HEADACHE		.091	.026
	DPCH		.174	.000
30	DYSP	0.848	.488	.000
	HEADACHE		.063	.041
	CHEP		.508	.000
31	DYSP	0.792	.626	.000
	CYAN		-.368	.000
	WFAT		.147	.000
32	DYSP	0.791	.622	.000
	CYAN		-.296	.000
	DPCH		.152	.000
33	DYSP	0.867	.452	.000
	CYAN		-.182	.000

	CHEP		.441	.000
34	DYSP	0.726	.734	.000
	WFAT		.046	.275
	DPCH		.187	.000
35	DYSP	0.853	.448	.000
	WFAT		.099	.002
	CHEP		.539	.000
36	DYSP	0.851	.454	.000
	DPCH		.099	.006
	CHEP		.494	.000
37	FEVE	0.413	-.024	.739
	HEADACHE		.081	.188
	CYAN		-.612	.000
38	FEVE	0.315	-.504	.000
	HEADACHE		.111	.097
	WFAT		.419	.000
39	FEVE	0.396	-.299	.000
	HEADACHE		.058	.352
	DPCH		.496	.000
40	FEVE	0.724	-.078	.086
	HEADACHE		.048	.255
	CHEP		.811	.000
41	FEVE	0.559	-.182	.006
	CYAN		-.611	.000
	WFAT		.413	.000
42	FEVE	0.543	-.053	.407

	CYAN		-.487	.000
	DPCH		.388	.000
43	FEVE	0.759	.023	.623
	CYAN		-.261	.000
	CHEP		.711	.000
44	FEVE	0.511	-.442	.000
	WFAT		.371	.000
	DPCH		.462	.000
45	FEVE	0.781	-.199	.000
	WFAT		.271	.000
	CHEP		.757	.000
46	FEVE	0.754	-.091	.031
	DPCH		.204	.000
	CHEP		.719	.000
47	HEADACHE	0.551	.115	.032
	CYAN		-.689	.000
	WFAT		.374	.000
48	HEADACHE	0.545	.061	.255
	CYAN		-.510	.000
	DPCH		.383	.000
49	HEADACHE	0.761	.044	.258
	CYAN		-.245	.000
	CHEP		.706	.000
50	HEADACHE	0.364	.147	.021
	WFAT		.230	.000
	DPCH		.518	.000

51	HEADACHE	0.758	.085	.031
	WFAT		.206	.000
	CHEP		.821	.000
52	HEADACHE	0.750	0.052	.188
	DPCH		0.199	.000
	CHEP		0.748	.000
53	CYAN	0.647	-.592	.000
	WFAT		.334	.000
	DPCH		.348	.000
54	CYAN	0.825	-.333	.000
	WFAT		.269	.000
	CHEP		.649	.000
55	CYAN	0.786	-0.234	.000
	DPCH		0.184	.000
	CHEP		0.633	.000
56	WFAT	0.782	.196	.000
	DPCH		.198	.000
	CHEP		.745	.000

The following equation shows that the parameters Dyspnea (DYSP), Cyanosis (CYAN) and Chest Pain (CHEP) is carrying the highest value of R-Square 0.867 as follows:

$\bar{Y} = 0.498 + (0.452 \times DYSP) + (-0.182 \times CYANIS) + (0.441 \times CHEP)$

From the above symptoms of Dyspnea, Cyanosis and Chest Pain which predict congenital heart disease prediction: has been predicted with 88.7% accuracy level.

4.9 MLR WITH FOUR PREDICTORS

This is going to develop the model in which four usability elements are the value of R square is designed with all conceivable combinations of IVs which are usability factors and researcher have come to know that all of them are statistically significant at the value of 0.000. Table 4.4 gives Multiple Linear Regression results where all of them being statistically significant. We have a total 70 combinations in total and we have 39 combinations were non-significant. Table 4.5 illustrates the results of regression which is based on four value combinations.

Table 4.5: MLR Results with Four Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
1	EMPT	0.835	0.418	0.000
	DYSP		0.565	0.000
	FEVE		-0.072	0.051
	HEADACHE		0.020	0.549
2	EMPT	0.846	0.379	0.000
	DYSP		0.538	0.000
	FEVE		-0.027	0.481
	CYAN		-0.129	0.004
3	EMPT	0.850	0.408	0.000
	DYSP		0.532	0.000
	FEVE		-0.128	0.001
	WFAT		0.122	0.001
4	EMPT	0.850	0.402	0.000
	DYSP		0.516	0.000
	FEVE		-0.075	0.034
	DPCH		0.124	0.001
5	EMPT	0.873	0.228	0.000

	DYSP		0.461	0.000
	FEVE		-0.066	0.043
	CHEP		0.330	0.000
6	EMPT	0.847	0.382	0.000
	DYSP		0.534	0.000
	HEADACHE		0.025	0.423
	CYAN		-0.141	0.001
7	EMPT	0.840	0.467	0.000
	DYSP		0.529	0.000
	HEADACHE		0.033	0.312
	WFAT		0.078	0.018
8	EMPT	0.846	0.439	0.000
	DYSP		0.506	0.000
	HEADACHE		0.022	0.480
	DPCH		0.123	0.001
9	EMPT	0.871	0.253	0.000
	DYSP		0.451	0.000
	HEADACHE		0.035	0.228
	CHEP		0.337	0.000
10	EMPT	0.859	0.377	0.000
	DYSP		0.482	0.000
	CYAN		-0.196	0.000
	WFAT		0.128	0.000
11	EMPT	0.857	0.372	0.000
	DYSP		0.486	0.000
	CYAN		-0.137	0.001

	DPCH		0.119	0.001
12	EMPT	0.879	0.199	0.000
	DYSP		0.434	0.000
	CYAN		-0.129	0.000
	CHEP		0.325	0.000
13	EMPT	0.852	0.456	0.000
	DYSP		0.473	0.000
	WFAT		0.085	0.008
	DPCH		0.131	0.000
14	EMPT	0.879	0.263	0.000
	DYSP		0.413	0.000
	WFAT		0.104	0.000
	CHEP		0.353	0.000
15	EMPT	0.876	0.261	0.000
	DYSP		0.419	0.000
	DPCH		0.095	0.004
	CHEP		0.314	0.000
16	EMPT	0.677	0.680	0.000
	FEVE		0.086	0.121
	HEADACHE		0.008	0.862
	CYAN		-0.257	0.000
17	EMPT	0.685	0.735	0.000
	FEVE		-0.109	0.051
	HEADACHE		0.008	0.869
	WFAT		0.226	0.000
18	EMPT	0.705	0.687	0.000

	FEVE		-0.017	0.723
	HEADACHE		-0.007	0.870
	DPCH		0.272	0.000
19	EMPT	0.763	0.346	0.000
	FEVE		-0.007	0.871
	HEADACHE		0.021	0.597
	CHEP		0.567	0.000
20	EMPT	0.732	0.583	0.000
	FEVE		-0.023	0.664
	CYAN		-0.305	0.000
	WFAT		0.261	0.000
21	EMPT	0.730	0.591	0.000
	FEVE		0.060	0.237
	CYAN		-0.221	0.000
	DPCH		0.252	0.000
22	EMPT	0.784	0.280	0.000
	FEVE		0.059	0.194
	CYAN		-0.201	0.000
	CHEP		0.536	0.000
23	EMPT	0.747	0.620	0.000
	FEVE		-0.125	0.013
	WFAT		0.227	0.000
	DPCH		0.273	0.000
24	EMPT	0.807	0.277	0.000
	FEVE		-0.125	0.005
	WFAT		0.237	0.000

	CHEP		0.568	0.000
25	EMPT	0.790	0.329	0.000
	FEVE		-0.019	0.648
	DPCH		0.187	0.000
	CHEP		0.494	0.000
26	EMPT	0.732	0.585	0.000
	HEADACHE		0.021	0.620
	CYAN		-0.314	0.000
	WFAT		0.256	0.000
27	EMPT	0.727	0.578	0.000
	HEADACHE		-0.006	0.886
	CYAN		-0.195	0.000
	DPCH		0.257	0.000
28	EMPT	0.782	0.258	0.000
	HEADACHE		0.020	0.598
	CYAN		-0.175	0.000
	CHEP		0.543	0.000
29	EMPT	0.737	0.682	0.000
	HEADACHE		0.011	0.792
	WFAT		0.182	0.000
	DPCH		0.267	0.000
30	EMPT	0.799	0.332	0.000
	HEADACHE		0.039	0.289
	WFAT		0.190	0.000
	CHEP		0.570	0.000
31	EMPT	0.790	0.334	0.000

	HEADACHE		0.015	0.675
	DPCH		0.186	0.000
	CHEP		0.495	0.000
32	EMPT	0.782	0.510	0.000
	CYAN		-0.285	0.000
	WFAT		0.245	0.000
	DPCH		0.246	0.000
33	EMPT	0.837	0.201	0.000
	CYAN		-0.266	0.000
	WFAT		0.245	0.000
	CHEP		0.531	0.000
34	EMPT	0.806	0.253	0.000
	CYAN		-0.164	0.000
	DPCH		0.178	0.000
	CHEP		0.475	0.000
35	EMPT	0.824	0.327	0.000
	WFAT		0.183	0.000
	DPCH		0.183	0.000
	CHEP		0.496	0.000
36	DYSP	0.786	0.697	0.000
	FEVE		-0.103	0.023
	HEADACHE		0.058	0.121
	CYAN		-0.241	0.000
37	DYSP	0.777	0.739	0.000
	FEVE		-0.281	0.000
	HEADACHE		0.062	0.109

	WFAT		0.152	0.000
38	DYSP	0.774	0.711	0.000
	FEVE		-0.213	0.000
	HEADACHE		0.051	0.186
	DPCH		0.162	0.000
39	DYSP	0.858	0.498	0.000
	FEVE		-0.109	0.001
	HEADACHE		0.046	0.133
	CHEP		0.463	0.000
40	DYSP	0.810	0.624	0.000
	FEVE		-0.171	0.000
	CYAN		-0.280	0.000
	WFAT		0.190	0.000
41	DYSP	0.800	0.633	0.000
	FEVE		-0.112	0.009
	CYAN		-0.228	0.000
	DPCH		0.150	0.000
42	DYSP	0.869	0.462	0.000
	FEVE		-0.051	0.149
	CYAN		-0.154	0.000
	CHEP		0.430	0.000
43	DYSP	0.793	0.653	0.000
	FEVE		-0.288	0.000
	WFAT		0.164	0.000
	DPCH		0.178	0.000
44	DYSP	0.878	0.440	0.000

