MUSTAFA AHMED

INTEGRATION OF DIAGNOSTIC TECHNOLOGY FOR HEART DISEASE (MYOCARDIAL INFARCTION) USING ARTIFICIAL INTELLIGENCE TO IMPROVE CARDIOVASCULAR HEALTHCARE

MASTER THESIS

2024

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF BIOMEDICAL ENGINEERING

INTEGRATION OF DIAGNOSTIC TECHNOLOGY FOR HEART DISEASE (MYOCARDIAL INFARCTION) USING ARTIFICIAL INTELLIGENCE TO IMPROVE CARDIOVASCULAR HEALTHCARE

M.Sc. THESIS

Mustafa Jamal AHMED

NICOSIA

NOVEMBER, 2024

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF BIOMEDICAL ENGINEERING

INTEGRATION OF DIAGNOSTIC TECHNOLOGY FOR HEART DISEASE (MYOCARDIAL INFARCTION) USING ARTIFICIAL INTELLIGENCE TO IMPROVE CARDIOVASCULAR HEALTHCARE

M.Sc. THESIS

Mustafa Jamal AHMED

SUPERVISOR

ASSOC.PROF.DR. ABDULLAHI UMAR IBRAHIM
CO-SUPERVISOR
ASSOC.PROF.DR. SULEYMAN ASIR

Nicosia

NOVEMBER, 2024

Approval

We certify that we have read the thesis submitted by **MUSTAFA JAMAL AHMED AHMED** titled **"INTEGRATION OF DIAGNOSTIC TECHNOLOGY FOR HEART DISEASE (MYOCARDIAL INFARCTION) USING ARTIFICIAL INTELLIGENCE TO IMPROVE CARDIOVASCULAR HEALTHCARE"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Master of Biomedical Engineering.

Examining Committee	Name-Surname	Signature
Head of the Committee:	Asst. Prof. Dr. OMID MIRZAEI	
Committee Member:	Asst. Prof. Dr. HUZAIFA UMAR	
Supervisor:	Assoc. Prof. Dr. Abdullahi Umar Ibrahim	
Co-supervisor:	Assoc. Prof. Dr. Süleyman AŞIR	

Approved by the Head of the Department

02 /12/2024

....

Assoc. Prof. Dr. Süleyman AŞIR

Head of Department

Approved by the Institute of Graduate Studies

02 /12/2024

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

Declaration and Ethical Concord

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Mustafa Jamal Ahmed

…. /12/2024

Acknowledgments

In the name of Allah, the Most Benevolent, the Most Merciful, I begin by expressing my immense gratitude to Allah, the One and Almighty Creator, for granting me the determination to complete this research study. I also want to extend my sincere appreciation and indebtedness to my supervisor, Supervisor **Assoc.Prof.Dr. Abdullahi Umar Ibrahim** and to my Co-supervisor **Assoc. Prof. Dr. Süleyman AŞIR** invaluable guidance, excellent ideas, and unwavering support throughout the years that have made this research possible. Despite their busy schedules and precious time, they were always available to offer me endless inspiration, kindness, and friendly advice during my research.

I would also like to express my gratitude to all the students and staff who assisted me in many ways during my research work, particularly those at the NEU. I am especially grateful for the support and encouragement of my friends and colleagues.

Lastly, I would like to give a special thanks to my father, **Mr. Jamal Ahmed**,

My mother, **Mrs. Athraa Azeez**, and all of my friends for their continued encouragement throughout my study. Without the support of my family, I would not be where I am today. It is my great honor to dedicate this thesis to my family

Mustafa Jamal Ahmed

Abstract

INTEGRATION OF DIAGNOSTIC TECHNOLOGY FOR HEART DISEASE (MYOCARDIAL INFARCTION) USING ARTIFICIAL INTELLIGENCE TO IMPROVE CARDIOVASCULAR HEALTHCARE

Ahmed Mustafa Jamal

MS.C, Department of Biomedical Engineering

November, 2024, 114 pages

Early diagnosis of diseases is of utmost importance, especially serious diseases that may lead to death, such as heart diseases. Diagnosis of diseases such as heart diseases, especially myocardial infarction, has been a major challenge for the health care sector in many countries, especially developing countries, where diagnostic tools and factors are very limited. The methods used to diagnose heart diseases always depend on the doctor's experience. We may take into consideration that heart diseases are among the most difficult diseases to diagnose because the human heart is affected by the patient's psychological state as well. The patient may be stressed and the patient's readings may be very high, such as high blood pressure or unstable heartbeats. This may lead to a wrong diagnosis, especially taking into account the experience of the doctor who diagnoses the case.

These reasons have become one of the motivations for developing diagnostic tools, especially early diagnosis, to find quick solutions, as the healthcare sector resorts to the most experienced doctors to diagnose complex cases such as heart diseases. This may be a limited number of people with experience in healthcare, which causes a delay in diagnosis and resorting to sending cases to different places for a more accurate diagnosis.

This challenge can be solved by using computer detection methods through pre-trained models based on artificial intelligence, which learn from the characteristics of patients with previous diseases and healthy patients to distinguish between new patients whether they are healthy or infected with one of the diseases

In this paper, we describe the pathological condition of myocardial infarction by creating pre-trained models to make a more accurate diagnosis. We trained the models using two types of data. The first type is image data that includes ECG images

that can be obtained from the emergency department upon immediate admission of the patient, which includes patients with myocardial infarction and ECG images of healthy patients who do not suffer from any medical disease. The second type that we trained the models on is patient symptom data such as the patient's age, gender, whether he suffers from chest pain, resting blood pressure, blood cholesterol, blood sugar, highest heart rate, and Number of major vessels. Where we trained the patient's symptoms on 5 different algorithms (RandomFores, SVC, KNN, Decision Tree, XGBoost) and we also trained the ECG images on 5 algorithms (RESNET 18 ALEXNET, LENET, CNN, Hybrid Model, CNN+SVM) And on this basis, a comparison is made between the models and which model performs better than the other.

The data was obtained through the patient's case during his admission to the hospital from the Kaggle website. The digital data contains 14 columns for the patient's label or symptoms and 1000 rows representing the number of patients, and the image data (electrocardiogram images) contains 10506 images for patients with myocardial infarction and 4046 for healthy patients, i.e. the total number of images is 14552 images.

For the classification of patient symptoms data (NUMERICAL DATASET (based on the number of complete data where classification was done using different training and testing ratios, First classification 20% for testing and 80% for training, Second classification 30% for testing and 70% for training, Third classification 40% for testing 60% for training

The model was evaluated based on several criteria (Accuracy, Sensitivity, Specificity, F1 score). Accuracy The model achieved the first classification to accuracy (98% 98.4% 97.3% 96.1% 97.5%), The second model achieved accuracy (99% 98.3% 97.1% 97.7% 97%), The third model (96.8% 98.2% 97.2% 96% 96.8%)

The model achieved Sensitivity, In the first model (99.1% 99.1% 95.7% 98.3% 99.1), And the second model (99.4% 98.8% 97.5% 98.8% 96.9%), The third model (97.7% 99.1% 96.8% 95.5% 98.2%), And the model also achieved in the performance evaluation Specificity, The first model (96.4 95.2% 97.6% 94% 95.2), The second model (97.8% 97.1% 97.1% 96.4% 97.1%) The third model (95.5% 95.5% 96.6% 96.6% 94.9%).

And the model was also evaluated on the F1 score, And the first model achieved (99.2% 97.9% 97% 97.1% 97.9%), Model 2 (98.8% 98.2% 96.6% 97.2% 97.2%), Model 3 (97.1% 97.8% 97.1% 94.8% 97.1%)

To classify the ECG images, the data split was used as follows (70% training and 30% testing), No of epochs 10 and Batch Size 32 were used. The models were trained on 5 different algorithms (RESNET 18, ALEXNET, LENET, CNN, Hybrid Model, CNN+SVM)

The model was evaluated based on (Training ratio, Batch Size, No of epochs, Accuracy test, Accuracy train, Sensitivity Specificity, Achieved mean square error), The model achieved Accuracy test (99% 95.89 98.79 98% 98), The highest accuracy was obtained from RESNET 18, The model achieved Accuracy train (99% 96.79 99 97.67% 97.8) also the highest value obtained from RESNET 18, In Sensitivity (1 94.80 98.39 97.01% 98.29%). And in Specificity (1 96.31 99.05 98.78% 97.03%). Specificity (1 96.31 99.05 98.78% 97.03%),Achieved mean square error (0 0.0411 0.0001278 0.0171 0.03) Overall, RESNET 18 is the best model in terms of performance evaluation values for the model

Our findings support the idea that medical picture classification using RESNET 18 models can be done more precisely and accurately. These models can now be used as a myocardial infarction diagnosis confirmation system, increasing the number of missed diagnoses and providing a substitute to lessen the burdensome and tiresome task that hospital cardiologist's face.

Key Words: myocardial infarction, Artificial Intelligence, Machine Learning, deep learning, Healthcare

ÖZET

KALP HASTALIKLARININ (MİYOKARD İNFARKTÜSÜ) TEŞHİS TEKNOLOJİSİNİN YAPAY ZEKA KULLANILARAK ENTEGRE EDİLMESİ VE KARDİYOVASKÜLER SAĞLIK HİZMETLERİNİN İYİLEŞTİRİLMESİ

Ahmed Mustafa Jamal

Yüksek Lisans, Biyomedikal Mühendisliği Bölümü

Kasım 2024, 114 Sayfa

Hastalıkların, özellikle ölüme yol açabilecek kalp hastalıkları gibi ciddi hastalıkların erken teşhisi son derece önemlidir. Kalp hastalıklarının, özellikle miyokard enfarktüsünün teşhisi, teşhis araçlarının ve faktörlerin çok sınırlı olduğu birçok ülkede, özellikle gelişmekte olan ülkelerde sağlık sektörü için büyük bir zorluk teşkil etmiştir. Kalp hastalıklarının teşhisinde kullanılan yöntemler her zaman doktorun deneyimine bağlıdır. Kalp hastalıklarının teşhisinin en zor hastalıklardan biri olduğunu göz önünde bulundurabiliriz, çünkü insan kalbi hastanın psikolojik durumundan da etkilenmektedir. Hasta stresli olabilir ve hastanın ölçümleri, yüksek tansiyon veya düzensiz kalp atışları gibi çok yüksek olabilir. Bu durum, özellikle teşhis eden doktorun deneyimi dikkate alındığında yanlış bir teşhise yol açabilir.

Bu nedenler, özellikle erken teşhis için teşhis araçlarının geliştirilmesi yönünde motivasyon sağlamıştır. Sağlık sektörü, kalp hastalıkları gibi karmaşık vakaların teşhisi için en deneyimli doktorlara başvurmaktadır. Ancak bu, sağlık alanında deneyimli insan sayısının sınırlı olması nedeniyle teşhiste gecikmelere ve vakaların daha doğru bir teşhis için farklı yerlere gönderilmesine neden olabilir. Bu zorluk, geçmiş hastalık vakalarından ve sağlıklı hastalardan öğrenen, önceden eğitilmiş yapay zeka tabanlı modeller kullanılarak bilgisayar destekli tespit yöntemleri ile çözülebilir. Bu yöntemler, yeni hastaların sağlıklı mı yoksa bir hastalık ile mi enfekte olduğunu ayırt edebilir.

Bu çalışmada, miyokard enfarktüsünün patolojik durumunu daha doğru bir teşhis yapmak için önceden eğitilmiş modeller oluşturarak tanımlıyoruz. Modelleri iki tür veri kullanarak eğittik. İlk veri türü, hastanın acil servise kabulünde elde edilebilecek EKG görüntülerini içeren ve miyokard enfarktüsü olan hastaların yanı sıra herhangi bir hastalığı olmayan sağlıklı hastaların EKG görüntülerini içeren görüntü

verisidir. Modelleri eğittiğimiz ikinci veri türü ise hastanın yaşı, cinsiyeti, göğüs ağrısı olup olmadığı, dinlenme kan basıncı, kan kolesterolü, kan şekeri, en yüksek kalp hızı ve ana damar sayısı gibi hastanın semptom verileridir. Hastanın semptomlarını beş farklı algoritma (Random Forest, SVC, KNN, Karar Ağacı, XGBoost) ile eğittik ve EKG görüntülerini beş algoritma (RESNET 18, ALEXNET, LENET, CNN, Hybrid Model, CNN+SVM) ile eğittik. Bu doğrultuda modeller arasında karşılaştırma yapılmış ve hangi modelin diğerlerinden daha iyi performans gösterdiği değerlendirilmiştir.

Veriler, hastanın hastaneye kabulü sırasında Kaggle web sitesinden alınmıştır. Dijital veriler, hasta etiketi veya semptomları için 14 sütun ve hasta sayısını temsil eden 1000 satır içermektedir. Görüntü verisi (elektrokardiyogram görüntüleri) ise 10.506 miyokard enfarktüsü olan hasta görüntüsü ve 4046 sağlıklı hasta görüntüsü içeriyor, yani toplam görüntü sayısı 14.552'dir.

Hasta semptom verilerinin sınıflandırılması için (TAM SAYISAL VERİ SETİ), farklı eğitim ve test oranlarına göre sınıflandırma yapılmıştır:
İlk sınıflandırma: %20 test ve %80 eğitim için, İkinci sınıflandırma: %30 test ve %70 eğitim için. Üçüncü sınıflandırma: %40 test ve %60 eğitim için
Model, birkaç kriter temelinde değerlendirilmiştir (Doğruluk, Duyarlılık, Özgüllük, F1 skoru). İlk sınıflandırmada model doğruluk oranları (98%, 98.4%, 97.3%, 96.1%, 97.5%) elde etti. İkinci sınıflandırmada doğruluk oranları (99%, 98.3%, 97.1%, 97.7%, 97%) ve üçüncü sınıflandırmada ise (96.8%, 98.2%, 97.2%, 96%, 96.8%) oranları elde edilmiştir.

EKG görüntülerinin sınıflandırılması için veriler %70 eğitim ve %30 test oranına göre ayrılmış, 10 epoch ve 32 batch size kullanılarak eğitilmiştir. Modeller beş farklı algoritma (RESNET 18, ALEXNET, LENET, CNN, Hybrid Model, CNN+SVM) üzerinde eğitilmiştir. Model (Eğitim oranı, Batch Size, Epoch sayısı, Test doğruluğu, Eğitim doğruluğu, Duyarlılık, Özgüllük, Ortalama kare hatası) temelinde değerlendirilmiştir.

Genel olarak, RESNET 18, modelin performans değerlendirme değerleri açısından en iyi modeldir. Sonuçlarımız, RESNET 18 modellerinin yüksek doğruluk ve hassasiyetle tıbbi görüntülerin sınıflandırılmasında kullanılabileceği görüşüyle uyumludur. Bu modeller, miyokard enfarktüsü teşhisinde bir doğrulama sistemi olarak hizmet edebilir, yanlış teşhisleri en aza indirir ve hastanelerde Kardiyologların karşılaştığı yoğun iş yükünü hafifletmek için bir alternatif sunabilir.

Anahtar Kelimeler: miyokard enfarktüsü, Yapay Zekâ, Makine Öğrenimi, Derin Öğrenme, Sağlık Hizmetleri

Table of Contents

Contents

List Of Tables

Chapter II

Chapter III

Chapter IV

List Of Figures

Chapter I

Chapter II

Chapter III

Chapter IV

List of Abbreviations

Abbreviation	Full Form
ECG	Electrocardiography
ML	Machine Learning
DL	Deep Learning
DT	Decision Trees
RF	Random Forest
SVM	Support Vector Machines
KNN	k-nearest neighbors
MLP	Multi-Layer Perceptron
CNN	Convolutional Neural Networks
CDSS	Clinical Decision Support Systems
PCA	Principal Component Analysis
DWT	Discrete Wavelet Transform
ROC	Receiver Operating Characteristic
NB	Naive Bayes
MIT-BIH	Massachusetts Institute of Technology
CHF	Congestive Heart Failure
AF	Atrial Fibrillation
Ml	Myocardium
CVD	Cardio Vascular Disease
CCF	Congestive Cardiac Failure
AI	Artificial Intelligence
NN	Neural Network
AC	Accuracy
WHO	World Health Organization
ROC	Receiver Operating Characteristic
BP	Blood Pressure

Chapter I

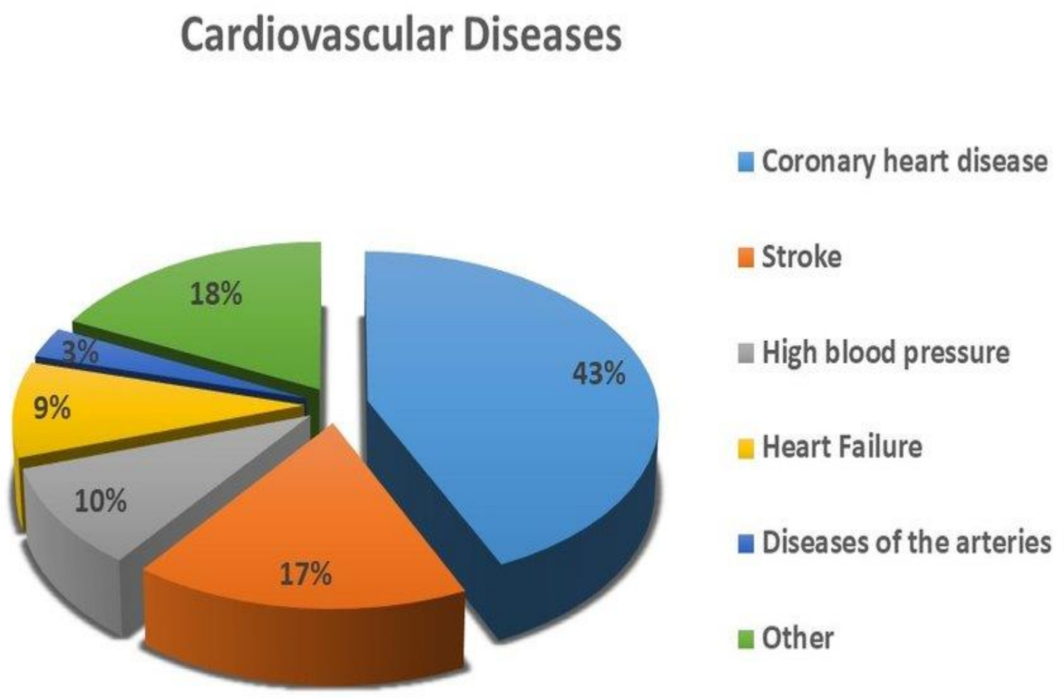
Introduction

1.0 Overview

Cardiovascular diseases (CVD) rank among the foremost causes of mortality worldwide. Prominent cardiovascular diseases (CVDs) include coronary heart disease, cerebrovascular illness, peripheral arterial disease, rheumatic heart disease, and congenital heart disease. The World Health Organization (WHO) reports that approximately 17.9 million individuals succumb annually to cardiovascular problems and heart disease. Over eighty percent of cardiovascular disease fatalities result from heart attacks and strokes. Unhealthy diets, insufficient physical activity, alcohol misuse, and tobacco consumption are potential risk factors that exacerbate cardiovascular problems. (Ahsan & Siddique, 2022)

Figure 1

Deaths Caused by Different Cardiovascular Diseases (Tasneem et al., 2021)



Cardiovascular disorders constitute the primary cause of mortality globally. In 2019, over 17.9 million individuals succumbed to cardiovascular illnesses, representing 32% of global mortality, with 85% of these fatalities resulting from heart attacks and strokes. Cardiovascular Diseases (CVDs), no date.