	FEVE		-0.182	0.000
	WFAT		0.165	0.000
	CHEP		0.474	0.000
45	DYSP	0.863	0.466	0.000
	FEVE		-0.118	0.000
	DPCH		0.099	0.004
	CHEP		0.442	0.000
46	DYSP	0.786	0.697	0.000
	HEADACHE		-0.103	0.023
	CYAN		0.058	0.121
	WFAT		-0.241	0.000
47	DYSP	0.795	0.623	0.000
	HEADACHE		0.064	0.077
	CYAN		-0.288	0.000
	DPCH		0.147	0.000
48	DYSP	0.869	0.453	0.000
	HEADACHE		0.051	0.077
	CYAN		-0.177	0.000
	CHEP		0.436	0.000
49	DYSP	0.735	0.727	0.000
	HEADACHE		0.096	0.020
	WFAT		0.055	0.190
	DPCH		0.180	0.000
50	DYSP	0.858	0.447	0.000
	HEADACHE		0.071	0.018
	WFAT		0.105	0.001

	CHEP		0.530	0.000
51	DYSP	0.855	0.456	0.000
	HEADACHE		0.059	0.052
	DPCH		0.095	0.007
	CHEP		0.486	0.000
52	DYSP	0.815	0.551	0.000
	CYAN		-0.355	0.000
	WFAT		0.157	0.000
	DPCH		0.162	0.000
53	DYSP	0.889	0.376	0.000
	CYAN		-0.242	0.000
	WFAT		0.165	0.000
	CHEP		0.452	0.000
54	DYSP	0.873	0.422	0.000
	CYAN		-0.179	0.000
	DPCH		0.093	0.005
	CHEP		0.420	0.000
55	DYSP	0.861	0.411	0.000
	WFAT		0.105	0.001
	DPCH		0.106	0.002
	CHEP		0.515	0.000
56	FEVE	0.558	-0.164	0.015
	HEADACHE		0.081	0.129
	CYAN		-0.608	0.000
	WFAT		0.405	0.000
57	FEVE	0.546	-0.042	0.513

	HEADACHE		0.055	0.310
	CYAN		-0.487	0.000
	DPCH		0.384	0.000
58	FEVE	0.761	0.032	0.500
	HEADACHE		0.048	0.223
	CYAN		-0.261	0.000
	CHEP		0.708	0.000
59	FEVE	0.512	-0.427	0.000
	HEADACHE		0.061	0.280
	WFAT		0.364	0.000
	DPCH		0.462	0.000
60	FEVE	0.784	-0.186	0.000
	HEADACHE		0.051	0.172
	WFAT		0.265	0.000
	CHEP		0.759	0.000
61	FEVE	0.756	-0.083	0.050
	HEADACHE		0.038	0.334
	DPCH		0.202	0.000
	CHEP		0.718	0.000
62	FEVE	0.663	-0.180	0.002
	CYAN		-0.495	0.000
	WFAT		0.371	0.000
	DPCH		0.353	0.000
63	FEVE	0.828	-0.081	0.052
	CYAN		-0.292	0.000
	WFAT		0.284	0.000

	CHEP		0.644	0.000
64	FEVE	0.786	0.010	0.823
	CYAN		-0.239	0.000
	DPCH		0.184	0.000
	CHEP		0.635	0.000
65	FEVE	0.813	-0.198	0.000
	WFAT		0.262	0.000
	DPCH		0.201	0.000
	CHEP		0.669	0.000
66	HEADACHE	0.649	0.079	0.097
	CYAN		-0.580	0.000
	WFAT		0.331	0.000
	DPCH		0.347	0.000
67	HEADACHE	0.827	0.060	0.073
	CYAN		-0.323	0.000
	WFAT		0.266	0.000
	CHEP		0.650	0.000
68	HEADACHE	0.787	0.037	0.313
	CYAN		-0.231	0.000
	DPCH		0.183	0.000
	CHEP		0.631	0.000
69	HEADACHE	0.788	0.068	0.065
	WFAT		0.196	0.000
	DPCH		0.194	0.000
	CHEP		0.739	0.000
70	CYAN	0.847	-0.311	0.000

	WFAT		0.256	0.000
	DPCH		0.173	0.000
	CHEP		0.585	0.000

The following equation shows that the best regression model as the previous values obtained in the 3 usability factors DYSP, CYAN and CHEP whereas Weakness of Fatigue (WFAT) is added into it as fourth value and giving the best R-Square here 0.889 even greater than the previous one it means R-Square has been increased by adding WFAT and Provides the following regression equation: Y = 0.344+ (0.37×DYSP) + (-0.242× CYAN) + (0.165× WFAT) + (0.452× CHEP) (4.7)

Dyspnea, Cyanosis and Chest Pain with Weakness and Fatigue are the symptoms that has been.

Predicted with disease of pulmonary embolism.

4.10 MLR WITH FIVE PREDICTORS

This creates the model by combining five independent variables of usability factors with a dependent variable of user rating. The R Square value is calculated using all possible combinations of predictors, with most of them statistically relevant at 0.000. Only statistically relevant combinations are included in MLR findings in Table 4.6. There was a total of four variations, four of which yielded no statistically significant results. The MLR results with five predictors are presented in" Table 4.6.

Table 4.6 MLR Results with Five Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
1	EMPT	0.847	.374	.000
	DYSP		.539	.000
	FEVE		-.024	.541

	HEADACHE		.023	.472
	CYAN		-.130	.004
2	EMPT	0.850	.404	.000
	DYSP		.533	.000
	FEVE		-.126	.001
	HEADACHE		.023	.470
	WFAT		.122	.000
3	EMPT	0.850	.399	.000
	DYSP		.517	.000
	FEVE		-.073	.041
	HEADACHE		.015	.639
	DPCH		.123	.001
4	EMPT	0.874	.221	.000
	DYSP		.462	.000
	FEVE		-.062	.058
	HEADACHE		.028	.327
	CHEP		.332	.000
5	EMPT	0.863	.349	.000
	DYSP		.492	.000
	FEVE		-.079	.041
	CYAN		-.168	.000
	WFAT		.149	.000
6	EMPT	0.857	.362	.000
	DYSP		.493	.000
	FEVE		-.030	.429
	CYAN		-.123	.004

	DPCH		.120	.001
7	EMPT	0.880	.192	.000
	DYSP		.439	.000
	FEVE		-.023	.514
	CYAN		-.119	.003
	CHEP		.324	.000
8	EMPT	0.864	.385	.000
	DYSP		.476	.000
	FEVE		-.134	.000
	WFAT		.134	.000
	DPCH		.136	.000
9	EMPT	0.890	.196	.000
	DYSP		.414	.000
	FEVE		-.131	.000
	WFAT		.148	.000
	CHEP		.357	.000
10	EMPT	0.879	.225	.000
	DYSP		.430	.000
	FEVE		-.066	.036
	DPCH		.096	.003
	CHEP		.309	.000
11	EMPT	0.861	.368	.000
	DYSP		.484	.000
	HEADACHE		.034	.261
	CYAN		-.196	.000
	WFAT		.131	.000

12	EMPT	0.857	.367	.000
	DYSP		.488	.000
	HEADACHE		.021	.496
	CYAN		-.136	.001
	DPCH		.118	.001
13	EMPT	0.880	.189	.000
	DYSP		.436	.000
	HEADACHE		.033	.236
	CYAN		-.128	.000
	CHEP		.327	.000
14	EMPT	0.853	.449	.000
	DYSP		.475	.000
	HEADACHE		.028	.362
	WFAT		.087	.006
	DPCH		.130	.000
15	EMPT	0.880	.252	.000
	DYSP		.411	.000
	HEADACHE		.043	.126
	WFAT		.104	.000
	CHEP		.359	.000
16	EMPT	0.877	.358	.000
	DYSP		.428	.000
	HEADACHE		-.194	.000
	DPCH		.137	.000
	CHEP		.130	.000
17	EMPT	0.872	.352	.000

	DYSP		.430	.000
	CYAN		-.200	.000
	WFAT		.143	.000
	DPCH		.129	.000
18	EMPT	0.898	.170	.000
	DYSP		.366	.000
	CYAN		-.193	.000
	WFAT		.152	.000
	CHEP		.352	.000
19	EMPT	0.886	.198	.000
	DYSP		.404	.000
	CYAN		-.126	.000
	DPCH		.092	.004
	CHEP		.305	.000
20	EMPT	0.886	.262	.000
	DYSP		.374	.000
	WFAT		.107	.000
	DPCH		.103	.001
	CHEP		.334	.000
21	EMPT	0.732	.580	.000
	FEVE		-.021	.702
	HEADACHE		.019	.653
	CYAN		-.307	.000
	WFAT		.262	.000
22	EMPT	0.730	.591	.000
	FEVE		.060	.241

	HEADACHE		.001	.989
	CYAN		-.221	.000
	DPCH		.252	.000
23	EMPT	0.784	.274	.000
	FEVE		.063	.169
	HEADACHE		.027	.483
	CYAN		-.203	.000
	CHEP		.539	.000
24	EMPT	0.747	.620	.000
	FEVE		-.125	.013
	HEADACHE		.001	.983
	WFAT		.227	.000
	DPCH		.273	.000
25	EMPT	0.807	.274	.000
	FEVE		-.117	.008
	HEADACHE		.029	.412
	WFAT		.232	.000
	CHEP		.572	.000
26	EMPT	0.791	.326	.000
	FEVE		-.017	.685
	HEADACHE		.014	.715
	DPCH		.187	.000
	CHEP		.495	.000
27	EMPT	0.783	.495	.000
	FEVE		-.047	.333
	CYAN		-.270	.000

	WFAT		.259	.000
	DPCH		.250	.000
28	EMPT	0.838	.184	.002
	FEVE		-.049	.246
	CYAN		-.250	.000
	WFAT		.259	.000
	CHEP		.534	.000
29	EMPT	0.807	.265	.000
	FEVE		.045	.299
	CYAN		-.183	.000
	DPCH		.175	.000
	CHEP		.472	.000
30	EMPT	0.834	.258	.000
	FEVE		-.129	.002
	WFAT		.231	.000
	DPCH		.188	.000
	CHEP		.498	.000
31	EMPT	0.782	.506	.000
	HEADACHE		.015	.683
	CYAN		-.286	.000
	WFAT		.246	.000
	DPCH		.246	.000
32	EMPT	0.838	.189	.001
	HEADACHE		.041	.210
	CYAN		-.266	.000
	WFAT		.249	.000

	CHEP		.535	.000
33	EMPT	0.806	.249	.000
	HEADACHE		.014	.693
	CYAN		-.163	.000
	DPCH		.178	.000
	CHEP		.476	.000
34	EMPT	0.825	.317	.000
	HEADACHE		.032	.344
	WFAT		.186	.000
	DPCH		.182	.000
	CHEP		.500	.000
35	EMPT	0.859	.192	.000
	CYAN		-.252	.000
	WFAT		.240	.000
	DPCH		.169	.000
	CHEP		.468	.000
36	DYSP	0.814	.621	.000
	FEVE		-.160	.000
	HEADACHE		.061	.083
	CYAN		-.279	.000
	WFAT		.191	.000
37	DYSP	0.802	.632	.000
	FEVE		-.102	.019
	HEADACHE		.050	.164
	CYAN		-.228	.000
	DPCH		.147	.000