Lately, a lot of statistics, reports, and models. This dispersion provides a significant advantage, as it allows for real-time results in a variety of situations. Despite extensive research in all fields, there are many possibilities that could help save the lives of millions of people. Myocardial infarction (MI) It is the narrowing of one or more of the arteries that supply blood to the heart muscle) Myocardium (This results from the build-up of cholesterol and a blockage, and can lead to a (heart attack). (i.e., complete blockage). Which can obstruct blood flow. Myocardial infarction (MI) or stroke occurs when a blood clot obstructs blood flow, which can lead to a heart attack (American Heart Association, 2022).

Figure 2

Myocardial Disease (Jikui et al., 2021)

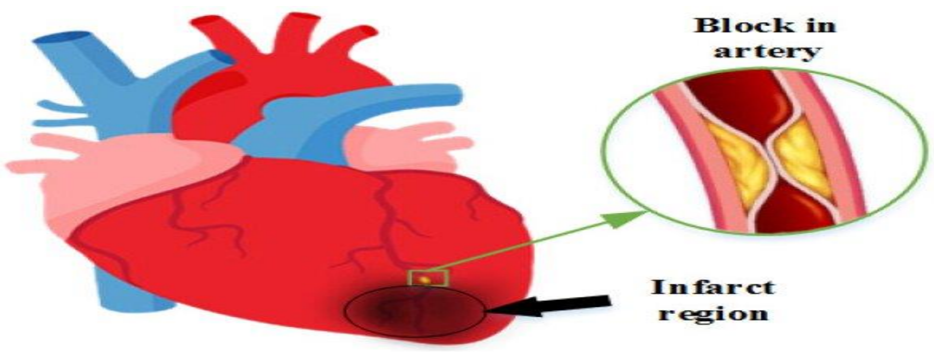


Figure 3

Type of Heart Disease (Set Different Types Heart Disease Collection Stock Vector (Royalty Free) 2468152079 | Shutterstock, n.d.)

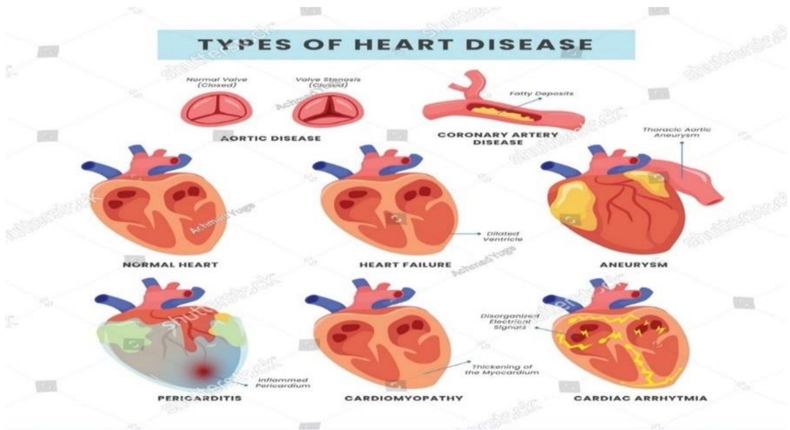
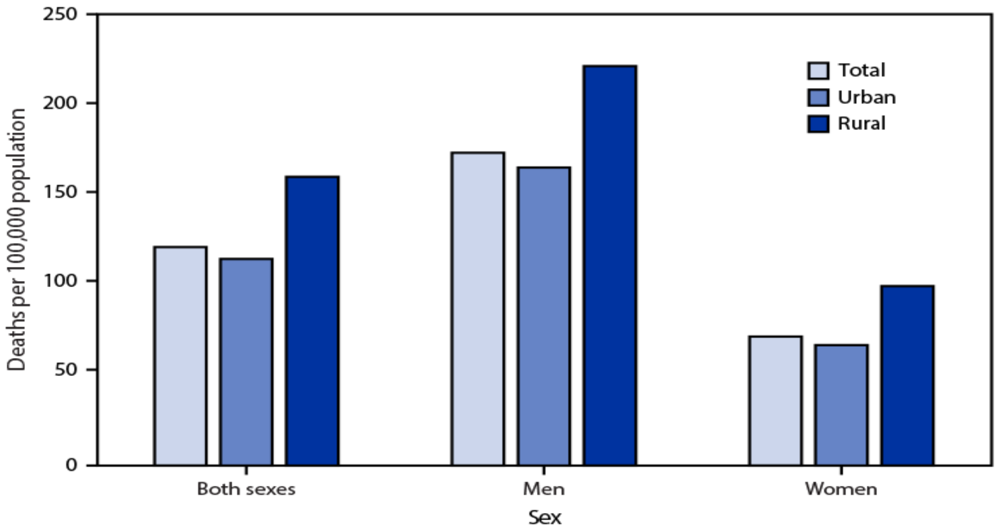


Figure 4

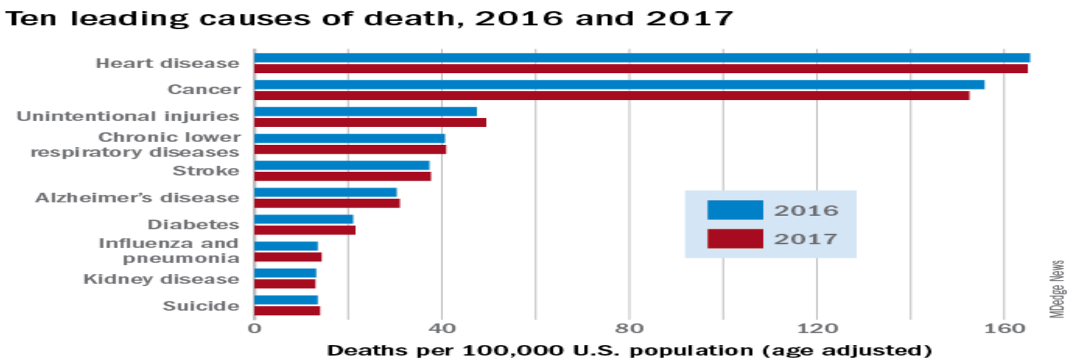
Age-Adjusted Heart Disease Death Rates by Sex and Urbanization Level in Adults Ages 45–64 (National Vital Statistics System, United States, 2019)



Heart disease is the most famous disease in the world that kills human lives in the world. In 2019, a predestined 17.9 million deaths were due to (ML), representing approximately 32% of all global deaths. It is noteworthy that 85% of these deaths were due to heart attacks and strokes, a large proportion of which occurred in low- and middle-income countries (WHO, 2022). To reduce the risk of heart attacks, lifestyle factors must be addressed, requiring patients to undergo vital tests such as cholesterol levels, electrocardiograms, blood pressure, and blood sugar levels. Analyzing these tests alongside existing patient data can be challenging for medical practitioners, especially when considering many complex factors (Anooj, 2022; Hedeshi & Abadeh, 2023).

Figure 5

In the United States, Heart Disease is Still the top Cause of Death. Depending on the illness, the number of deaths (Heart Disease Remains the Leading Cause of Death in U.S. | MDedge, n.d.)



Rapid medical intervention may be necessary in cases of heart attacks to prevent heart damage and save human lives. The use of artificial intelligence applications in medicine has increased in recent times, which has helped in early intervention and risk reduction (Ali & Mehdi, 2022). Advanced expert systems can predict a patient's likelihood of developing many heart diseases, which supports rapid and early intervention. Classification aims to identify patterns to predict categories of unknowns, which increases the accuracy of medical diagnoses (Vipul, 2023).

Situations where AI Enhances Diagnostic Precision Improving the accuracy of disease diagnoses via the development of a complex system allows for more prompt and precise medical treatment.

The technology helps doctors and nurses make quick, well-informed choices, which improves the quality of treatment patients get.

Using patient data and machine learning algorithms, predictive analysis can determine the probability of diseases, which allows for early treatment and prevention. (Mahmoud et al., 2022)

Figure 6

AI in Heart Disease (Armoundas et al., 2024)

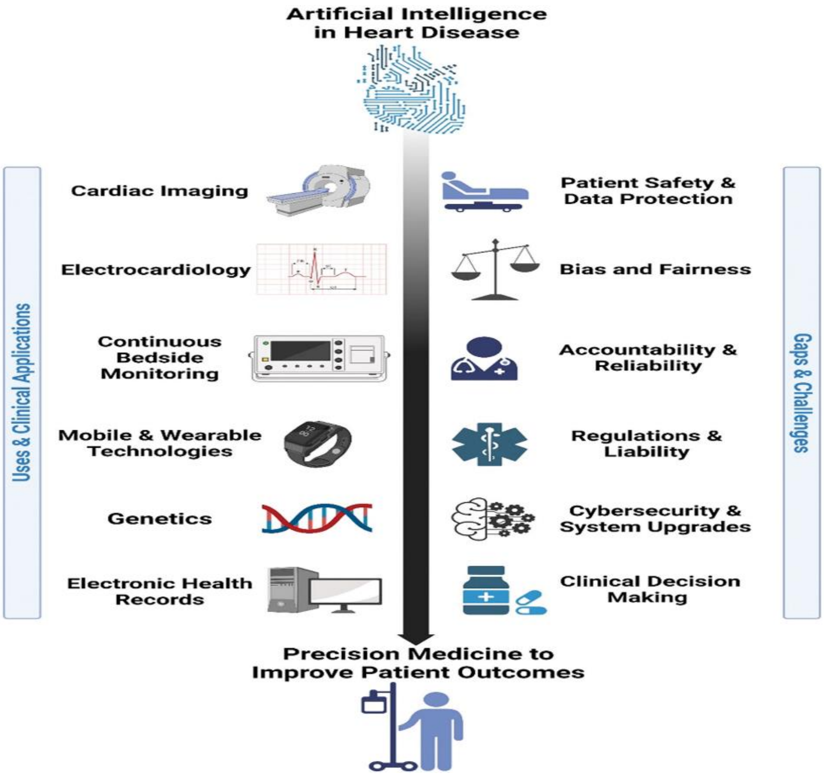
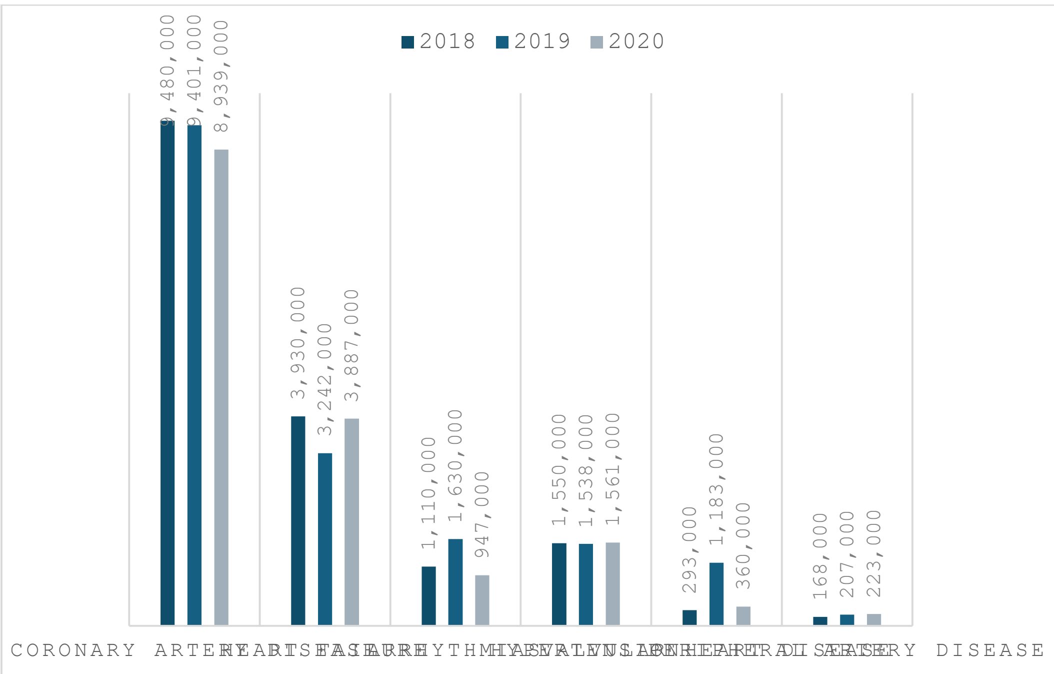


Figure 7

Cardiovascular Diseases Mentioned, According to the Global Burden of Disease Study (Cardiovascular diseases, n.d.)



Artificial intelligence systems such as genetic algorithms, neural networks, fuzzy logic, and neuro-fuzzy systems have revolutionized disease diagnosis by diagnosing and identifying intractable medical conditions (Zaptron, 2022). These systems combine the interpretation capabilities of fuzzy systems with the learning ability of neural networks, creating a hybrid intelligent system that takes advantage of the strengths of both technologies (Mehdi et al., 2023). This thesis aims to create an artificial intelligence system that relies on two types of data for diagnosis. For the patient, one type of data may be insufficient in accurately diagnosing any disease, as some symptoms are common to many diseases, especially heart disease.

In addition, the diagnosis of heart disease has many external influences, which makes the process of diagnosing the heart difficult, and it depends on integrating techniques. Many artificial intelligence systems, deep learning, and neural networks are used to obtain high accuracy in diagnosing heart diseases, which reduces the error rate. and the study highlights the potential of deep learning and machine learning techniques in transforming traditional diagnostic methods, offering a more objective and data-driven approach to medical assessments.

1.1 Problem Statement

Artificial intelligence and machine learning possess significant promise to improve healthcare, particularly in cardiovascular medicine. Despite comprehensive research in this domain, significant potential exists to leverage publicly accessible information and enormous data sets to enhance disease detection, risk assessment, and individualized treatment strategies. (Sultan et al. 2022).

Cardiovascular diseases (CVD):

Cardiovascular diseases rank among the most serious non-communicable diseases, substantially contributing to rising mortality rates. Reports indicate that these diseases caused around 17.9 million fatalities in 2019, accounting for almost one-third of all recorded deaths worldwide. Zheng et al., 2024. Cardiovascular disorders are intricately associated with other conditions, notably atherosclerosis, a disorder characterized by the accumulation of lipids and cholesterol on arterial walls, resulting in their constriction. This constriction might result in thrombosis, potentially leading to myocardial infarction or cerebrovascular accident, as it obstructs blood supply to essential tissues. The ramifications of these diseases extend beyond national boundaries, resulting in millions of fatalities each year globally, underscoring the pressing necessity to enhance prevention and intervention techniques to mitigate the incidence of these diseases and their detrimental effects on public health. (Münzel et al., 2020)

1.2 Challenges in Diagnosis and Treatment:

Appraising patients' case data and covering tests like cholesterol levels, ECGs, and blood pressure readings is a considerable challenge for healthcare providers. Making precise medical decisions in instances of MI Imposes fast response times since Timely medical intervention is critical to prevent damage and preserve lives.

Current diagnostic solutions, however, have challenges in properly integrating varied and complicated data sets. This necessitates the advancement of novel technologies that boost the data analysis process, facilitating precise diagnoses and more efficacious health care while strengthening physicians' capacity to make judgments based on inclusive and appropriate evidence. (Khera et al. 2024).

The role of Artificial intelligence:

AI and ML technologies provide powerful solutions by triggering the study of extensive datasets and assisting in the identification of patterns and biomarkers linked to cardiac disease. These technologies enhance the precision and efficacy of cardiac imaging analysis, thereby aiding clinical decision-making. AI-motivated predictive models improve risk evaluation, assisting physicians in formulating the best treatment strategies for patients. Through wide-ranging data analysis, these technologies can deliver novel insights that improve treatment outcomes and augment healthcare. (Adam, 2023)

Ethical and applied challenges:

The integration of AI technologies into healthcare is an Innovative step, but it poses a set of complex challenges that must be handled seriously. The most notable of these challenges is data privacy issues, as AI systems deal with vast amounts of patient information, which requires strict measures to protect this data from breach or unauthorized use.

In addition, there is a need to adhere to applicable regulations, as laws vary between countries concerning the use of health data. Healthcare organizations must ensure that AI technologies are compliant with these regulations to protect patients' rights and ensure that they are not exposed to additional risks.