38	DYSP	0.870	.462	.000
	FEVE		-.042	.233
	HEADACHE		.046	.116
	CYAN		-.155	.000
	CHEP		.428	.000
39	DYSP	0.795	.652	.000
	FEVE		-.278	.000
	HEADACHE		.053	.152
	WFAT		.165	.000
	DPCH		.175	.000
40	DYSP	0.880	.440	.000
	FEVE		-.174	.000
	HEADACHE		.048	.088
	WFAT		.166	.000
	CHEP		.471	.000
41	DYSP	0.865	.467	.000
	FEVE		-.110	.001
	HEADACHE		.041	.167
	DPCH		.096	.005
	CHEP		.440	.000
42	DYSP	0.830	.548	.000
	FEVE		-.173	.000
	CYAN		-.267	.000
	WFAT		.200	.000
	DPCH		.163	.000
43	DYSP	0.896	.386	.000

	FEVE		-.109	.001
	CYAN		-.193	.000
	WFAT		.191	.000
	CHEP		.431	.000
44	DYSP	0.875	.433	.000
	FEVE		-.053	.126
	CYAN		-.150	.000
	DPCH		.094	.004
	CHEP		.409	.000
45	DYSP	0.887	.401	.000
	FEVE		-.185	.000
	WFAT		.173	.000
	DPCH		.111	.000
	CHEP		.448	.000
46	DYSP	0.817	.549	.000
	HEADACHE		.072	.036
	CYAN		-.348	.000
	WFAT		.161	.000
	DPCH		.158	.000
47	DYSP	0.893	.376	.000
	HEADACHE		.060	.023
	CYAN		-.238	.000
	WFAT		.168	.000
	CHEP		.447	.000
48	DYSP	0.875	.424	.000
	HEADACHE		.047	.095

	CYAN		-.174	.000
	DPCH		.090	.006
	CHEP		.416	.000
49	DYSP	0.866	.411	.000
	HEADACHE		.067	.022
	WFAT		.110	.000
	DPCH		.103	.003
	CHEP		.508	.000
50	DYSP	0.897	.340	.000
	CYAN		-.241	.000
	WFAT		.170	.000
	DPCH		.103	.001
	CHEP		.430	.000
51	FEVE	0.666	-.169	.004
	HEADACHE		.070	.134
	CYAN		-.497	.000
	WFAT		.379	.000
	DPCH		.343	.000
52	FEVE	0.830	-.072	.087
	HEADACHE		.051	.123
	CYAN		-.292	.000
	WFAT		.285	.000
	CHEP		.640	.000
53	FEVE	0.787	.017	.699
	HEADACHE		.039	.292
	CYAN		-.239	.000

	DPCH		.182	.000
	CHEP		.633	.000
54	FEVE	0.814	-.191	.000
	HEADACHE		.042	.231
	WFAT		.262	.000
	DPCH		.198	.000
	CHEP		.668	.000
55	FEVE	0.852	-.093	.018
	CYAN		-.270	.000
	WFAT		.281	.000
	DPCH		.178	.000
	CHEP		.570	.000
56	HEADACHE	0.850	.053	.090
	CYAN		-.307	.000
	WFAT		.260	.000
	DPCH		.170	.000

The following equation shows that EMPT, DYSP, "CYAN, WFAT and CHEP are the five attributes model also needs a single computation. It has an R^2 value of 0.918 and uses five predictors namely.

$\bar{Y}$ = 0.113 + (0.170×EMPT) + (0.366×DYSP) + (-0.193×CYAN) + (0.152×WFAT) + (0.352×CHEP)

Dyspnea, Cyanosis and Chest Pain with weakness and Fatigue and Emptysis are the symptoms leading to Heart Valve and Pericardial diseases.

4.11 MLR WITH SIX PREDICTORS

Seven independent variables and one dependent variable are used to form the model. R2 is calculated by combining all available predictors, many of which are not statistically significant at the 0.05 level. A combination of seven variables in Table 4.7 yields no statistically significant outcomes. There was just one meaningful statistical combination discovered.

Table 4.7: MLR Results with Six Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
1	EMPT	0.864	.343	.000
	DYSP		.492	.000
	FEVE		-.076	.053
	HEADACHE		.028	.355
	CYAN		-.169	.000
	WFAT		.150	.000
2	EMPT	0.857	.358	.000
	DYSP		.494	.000
	FEVE		-.027	.477
	HEADACHE		.018	.556
	CYAN		-.124	.004
	DPCH		.119	.001
3	EMPT	0.881	.184	.000
	DYSP		.440	.000
	FEVE		-.018	.609

	HEADACHE		.031	.267
	CYAN		-.120	.002
	CHEP		.327	.000
4	EMPT	0.864	.343	.000
	DYSP		.492	.000
	FEVE		-.076	.053
	HEADACHE		.028	.355
	WFAT		-.169	.000
	DPCH		.150	.000
5	EMPT	0.891	.188	.000
	DYSP		.415	.000
	FEVE		-.127	.000
	HEADACHE		.033	.222
	WFAT		.149	.000
	CHEP		.360	.000
6	EMPT	0.880	.219	.000
	DYSP		.431	.000
	FEVE		-.063	.048
	HEADACHE		.024	.393
	DPCH		.094	.004
	CHEP		.312	.000

7	EMPT	0.876	.328	.000
	DYSP		.437	.000
	FEVE		-.086	.021
	CYAN		-.164	.000
	WFAT		.160	.000
	DPCH		.133	.000
8	EMPT	0.902	.140	.002
	DYSP		.376	.000
	FEVE		-.083	.012
	CYAN		-.163	.000
	WFAT		.175	.000
	CHEP		.354	.000
9	EMPT	0.886	.190	.000
	DYSP		.410	.000
	FEVE		-.025	.465
	CYAN		-.115	.003
	DPCH		.092	.003
	CHEP		.305	.000
10	EMPT	0.898	.190	.000
	DYSP		.377	.000
	FEVE		-.135	.000

	WFAT		.156	.000
	DPCH		.107	.000
	CHEP		.335	.000
11	EMPT	0.873	.351	.000
	DYSP		.430	.000
	HEADACHE		.030	.305
	CYAN		-.195	.000
	WFAT		.139	.000
	DPCH		.129	.000
12	EMPT	0.900	.156	.001
	DYSP		.367	.000
	HEADACHE		.044	.086
	CYAN		-.193	.000
	WFAT		.156	.000
	CHEP		.356	.000
13	EMPT	0.886	.189	.000
	DYSP		.406	.000
	HEADACHE		.029	.285
	CYAN		-.126	.000
	DPCH		.090	.004
	CHEP		.308	.000

14	EMPT	0.888	.250	.000
	DYSP		.376	.000
	HEADACHE		.039	.153
	WFAT		.110	.000
	DPCH		.101	.001
	CHEP		.338	.000
15	EMPT	0.906	.168	.000
	DYSP		.331	.000
	CYAN		-.192	.000
	WFAT		.158	.000
	DPCH		.101	.000
	CHEP		.332	.000
16	EMPT	0.784	.493	.000
	FEVE		-.045	.354
	HEADACHE		.012	.762
	CYAN		-.271	.000
	WFAT		.259	.000
	DPCH		.249	.000
17	EMPT	0.839	0.174	.004
	FEVE		-0.044	.300
	HEADACHE		0.037	.255

	CYAN		-0.252	.000
	WFAT		0.261	.000
	CHEP		0.538	.000
18	EMPT	0.80 8	.261	.000
	FEVE		.048	.273
	HEADACHE		.019	.592
	CYAN		-.184	.000
	DPCH		.174	.000
	CHEP		.474	.000
19	EMPT	0.835	.253	.000
	FEVE		-.126	.002
	HEADACHE		.022	.508
	WFAT		.232	.000
	DPCH		.187	.000
	CHEP		.500	.000
20	EMPT	0.861	.171	.002
	FEVE		-.062	.113
	CYAN		-.232	.000
	WFAT		.258	.000
	DPCH		.173	.000
	CHEP		.471	.000

21	EMPT	0.860	.182	.001
	HEADACHE		.035	.252
	CYAN		-.253	.000
	WFAT		.243	.000
	DPCH		.168	.000
	CHEP		.472	.000
22	DYSP	0.827	0.547	.000
	FEVE		-0.163	.000
	HEADACHE		0.052	.116
	CYAN		-0.267	.000
	WFAT		0.201	.000
	DPCH		0.16	.000
23	DYSP	0.899	.385	.000
	FEVE		-.100	.002
	HEADACHE		.048	.062
	CYAN		-.193	.000
	WFAT		.192	.000
	CHEP		.428	.000
24	DYSP	0.877	.433	.000
	FEVE		-.045	.196
	HEADACHE		.042	.146

	CYAN		-.150	.000
	DPCH		.091	.005
	CHEP		.407	.000
25	DYSP	0.888	.401	.000
	FEVE		-.177	.000
	HEADACHE		.043	.114
	WFAT		.173	.000
	DPCH		.109	.001
	CHEP		.446	.000
26	DYSP	0.905	.349	.000
	FEVE		-.113	.000
	CYAN		-.189	.000
	WFAT		.198	.000
	DPCH		.107	.000
	CHEP		0.407	.000
27	DYSP	0.900	.341	.000
	HEADACHE		.056	.029
	CYAN		-.236	.000
	WFAT		.174	.000
	DPCH		.100	.001
	CHEP		.425	.000

28	FEVE	0.854	-.085	.031
	HEADACHE		.043	.166
	CYAN		-.270	.000
	WFAT		.281	.000

From above tables summarize the results of applying forward stepwise MLR on our dataset. Six attributes model also needs a single computation. It has an R^2 value of 0.916 and uses six predictors which are EMPT, DYSP, CYAN, WFAT, DPCH and" CHEP.

4.12 MLR WITH SEVEN PREDICTORS

The model is constructed by mixing seven independent variables with a single dependent one. R2 is calculated using all potential predictor combinations, many of which are not statistically significant at the 0.05 level. A combination of seven variables, shown in Table 4.7, has no statistically significant outcomes. Statistical significance has been observed in just one combination so far.

Table 4.8 MLR Results with Seven Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
1	EMPT	0.877	.324	.000
	DYSP		.439	.000
	FEVE		-.083	.027
	HEADACHE		.023	.425
	CYAN		-.165	.000
	WFAT		.161	.000

	DPCH		.132	.000
2	EMPT	0.903	.131	.005
	DYSP		.376	.000
	FEVE		-.078	.019
	HEADACHE		.038	.137
	CYAN		-.165	.000
	WFAT		.176	.000
	CHEP		.357	.000
3	EMPT	0.887	.184	.000
	DYSP		.411	.000
	FEVE		-.020	.546
	HEADACHE		.027	.324
	CYAN		-.116	.003
	DPCH		.091	.004
	CHEP		.307	.000
4	EMPT	0.899	.184	.000
	DYSP		.378	.000
	FEVE		-.131	.000
	HEADACHE		.029	.274
	WFAT		.157	.000
	DPCH		.106	.000

	CHEP		.338	.000
5	EMPT	0.910	.136	.002
	DYSP		.340	.000
	FEVE		-.088	.006
	CYAN		-.160	.000
	WFAT		.182	.000
	DPCH		.105	.000
	CHEP		.333	.000
6	EMPT	0.907	.155	.001
	DYSP		.332	.000
	HEADACHE		.040	.105
	CYAN		-.192	.000
	WFAT		.161	.000
	DPCH		.099	.001
	CHEP		.336	.000
7	EMPT	0.862	.163	.003
	FEVE		-.058	.141
	HEADACHE		.030	.324
	CYAN		-.234	.000
	WFAT		.259	.000
	DPCH		.171	.000

	CHEP		.474	.000
8	DYSP	0.906	.349	.000
	FEVE		-.105	.001
	HEADACHE		.044	.080
	CYAN		-.189	.000
	WFAT		.198	.000
	DPCH		.104	.000
	CHEP		.405	.000

The equation is given below for the suggested model with R-Square 0.926

Y = 0.078+ (0.136× EMPT) + (0.340× DYSP) + (-0.088× FEVE)

+ (-0.160× CYAN) + (0.182×WFAT) + (0.105× DPCH) + (0.333× CHEP)

Chest pain and fever has been overlapping symptoms with has been overlapping symptoms leading to Heart Attack disease.