Therefore, healthcare providers and developers should work to foster dialogue among various collaborators, including ethicists, regulators, and healthcare representatives, to ensure that AI is used in a way that benefits society and increases Dependence between patients and caregivers. Eventually, it is essential to develop a Wide-ranging framework that enhances the benefits of AI in healthcare while putting in place Controlled controls to mitigate potential risks, which contributes to improving the quality of healthcare and effectively enhancing patient outcomes. (Khera et al., 2024)

1.3 Aims and Objectives of the Study

1. Create a dynamic and flexible system for prediction and diagnosis of myocardial infarction (MI): Aims to develop a system that combines accuracy and flexibility in diagnosing the cases related to myocardial infarction, enhancing clinical decision-making.

2. Predicting myocardial infarction (MI): Available patient datasets are used to identify the most fitting algorithms used in diagnosis, which assist improve the accuracy of predicting the occurrence of myocardial infarction.

3. Creating predictive models: Ordinary features are inferred from approved datasets, which contributes to improving predictive analysis and developing accurate models that reflect the health condition of patients.

4. Aiding healthcare staff: The research aims to support medical staff and healthcare providers in making accurate and instant diagnostic decisions, which enhances Swift and accurate diagnosis of patients, which means avoiding human errors caused by

5.ECG signal analysis: DL techniques will be used to detect myocardial infarction cases by analyzing electrocardiogram data. This includes comparing and analyzing signal analysis algorithms to determine the best methods for training models.

6.Symptom Analysis: The research aims to develop machine learning algorithms to analyze symptoms associated with myocardial infarction, helping to enhance the ability to predict and diagnose this critical health condition early.

1.4 Significance of the Research

The reason for the importance of this study is it engages ML and DL techniques to diagnose and treat heart ailments, specifically myocardial infarction (MI).

Key considerations include:

Improving diagnostic accuracy: An intelligent and adaptive system that evaluates cardiac electrical signals (ECG) using DL methodologies can achieve the accuracy that the existing approaches for identifying myocardial infarction lack. This enables medical personnel to punctually mitigate the severity of harm from a missed MI diagnosis.

The approach enhances healthcare efficiency and reduces the probability of diagnostic errors by aiding medical professionals in making prompt and educated decisions based on real-time data analysis. Analyzing patient data with machine-learning algorithms enables the prediction of potential myocardial infarction (MI) cases, facilitating preemptive intervention.

Illness symptom analysis: Clinical symptoms may be very accurately analyzed using trained models to identify illness differentiating features, leading to improved clinical diagnosis.

Methods compared and analyzed: To improve the system's overall performance, the research compares and analyzes various methods and algorithms used to analyze ECG data, and then uses the most efficient ones to train the model.

1.5 THESIS Structure

This thesis is divided into five chapters, the majority of which are devoted to the design, optimization, the myocardial disease (ML) by using ai concept. The thesis structure is organized as follows.

In this chapter (Chapter 1), we have explained an overview and Introduction of the main topic and clarified the research problem, the aim, the object, and the method of achieving these goals.

In Chapter 2, review the history of MLT and DL technologies and the heart disease (ml) state of the art in rectenna designs for these purposes through a comprehensive literature review. It deals with two important sections: researchers who used ECG images, researchers who used numerical data (symptoms), and researchers who used both techniques together.

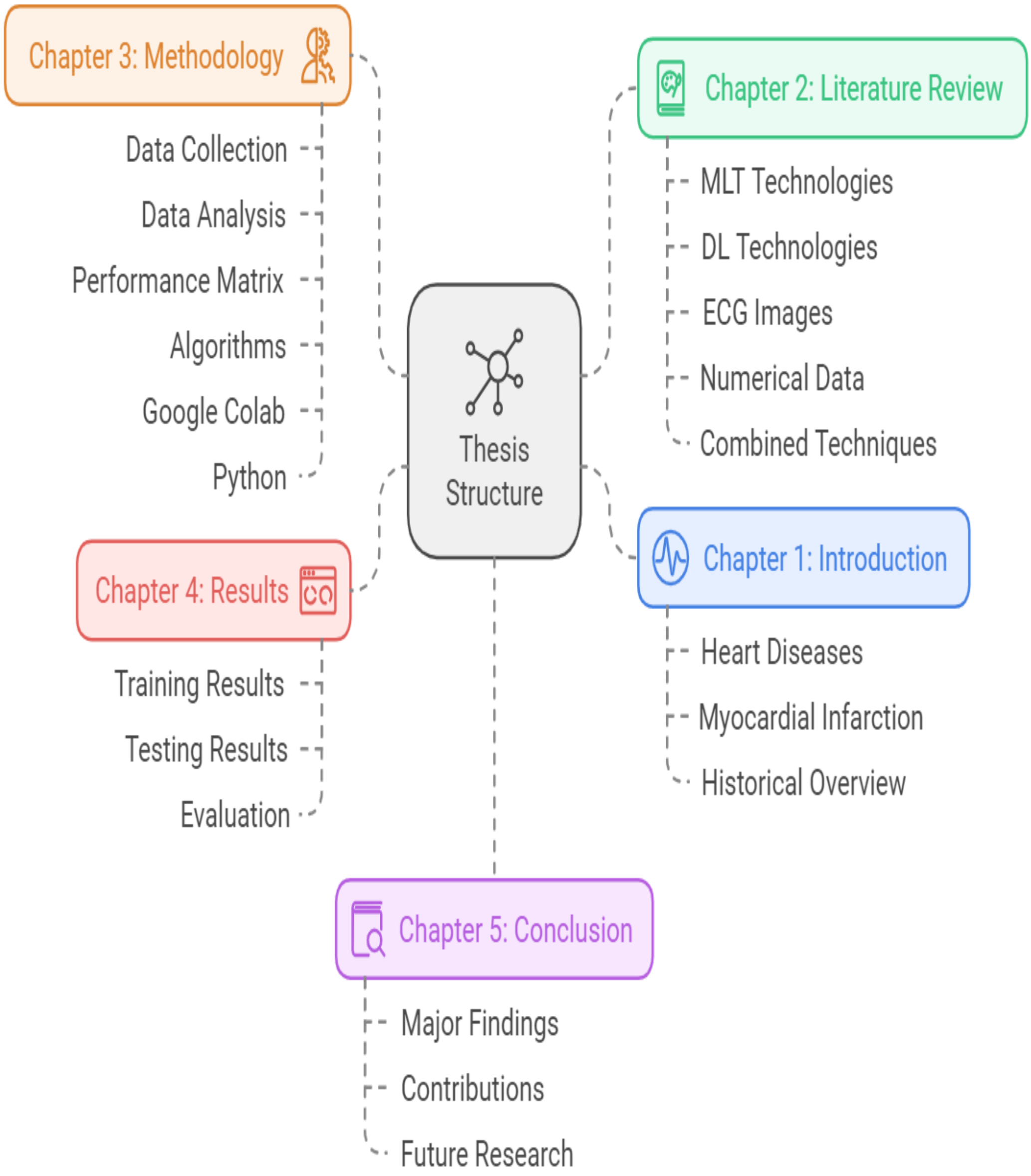
Chapter 3 of this thesis outlines the methodology for designing the Artificial Intelligence Model. It makes decisions based on the combined result of the two types of data, taking into account the performance matrix. Moreover, the section elaborates on the data collection methods and analysis procedures employed.

Chapter 4 of this thesis details the design and testing of an Artificial Intelligence (AI) model suitable for the investigated system. To meet the system's diverse requirements, it has been suggested to use two types of data. For each type of data, five types of algorithms have been employed. The Google Colab environment, along with the Python programming language, is utilized to train the model.

Chapter 5 concludes the thesis and highlights the major findings and contributions of the research. It also identifies potential areas for further investigation. And the figure 8 shows diagram of thesis structure

Figure 8

Thesis Structure



CHAPTER II
Literature Review

2.0 Overview

The medicine and healthcare sector has been rapidly growing and advancing. The application of data-driven, robust, and efficient machine learning (ML) to deep learning (DL) technologies has launched and influenced this progress. Machine learning in the medical sector is advancing rapidly, transforming medicine, and enhancing the experiences of clinicians and patients. Machine learning technologies have progressed into data-intensive deep learning methodologies, which exhibit more robustness and efficiency in managing medical data. (Chakraborty et al., 2024) This article examines essential data-driven elements of machine intelligence in healthcare.

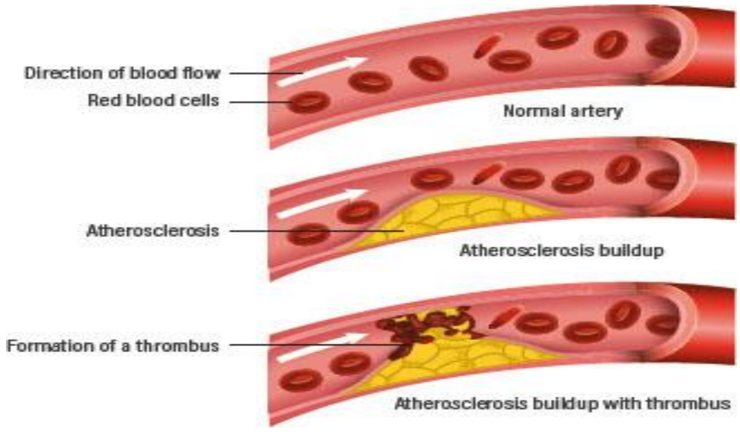
2.1 Clinical Background

Acute myocardial infarction (AMI) is a leading cause of death worldwide. Notwithstanding progress in treatments and interventional techniques, numerous patients continue to experience severe adverse cardiovascular events (MACEs) following acute myocardial infarction (AMI).

Evaluating the risk of these occurrences is essential for enhancing patient care and tailoring effective treatment. In this context, machine learning (ML) approaches have emerged as potent instruments for analyzing extensive medical data and forecasting risks. This study provides a comprehensive assessment of different machine learning algorithms to determine their efficacy in predicting negative cardiovascular outcomes in patients with acute myocardial infarction (AMI). (Xiao et al., 2022)

Figure9

Acute Coronary Syndrome (Norton, 2017)

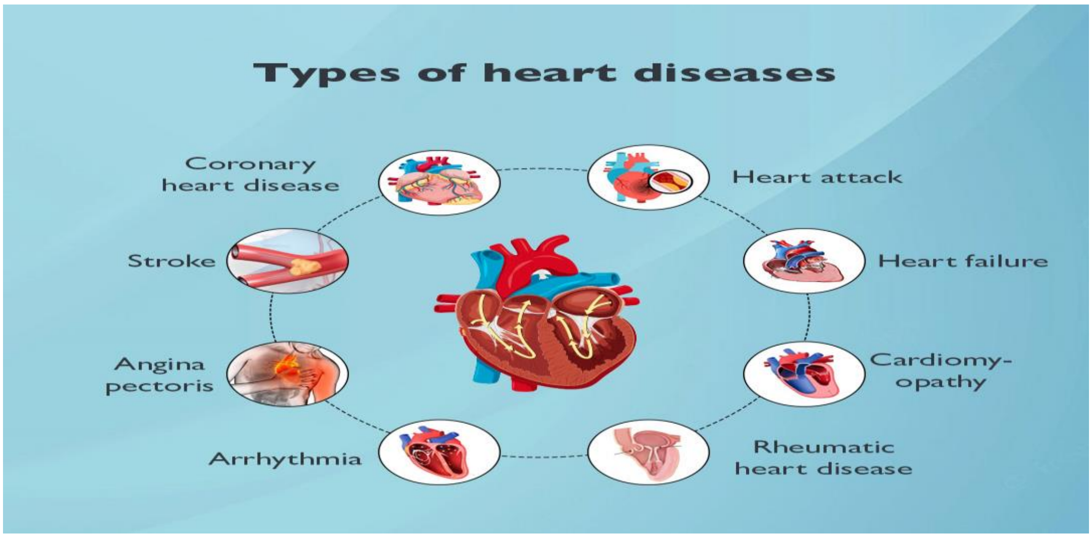


2.2 Heart Disease and Types

Cardiovascular disease, sometimes known as heart disease, is a broad group of disorders that affect the structure and function of the heart. According to the World Health Organization, these diseases account for approximately 30% of all (WHO), making them a leading cause of death. Heart problems include a range of conditions, such as arrhythmias, congenital heart defects, and diseases resulting from blocked arteries, which significantly impact an individual's health and quality of life., heart failure, and coronary artery disease. In order to lower death rates and improve patients' quality of life, it is essential to recognize cardiac disease early and treat it effectively. (Jindal et al., 2021)

Figure 10

Heart Disease Types (Heart Disease Risk Factors In Young Adults | Sprint Medical, n.d.)

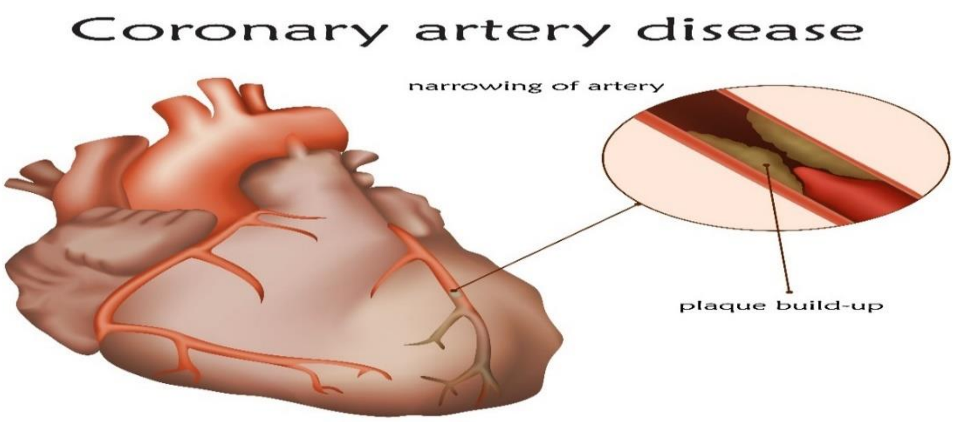


Heart disease diagnostics and therapies have both been greatly improved by recent developments in medical technology and scientific inquiry. Ongoing hurdles are caused by the disease's intricacy and heterogeneity, however. This chapter gives a general outline of the many forms of cardiac disease, as well as its symptoms, potential causes, and contemporary approaches to diagnosis and therapy. (Zhang et al., 2020) Heart disease may be classified into many categories, each with unique features and consequences. The primary categories comprise:

A prevalent kind of heart disease called coronary artery disease (CAD) is brought on by plaque accumulation in the coronary arteries, which carry blood to the heart muscle. Angina, myocardial infarctions, and a number of other issues could result from it.

Figure 11

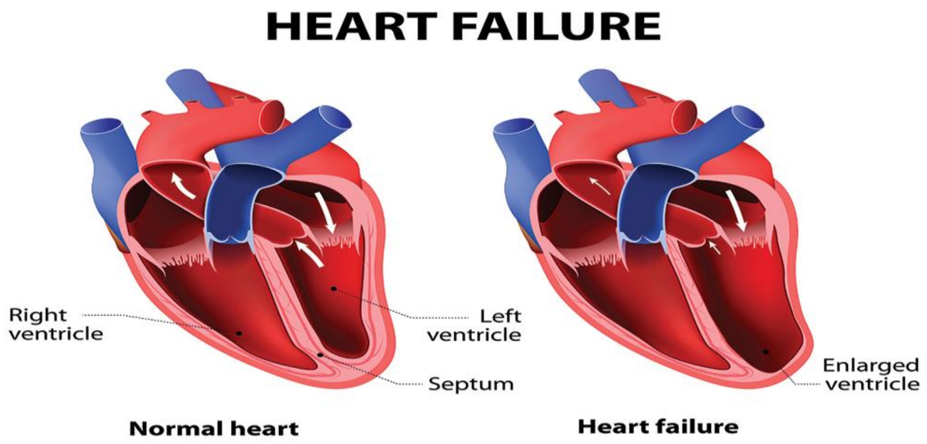
Coronary Artery Disease (Coronary Artery Disease - Lathrup Village, MI: Heart and Vein Center, n.d.) e



Heart failure, occasionally termed congestive heart failure, is a medical condition marked by the heart's incapacity to effectively pump blood to meet the body's demands. It may stem from various underlying conditions, including coronary artery disease (CAD), hypertension, and diabetes.

Figure 12

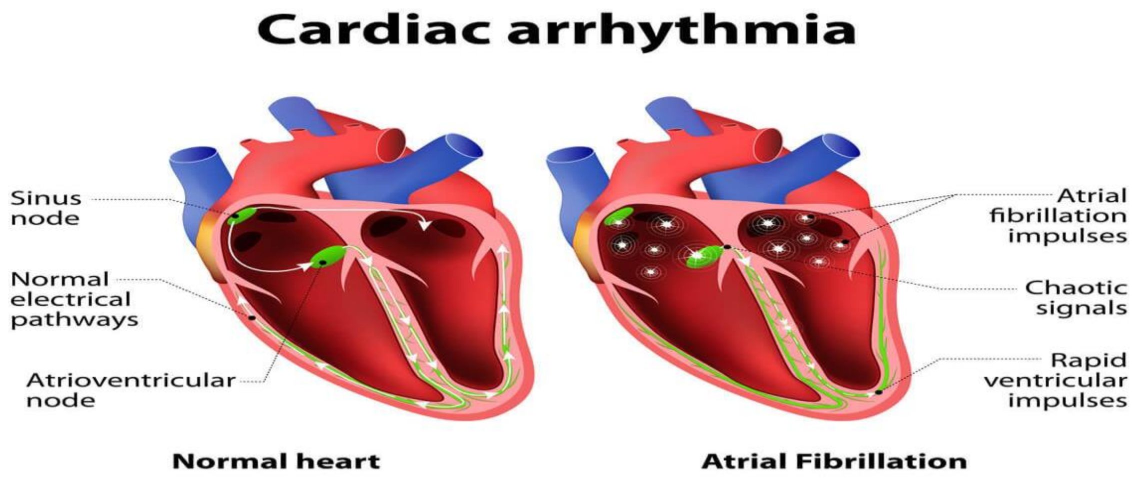
*Heart Failure (*What Is Heart Failure? – PB Cardiovascular*, n.d.)*



Arrhythmias: refer to abnormal heart rhythms that might manifest as excessively rapid, excessively slow, or irregular heartbeats. Typical examples are atrial fibrillation, ventricular tachycardia, and bradycardia. Arrhythmias may result in consequences such as stroke or abrupt cardiac arrest.