4.13 MLR WITH EIGHT PREDICTORS

When we applied the regression test with all usability factors as an independent variables and user rating and it is found that results are non-significant even R-Square is increased 0.911 which is consequently implying that headache is making this non-significant.

Table 4.9: MLR Results with Eight Predictors

Dependent Variable	Predictors (In Pair)	R-Square	Standardized Beta	Significance
User Rating	EMPT	0.911	.128	.004
	DYSP		.341	.000
	FEVE		-.083	.009
	HEADACHE		.034	.172
	CYAN		-.163	.000
	WFAT		.183	.000
	DPCH		.103	.000
	CHEP		.336	.000

Y = -3.34×10'-5+ (0.128× EMPT) × (0.341+ DYSP) + (-0.083× FEVE)
+ *(0.034*× HEADACHE) × *(-0.163*+ CYAN*)* × *(0.183**WFAT*)*
+ *(0.103*× DPCH*)* + *(0.336*× CHEP.

4.14 SUGGESTED MODEL FOR UAE HOSPITALS

After applying MLR, this study was able to get findings for all combinations of 2, 3, 4, 5, 6, and 7 predictor variables. Increase the number of variables gradually until the optimal model is obtained. Table 4.9 lists all the most effective regression models that produced significant findings. After collecting the results by using the statistical methods, the final results included the Predictors, the best model, Accuracy, and no of patients for each case has been shown in Table 4.10.

Table 4.10 Summary of MLR Based Models

Sr. No	Predictors	Prediction Model	Squared
1	CHEP	$\overline{Y}$ = 0.347 + (0.846× CHEP)	O.715

2	CHEP, DYSP	Y = -0.248+ (0.487× DYSP) + (0.517× CHEP)	0.844
3	)YSP, CYAN and CHEP	$\overline{Y}$ = 0.498+ (.452× DYSP) + (-0.182× CYANIS) + (0.441× CHEP)	0.867
4	′SP, CYAN, CHEP and WFAT	*Y = 0.344+ (0.376× DYSP) + (-0.242×* CYAN*) + (0.165×* WFAT*) + (0.452×* CHEP)	0.889
5	IPT, DYSP, CYAN, WFAT and CHEP	$\overline{Y}$ = 0.113 + (0.170× EMPT) + (0.366× DYSP) + (-0.193*CYAN) + (0.152× WFAT) + (0.352× CHEP)	0.898
6	IPT, DYSP, CYAN, WFAT, DPCH and CHEP	Y= -0.024+ (0.168× EMPT) + (0.331+ DYSP) + (-0.192+ CYAN) + (0.158+ WFAT) + (0.101× DPCH) + (0.332× CHEP)	0.906
7	IPT, DYSP, FEVE, CYAN, WFAT, DPCH and CHEP	Y = 0.078+ (0.136× EMPT) + (0.340+ DYSP) + (-0.088+ FEVE) + (-0.160+ CYAN) + (0.182× WFAT) + (0.105× DPCH) + (0.333× CHEP)	0.910
8	IPT, DYSP, FEVE, CYAN, WFAT, DPCH, HEAD and CHEP	Y= -3.34× 10'-5+ (0.128× EMPT) + (0.341+ DYSP) + (-0.083+ FEVE) - (0.034× HEADACHE) + (-0.163+ CYAN) + (0.183× WFAT) + (0.103× DPCH) + (0.336× CHEP)	0.911

In the MLR model with seven predictors, the larger value of the R-Square is 0.911, as shown at Sr. No. 7. The R-Square, on the other hand, shows just a 0.001 rise when the eight

variables are included in the regression model, but the R-Square shows a higher change in prior models. As a result, the final model has seven variables excluding headaches.

$\bar{Y}$ = -0.078+ (0.145× EMPT) + (0.395× DYSP) + (-0.082× FEVE) + (-0.128× CYAN) + (0.190× WFAT) + (0.118× DPCH) + (0.376× CHEP)

Figure 4.4 shows the results of the accuracy for the suggested models as follows.

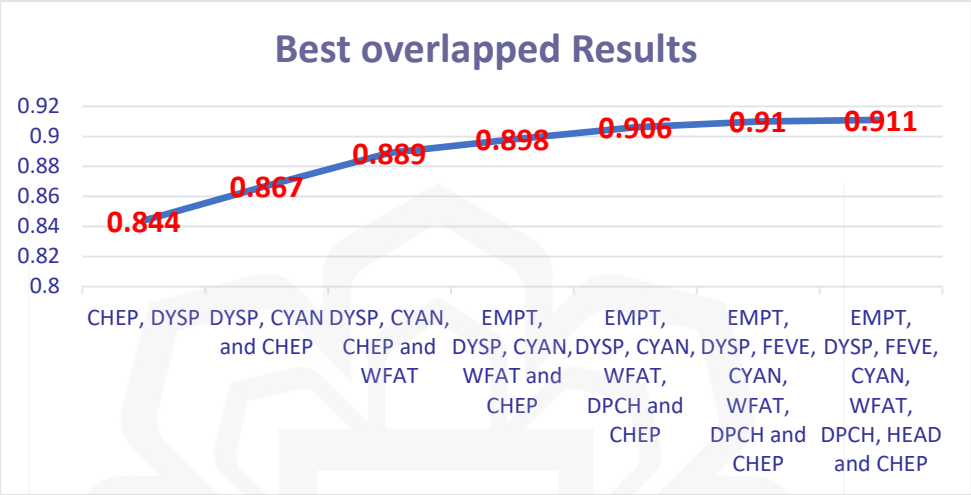


Figure 4.4 Best Overlapped symptom

4.15 COMPARISON OF SUGGESTED MODEL (HYBRID VS. RUF)

Using elements refined by statistics methodologies, an optimal hybrid model with eight usability aspects was determined in this study, and they are as follows: The independent variables/factors in the hybrid model that influenced the dependent variable's rating regressively were Emptysis, Diastolic Blood Pressure, Fever, Cyanosis, weakness or Fatigue, chest discomfort pressure, and Chest pain. Our suggested RUF (Rating of Usability Factor) is compared to the HYBRID model in the next section. An eight-factor Hybrid Model Rating, an RUF Model, and the MLR and SLR techniques are used to finish the RUF Model.

Users' ratings of the hybrid model were shown to be non-significantly predicted by the model's R2 value.

The regression equation is $\bar{Y}$ = 0.347 + (0.846*CHEP)

Predicted value of user rating (dependent variable) is denoted by Y, whereas rating provided by the hybrid model is represented by OR (independent variable).

When it comes to forecasting user rating apps' usability characteristics, the RUF Model (based on 7 elements) performs better than other models, according to the coefficient of determination (R2 square). According to the seven usability criteria, RUF explained 91% of variance (R2= 0.910), whereas Hybrid explained 91% of variance (R2 Square= 0.911).

Predicted value of user rating (dependent variable) is denoted by Y, whereas rating provided by the hybrid model is represented by OR (independent variable).

When it comes to forecasting user rating apps' usability characteristics, the RUF Model (based on 7 elements) performs better than other models, according to the coefficient of determination (R2 square). According to the seven usability criteria, RUF explained 91% of variance (R2= 0.910), whereas Hybrid explained 91% of variance (R2 Square= 0.911).

4.16 CHAPTER SUMMARY

This chapter explained the steps for implementing the methodology starting with finding the weights for influential index using Gini-Entropy Index and got the highest effectiveness results of RF, from the findings previously, tis chapter explained how to detect the overlapping in symptoms by analyzing the symptoms using Regression Model and based on the results discussed the best model with highest accuracy and lowest error. Each model compared with other models based on No of Predictors. R squared and accuracy and got total no of patients in each model using statistical methods.

CHAPTER FIVE

ASSESSMENT AND VALIDATION OF PROPOSED MODEL

5.1 INTRODUCTION

This chapter "provides for evaluating model values by using K fold cross validation with PRED(X) and MMRE, as well as summarized k-fold models with R-square values. In the proposed study the dataset was split into two parts, 20% test and 80% for training, this study used 621 samples for testing and 2000 samples for training, this study used 10th k fold cross validation to test the performance of the model.

5.2 VALIDATION PROCESS

Cross-validation K-fold (Rodríguez, 2010) is a non-exhaustive model validation technique. This technique uses the performance of the predicted model on an autonomous data set. The dataset is divided into the same number of folds depending on the data size. Data calibration and validation doubles their data points through k-fold cross-validation. The training segment contains (k-1) and the test section contains (k-1) for each interaction data set (1). It was divided into 8 folds, each containing 21 answers, due to the large number of replies in our data collection. It was predetermined the sequence of the answers in the folds. In each iteration, eight folds were used to calibrate the model, and the rest were used to validate the model.

5.3 K-FOLD CROSS STEPS VALIDATION

K -Fold Cross Validation can be calculated as follows:

- ✓ The dataset is split into training and test dataset.
- ✓ The training dataset is then split into K-folds.
- ✓ Out of the K-folds, (K-1) fold is used for training.
- ✓ 1-fold is used for validation
- ✓ The model with specific hyper parameters is trained with training data (K-1 folds) and validation data as 1-fold. The performance of the model is recorded.
- ✓ The above steps (step 3, step 4, and step 5) is repeated until each of the k-fold got used for validation purpose. Therefore, it is called k-fold cross-validation.
- ✓ Finally, the mean and standard deviation of the model performance is computed by taking all of the model scores calculated in step 5 for each of the K models.
- ✓ Step 3 to Step 7 is repeated for different values of hyper parameters.
- ✓ Finally, the hyper parameters which result in the most optimal mean and the standard value of model scores get selected.
- ✓ The model is then trained using the training data set (step 2) and the model performance is computed on the test data set (step 1).
- ✓ In the proposed study, Dataset was divided into 2 parts 20% for test and 80% for training which is the data values in all folds are divided into eight fixed groups. During each fold, each category comprises a total of 20 instances, i.e., 1–21 and 22–42. For fold K = 1 to K= 10 It has been observed that even after dividing data into 10 folds each fold has been found to be significant which proves that our dataset is correct and idea, even each of its value in the K fold prediction we identified all data sets are significant values.