Figure 13

Arrhythmias (Why Arrhythmia Occurs？Top 5 Things You Should Know About Arrhythmia – Wellue, n.d.)

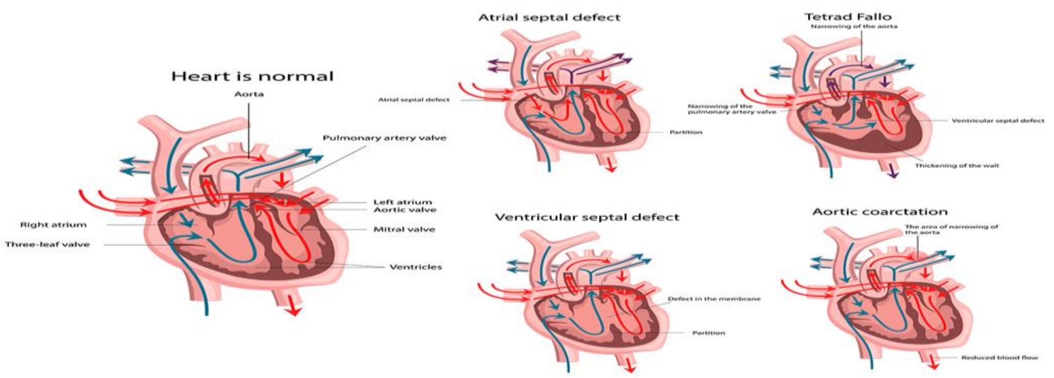


Congenital heart: defects refer to inherent structural abnormalities of the heart that are present from birth. They have the potential to impact the myocardium, cardiac valves, and vasculature. Certain flaws might pose a risk to life and need urgent correction.

Heart Valve Disease: refers to the impairment of one or more valves in the heart, leading to disruptions in the circulation of blood inside the heart. The condition might arise due to diseases, the natural process of aging, or congenital anomalies. (Bemtgen et al., 2022)

Figure 14

Congenital Heart Disease (Congenital Heart Disease Types, Symptoms, and Treatment | Dr. Raghu, n.d.)



2.3 Myocardial Infarction

In order to increase the global use of myocardial infarction diagnosis criteria and standardize them, the World Health Organization and significant national and international cardiac associations created and approved the Universal Definition of Myocardial Infarction. According to the current definition, when there is acute myocardial damage accompanied by clinical evidence of myocardial ischemia, the phrase "acute myocardial infarction" should be used. To diagnose myocardial infarction, several requirements must be fulfilled, such as an increase or decrease in the level of cardiac troponin in the blood, with at least one value or above. (Lindahl & Mills, 2023)

Cardiovascular illnesses are the largest cause of mortality globally, with ischaemic heart disease responsible for 49.2% of deaths in 2019. Typically, a thrombus obstructing an artery or a bypass graft triggers an acute myocardial infarction, characterized by a sudden decrease in blood supply to the myocardium, leading to heart failure and mortality. (Q. Zhang et al., 2022)

Body weakness is a prevalent symptom among the aged that causes physical manipulation, tiredness, reduced body activity, weight loss, and decreased walking speed. It is also a typical result of cardiovascular disorders. Most of the independent cardiovascular (CV) risk factors linked to insulin include obesity, sedentary lifestyle, hypertriglyceridemia, or inflammatory indicators (such as high-sensitivity C-reactive protein [hs-CRP]). Smoking was identified as a distinct cardiovascular risk factor in 80% of Myocardial Infarction patients under the age of 55.(Salari et al., 2023)

2.3.1 Type of MI

Type 1 and type 2 are the two primary forms of MYL disease . The main characteristic of type 1 CAD is the rupture or erosion of atherosclerotic plaque, which results in the development of an occlusive or partly occlusive thrombus. On the other hand, an oxygen supply-demand imbalance—which may or may not have an atherosclerotic component—is the primary characteristic of a type 2 MI.5. While NSTEMI is more prevalent for both forms of MI, both type 1 and type 2 MI may manifest as STEMI or NSTEMI.16–18 Although the relative proportions of type 1 and type 2 NSTEMI vary depending on the clinical situation, the majority of NSTEMI pathophysiology is consistent with type 1 MI.(Wereski et al., 2022)

2.3.2 Symptoms

Cardiac discomfort (angina): Sensations of discomfort, pressure, or pain in the thoracic region.

Dyspnea: Experiencing respiratory distress during physical exertion or when at rest.

Tiredness: Experiencing fatigue or lack of energy that is out of the ordinary.

Cardiac arrhythmia: Arrhythmia or tachycardia.

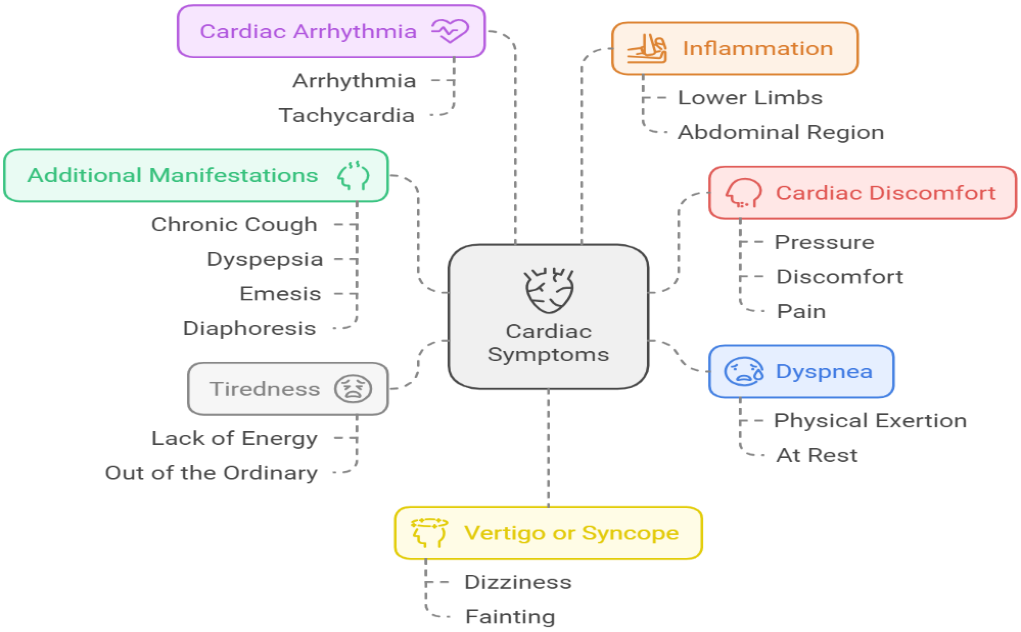
Inflammation: Swelling in the lower limbs or abdominal region.

Vertigo or syncope: Experiencing dizziness or fainting.

Additional manifestations: Chronic cough, dyspepsia, emesis, or diaphoresis. (Brynja Ingadóttir et al., 2024)

Figure 15

Cardiac Symptoms



2.3.3 Causes and Risk Factors

The cardiovascular disease prevention system, especially Myocardial Infarction, should traditionally be based on the identification of variable risk factors in the elderly age groups, i.e., the presence of such acute forms and conditions of Myocardial Infarction, each of which increases the likelihood of death from cardiovascular disease.(D.B & F.A., 2022a)

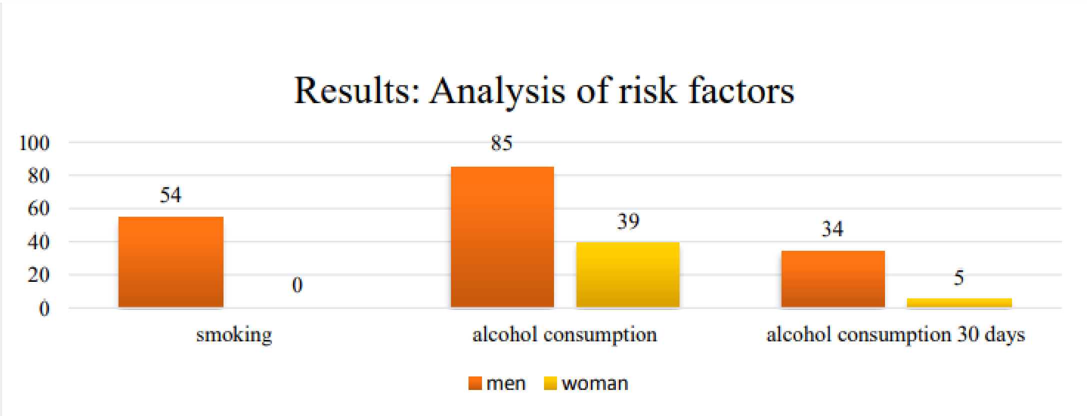
Young myocardial infarction patients have several risk factors, such as inflammatory eccentric atherosclerotic plaques, a greater prevalence of obesity and tobacco use, and more risk factors for healthy lifestyle choices, such as alcohol use and inactivity. Young MI patients are more likely to be male, have a family history of FCHL, and have greater levels of Lp(a) than elderly patients, where MI is more common in both men and women. Furthermore, the use of AASs, cocaine, and cannabis are risk factors for MI in young individuals. Younger and older MI patients have also been shown to have different genomes, particularly in the coagulation and lipid metabolism pathways. The proportional roles of gene circuits linked to vascular, cellular proliferation, inflammation, lipid metabolism.

Alternatively, additional yet-to-be-discovered paths may yield significant insights. A heightened likelihood of early-onset myocardial infarction correlates with mutations in familial hypercholesterolemia and elevated polygenic scores. The different pathophysiology and risk factor profiles of younger and older myocardial infarction patients may help find younger people who are more likely to have a recurrence and guide primary and secondary prevention efforts. (Sagris et al., 2022)

In a study titled "DETERMINE THE VALUE OF RISK FACTORS FOR MYOCARDIAL INFARCTION," researchers examined 54 male patients, was 17.2% of factory-produced smokers. A total of 124 patients (85 men and 39 women) were alcohol users (egg, beer, wine, vodka, and cognac). Of them, 106 had drunk alcohol over the previous 12 months, and 39 of the 106 patients (34 men and 5 women) had taken alcohol within the previous 30 days. The figure1.9 show 39.5% compared to 314 patients and 31.5% compared to those who consumed 124 alcoholic beverages.(D.B & F.A., 2022b)

Figure 16

Risk Factors in Men and Woman (D.B & F.A., 2022b)



2.4 Diagnosis

Heart disease includes a range of disorders that impact the heart's structure and function, such as coronary artery disease, heart failure, and arrhythmias. Effective treatment strategies are crucial for managing symptoms, improving quality of life, and reducing the risk of serious complications. Here are some of the primary approaches to treating heart disease:

2.4.1 Medical Devices

Pacemakers are important medical innovations that aim to control the heart rhythm using electrical signals. These devices are used to treat arrhythmias and bradycardia, and are surgically implanted inside the body to monitor and correct abnormal rhythms. These devices regulate heart rhythms by delivering electrical shocks, which helps correct serious irregularities in the heartbeat, and this is vital to avoid sudden cardiac arrest in high-risk individuals.

On the other hand, left ventricular assist devices are used to support the function of the heart’s left ventricle, helping it pump blood efficiently. These devices are often a temporary measure for severe heart failure, until the patient can undergo a heart transplant.

Stents are used to maintain blood flow in the coronary arteries. They play a key role in treating narrowed or blocked arteries, which improves blood circulation and helps improve heart health.

Artificial heart valves are designed to replace faulty heart valves. These devices are used to treat valve disorders such as stenosis or regurgitation, improving the efficiency of the cardiovascular system.

Finally, continuous glucose monitors and insulin pumps are vital tools in diabetes management. These devices aim to track blood glucose levels and regulate insulin delivery, helping to control diabetes, a condition that increases the risk of heart disease. (Armoundas et al., 2024)

2.4.2. Medical Procedures and Surgeries

Coronary angioplasty is a vital medical procedure that aims to open narrowed or blocked coronary arteries. This procedure often involves placing a stent inside the artery to maintain continuous blood flow.

The goal of coronary artery bypass grafting (CABG), a surgical technique, is to improve the blood supply to the heart muscle by establishing a new route for blood flow to the heart that avoids the clogged or constricted arteries.

In the context of treating heart rhythm disorders, pacemakers, including implantable cardioverter defibrillators (ICDs), play an important role in managing these conditions. These devices are used to send electrical pulses that aim to regulate the heartbeat and restore a normal rhythm.

With heart valves, valve repair or replacement surgery is performed to deal with damaged valves, as this procedure aims to ensure proper blood flow through the heart.

In cases of severe heart failure, a heart transplant is the final solution, where the diseased heart is replaced with a healthy heart from a donor, which restores life to people with critical heart conditions.

2.4.3 Alternative Therapies

Cardiac Rehabilitation: Rehabilitation Programs: Comprehensive programs involving exercise training, education, and counseling to help patients recover and improve their heart health.

Stress Management: Techniques: Incorporate stress-reducing practices such as meditation, yoga, and deep
breathing exercises. (Edwards & Arya, 2024)

Lifestyle Changes

Diet: A heart-healthy diet should emphasize foods high in fruits, vegetables, whole grains, lean meats, and healthy fats (such those found in nuts, avocados, and olive oil), Limit Sodium and Sugar: Reduce salt and sugar intake to manage blood pressure and weight.

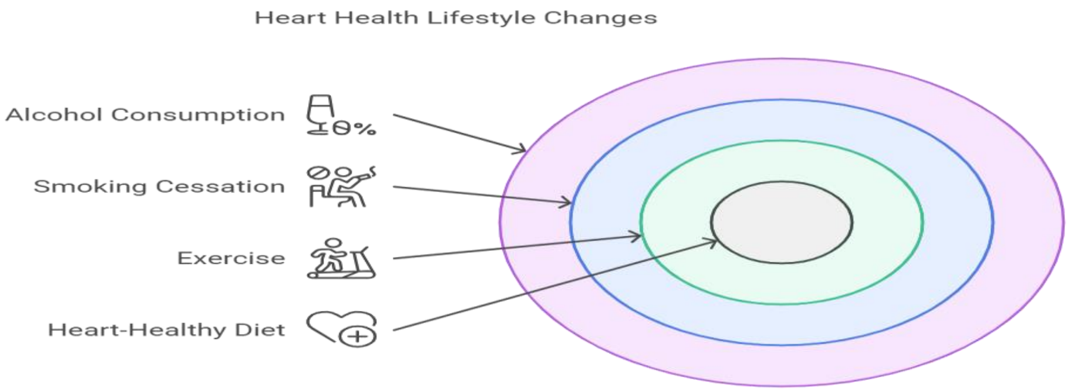
Exercise: Regular Physical Activity: Perform muscle-strengthening exercises two or more days a week in addition to 150 minutes of moderate-intensity aerobic exercise or 75 minutes of intense activity every week.

Smoking Cessation: Quit Smoking: Avoid smoking and use of tobacco products, which significantly increase the risk of heart disease.

Alcohol Consumption: Moderate Alcohol Intake: Keep alcohol intake within reasonable bounds, which are one drink for women and two for men per day.

Figure 17

Heart Healthy Lifestyle Changes



2.5 Technical Solutions

2.5.1 Medical Devices:

Pacemakers and Implantable Cardioverter Defibrillators (ICDs) are medical devices that are used to regulate abnormal heart rhythms and avoid the occurrence of sudden cardiac arrest.

LVADs are used to augment cardiac function in patients with severe heart failure. Stents and heart valve prostheses are medical devices used to maintain the patency of arteries and replace dysfunctional heart valves.

State-of-the-art Imaging Techniques: Echocardiography and MRI provide comprehensive visualization of cardiac anatomy and physiology.

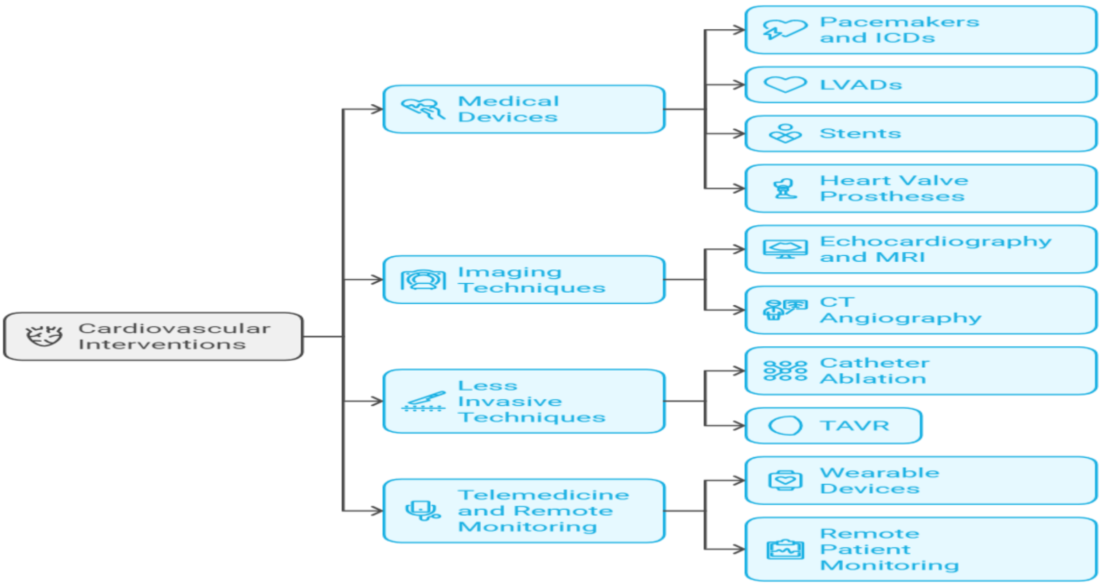
CT Angiography is a medical imaging technique used to get visual representations of the coronary arteries. Less invasive medical techniques: Catheter ablation is a medical procedure used to treat arrhythmias. Transcatheter Aortic Valve Replacement (TAVR) is a procedure that allows for the replacement of aortic valves without the need for open-heart surgery.

Telemedicine and remote monitoring: Wearable devices: Track and analyze heart rate and rhythm.

Remote Patient Monitoring: Monitor cardiovascular health from the comfort of your own home. (Gennaro Tartarisco et al., 2024)

Figure 18

Technical Solutions as Medical Devices

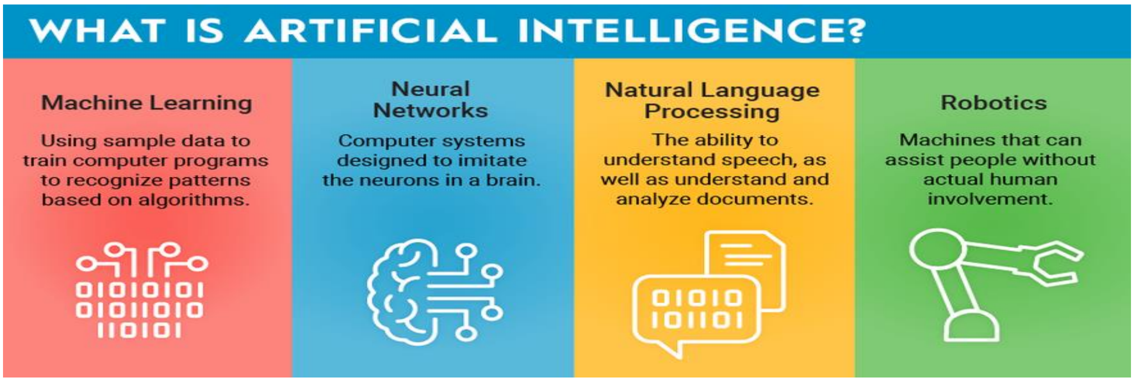


2.5.2 Artificial Intelligence

Artificial Intelligence (AI) denotes the emulation of human intelligence in computers designed to think and learn in a manner analogous to humans. AI systems employ algorithms and computer models to execute activities including problem-solving, decision-making, and pattern recognition. The domain includes multiple sub-disciplines, such as machine learning, natural language processing, robotics, and computer vision. AI has transformative potential across numerous industries, from healthcare and finance to transportation and entertainment, driving efficiency, innovation, and the development of new technologies. (Alamoodi et al., 2020)

Figure 19

What is Artificial Intelligence (What Is Artificial Intelligence? | The Motley Fool, n.d.)



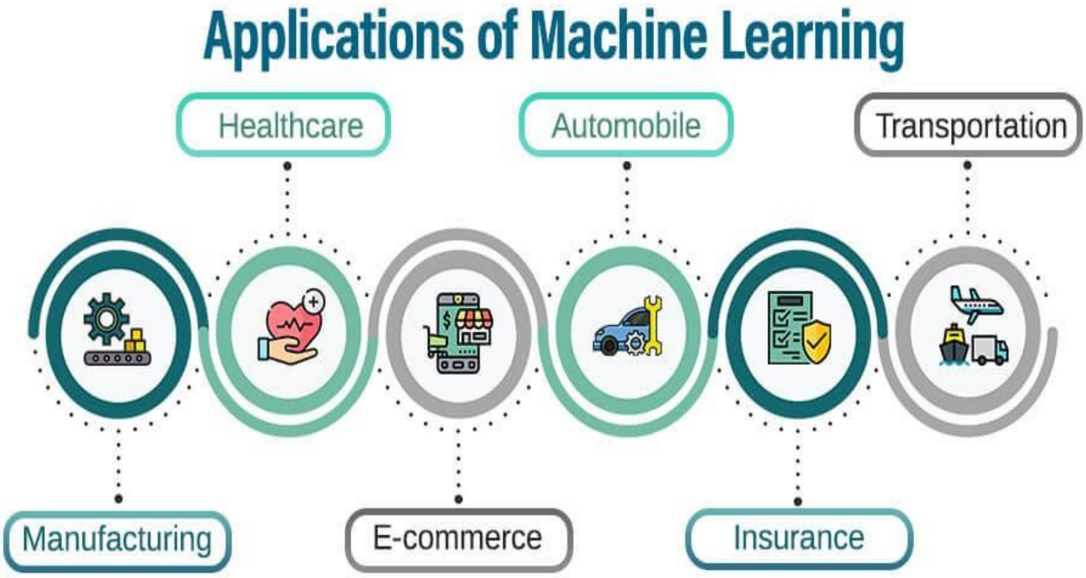
2.5.2.1 Machine learning

Machine learning is a subset of artificial intelligence that seeks to enable computers to acquire the capability to execute particular tasks autonomously, without direct human programming. This technique relies on the construction of models that acquire knowledge from data and render choices or predictions upon the availability of additional data. Deep Learning (DL) represents an advancement of Machine Learning (ML) that employs a multi-layered architecture known as a Neural Network (ANN). Deep learning techniques necessitate diminished human intervention due to the autonomous extraction of information.

Nonetheless, a significant distinction from other ML techniques is that deep learning necessitates extensive data to function effectively. Although ML and deep learning are modern concepts, they have their roots in Arthur Samuel's 1952 computer learning program and Frank Rosenblatt's 1957 proposal of the first neural network. 2 Since the 1990s, advancements in both ML and deep learning have been substantial, primarily attributable to the increase in computational power and the accessibility of vast datasets. Numerous ML methodologies are available to address various issues. The world has witnessed extensive medical uses of ML.(Méndez et al., 2023)

Figure 20

Application of Machine Learning (Machine Learning Applications and Examples - IABAC, n.d.)



2.5.2.1.1 Supervised Learning

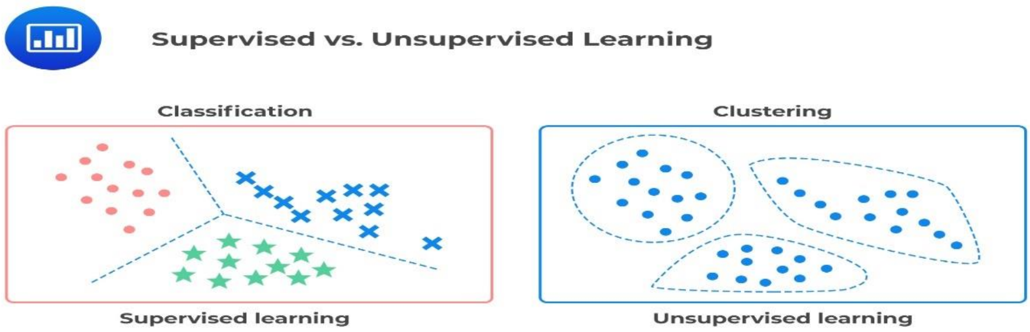
The machine learning challenge of determining a function that links an input to an output using the given input-output pairs is known as supervised learning. It uses labeled training data, such as training examples, to infer a function. Algorithms for supervised machine learning need outside direction. The input dataset is divided into training and testing sets. The training dataset contains a dependent variable that requires prediction or classification. All algorithms (Mahesh, 2019)

2.5.2.1.2 Unsupervised Learning

Unsupervised learning is a significant category of algorithms that reveals the intrinsic structure of data without external supervision, unlike supervised learning. Finding similarities between a set of examples (known as clustering), determining the distribution of the data (known as density estimation), or reducing the number of dimensions (known as subspace learning) are common unsupervised learning problems (Tyagi et al., 2022).

Figure 21

Supervised and Unsupervised Machine Learning (Supervised Machine Learning, Unsupervised Machine Learning, and Deep Learning - CFA, FRM, and Actuarial Exams Study Notes, n.d.)



2.5.2.2 Deep Learning

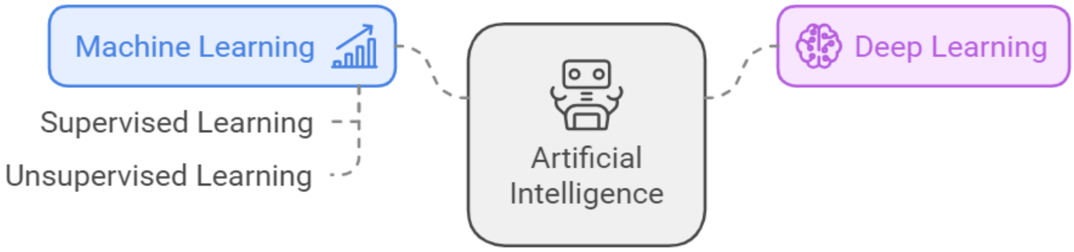
A kind of machine learning called "deep learning" uses multi-layer artificial neural networks to simulate how the human brain analyzes information and makes decisions. Deep learning techniques are distinguished by their capacity to acquire knowledge from vast quantities of data and identify intricate patterns that would be undetectable using conventional approaches. Neural networks are composed of layers of nodes (neurons) that progressively acquire more intricate representations of the input data. The technique described is referred to as "hierarchical learning," in which information is systematically constructed from one layer to the subsequent one.

Deep learning has been widely used in a number of domains, such as recommendation systems, computer vision, natural language processing, and speech recognition. Deep neural networks can examine photos to accurately identify things inside them or interpret and evaluate textual information similarly to human comprehension.

Two primary elements enhance the effectiveness of deep learning: the accessibility of extensive datasets and the utilization of sophisticated graphics processing units (GPUs), which enable expedited and more efficient training of deep models. Furthermore, the development of novel algorithms and optimization methods for neural networks has significantly enhanced the efficacy of these models. (K. Li et al., 2024)

Figure 22

Artificial Intelligence



2.5.2.3 Optimizer

An optimizer is an algorithm or mathematical method employed to modify the parameters of a model or function to minimize or maximize a specified objective function. The principal objective of an optimizer is to ascertain the optimal solution, which may represent either the maximum or least value of the function, contingent upon the context.

Basic Categories of Optimizers

Optimizers can be classified into several categories based on the specific problem they aim to address and the methodology adopted. Among the basic categories are:

Gradient-based optimizers: Gradient-based gradient descent methods are common algorithms in the field of machine learning, and are particularly useful when studying neural networks. This procedure relies on the gradient of the loss function to make iterative adjustments to the model parameters, with the goal of minimizing the amount of error and improving the overall performance of the model.

Alternative versions of optimizers include techniques such as non-gradient-based optimizers, which are used in certain cases where the objective function may be discontinuous or contain inconsistencies in behavior, making the use of gradient methods ineffective. Alternate versions:

Stochastic Gradient Descent (SGD) is an optimization algorithm that updates the parameters by considering just one sample, or a small subset of samples, at each iteration.

A technique called Mini-Batch Gradient Descent finds a middle ground between batch gradient descent and stochastic gradient descent (SGD). It entails updating the settings using a tiny selection of data.
Momentum: Enhances convergence by incorporating a portion of the prior update into the current update.

Adam, an abbreviation for Adaptive Moment Estimation, is an optimization technique that integrates the advantages of AdaGrad, which adapts learning rates, and RMSProp, which uses momentum. Adam adjusts learning rates by considering both the first and second moments of the gradients.

2. Optimizers of the second order

Newton's Method employs the second derivative, commonly referred to as the Hessian matrix, to ascertain the minimum value of a specified function. It exhibits a superior convergence rate relative to gradient-based methods; nonetheless, it necessitates greater computational resources.
Quasi-Newton methods, including BFGS, approximate the Hessian matrix to reduce computational complexity while retaining the efficiency of Newton's method.

3. Optimizers that do not rely on derivatives

Genetic Algorithms (GA) are a computer method that employs concepts of natural selection and genetics to address optimisation challenges. Natural selection shapes these optimisers, providing a foundation for the evolution of a population of solutions. They select the most appropriate people for reproduction and mutation to investigate the solution space.

Simulated Annealing is a computational method that replicates the annealing process utilised in metallurgy. It entails heating a substance and subsequently cooling it gradually to ascertain a low-energy state, signifying an optimal solution.

Particle Swarm Optimization (PSO) is a computational technique that simulates a group of particles navigating through a given solution space. Each particle adjusts its location depending on its own experience and the experiences of its nearby particles.

4. Optimizers based on constraints

Linear Programming (LP) is a method used to solve optimization problems in which both the objective function and constraints are linear. Commonly used in the field of operations research.

Quadratic Programming (QP) is a mathematical optimization technique that is similar to Linear Programming (LP), but it allows for the inclusion of quadratic components in the target function.

Integer Programming is a kind of linear programming that has the constraint that some or all variables must be integers.

5. Bayesian Optimization

Utilizes probabilistic models to construct a surrogate of the goal function and determines the subsequent points to assess by effectively managing the trade-off between exploration and exploitation. It is especially beneficial for optimizing functions that are costly to assess, such as hyperparameter tweaking in machine learning.

Practical uses of optimizers

Optimizers play a vital role in several domains, encompassing: Machine Learning is the process of training models by minimizing the loss function in order to enhance the accuracy of the model. Operations Research involves the resolution of logistical, production, and resource allocation issues.

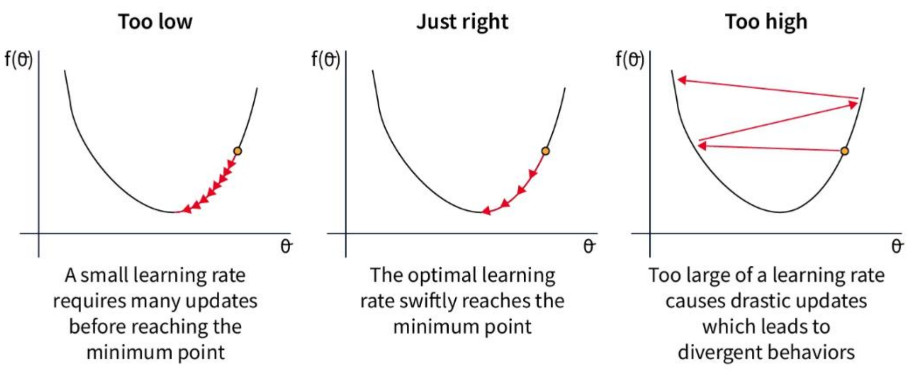
Engineering involves the process of design optimization, which focuses on lowering weight while ensuring the maintenance of strength.

Economics involves the identification of equilibrium points in markets and the optimization of investment portfolios.

Every optimizer has its own set of advantages and disadvantages, and the selection of an optimizer is often based on the particular issue, the characteristics of the objective function, and computing limitations.(Soydaner, n.d.)

Figure 23

Optimizer Effect (Optimizers in Deep Learning - Scaler Topics, n.d.)



2.6 Previous Works

Heart Disease Datasets were used by Tama et al. in 2020. The datasets used to forecast cardiac disease are sourced from publicly available repositories. The data type employed in the investigation is numeric. (Datasets from Z-Alizadeh Sani, Statlog, Cleveland, and Hungary).

Alshehri & Alharbi, n.d. The study used three datasets StatLog UCI Dataset: 13 features, 270 cases.Z-Alizadeh Sani Dataset: 54 features, 303 cases. Cardiovascular Disease (CVD) Dataset: 11 features, 70,000 cases.

Sharma et al. (2020) conducted an investigation of deep neural networks utilizing the Cleveland heart disease dataset. This research has produced many DNN models, each exhibiting distinct features. The primary objective is to develop a model that optimally predicts the diagnosis of cardiovascular heart disease in a patient. We employ diverse optimization algorithms to minimize the loss function, utilize various weight initialization methods to set the model parameters, and implement varying quantities of hidden layers of neurons to achieve optimal outcomes from the study. We assess the efficacy of each neural network model using performance metrics such as accuracy, sensitivity, specificity, and precision.

Hossen et al. (2021) created a computer-aided diagnostic method utilizing many factors as input. We subsequently compared the accuracy, precision, recall, and F1 scores of the three classification systems. Decision Trees, Random Forests, and Logistic Regression are the most effective classification techniques for predicting heart disease. Through the utilization of several methodologies, accuracy has attained 92%.

Akella & Akella (2021) conducted comparison research utilizing six distinct machine learning algorithms on the Cleveland dataset to forecast disease outcomes. The aim is to introduce an open-source machine learning system that can detect coronary artery disease effectively. The present study assessed six machine learning algorithms: k-closest neighbor, neural network, support vector machine, nearest neighbor, random forest, regression tree, and linear regression. All exhibited commendable performance, achieving accuracies exceeding approximately 80%, with the nearest neighbor algorithm surpassing 93% accuracy. We compare the accuracy, recall, F1 score, and area under the curve-receiver operating (AUC-ROC) of the machine learning models.

Garg et al. (2021) employed various algorithms, including K-Nearest Neighbor (K-NN) and Random Forest. Following the application of multiple algorithms, the algorithms employed in this experiment have demonstrated commendable performance utilizing the available qualities. We calculate the accuracy of the K-Nearest Neighbor (K-NN) to be 86.885%, and the Random Forest to be 81.967%.

Xiao et al., 2022examined the value of prediction using machine learning (ML) techniques in patients with acute myocardial infarction (AMI), indicating the importance of identifying optimal models for predicting major adverse cardiovascular events (MACEs). Several ML models were used, including random forest (RDF), which proved to be superior in prediction compared to traditional methods such as logistic regression (LR). Five independent predictors of MACEs were identified, including Killip classification, medication compliance, age, and creatinine and cholesterol levels, which enhances the accuracy of prediction and individualized patient treatment.

Aguru et al., 2022 want to improve the LeNet-5 model so that it can better predict cardiovascular diseases using a set of ECG images from patients. The Ch. Pervaiz Elahi Institute of Cardiology in Multan, Pakistan, developed an ECG image dataset of cardiac patients to evaluate LeNet-5 and a modified variant, aiming to aid the scientific community in studying cardiovascular disorders. Accuracy, precision, recall, and F1-score are the four evaluation measures we used to compare LeNet-5 and a modified version. According to the experimental findings, LeNet-5's accuracy as a transfer model was 89.24%, while a modified version of the model reached 98.38% accuracy.

Almulihi et al., 2022 The study used two datasets on heart disease for their investigation. More precisely, the initial dataset consisted of 18 distinct characteristics used as the class label to forecast heart disease. These characteristics were in the form of numerical data. The second dataset consisted of 14 numeric characteristics and four categorical features. Before analysis, duplicates in the categorical features (image data).