Table 5.1 Summarized K-FOLD Predicted Model

Folds	**Prediction Models**	**R^2**	**MMRE**
CV 1.	Y (K-1) = .121 + (0.119×EMPT) + (0.350×DYSP) + (-0.107×FEVE) + (-0.157×CYAN) + (0.186×WFAT) + (0.099×DPCH) + (0.339×CHEP)	0.914	0.1523
CV2.	Y (K-2) = .079 + (0.100×EMPT) + (0.388×DYSP) +	0.915	0.0705

	(-0.084×FEVE) + (-0.160×CYAN) + (0.155×WFAT) + (0.132×DPCH) + (0.332×CHEP)		
CV3.	Y (K-3) = .212 + (0.097×EMPT) + (0.320×DYSP) + (-0.094×FEVE) + (-0.195×CYAN) + (0.197×WFAT) + (0.108×DPCH) + (0.374×CHEP)	0.908	0.0486
CV5	Y (K-2) = .079 + (0.100×EMPT) + (0.388×DYSP) + (-0.084×FEVE) + (-0.160×CYAN) + (0.155×WFAT) + (0.132×DPCH) + (0.332×CHEP)	0.915	0.0705
CV6.	Y (K-4) = .050 + (0.123×EMPT) + (0.360×DYSP) + (-0.089×FEVE) + (-0.160×CYAN) + (0.174×WFAT) + (0.101×DPCH) + (0.336×CHEP)	0.905	0.0471
CV7.	Y (K-5) = .089 + (0.141×EMPT) + (0.327×DYSP) + (-0.091×FEVE) + (-0.160×CYAN) + (0.193×WFAT) + (0.105×DPCH) + (0.321×CHEP)	0.911	0.0509
CV8	Y (K-6) = -.001 + (0.155×EMPT) + (0.338×DYSP) + (-0.070×FEVE) + (-0.146×CYAN) + (0.173×WFAT) + (0.088×DPCH) + (0.326×CHEP)	0.914	0.0451
CV9.	Y (K-7) = -.004 + (0.160×EMPT) + (0.322×DYSP) + (-0.084×FEVE) + (-0.140×CYAN) + (0.178×WFAT) + (0.109×DPCH) + (0.326×CHEP)	0.909	0.0408
CV10	Y (K-8) = .081 + (0.196×EMPT) + (0.311×DYSP)	0.910	0.0558

Figure 5.1 shows the results of 10th Fold R Squared for different levels of parameters, the highest value = 0. 915 with seven predictors based on the results of table 5.1.

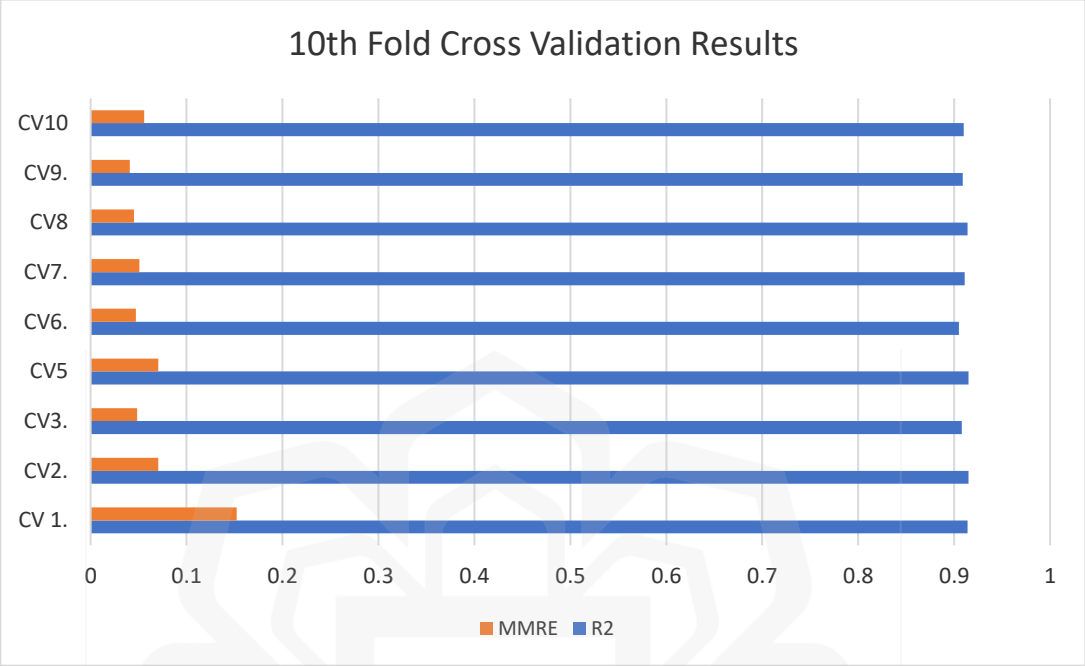


Figure 5.1 10th-Fold Cross Validation Results

As a results in table 5.1 the best results were in CV9 since the MMRE value was the = 0.048, and R squared= 0.909 with training 7 overlapped predictors (RF) which they are: EMPT, DSYP, FEVE, CYAN, WFAT, DPBH, and CHEP, which means that our suggested model GERM is reliable and correct in performance. Which means the strongest correlation between the RF was in level 9 which supports the reliability and efficiency of the model. Table 5.2 explains the experimental results for the deep learning model based on MLP using the suggested parameters as explained previously as follows:

Table 5.2 Experimental Results for the deep learning model based on MLP.

Epoch No.	Loss	Accuracy	Validation Loss	Validation Accuracy
1	1.6093	0.6313	0.8935	0.8117
2	0.6801	0.8380	0.5132	0.8738
3	0.4693	0.8391	0.4027	0.8955
	0.4693	0.8710	0.4027	0.8955

4	0.3926	0.8831	0.3515	0.9038
5	0.3541	0.8901	0.3235	0.8911
6	0.33021	0.8901	0.3082	0.9005
7	0.3133	0.8912	0.2963	0.9012
8	0.3003	0.8944	0.2846	0.9083
9	0.2895	0.8945	0.2744	0.9000
10	0.2790	0.8988	0.2668	0.9032

The Baseline Error for MLP Model is 7.68% and the highest accuracy is 0.8988 which means that our hybrid model GERM performed the efficiency with highest accuracy 91% and less error rate less than 0.001 as shown in figure. Figure 5.2 shows a comparison between the performances of GERM and MLP Model

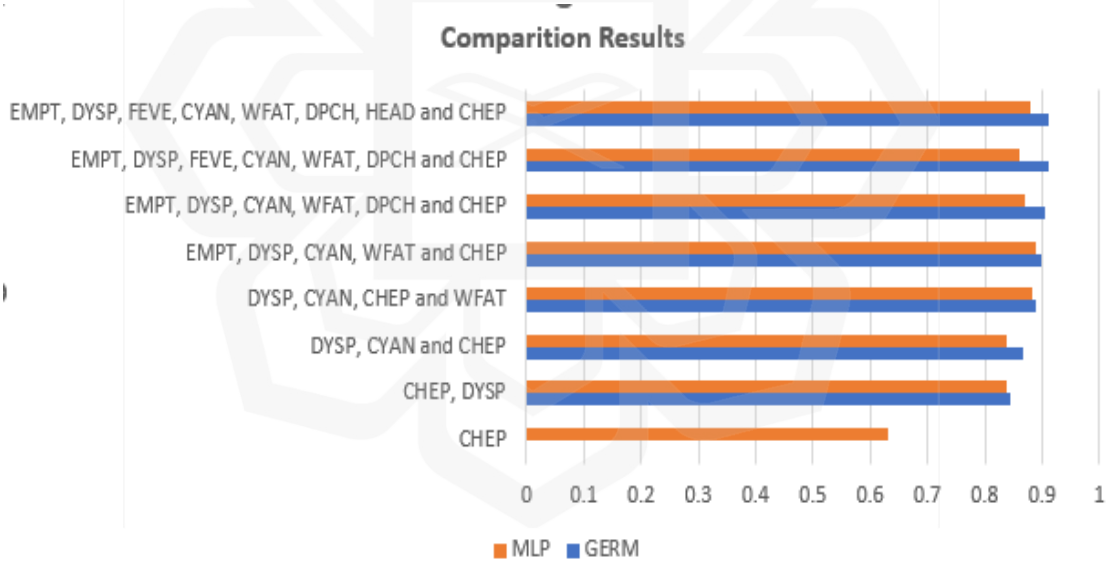


Figure 5.2 Comparison of Accuracy Results for GERM and MLP Models

5.4 EVALUATING THE CVD PREDICTION

In this study after the most influential indexes for the Risk Factors have been analyzed, the right decision depends on detecting the known and hidden patterns in the dataset, this will be satisfied by the rule analysis of the most influential RF, and the values are calculated based on the Matrix in Table as follows:

Table 5.3 Calculating Results Based on Matrix

Estimated/Real samples	**Negative Class**	**Positive Class**
Negative Class	TN	FP
Positive Class	FN	TP

TP= No. of truly identified Positive values (Risk of CVD)

FP = No. of falsely identified normal values (No Risk of CVD)

TN = No truly identified normal value (non-attack)

FN= No. of falsely identified positive samples with (Risk of CVD)

By submitting the values to the following equations:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Recall} = \frac{TP+TN}{FP+TN}$$

Table 5.4 Shows No. of observed patients regarding to overlapped symptoms (RF) starting from the highest effective (RF) on patients which is Diastolic Blood Pressure with High Rate (DBPH) with other RF. Table 5.3 has shown the details of Frequency and redundancy of esteemed and real dataset using Clutter Matrix as explained before, this is called the Penetration Detection since will detect the hidden pattern beside the known as follows:

Table 5.4 Detection the Patterns in the Dataset using Clutter Matrix

Detection Rule	Frequency in no. of samples (Patients)
Diastolic Blood Pressure with High rate (DBPH) more than 120 mg, CYAN and CHEP increase the risk of CVD	observed= 176 patients Accuracy= 84%
Normal DBPH(Systolic), normal weight BMI, and no history of smoking CHEP, DYSP	observed= 480 patients Accuracy= 86%
Normal systolic blood pressure and normal weight are not associated with the CVD	observed= 232 patients Accuracy = 88%
Higher systolic blood pressure than 120 mmHg, above normal weight, and above-normal diastolic blood pressure, DYSP, CYAN, CHEP and WFAT are associated with the risk of CVD.	observed = 644 patients Accuracy= 0.89%
Blood pressure of 120-139 mmHg, absence of HEAD, and high BMI i EMPT, DYSP, FEVE, CYAN, WFAT, DPCH, HEAD and CHEP increase the risk CVD	observed in 628 patients Accuracy = 91.5%
Higher Diastolic blood pressure is associated with the risk of CVD	Observed in 524 patients Accuracy= 88%

5.5 COMPARATIVE ANALYSIS WITH OTHER CLASSIFIERS

In this section, machine learning offers interpretability and comparison between GERM Model which can work well with larger datasets, and more attributes exceeded 16 attributes and satisfied a higher accuracy result. Deep learning, on the other hand, excels in tasks that involve vast amounts of data and complex patterns but often sacrifices interpretability. The choice between ML and DL depends on the specific problem, available data, and computational resources. The following table illustrates the results of another classifiers for Heart- CVD prediction the details were as follows: KNN classifier with 14 attributes, accuracy = 70%, KNN with 10 attributes, accuracy = 71%, Decision Trees(ID3) using 10 attributes, accuracy = 91%, Decision Trees (ID3) using 14 attributed , accuracy= 88%,

Gaussian NB using 14 attributes, accuracy = 88%, Logistic Regression using 14 attributes, accuracy = 89% all the details are shown in figure 5.3 as follows:

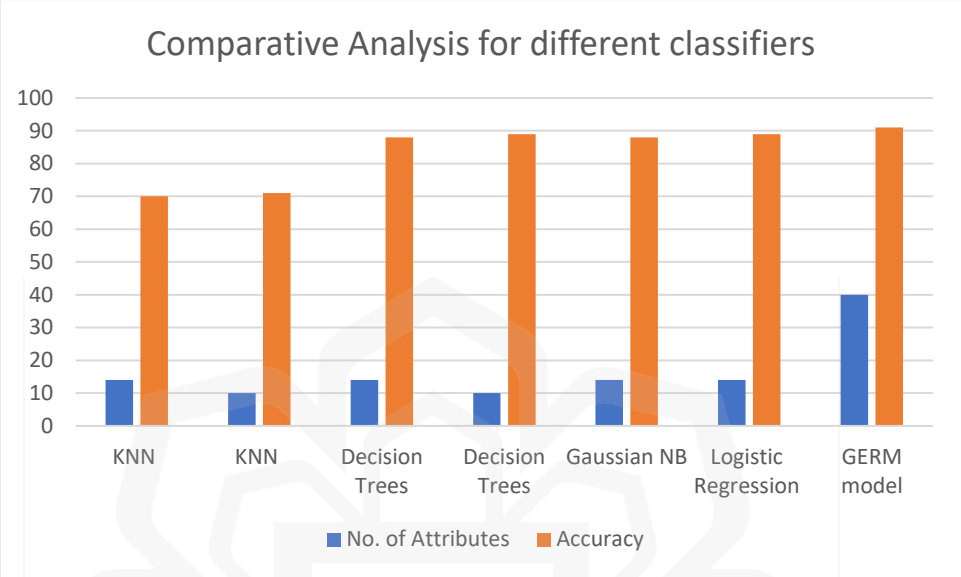


Figure 5.3 Comparative Analysis for Different Classifiers of Heart- CVD Prediction

5.6 CHAPTER SUMMARY

This chapter presented the performance of the suggested (GERM) model by using k-Fold Cross Validation the dataset was split into two parts and 10^{th} Fold Cross validation was used for split the data into 10 equal parts, summary of the Random K-Fold Predicted Model was shown that the accuracy value for Cross Validation was = 91% which proved the Reliability and Validity of the proposed model , This chapter provided a comparison for accuracy results between GERM and Deep Learning MLP model which proved that the suggested model has a higher accuracy with less error than Deep Learning model based on MLP.

CHAPTER SIX

CONCLUSION AND FUTURE WORK

6.1 INTRODUCTION

This chapter provides the conclusion results for the proposed Model GERM and the effectiveness of the highest influential indexes on CVD patients in UAE and suggests some ways for developing the suggested system and the limitations of the study and future directions.

6.2 CONCLUSIONS

In this work medical online Dataset involved 40 attributes downloaded from http://www.figshare.con for 2621 people (male and female). Dataset divided into 20%test and 80% train were preprocessed by filtering techniques removing the redundant and noisy data, mean values were given to the missing data, to avoid over fitting, Cross Validation process was used, hidden pattern had been detected by Penetration Detection analysis using Clutter Matrix. Statistical methods, Deep Learning, Gini-Entropy Index, Dimensional Reduction, parameter Analysis, Correlation Pearson Matrix, Decision Trees, MLP, Back propagation algorithm Deep Learning Techniques, Regression Analysis were used in parameters (RF) Analysis and CVD prediction which programmed and analyzed by Python, Minitab, and SPSS. Each person has 40 attributes regarding heart problems different statements.

The main conclusions drawn from this work are presented in the following.

- In this study, the statistical results showed that the most significant Factors for CVD patients in UAE community which were achieved by calculating the most influential indexes using Gini-Entropy Index algorithm were as follows: DYSP, CHEP, FEVE, CYAN, WFAT, EMPT and DPBH provided the strongest effects on CVD patients in UAE community with accuracy of 91%. The accuracy results for all the levels were as follows:

Level Number One:

DYSP and CHEP = 84.4%

Level Number Two:

DYSP, CHEP and CYAN is explaining 86.7%

Level Number Three:

DYSP, CHEP, CYAN and WFAT explaining 88.9%

Level Number Four:

DYSP, CHEP, CYAN, WFAT and EMPT 89.8%

Level Number Five:

DYSP, CHEP, CYAN, WFAT, EMPT and DPCH explaining 90.6%

Level Number Six: (Ideal Model)

DYSP, CHEP, CYAN, WFAT, EMPT, DPCH and FEVE explaining 91%.

- It is very important to detect the positive and negative correlations for the Risk Factors (RF) of Cardiovascular Disease (CVD) in the proposed Model to discover the trends of data, and extract the important features by which were used as the basic parameters for detecting the overlapped symptoms in the next stage, the statistical analysis showed that:

- Age Factor showed that, 60% of the patients were male, and 40% were female. Therefore, CVD more commonly affects men than women in UAE community. BMI/Weight Factor showed that 30% of the patients had normal weight, 70% were overweight, and Therefore CVD commonly effects obese patients in UAE community. Cholesterol Factor: According to the obtained data, 30% of the patients had lower cholesterol levels than 200, while 53% had cholesterol levels of 200-239 (borderline level). The other patients had higher cholesterol levels than 240, and 99% of these patients' experienced CVD. In addition, 53% of the patients had a family history of cardiovascular diseases, while no such history was reported in 47%. However. Among the patients, 58% had no history of physical exercise, and this category includes the majority of cardiovascular patients. Moreover, 99% of the patients with high diastolic blood pressure had cardiovascular diseases. − Statistical data indicated CVD even in the patients without a history of hereditary heart failure or smoking habits. Heart failure has also been reported in patients with low cholesterol. Therefore, it could be inferred that statistical data alone

may not provide accurate data on the risk factors and diagnosis of cardiovascular diseases, and data mining techniques are required due to overlapping factors.

- The profound learning model detected the overlapping in cardiovascular diseases predictors (symptoms) by using Deep learning technique based on MLP and Regression Analysis and compared the values of R squared in pairs of 2, 3, 4,5,6,7 and 8 Independent Variables (IV), the results of detecting the overlapped in symptoms by Deep Learning model based on MLP = 0.89% while GERM achieved an accuracy = 91% with 40 attributes and demographic characteristics.
- This study decided to use Multiple Linear Regression to construct this by making sets of these components reliant on every conceivable combination of a safe modelling model, 10th k fold pairs were formed, and 21 out of 29 sets were found to be significant, with R-Square values in the great range with sig 0.000 R-Squares. Once again, the R-Square for DYSP and CHEP is the highest at 0.844, accounting for 84.4% of the difference in outcomes. Indicating which factors are the most important when it comes to safeguarding users. a triple pair, has recently been introduced on the crossover stage. DYSP, CHEP, and CYAN together explain 86.7% of the model's variation. The model's R Square has been found to have been improved by CYAN. Multiple Linear Regressions have been used in the same way for four and five regressions. In the fourth matched relapse, the R-Square is 0.889, with the expansion of 88.9 percent of the equation clarified, and in the fifth, the R-Square is 0.898 with the expansion of clarifying 0.9 percent more than the fourth matched relapse, which is a significant improvement. As a result of including WFAT and Productive, the models explain ability increased. When DPCH is added to the pair, only five of the 28 mixes have strong R-Square upsides of 0.906 and represent 90.6% of the model, due to the specialist's use of six matching relapses.
- Furthermore, it has been shown that the number of important blends decreases when the matching approach is enlarged. For example, just one set of EMPT, DYSP, FEVE, CYAN, WFAT, DPCH, and CHEP are important in the seventh combined various relapse. As a result, the model's R-Square is enlarged to 0.910, making it clearer than the previous one.

- Using K-fold in which testing is done other accuracy is acceptable and relatively high at the 7th level, and accuracy =91%

6.3 REJECTED MODEL

The R-Square increases by only 0.1 percent when the MLR is applied to all eight components, but the condition is no longer important. The sixth model is the best option. MRE was applied to the remaining 21-147 with Constants of 0-21 K-FOLDs and the equivalent has been finished with irregular information and it was demonstrated that all of the K-FOLDs are 0-21. We were able to isolate the data in this manner. A Pred(X) of 0.25 or less was found in just one of the 336 MRE esteems, yet MMREs had the right attributes for every one of them, proving the veracity of our data. According to our data our arranged MMRE values are 0.1523, 0.0705, 0.0486, 0.0471, 0.0509, 0.0451, 0.0408 and 0.0558, some values are less than 0.05, we have applied grand MRRE and found 0.05 proving that our dataset is free of mean relative error.

6.4 FUTURE DIRECTIONS

Study promises to better integrate clinical information and dress code infrequently inpatient forms of illness, as well as bridging the chance to omit investigation and bed phenotypes as well as pharmacological availability. In order to overcome current limits, especially restricting notions, progress in technology may be necessary. Considering the design and modeling, characteristics defining limitations, and techniques to combat overfitting in cardiovascular, the literature presently exists, cardiovascular in particular, is judged to be fit to benefit from the current research skills. This is since merging routine health records with lifestyle guidelines saves data streams. If privacy and data security issues are addressed, this combination may provide the groundwork for a "medical Internet of Things (IoT)" that can continuously monitor and observe illnesses, risk factors, and early indications of probable harm. Further research can be done on using the worldwide population to increase the scope of the study. As the DL results are more accurate with the

increasing number of samples, so a worldwide data will provide more accuracy. The proposed model can be improved to detect the overlapping symptoms under the conditions of Covid 19, since RF will be affected and modified.

6.5 LIMITATION OF THE STUDY

This study didn't consider the length of stay for patients in UAE hospitals since this issue will update the demographic results and increase the complexity for finding the influential indexes for them, proposed system. The collected dataset didn't consider the effects of Covid 19 conditions in modifying the Demographic and characteristics for UAE patients since the dataset was served and collected during the last 3 years.

6.6 CHAPTER SUMMARY

This chapter provided details about it and curacy for each overlapped Risk Factors (RF) and discussed the effects of highest influential indexes of RF on patients, compare the results with Deep learning MLP model, and explained which model considered to be the optimal one and the reasons behind it, and discussed the future directions and how to improve the suggested model in future with its limitations.