Tippannavar et al., 2022 discusses using a 2D Convolutional Neural Network (CNN) to detect heart disease from ECG signals with high accuracy. The study used the MIT BIH arrhythmia database and achieved a 98.10% accuracy on the test set. This method was successful without needing manual preprocessing of the ECG data. The CNN outperformed other methods like SVM, RNN, and FFNN, making it highly effective for heart disease classification.

Manolkar and Gawande, 2023 Discovered. Researchers have employed machine learning and deep learning techniques to analyze ECG readings. Utilizing deep learning and machine learning methodologies in ECG analysis enables researchers and healthcare practitioners to improve the precision, efficiency, and diagnostic proficiency of heart disease identification. These methodologies provide the possibility for early identification, tailored therapies, and enhanced patient results in cardiology. A thorough examination of the current methodologies was conducted. The various strategies were analyzed, contrasted, and evaluated. The analysis indicates that the most efficient algorithm is employed to train the model on the MIT-BIH Arrhythmia Database.

Lida Kermanidis et al., 2024 This paper presents Paper addresses heart issues with deep learning on ECG signals. Combines LSTM and CNN models for arrhythmia and heart failure prediction. Outperformed existing methods in classifying ECG data from databases. And focus on ECG signal classification using CNN and LSTM models. Addresses challenges in early detection of heart issues using AI. Combines LSTM networks with CNNs for predicting arrhythmias. And the researchers found Proposed deep-learning model outperformed existing methods in ECG signal classification. Achieved high accuracy of 97.6% and 99.20% in two classification scenarios.

C. J. Zhang et al., 2024 in this paper Heart failure classification using deep learning on ECG spatiotemporal features. Focus on NYHA functional classification model for heart failure diagnosis. Utilizes MIMIC-III database for ECG signal analysis in heart failure. the method use Deep learning with integrating attention mechanism for heart failure classification. Traditional shallow machine learning methods for heart failure classification limitations. CNN-LSTM-SE model outperformed other models in heart failure classification. Model achieved high accuracy (99%), specificity, sensitivity, and positive predictive value. The proposed model showed superior performance with 12-second ECG signal segments.

Altaf et al., 2024, utilize ECG signals, image processing, and deep learning methodologies to enhance personality detection. The data employed in this work is sourced from a public repository, ASCERTAIN, which contains a collection of physiological signals, including ECG recordings, acquired from 58 subjects exposed to various video stimuli classified by arousal and valence degrees. The complex interaction between physiological signals and emotional responses forms the basis of our investigation into personality traits. Our process entails generating spectrograms from ECG signals through meticulously calibrated window sizes. Convolutional Neural Networks (CNNs), especially the ResNet-18 architecture and the visual transformer architecture, are the building blocks for personality trait classification and feature extraction. Our primary contribution lies in employing spectrograms derived from physiological data (ECG) and training extensive deep-learning models to precisely characterize personality traits.

Adam et al. (n.d.) elucidate the impact of artificial intelligence, machine learning, and deep learning on the evolution of cerebral and cardiac healthcare. We analyze the current state of AI-driven innovations from a multidisciplinary perspective, identify significant obstacles and opportunities, and provide solutions to address barriers to adoption and execution. By encouraging interdisciplinary collaboration and communication among stakeholders, including clinicians, researchers, policymakers, and industry partners, we hope to harness the transformative potential of AI and ML to improve patient outcomes, advance healthcare delivery, and usher in a new era of precision medicine and personalized healthcare for brain and heart care. The researcher developed a pre-training model, employed various methods, and compared the outcomes to identify the optimal model regarding accuracy and sensitivity. This study forecasted cardiovascular risk using a variety of machine learning techniques. Based on predictor factors, the chance of a binary outcome (such as the incidence or non-occurrence of a cardiovascular event) was simulated using logistic regression, a conventional statistical technique. A resilient prediction model was created by combining many decision trees using the random forest, an ensemble learning technique. The ability of deep learning neural networks, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), to recognize complex patterns in the data was investigated.

Nazir et al., n.d. is dedicated to investigating the revolutionary capabilities of AI, ML, and DL in the domains of brain and cardiovascular health. We want to clarify the mechanisms by which cognitive computing influences medical decision-making, patient management, and population health through the integration of innovative research findings, clinical insights, and technical improvements. We aim to tackle significant obstacles and opportunities in utilizing AI-driven methods to enhance outcomes for patients with neurological and cardiovascular disorders through a multidisciplinary perspective.

2.6.1 Numerical Data

Table 1

Numerical Data Previous Works

Reference	Model	Dataset	Preprocessing	Classes	Sensitivity	Specificity	f1-score	Accuracy
(Hossen et al., 2021)	Random Forest, Decision Tree, and Logistic Regression	UCI Cleveland dataset	Data cleaning , Missing Data Analysis Standard Scalling	2(cardiac disease and non-cardiac disease)	86.96% 73.91% 92.31%	73.91% 69.57% 91.89%	82 72 92	80 72 92
(Sharma et al., 2020)	DNN	Cleveland Heart disease dataset	Exploratory Data Analysis Categorical Encoding	2 heart disease	81.03%	82.81%	81.03%	81.96%
(Tama et al., 2020)	Random Forest Gradient Boosting Machine XGBoost	Z-Alizadeh Sani dataset Statlog dataset Cleveland dataset Hungarian dataset	Correlation-based Feature Selection, Particle Swarm Optimization	2(Coronary Heart Disease)				in Statlog 78.90% In Hungarian 91.13% Z-Alizadeh Sani 87.65%
(Akella & Akella, 2021)	Generalized linear model Decision tree Random forest Support-vector machine Neural network k-Nearest neighbor	UCI Cleveland dataset	Remove missing values, Normalizing, correlation matrix	coronary artery disease and normal	80% 74.47% 82.61% 79.59% 93.80% 78.72%		87.86% 79.70% 87.51% 86.62% 89.84% 84.19%	87.64% 79.78% 87.64% 86.52% 93.03% 84.27%
(Xiao et al., 2022)	LR DT Random Decision	Zhuzhou Central Hospital	Dealing with missing values	2 Myocardial Infarction And normal			56.5% 53.2$ 50.3% 57%	72.1% 64.4% 73.3% 71.7%

	Forest (RDF) Naive Bayes (NB) SVM						48% 45.3% 10.3%	74.9% 73.7% 66.3%
(Garg et al., 2021)	K-NN Random Forest	Kaggle	analyze the data correlation matrix	2(Heart disease)	85.19% 78.57%	88.24% 84.85%	85.19 % 80%	86.88% 81.96%
(Alshehri & Alharbi, n.d.)	AdaBoost SVM Decision Tree Random Forest	StatLog UCI Z-Alizadeh Sani from UCI CVD dataset	Normalization	2(CVD)	RF 82% RF 92% RF 70%	RF 82% RF,Adaboot 92% RF 70%	RF 84% RF,adaboost 91% RF 71%	STATLOG RF 84% Z-Alizadeh RF 91% CVD dataset Svm,Rf,adaboost 72%
Adam et al. (2024)	LR RF Deep Learning	electronic health records (EHRs)	feature selection				80% 85% 90%	80% 85% 90%
(Nazir et al., 2024.)	LR RF Deep Learning Neural Network	PubMed, Scopus, Web of Science, and IEEE Xplore,		cardiovascular diseases			78% 81% 85%	80% 82% 85%

2.6.1.1 Limitation of Previous Work

Although previous studies have shown that researchers have made significant contributions to the field and achieved valuable results, this research is not without limitations. There are often factors that affect the accuracy or generalizability of the results. These limitations include aspects such as limited sample size, insufficient data diversity, or reliance on certain techniques that may not be appropriate for all cases.

It is important to acknowledge these limitations to clarify the limits of the conclusions that can be drawn, and to provide directions for future research to overcome these challenges and achieve more comprehensive and accurate results. and they used Several machine learning algorithms, such as Random Forest, Decision Tree, Logistic Regression, Deep Neural Networks (DNN), and Support Vector Machines

(SVM), have been used for cardiac disease classification using various datasets. The UCI Cleveland cluster has been widely used along with other clusters such as Kaggle, Statlog, and Hungarian.

The results showed that the sensitivity of the model's Highest score ever achieved 91.13%, and the accuracy achieved 93.80%, reflecting variability in performance depending on the algorithm used and the dataset.

2.6.2 Images Data

Table 2

Images Data Previous Works

Reference	**Model**	**Dataset**	**Preprocessing**	**Classes**	**Sensitivity**	**Specificity**	**f1-score**	**Accuracy**
(Mahmoud et al., 2022)	LeNet-5 Modified Version of LeNet-5	Ch.Pervaiz Elahi Institute of Cardiology Multan	Resize images, MaxPooling Dropout	4(Myocardial Infarction Patients, abnormal heartbeat, history of (MI), Normal ECG			88% 99%	89.24% 98.38%
(Tippannavar et al., 2022)	CNN	MIT BIH arrhythmia database's ECG records	Convert data to 2D images	7(APC, LBBB, NOR, PAB, PVC, RBBB, VEB)			98%	98.1%
(Manolkar & Gawande, 2023)	DNN	MIT-BIH Arrhythmia Database	removing noise, filtering, normalizing	2(Chronic Heart Failure and normal)	93.96 %	98.70%	95.49 %	96.9%
(Lida Kermanidis et al., 2024)	FFT, CNN–LSTM	PhysioNet dataset	FFT, ECG signal segmentation	3(ARR, CHF, NSR)	ARR: 99.09 %, CHF: 99.87 %, NSR:	ARR: 99.19%, CHF: 99.88%, NSR: 99.35%	ARR: 98.89 %, CHF: 99.67 %, NSR:	97.4% 98.90%

					89.01 %		92.02 %	
(Zhang et al., 2024)	CNN-LSTM SE,	MIMIC-III	ECG signal segmentation , Z-Score	2(CHF, normal)	99.03 %	99.64%		In 12 second 99.09%
(Altaf et al., 2024)	ResNet 18	ASCER TAIN dataset	STFT	encompassing extra-version, neuroticism, agreeableness, conscientiousn ess, and openness			90%	98%

2.6.2.1 Limitation of Previous Work

Although previous studies such as Mahmoud et al., 2022 and Tippannavar et al., 2022 have achieved high results using models such as LeNet-5 and CNN on ECG datasets, there are several limitations to consider. First, most of these studies used specific datasets such as MIT BIH Arrhythmia or PhysioNet, which may limit the generalization of the results to other datasets or more complex real-world scenarios. Additionally, many established models such as CNN-LSTM and ResNet18 rely heavily on data pre-optimization through techniques such as scaling and partitioning, which may improve performance on carefully curated data but reduce the efficiency of the models when dealing with real-world data that may be more complex or contain more noise. Furthermore, some studies rely on binary classifications or a limited number of classes and one type of dataset, which may not cover the full complexity of heart disease.

In recent studies, researchers had presented a variety of models for analyzing ECG data and diagnosing heart diseases. (Mahmoud et al., 2022) used a modified version of LeNet-5 to analyze ECG images from the Heart Institute, Multan, and achieved an accuracy of up to 88% in classifying four categories of heart conditions. In contrast, (Tippannavar et al., 2022) used a CNN to convert ECG data into 2D pictures with 98% accuracy. Manolkar & Gawande (2023) employed deep neural networks (DNN) on MIT-BIH data to analyze chronic heart failure cases with 95.49% accuracy. In addition, Lida Kermanidis et al. (2024) employed FFT and CNN-LSTM on the PhysioNet dataset to categorize ARR, CHF, and NSR categories with excellent accuracy. Zhang et al. (2024) analyzed MIMIC-III data with CNN-LSTMSE and

99.03% accuracy. Finally, Altaf et al. (2024) used ResNet18 to assess ECG data for personality traits with 90%–98% accuracy. More experiments with various datasets are necessary to improve the generalizability of these models, which excel at evaluating ECG data and identifying heart disorders.

2.6.3 Hybrid Model

Table 3

Hyper Model Previous Works

Reference	Model	Dataset	Preprocessing	Classes	Accuracy
(Almulihi et al., 2022)	ML Algorithms SVM,Logistic Regression (LR) , Nave Bayes (NB) , Decision tree , RandomForest (RF) , and K-nearest Neighbors (k-NN). The Hybrid Model CNN-LSTM and CNN-GRU	1-heart disease (kaggle) 2-Cleveland dataset	removing duplicate records and encoding categorical data	2(Heart disease, Healthy)	Dataset1 Ml 75% 75% 67.28% 60.87% 73.16 Hybrid models 76.22 75.63 Dataset2 ML 86.34% 73.17% 82.44% 64.88% 66.34% Hybrid models 89.76% 88.29%

2.6.3.1 Limitation of Previous Work

Although (Almulihi et al., 2022) presented effective hybrid models such as CNN-LSTM and CNN-GRU for analyzing heart disease using two datasets from Kaggle and Cleveland, there are some limitations that should be taken into consideration. First, removing duplicate records and encoding categorical data are important data processing steps, but they may result in the loss of some subtle information that may be useful in improving the accuracy of the model. Also, focusing on only two classifications (heart disease and heart health) reduces the complexity of the problem, which may affect the ability of models to generalize in more complex environments or across multiple cases.

Moreover, the dependence of performance on data quality and the difference between the two datasets used shows that models may be more sensitive to the quality and organization of the data than their intrinsic strength. The results of the study showed that traditional machine learning models such as SVM and RandomForest achieved modest accuracy on the first dataset (Kaggle), with accuracy ranging from 60.87% to 75%. Hybrid models such as CNN-LSTM and CNN-GRU showed a slight improvement with an accuracy of 76.22%. For the second dataset (Cleveland), traditional models achieved a higher accuracy of 86.34% using SVM, while hybrid models performed even better, with an accuracy of 89.76%. This improvement suggests that hybrid models may be more effective in handling complex data.

Chapter III
Methodology

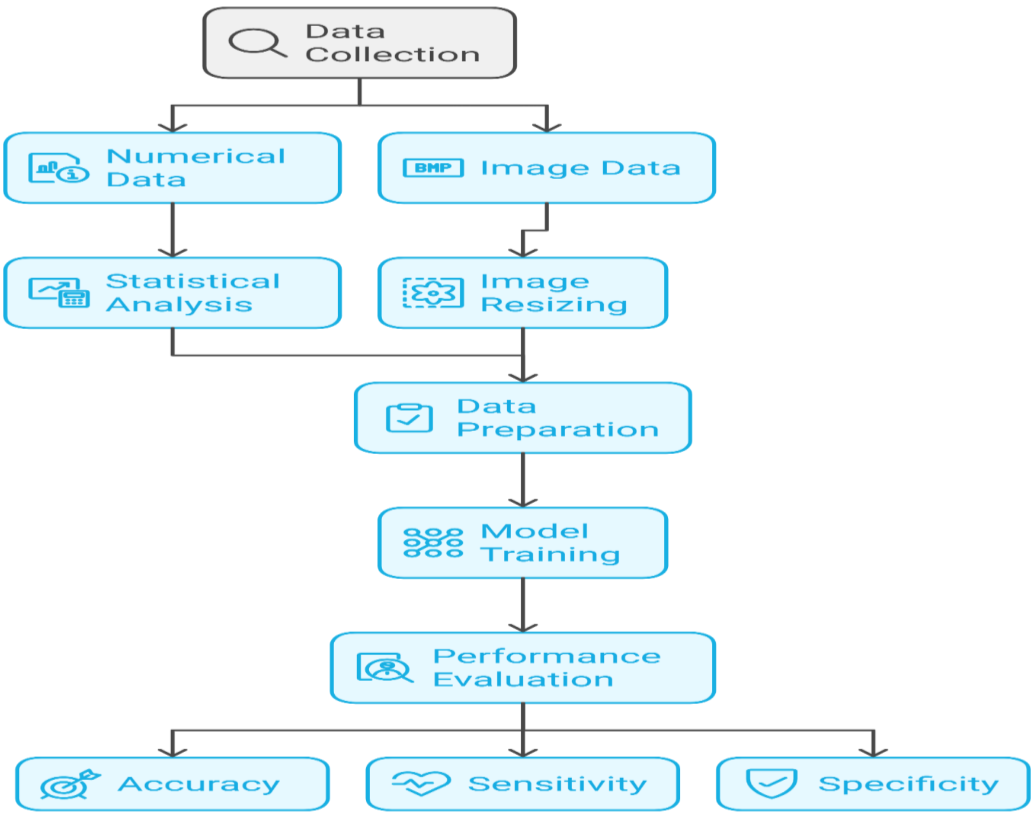
3.0 Overview

In this research, the methodology used is explained through a flow chart, and the steps of the work can be summarized as follows:First, we collected data from the GitHub and Kaggle websites, and this data included two types: numerical data and image data. For the image data, we reduced the size of the images to be suitable for the algorithms used in the research. As for the numerical data, we conducted a statistical description to clarify the relationship between the different variables, which helps us identify the most related variables to facilitate the model training process.

The model was trained using the Python programming language, with the use of the Google Colab environment. Performance was evaluated based on criteria such as accuracy, sensitivity, and specificity. In figure 24 show the flow chart of the methodology.

Figure 24

Data Collection



3.1 Data Collection

3.1.1 Numerical Data (Symptoms)

The data was obtained through the website Kaggle, which describes the patient's symptoms. The data consists of 14 columns (feature or labels) and 1000 rows representing the number of patients. A specialist in the emergency department collects the data, which contains the health condition and symptoms that the patient suffers from. This table shows the most important symptoms worked on in the research.(*Myocardial Infarction Complications*, n.d.)

Table 4
Definition Labels of Numerical Data (symptoms(Myocardial Infarction Complications, n.d.)

No	Description	Code	Unit	Type of the Data
1	Patient ID Number	patient	Number	Numeric
2	Age	age	By Years	Numeric
3	Gender	gender	Male = 1, Female = 0	Binary
4	Chest pain type	chest pain	0, 1, 2, 3 Value 0: typical angina Value 1: atypical angina Value 2: non-anginal pain Value 3: asymptomatic	Nominal
5	Resting blood pressure	Resting blood pressure	94-200 (in mm HG)	Numeric
6	Serum cholesterol	serum cholesterol	126-564 (in mg/dl)	Numeric
7	Fasting blood sugar	Fasting blood sugar	0, 1 > 120 mg/dl 0 = False, 1 = True	Binary
8	Resting electrocardiogram results	resting electro	0, 1, 2 Value 0: normal Value 1: ST-T wave abnormality Value 2: left ventricular hypertrophy	Nominal
9	Maximum heart rate achieved	max heart rate	71-202	Numeric
10	Exercise induced angina	exercise angina	0, 1 (0 = No, 1 = Yes)	Binary
11	Old peak = ST	Old-peak	0-6.2	Numeric
12	Slope of the peak exercise ST segment	slope	1, 2, 3 1 = upsloping, 2 = flat, 3 = downsloping	Nominal
13	Number of major vessels	no of major vessels	0, 1, 2, 3	Numeric
14	Classification	Output (target)	0, 1 0 = Absence of Heart Disease 1 = Presence of Heart Disease	Binary

The image 25 shows the description of the data, as this process showed the number of samples (patients), the percentage and type of chest pain, and the number of female and male samples, and the target (normal or MI). It became clear that this process is important for understanding the data being dealt with in the model.

Figure 25

Data Information

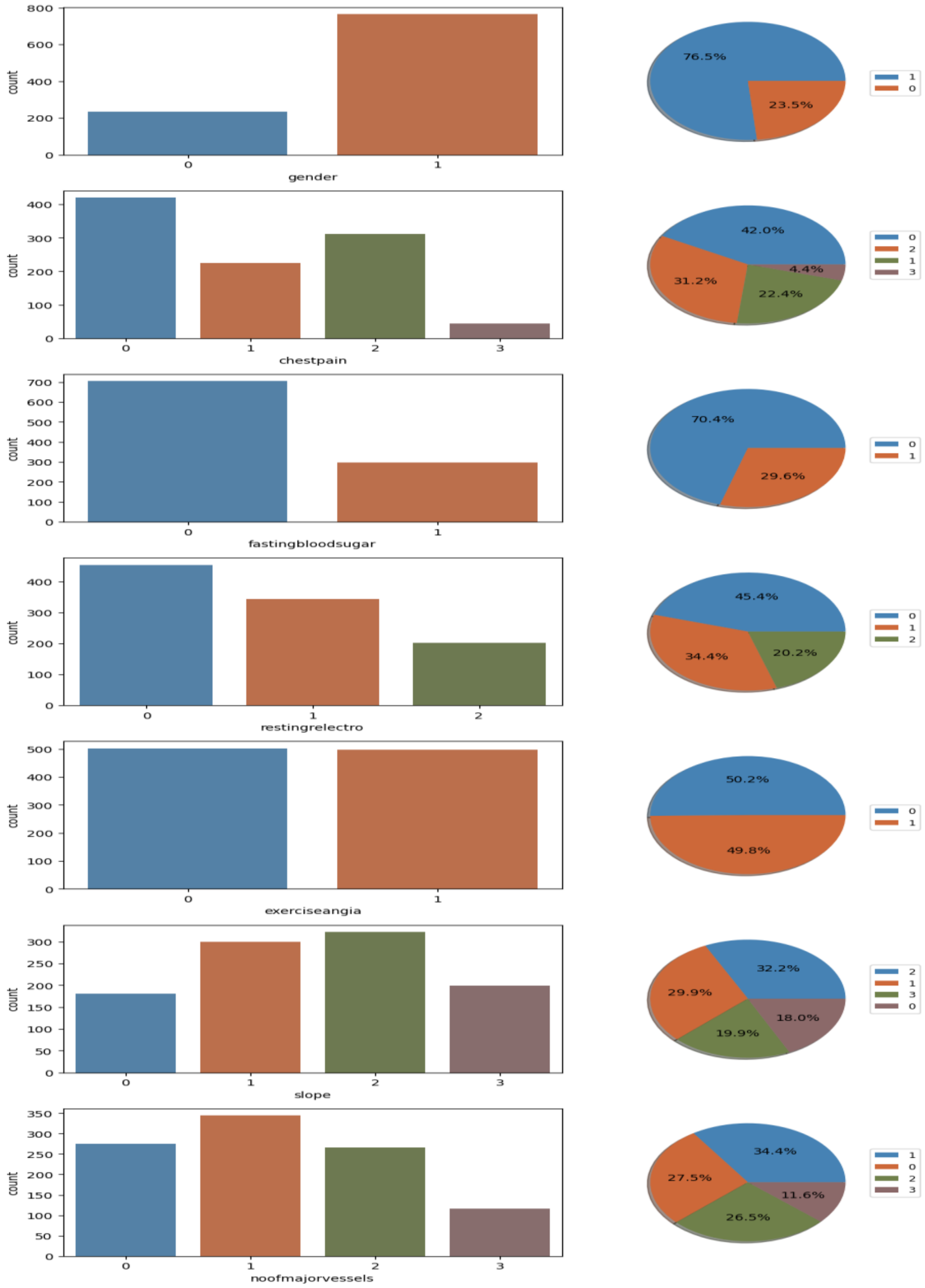


Table 5

Data Description

Feature	Description
Patient ID	Unique identifier for each patient.
Age	Ranges from 20 to 80 years, with a mean age of 49.24 years.
Gender	Represents the gender status, showing a significant proportion of males.
Chest Pain	Recorded on a scale from 0 to 3.
Resting Blood Pressure	Ranges from 94 to 200 mmHg, with a mean of 151.75 mmHg.
Serum Cholesterol	Ranges from 0 to 602 mg/dL, with a mean of 311.45 mg/dL.
Fasting Blood Sugar	Recorded as a binary value (0 or 1), where 1 indicates elevated fasting blood sugar.
Resting Electrocardiogram	A binary value representing the presence or absence of changes in heart response.
Max Heart Rate	Ranges from 71 to 202 beats per minute, with a mean of 145.48 bpm.
Exercise Angina	A binary value indicating the presence of chest pain during exertion.
Old Peak	Ranges from 0 to 6.2, with a mean of 2.71.
Slope	Represents the slope of the exercise graph, ranging from 1 to 3.
Number of Major Vessels	Ranges from 0 to 3.
Target	Indicates whether the patient has heart disease (0 or 1).

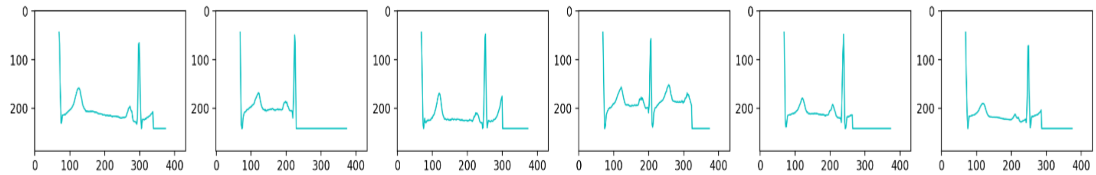
3.1.2 Image Data

The ECG contains many distinct physiological data that can be used to identify and encode a wide range of medical disorders. Since ionic currents are frequently affected very early in many disease processes, adding artificial intelligence (AI) to a standard ECG, a common, inexpensive test that doesn't require bodily fluids or reagents, transforms it into a powerful diagnostic screening tool that may also allow monitoring and assessment of response to therapy. (Attia and others, 2021).

The data was obtained through the website Kaggle, where this data describes the patient's ECG test. The data is collected by the emergency department specialist, containing the ECG condition of the patient. The data set is classified into two type, normal and abnormal (myocardial). This figure shows the sample from the ECG picture(*Ecg_image_data*, n.d.)

Figure 26

The ECG Sample (Normal and Abnormal) (Ecg_image_data, n.d.)



3.2 Data Processing

3.2.1 Numerical Data (Symptoms)

After collecting the symptom data, we perform a statistical description to understand the data and its characteristics. This involves conducting a descriptive analysis of the data to determine the overall distribution of each variable. In this process, we calculate statistical measures such as the mean, median, standard deviation, and others that help describe the data and identify key features.

Next, we examine the correlation between variables. With values ranging from -1 to +1, correlation is a metric used to assess the direction and strength of a linear relationship between two variables. A value of +1 indicates a strong positive relationship, meaning both variables increase together. A value of -1 indicates a negative or inverse relationship, where one variable decreases as the other increases. Correlation helps to understand which variables have a strong relationship that can be relied upon to create the model.

3.2.1.1 Preprocessing Techniques

We used several techniques to understand and organize the data (data cleaning) before processing it and finding the relationship between the features. The following are the techniques used for (data cleaning):

1 Standard Scaler: It normalizes the data so that each feature exhibits a distribution with a mean of 0 and a standard deviation of 1. Support Vector Machines (SVM) and other models influenced by feature values primarily use it.

2-MinMaxScaler: This tool normalizes the data to a specified range, typically between 0 and 1, thereby preserving the original distribution. It is particularly useful for models that require feature values within a specific interval.

3-LabelEncoder: This tool converts categorical data into numerical values. For example, it will convert category qualities like "male" and "female" into numeric values, specifically 0 and 1. This strategy is crucial for handling categorical data in models that necessitate exclusively numeric inputs.

4- Train_test_split function divides the dataset into training and testing subsets. The training set is utilised to develop the model, while the test set is employed to assess the model's performance. We divided the data into various ratios of 20-80%, 30-70%, and 40-60% to identify the optimal model among them.

5-GridSearchCV: This tool identifies the model's optimal hyperparameters. This evaluates many configurations and subsequently identifies the optimal one based on the model's performance.

3.2.1.2 Correlation

Correlation coefficients between features shown in the heatmap:

Age: Moderate correlation with target (-0.23), indicating that age may have a small effect on the likelihood of developing the target condition. Weak to moderate correlation with chest pain (-0.06), meaning that age is not strongly associated with chest pain.

Gender: Moderate correlation with target (-0.04), meaning that gender does not have a strong effect on developing the target condition. Weak to moderate correlation with chest pain (0.5), which may indicate that there is some association between gender and chest pain.

chest pain: Relatively strong correlation with target (0.53), suggesting that chest pain may be an indicator of the target condition. Weak correlation with other features, except for a moderate correlation with number of Major Blood Vessels.

resting BP: Weak correlation with target (0.04), meaning that resting blood pressure is not strongly associated with the target condition. Weak to moderate correlation with some features such as age and cholesterol, indicating small associations.

serum cholesterol: Very weak correlation with target (0.02), meaning that cholesterol level has no significant effect on target condition. Weak to moderate correlation with age and resting blood pressure.

fasting blood sugar: Very weak correlation with target (0.04), indicating that fasting blood sugar does not significantly affect the incidence of the condition.Weak correlation with other features as well.

resting electro: Very weak correlation with target (0.14), indicating that resting ECG has a small effect. Weak correlation with other features.

max heartrate: Moderate negative correlation with target (-0.43), meaning that maximal heart rate is inversely correlated with target condition. Weak correlation with other features.

Exercise angina: Moderate correlation with target (0.49), meaning that exercise-induced angina has a moderate effect on target condition. Moderate to weak correlation with some features such as max heart rate.

old peak: Moderate correlation with target (0.43), indicating that ST slope may have a significant effect. Weak correlation with other features.

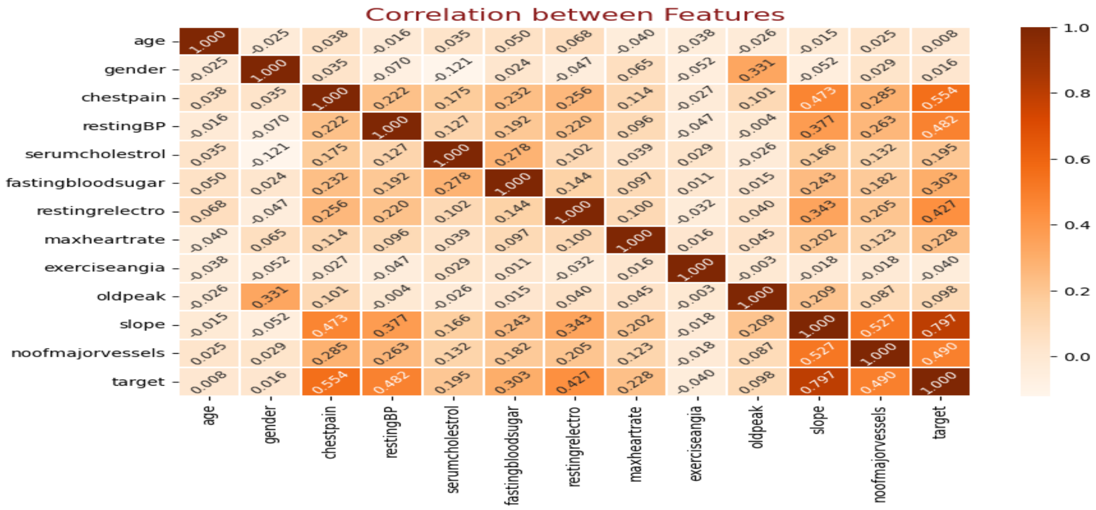
Slope: Slight correlation with target (0.3), indicating that slope does not have a strong effect. Weak correlation with most features.

No of major vessels: Relatively strong correlation with target (0.46), indicating that number of major blood vessels may be a good predictor.Weak to moderate correlation with some features.

Target: Shows various relationships with features, with the strongest correlations being with chest pain, number of major blood vessels, maxheartrate, and exercise angina. This gives a clear explanation that these symptoms are closely related to the patient having a myocardial infarction when he has chest pain, number of major blood vessels, maxheartrate, and exercise angina.

Figure 27

Correlation Between Features



3.2.2 Image Data

For myocardial, the images are labeled into 2 groups, positive (myocardial) and negative (normal) as shown in figure 3.5 ECG images, The images acquired range from 2 mega bites (KB) to 6KB, with pixel sizes (based on horizontal and vertical) 432X288. These images are too big to fit into several Models. However, to reduce them to 224X224 pixels size, we use Python code to reduce the size after resizing the images to the required pixel size the images are fed into Google Collab using the Python codes.

Figure 28

Images Label

positive (myocardial) negative (normal)

Table 6

Number of Images

Image type	Number of images	test	Train	total
Normal	4,046	808	3,238	14,552
Myocardial infarction	10,506	2,101	8,405	

3.2.2.1 Min_max_scalar

We used the function "min_max_scalar" normalized to the range [0, 1]: This step converts the pixel values to the range [0, 1] using the function min_max_scalar(), this function loads the image from its path using the Pillow library, and changes the images to (0, 1).

3.2.2.2 Images Resizing

We used the "process_image" function to reduce the size of the images. This function loads the image from its path using the Pillow library. It resizes the image to the dimensions specified in the size variable, which is (224, 224) pixels in this case. It uses Image.Resampling.LANCZOS (or Image.LANCZOS depending on the library version) to improve the quality of the image when it is reduced. It then converts the image to grayscale by selecting the gray channel (one channel). We selected one channel instead of three (RBG) channels to avoid extracting additional unreal features when applying the three filters. Finally, it converts the image to a NumPy array and reshapes it to the size (224, 224, 1).

3.3 Data Split

3.3.1 Numerical Data (Symptoms)

We divided the data into three sections, The first section: 20% test, 80% training, The second section: 30% test, 70% training, The third section: 40% test, 60% training, And find the best results.

3.3.2 Image Data

Splitting datasets is an essential step in the process of training and proving the effectiveness of AI models. Researchers recommend splitting data into different ratios such as 70:30, 75:25, and 80:20 to strike a balance between training and evaluation. The choice of split ratio often depends on the size and diversity of the data, with some studies preferring 80:20 to have more training data, while others argue that 70:30 gives models the opportunity to test on less-seen data, which improves their ability to generalize. In this context, we chose a 70:30 split to have a large test set that helps evaluate the model's performance on previously untrained data and gives us flexibility in analysis and development.

3.4 Model Training

We used the Google CoLab platform to train the model with 70% of the dataset; this cloud-based infrastructure is accessible via a web browser and provides users with robust computational capabilities. Google Colab offers access to CPUs, GPUs, and advanced TPUs, rendering it an optimal tool for expediting the training of computationally intensive models, such as those in deep learning. We initiated a GPU-optimized operating environment in this experiment to speed up training and reduce the time required to develop the model from the data.

We designated the remaining 30% of the data as the evaluation or test dataset, making it inaccessible to the model during the training phase. This assesses the model's efficiency and generalization capability by enabling the evaluation of its performance on new data not used during training. We refer to this technique as cross-validation, where we set aside a segment of the data to ensure the model not only memorizes the data but also acquires generalizable patterns from it.

This study employed a total of 10 models, training five distinct ones on two categories of data: numerical and image data. We use each pre-trained model on a designated dataset to experiment with various approaches and facilitate performance comparison. This approach of using multiple models allows for more accurate and diverse results, which enhances the accuracy and efficiency of the model in predicting and processing data.