REFERENCE

Alireza Mehrankia., Mohammad, Reza., & Kamal, Mirzaei. (2022). Prediction of Heart Attacks Using Biological Signals Based on Recurrent GMDH Neural Network. :10.1007/s11063-021-10667-8.

Al'Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., Pandey, M., Maliakal, G., Van Rosendael, A. R., & Beecy, A. N. J. E. h. j. (2019) Clinical applications

of machine learning in cardiovascular disease and its relevance to cardiac imaging. *40*(24), 1975-1986.

Alhadeethy, N. F. A., Zeki, A. M., & Shah, A. (2021) Deep Learning Model For Predicting And Detecting Overlapping Symptoms Of Cardiovascular Diseases In Hospitals Of UAE. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12*(14), 5212-5224.

Megha Bhushan, Akkshat Pandit, Ayush Gargr. (future) Machine learning and Deep Learning techniques for the analysis of heart disease . *56*(12), 1-52.

Bakator, M., Radosav, D. J. M. T., & Interaction. (2018) Deep learning and medical diagnosis: A review of literature. *2*(3), 47.

Baldi, P. J. A. r. o. b. d. s. (2018) Deep learning in biomedical data science. *1*, 181-205.

Bell, E., Bryman, A., & Harley, B. (2018). *Business research methods*: Oxford university press.

Maria Teresa, Martin Bayon, Carmen, & Jose.(2023) Heart disease risk prediction using deep learning techniques with feature augmentation.82:31759-31773

Ben-Haim, Y., & Tom-Tov, E. J. J. o. M. L. R. (2010) A Streaming Parallel Decision Tree Algorithm. *11*(2).

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., & Ballester, M. A. G. J. I. t. o. m. i. (2018) Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *, 37*(11), 2514-2525.

Saravana Srinivasan., Subathra G., Sandeep Kumar., Benjula Anbu., Prabhu Jayagobal., & Gemmachis Dalu. (2023) An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database.13588 .

Bishop, C. M. (2006). *Pattern recognition and machine learning*: springer.

Brennan, C. W., Verhaak, R. G., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., & Berman, S. H. J. C. (2013) The somatic genomic landscape of glioblastoma. *155*(2), 462-477.

Bright, B., Hicke, J., & Hudak, A. J. E. R. L. (2012) Landscape-scale analysis of aboveground tree carbon stocks affected by mountain pine beetles in Idaho. *7*(4), 045702.

Brijain, M., Patel, R., Kushik, M., & Rana, K. (2014) A survey on decision tree algorithm for classification.

Brunese, L., Martinelli, F., Mercaldo, F., & Santone, A. J. P. C. S. (2020) Deep learning for heart disease detection through cardiac sounds. *176*, 2202-2211.

Cao, Y., Liu, Z., Zhang, P., Zheng, Y., Song, Y., Cui, L. J. J. o. A. I., & Systems. (2019) Deep learning methods for cardiovascular image. *1*(1), 96-109.

Chaddha, A., Robinson, E. A., Kline-Rogers, E., Alexandris-Souphis, T., & Rubenfire, M. J. T. A. j. o. m. (2016) Mental health and cardiovascular disease. *129*(11), 1145-1148.

Chawla, N. V., & Davis, D. A. J. J. o. g. i. m. (2013) Bringing big data to personalized healthcare: a patient-centered framework. *28*(3), 660-665.

Chen, R., Mias, G. I., Li-Pook-Than, J., Jiang, L., Lam, H. Y., Chen, R., Miriami, E., Karczewski, K. J., Hariharan, M., & Dewey, F. E. J. C. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. *148*(6), 1293-1307.

Cheung, C. Y., Xu, D., Cheng, C.-Y., Sabanayagam, C., Tham, Y.-C., Yu, M., Rim, T. H., Chai, C. Y., Gopinath, B., & Mitchell, P. J. N. B. E. (2020) A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. 1-11.

Commandeur, F., Goeller, M., Betancur, J., Cadet, S., Doris, M., Chen, X., Berman, D. S., Slomka, P. J., Tamarappoo, B. K., & Dey, D. J. I. t. o. m. i. (2018) Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast CT. *37*(8), 1835-1846.

David, H. B. F. IMPACT OF ENSEMBLE LEARNING ALGORITHMS TOWARDS ACCURATE HEART DISEASE PREDICTION.

Demšar, J., Cu rk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., & Starič, A. J. t. J. o. m. L. r. (2013) Orange: data mining toolbox in Python. *14*(1), 2349-2353.

Ellis, G. K., Robinson, J. A., & Crawford, G. B. J. A. f. p. (2006) When symptoms of disease overlap with symptoms of depression. *35*(8).

Escárcega, R. O., Lipinski, M. J., Garcia-Carrasco, M., Mendoza-Pinto, C., Galvez-Romero, J. L., & Cervera, R. J. A. r. (2018) Inflammation and atherosclerosis: cardiovascular evaluation in patients with autoimmune diseases. *17*(7), 703-708.

Frank, E., Trigg, L., Holmes, G., & Witten, I. H. J. M. L. (2000) Naive Bayes for regression. *41*(1), 5-25.

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., & Fox, C. S. J. C. (2013) Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *127*(1), e6-e245.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1): MIT press Cambridge.

Grotzinger, J. P., Crisp, J., Vasavada, A. R., Anderson, R. C., Baker, C. J., Barry, R., Blake, D. F., Conrad, P., Edgett, K. S., & Ferdowski, B. J. S. s. r. (2012) Mars Science Laboratory mission and science investigation. *170*(1), 5-56.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.* Paper presented at the Proceedings of the IEEE international conference on computer vision.

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. J. J. o. d. i. (2019) Deep learning techniques for medical image segmentation: achievements and challenges. *32*(4), 582-596.

Hsu, C.-S., Wen, S.-H., Hung, J.-S., Liu, T.-T., Yi, C.-H., Lei, W.-Y., Pace, F., Chen, C.-L. J. D. d., & sciences. (2017) Overlap of dyspepsia in patients with gastroesophageal reflux disease: impact of clinical, metabolic, and psychosocial characteristics. *62*(4), 994-1001.

Hung, D.-Z., Yang, H.-J., Li, Y.-F., Lin, C.-L., Chang, S.-Y., Sung, F.-C., & Tai, S. C. J. P. o. (2015) The long-term effects of organophosphates poisoning as a risk factor of CVDs: a nationwide population-based cohort study. *10*(9), e0137632.

Ingre, B., & Yadav, A. (2015). *Performance analysis of NSL-KDD dataset using ANN.* Paper presented at the 2015 international conference on signal processing and communication engineering systems.

Isin, A., & Ozdalili, S. J. P. c. s. (2017) Cardiac arrhythmia detection using deep learning. *120*, 268-275.

Jonas, E., & Kording, K. P. J. P. c. b. (2017) Could a neuroscientist understand a microprocessor? *, 13*(1), e1005268.

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. J. J. (2011) Risk prediction models for hospital readmission: a systematic review. *306*(15), 1688-1698.

Karimi-Bidhendi, S., Arafati, A., Cheng, A. L., Wu, Y., Kheradvar, A., & Jafarkhani, H. J. J. o. C. M. R. (2020) Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *22*(1), 1-24.

Kohane, I. S. J. N. R. G. (2011) Using electronic health records to drive discovery in disease genomics. *12*(6), 417-428.

Kording, K. P., Benjamin, A., Farhoodi, R., & Glaser, J. I. (2018). *The roles of machine learning in biomedical science.* Paper presented at the Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2017 Symposium.

Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., Baber, U., Min, J. K., Tang, W. W., Halperin, J. L., & Narayan, S. M. J. E. h. j. (2019) Deep learning for cardiovascular medicine: a practical primer. 40(25), 2058-2073.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. J. J. o. t. A. C. o. C. (2017) Artificial intelligence in precision cardiovascular medicine. *69*(21), 2657-2664.

Kwon, J. m., Kim, K. H., Jeon, K. H., & Park, J. J. E. (2019) Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *36*(2), 213-218.

Lai, M. J. a. p. a. (2015) Deep learning for medical image segmentation.

Landahl, H., McCulloch, W. S., & Pitts, W. J. T. b. o. m. b. (1943) A statistical consequence of the logical calculus of nervous nets. *5*(4), 135-137.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. J. P. o. t. I. (1998) Gradient-based learning applied to document recognition. *86*(11), 2278-2324.

Lee, N., & Lings, I. (2008). *Doing business research: a guide to theory and practice*: Sage.

Levin, I., & Stokes, J. P. J. J. o. a. p. (1989) Dispositional approach to job satisfaction: Role of negative affectivity. *74*(5), 752.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. J. M. i. a. (2017) A survey on deep learning in medical image analysis. *42*, 60-88.

McCulloch, W. S., & Pitts, W. J. T. b. o. m. b. (1943) A logical calculus of the ideas immanent in nervous activity. *5*(4), 115-133.

Members, A. T. F., Elliott, P. M., Anastasakis, A., Borger, M. A., Borggrefe, M., Cecchi, F., Charron, P., Hagege, A. A., Lafont, A., & Limongelli, G. J. E. h. j. (2014) 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *35*(39), 2733-2779.

Mensah, G. A., & Collins, P. Y. J. G. h. (2015) Understanding mental health for the prevention and control of cardiovascular diseases. *10*(3), 221.

Milberger, S., Biederman, J., Faraone, S. V., Murphy, J., & Tsuang, M. T. J. T. A. j. o. p. (1995) Attention deficit hyperactivity disorder and comorbid disorder: Issues of overlapping symptoms.

Mirnezami, R., Nicholson, J., & Darzi, A. J. N. E. J. M. (2012) Preparing for precision medicine. *366*(6), 489-491.

Nature, 摇. J. (2015) Y, BENGIO Y, HINTON G. Deep learning. *521*(7553), 436-444.

Norris, R. M., White, H. D., Cross, D. B., Wild, C. J., & Whitlock, R. M. J. E. h. j. (1992) Prognosis after recovery from myocardial infarction: the relative importance of cardiac dilatation and coronary stenoses. *13*(12), 1611-1618.

Olier, I., & Vellido, A. J. N. n. (2008) Advances in clustering and visualization of time series using GTM through time. *21*(7), 904-913.

Organization, W. H. (2013). *Transforming and scaling up health professionals' education and training: World Health Organization guidelines 2013*: World Health Organization.

Oulhaj, A., Bakir, S., Aziz, F., Suliman, A., Almahmeed, W., Sourij, H., & Shehab, A. (2020) Agreement between cardiovascular disease risk assessment tools: An application to the United Arab Emirates population. *PLoS One, 15*(1), e0228031.

Oulhaj A, S. A. (2019) Data set from agreement between CVD risk assessment tools in UAE study.

Peili, Y., Xuezhen, Y., Jian, Y., Lingfeng, Y., Hui, Z., & Jimin, L. (2018). *Deep learning model management for coronary heart disease early warning research.* Paper presented at the 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA).

Pham, T., Tran, T., Phung, D., & Venkatesh, S. J. J. o. b. i. (2017) Predicting healthcare trajectories from medical records: A deep learning approach. *69*, 218-229.

Pittoli, F., Vianna, H. D., Barbosa, J. L. V., Butzen, E., Gaedke, M. Â., da Costa, J. S. D., dos Santos, R. B. S. J. T., & Informatics. (2018) An intelligent system for prognosis of noncommunicable diseases' risk factors. *35*(5), 1222-1236.

Platform, P. M. (2021) American Heart Association.

Polson, N. G., & Sokolov, V. O. J. W. S. S. R. O. (2014) Deep learning. 1-12.

Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. J. N. B. E. (2018) Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *2*(3), 158-164.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., & Nielsen, M. (2013). *Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network.* Paper presented at the International conference on medical image computing and computer-assisted intervention.

Psaty, B. M., Furberg, C. D., Kuller, L. H., Cushman, M., Savage, P. J., Levine, D., O'Leary, D. H., Bryan, R. N., Anderson, M., & Lumley, T. J. A. o. i. m. (2001) Association between blood pressure level and the risk of myocardial infarction, stroke, and total mortality: the cardiovascular health study. *161*(9), 1183-1192.

Quinlan, J. R. J. M. l. (1986) Induction of decision trees. *1*(1), 81-106.

Rajkumar, A., Ganesan, M., & Lavanya, R. (2019). *Arrhythmia classification on ECG using Deep Learning.* Paper presented at the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).

Rish, I. (2001). *An empirical study of the naive Bayes classifier.* Paper presented at the IJCAI 2001 workshop on empirical methods in artificial intelligence.

Rish, I., Hellerstein, J., & Thathachar, J. J. I. T. W. R. C. (2001) An analysis of data characteristics that affect naive Bayes performance. *30*, 1-8.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., & Roberts, D. J. R. S. o. E. (2008) Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *112*(5), 2272-2283.

Romiti, S., Vinciguerra, M., Saade, W., Anso Cortajarena, I., Greco, E. J. C. R., & Practice. (2020) Artificial Intelligence (AI) and Cardiovascular Diseases: An Unexpected Alliance. *2020*.

Rosengren, A., Hawken, S., Ôunpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A., Blackett, K. N., Sitthi-Amorn, C., Sato, H., & Yusuf, S. J. T. L. (2004) Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *364*(9438), 953-962.

Ruangkanokmas, P., Achalakul, T., & Akkarajitsakul, K. (2016). *Deep belief networks with feature selection for sentiment classification.* Paper presented at the 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS).

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. J. H. b. m. (2017) Deep learning with convolutional neural networks for EEG decoding and visualization. *38*(11), 5391-5420.

Schlesinger, D. E., & Stultz, C. M. J. C. T. O. i. C. M. (2020) Deep learning for cardiovascular risk stratification. *22*(8), 1-14.

Schmidhuber, J. J. N. n. (2015) Deep learning in neural networks: An overview. *61*, 85-117.

Seetharam, K., Kagiyama, N., Sengupta, P. P. J. E. r., & practice. (2019) Application of mobile health, telemedicine and artificial intelligence to echocardiography. *6*(2), R41-R52.

Shen, D., Wu, G., & Suk, H.-I. J. A. r. o. b. e. (2017) Deep learning in medical image analysis. *19*, 221-248.

Silberberg, J. S., Barre, P. E., Prichard, S. S., & Sniderman, A. D. J. K. i. (1989) Impact of left ventricular hypertrophy on survival in end-stage renal disease. *36*(2), 286-290.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., & Lanctot, M. J. n. (2016) Mastering the game of Go with deep neural networks and tree search. *529*(7587), 484-489.

Srivastava, K., Kumar Choubey, D., & Kumar, J. J. A. a. S. (2020) Implementation of Inventory Management System.

Thomas, H., Diamond, J., Vieco, A., Chaudhuri, S., Shinnar, E., Cromer, S., Perel, P., Mensah, G. A., Narula, J., & Johnson, C. O. (2018). Global atlas of cardiovascular disease 2000-2016: the path to prevention and control.

Ulbricht, T., & Southgate, D. J. T. l. (1991) Coronary heart disease: seven dietary factors. *338*(8773), 985-992.

Usama, M., Ahmad, B., Wan, J., Hossain, M. S., Alhamid, M. F., & Hossain, M. A. J. I. A. (2018) Deep feature learning for disease risk assessment based on convolutional neural network with intra-layer recurrent connection by using hospital big data. *6*, 67927-67939.

van den Tempel, N., Odijk, H., van Holthe, N., Naipal, K., Raams, A., Eppink, B., van Gent, D. C., Hardillo, J., Verduijn, G. M., & Drooger, J. C. J. I. J. o. H. (2018) Heat-induced BRCA2 degradation in human tumours provides rationale for hyperthermia-PARP-inhibitor combination therapies. *34*(4), 407-414.

Wang, J., Ding, H., Bidgoli, F. A., Zhou, B., Iribarren, C., Molloi, S., & Baldi, P. J. I. t. o. m. i. (2017) Detecting cardiovascular disease from mammograms with deep learning. *36*(5), 1172-1181.

Webb, G. I. J. E. o. m. l. (2010) Naïve Bayes. *15*, 713-714.

Wilson, A. G., Hu, Z., Salakhutdinov, R., & Xing, E. P. (2016). *Deep kernel learning.* Paper presented at the Artificial intelligence and statistics.

Wilson, A. G., & Izmailov, P. J. a. p. a. (2020) Bayesian deep learning and a probabilistic perspective of generalization.

Wilson, J. (2014). *Essentials of business research: A guide to doing your research project*: Sage.

Wu, Y., Benjamin, E. J., & MacMahon, S. J. C. (2016) Prevention and control of cardiovascular disease in the rapidly changing economy of China. *133*(24), 2545-2560.

Yan, H., Jiang, Y., Zheng, J., Peng, C., & Li, Q. J. E. S. w. A. (2006) A multilayer perceptron-based medical decision support system for heart disease diagnosis. *30*(2), 272-281.

Yan, Y., Zhang, J.-W., Zang, G.-Y., & Pu, J. J. J. o. g. c. J. (2019) The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine? *, 16*(8), 585.

Yanagimoto, H., Shimada, M., & Yoshimura, A. (2013). *Document similarity estimation for sentiment analysis using neural network.* Paper presented at the 2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS).

Yoo, Y., Tang, L. Y., Brosch, T., Li, D. K., Kolind, S., Vavasour, I., Rauscher, A., MacKay, A. L., Traboulsee, A., & Tam, R. C. J. N. C. (2018) Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *17*, 169-178.

Zhang, J., Li, B., Xiang, K., & Shi, X. J. a. p. a. (2019) Method of diagnosing heart disease based on deep learning ECG signal.

Zhang, Q., Zhou, D., & Zeng, X. (2017). *PulsePrint: Single-arm-ECG biometric human identification using deep learning.* Paper presented at the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON).

Zhu, J., Shen, B., Abbasi, A., Hoshmand-Kochi, M., Li, H., & Duong, T. Q. J. P. o. (2020) Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *15*(7), e0236621.

Zoni-Berisso, M., Lercari, F., Carazza, T., & Domenicucci, S. J. C. e. (2014) Epidemiology of atrial fibrillation: European perspective. *6*, 213.

APPENDIX A

LIST OF PUBLICATIONS AND PRESENTATIONS

Najwa, F. A., Akram, M. Z., & Asadullah S. (2021). Deep Learning Model for Predicting and Detecting Overlapping Symptoms of Cardiovascular Diseases in Hospitals of UAE. Turkish Journal of Computer and Mathematics Education. Vol. 12 No. 14(2021). E-ISSN: 5212-5224.

Najwa, F. A., Akram, M. Z., & Asadullah S. (2022). Image Dehazing Using Deep Learning Approach. 4th International Conference on Communication Engineering and Computer Science (CIC-COCOS'2022).

Najwa, F. A., Syeda S. M., Akram, M. Z., & Asadullah S. (2023). Deep Learning Approach for Bone Marrow Cell Detection and Classification on Whole Side Images. IEEE International Conference on Engineering Technologies and Applied Science (ICETAS).

Najwa, F. A., Syeda S. M., Akram, M. Z., & Asadullah S. (2023). Overlapped Symptoms Detection for Cardiovascular Disease Based on Deep Learning Model IEEE International Conference on Engineering Technologies and Applied Science (ICETAS).

APPENDIX B

10th-FOLD VALIDATION RESULTS

Data Point	Predictor	Coefficient	St. Error	Sig.	R2
01 to 21	(Constant)	.121	.182	.508	0.914
	Emptysis	.128	.049	.010	
	Learnability	.408	.045	.000	
	Fever	-.097	.030	.001	
	Cyanosis	-.125	.030	.000	
	Weakness and Fatigue	.190	.030	.000	
	Discomfort Pressure in Chest	.111	.034	.001	
	Chest Pain	.385	.051	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
22 to 42	(Constant)	.079	.172	.648	0.915
	Emptysis	.102	.046	.027	
	Learnability	.433	.041	.000	
	Fever	-.073	.029	.013	
	Cyanosis	-.118	.027	.000	
	Weakness and Fatigue	.155	.030	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
43 to 63	(Constant)	.212	.184	.250	0.908
	Emptysis	.103	.050	.041	
	Learnability	.372	.045	.000	

	Fever	-.086	.031	.007	
	Cyanosis	-.155	.030	.000	
	Weakness and Fatigue	.201	.031	.000	
	Discomfort Pressure in Chest	.123	.035	.001	
	Chest Pain	.404	.052	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
64 to 84	(Constant)	.050	.190	.793	0.905
	Emptysis	.131	.051	.011	
	Learnability	.419	.046	.000	
	Fever	-.084	.032	.010	
	Cyanosis	-.128	.030	.000	
	Weakness and Fatigue	.179	.032	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
85 to 105	(Constant)	.089	.181	.624	0.911
	EMPT	.150	.049	.002	
	DYSP	.383	.046	.000	
	FEVE	-.087	.031	.005	
	CYAN	-.133	.030	.000	
	WFAT	.212	.032	.000	
	DPCH	.120	.035	.001	
	CHEP	.365	.052	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
106 to 126	(Constant)	-.001	.182	.995	0.914
	EMPT	.168	.052	.002	
	DYSP	.401	.047	.000	
	FEVE	-.067	.031	.030	
	CYAN	-.118	.029	.000	
	WFAT	.190	.031	.000	
	DPCH	.101	.033	.003	
	CHEP	.372	.051	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
127 to 147	(Constant)	-.004	.179	.984	0.909
	EMPT	.176	.052	.001	
	DYSP	.377	.046	.000	
	FEVE	-.079	.032	.014	
	CYAN	-.118	.030	.000	
	WFAT	.196	.033	.000	
	DPCH	.123	.034	.000	
	CHEP	.368	.052	.000	

Data Point	Predictor	Coefficient	Std.Error	Sig.	R2
188-200	(Constant)	.081	.180	.655	0.910
	EMPT	.206	.053	.000	
	DYSP	.361	.045	.000	
	FEVE	-.076	.038	.049	

	CYAN	-.131	.034	.000	
	WFAT	.194	.030	.000	
	DPCH	.118	.033	.001	
	CHEP	.343	.054	.000	