3.4.1 Model for Numerical Dataset

3.4.1.1 Logistic Regression

It is a model used to estimate the probability of a given event occurring. Although its name is "regression", it is actually used to solve binary classification problems (where the outcome is one of two categories, such as "yes" or "no", "sick" or "not sick").(Li & Liu, 2022)

3.4.1.2 Support Vector Machine

The system has the capability to analyze a dataset and generate predictions for each entry based on its attributes. This method is a non-probabilistic linear binary classification strategy. Each output class is associated with or contingent upon a specific category. (Faieq & Mijwil, 2022)

3.4.1.3 K-Nearest Neighbors

KNN is a non-linear algorithm based on the principle of finding "nearest neighbors" to make a classification or prediction decision. The algorithm is based on comparing the new example (that you want to classify or predict) with pre-existing examples in the training set, based on certain distances, and choosing the class or value based on the "K nearest neighbors". The KNN algorithm uses neighborly classification as the predicted value of the new test sample.(Damayunita et al., 2022)

3.4.1.4 Decision Tree

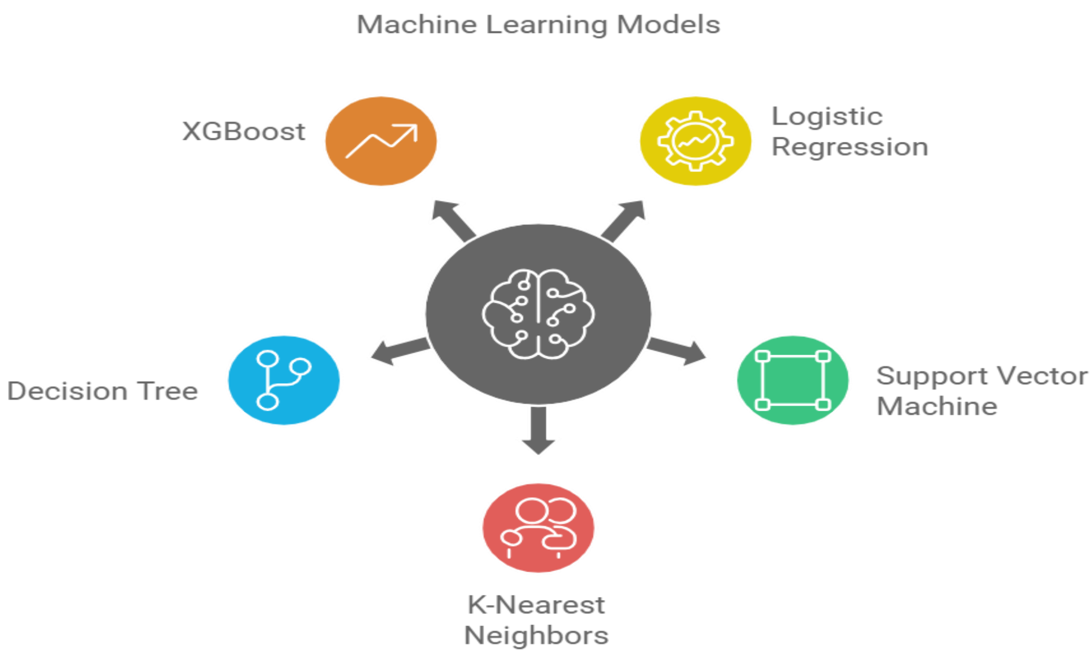
The Decision Tree is a machine learning technique that employs a tree-like framework for classification and regression tasks. This method addresses issues by illustrating decision rules as a tree, with each branch signifying a decision based on particular criteria. A decision tree operates by partitioning the dataset according to the most salient attribute to categorize the data into distinct classes. Each branch of the tree signifies a condition or rule that the data must fulfill. The outcomes of each branch are predictions or determinations made by the algorithm. (Damayunita et al. 2022)

3.4.1.4 XGBoost

Reliable pile load test data serve as the foundation for the XGBoost algorithm. The following elements are included in the current study: Our goal is to create a model that accurately captures the intricate link between the axial pile carrying capacity and the elements that influence it. (Amjad and others, 2022).

Figure 29

Machine Learning Model



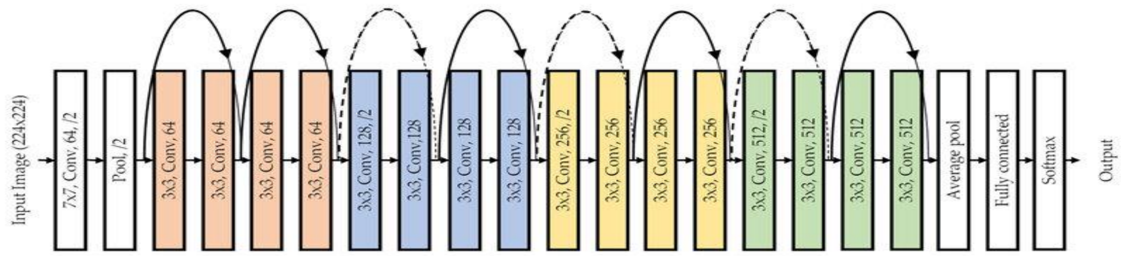
3.4.2 Model for image dataset

3.4.2.1 ResNet-18

It is a variant of deep neural networks created by a group of researchers at Microsoft in 2015. "ResNet" is an abbreviation for Residual Network, and 18 denotes the quantity of layers constituting this network. The convolution operation, a core element of a ResNet, extracts features by applying a filter, often referred to as a convolution kernel or feature detector, to the input data. The main thing that a ResNet does is move the filter of a convolution layer over the input image. This creates a feature map by multiplying and adding each element in the input image to the filter. (Wang et al., 2024)

Figure 30

ResNet-18 Structure (Brown et al., 2023)

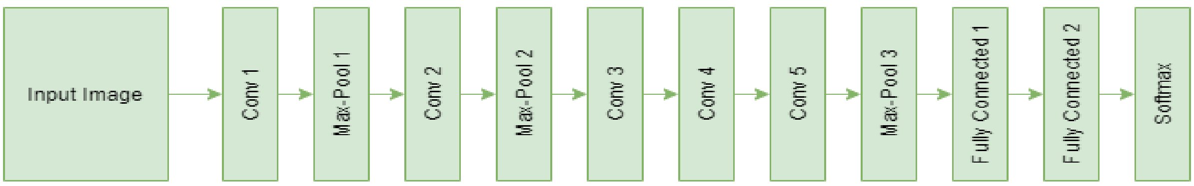


3.4.2.2 AlexNet

Alex Krizhevsky et al. created AlexNet, a deep neural network, in 2012. The goal was to sort photos into groups for the ImageNet LSVRC-2010 challenge, and it won. Chen et al. (2022) It has garnered substantial recognition in the field of computer vision because of its exceptional performance in the ImageNet image categorization competition. AlexNet has produced considerable progress relative to previous models and has markedly increased research interest in convolutional networks and their various applications. Singh et al. (2022) provide an illustration of the AlexNet architecture.

Figure 31

AlexNet Structure (ML | Primeiros Passos Com AlexNet – Acervo Lima, n.d.)



3.4.2.3 Hyper CNN

SVMs (Support Vector Machines) effectively integrate with CNNs (Convolutional Neural Networks) to enhance classification accuracy. The combination of CNN and SVM significantly enhances performance in picture classification tasks. CNN identifies critical attributes in images, whereas SVM categorizes them with high precision.

3.4.2.4 LeNet

LeNet is one of the first deep neural networks to be proven effective for computer vision tasks. The LeNet architecture is a convolutional neural network (CNN) specifically developed for image recognition tasks. Developed by Yann LeCun and colleagues in 1998, it has emerged as a core paradigm in deep learning. The design comprises convolutional layers, pooling layers, and fully linked layers. (Khairina et al., n.d.)

Figure 32

LeNet Structure (Key Deep Learning Architectures: LeNet-5 | by Max Pechyonkin | Medium, n.d.)

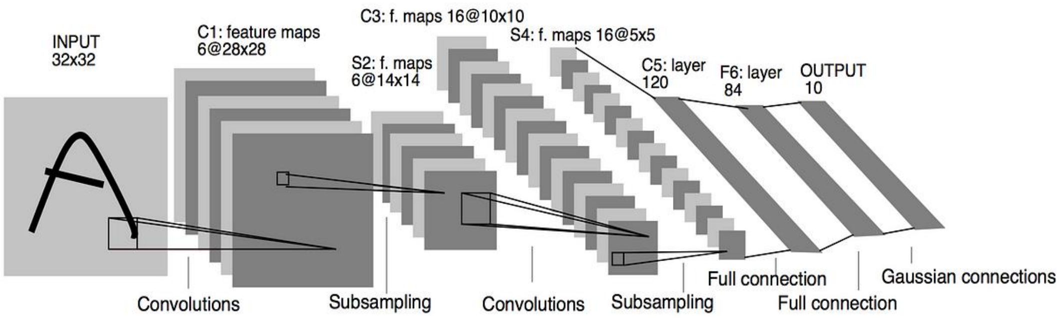
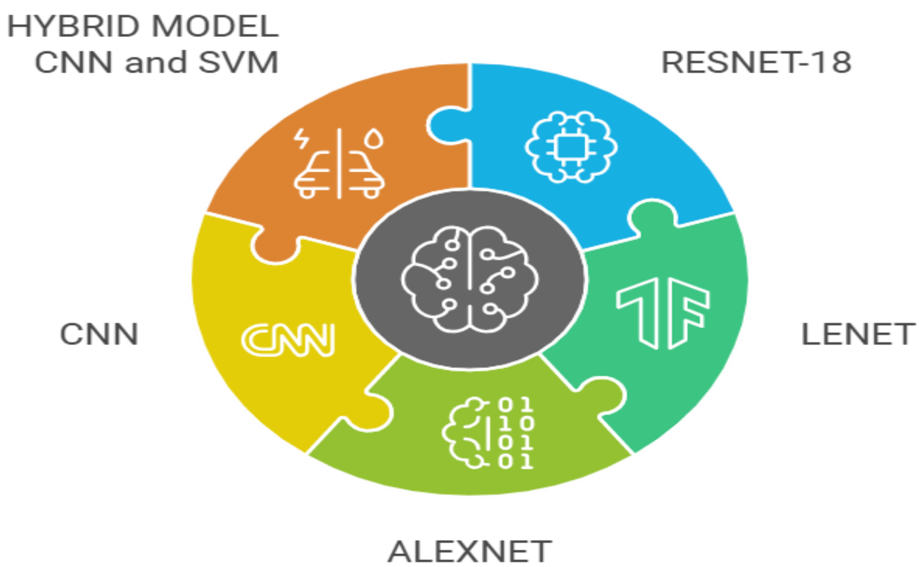


Figure 33

Overview of Image Classification Model



CHAPTER IV
Results

4.0 OVERVIEW

The model was trained according to the type of data used, as different algorithms were used for numerical data to get the best performance and also for image data. Each model was trained on a different number of layers and on different methods, and this difference enhances the efficiency of learning the model. model's parameters for training the
models include:

1. Batch size refers to the number of samples or data (such as images or examples) that the model processes at each training step before updating the model weights. In other words, instead of passing all the data in at once, the data is split into smaller batches (batches), and the model computes predictions and updates for each batch separately. 2. Epoch: This is calculated by dividing the total number of training photos by the product of the batch size and the number of iterations. 3. The number of iterations per epoch is determined by dividing the maximum by the number of epochs.

4.1 Myocardial Infarction

4.1.1 Numerical Data (Symptoms)

We trained the model on the data after we did the preprocessing process which includes features selection through correlation to get values of accuracy. 94.25**%**, 90.909%, 89.95%**.**

4.1.2.1 Algorithmic

4.1.2.1.1 Logistic Regression

It is a statistical method that uses a set of independent variables to predict binary outcomes. Despite its name, it is not a type of regression in the traditional sense, but is actually used as a classification method. Additionally, it is employed to translate predicted values into the likelihood that they fall into a particular class. The logistic function is given by:(G et al., 2022)

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

4.1.2.1.2 Support Vector Machine

SVM is a well-liked machine learning method for regression and classification. The Support Vector Machine (SVM) is notable for employing "interclass distance" to define the boundaries that differentiate multiple classes within the dataset. It seeks to identify the optimal hyperplane that separates several classes. The goal is to identify the class boundaries while maximizing the margin between the nearest points in each class. The subsequent expression delineates the mathematical formulation of the objective: (Ahmed et al., 2023)

$$\mathrm{f(x) = wTx + b}$$

4.1.2.1.3 K-Nearest Neighbors

KNN is a machine learning technique used for value estimation or data classification. The fundamental idea behind this technique is to identify a selected data point's closest neighbors (hyperplane) based on a certain distance. The following mathematical equation is an expression for it: (Ahmed et al., 2023)

$$\mathrm{y\hat{} = K1i = 1\sum Kyi}$$

4.1.2.1.4 Decision Tree

This machine-learning algorithm is utilized for classification and regression applications. We employ it to construct a model capable of making decisions based on a collection of rules derived from data. Applications for this machine learning algorithm include regression and classification. We utilize it to build a model that can make decision based on several dataset. (Shaik et al., 2023)

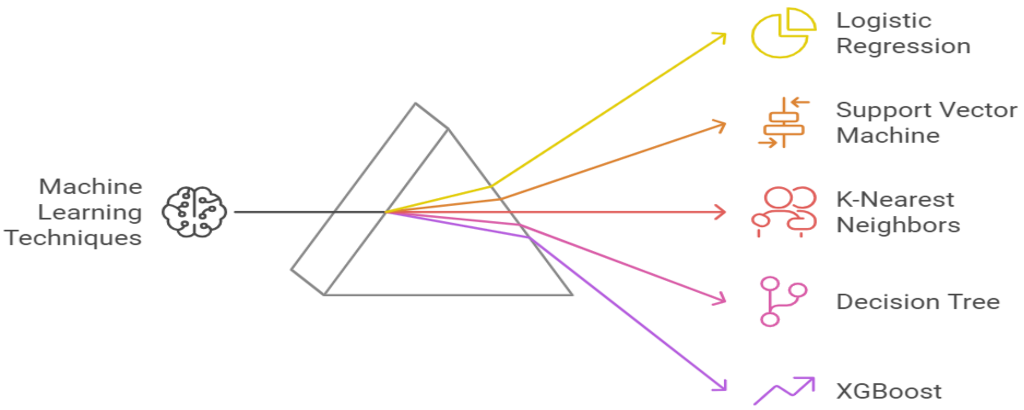
4.1.2.1.5 XGBoost

Extreme Gradient Boosting is a sophisticated machine-learning technique that is frequently used in prediction and classification because of its remarkable efficacy and efficiency. This strategy makes use of the gradient boosting technique, which involves building several basic models—typically decision trees—and combining or develop them to enhance performance. XGBoost

At first, a simple model is created that predicts the training results (usually a small decision tree). The errors generated by this initial model are then evaluated, focusing on examples that are difficult to predict correctly.(Arif Ali et al., 2023)

Figure 34

Machine Learning Techniques

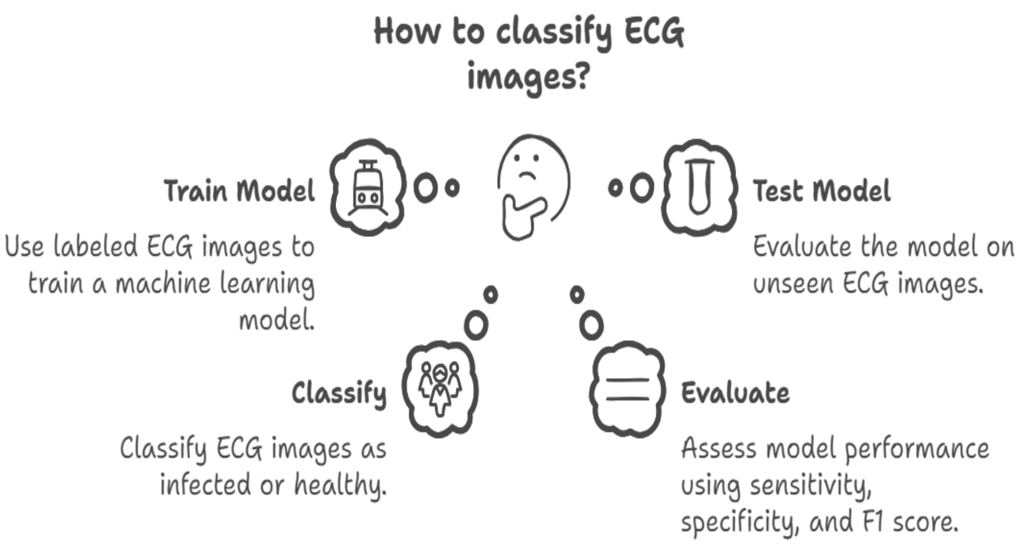


4.2.1 Image Data

The model's training resulted in a maximum of 10 completed epochs and 32 batch_size per epoch. According to the examination of the model's performance, it acquired a training accuracy 99%, a validation accuracy of 96.79%, a sensitivity of 98%, and a specificity of 98.69%. Figure 4.2 describes the obtained outcome.

Figure 35

Classification of Myocardial Infarction



4.2.1.2 Algorithmic architecture

We have actually used four different algorithms for each different architecture algorithm.

4.2.1.2.1 RESNET 18

We built and trained a Convolutional Neural Network (CNN) model based on the ResNet18 architecture, focusing on classifying ECG signals extracted as images.

ResNet18 Model:

We loaded the pre-trained ResNet18 model on a large dataset, and the final layer was modified to suit our task (binary classification).

The ResNet architecture contains several layers, including Conv2d, BatchNorm2d, and ReLU layers, which extract features from images, in addition to using skip connections to mitigate gradient vanishing problems.

Training: We used CrossEntropyLoss as the loss function, and Adam's algorithm as a method to improve the model.

We trained the model on the training data for 10 epochs, where the loss and accuracy were calculated in each epoch.

Evaluation: After training, the model was evaluated on the test data, where the accuracy was calculated and the confusion matrix and classification report were used to measure performance.

Measures such as: Accuracy (Test) 99%, Accuracy (Training) 99%, Sensitivity (Sensitivity) 1, Specificity (Percentage) 1, Mean Squared Error (Mean Squared Error) 0 were calculated.and the table show the Structure of ResNet-18 we used

Table 7

ResNet-18 Structure

Layer (type)	Output Shape
Conv2d-1	[-1, 64, 112, 112]
BatchNorm2d-2	[-1, 64, 112, 112]
ReLU-3	[-1, 64, 112, 112]

MaxPool2d-4	[-1, 64, 56, 56]
Conv2d-5	[-1, 64, 56, 56]
BatchNorm2d-6	[-1, 64, 56, 56]
ReLU-7	[-1, 64, 56, 56]
Conv2d-8	[-1, 64, 56, 56]
BatchNorm2d-9	[-1, 64, 56, 56]
ReLU-10	[-1, 64, 56, 56]
BasicBlock-11	[-1, 64, 56, 56]
Conv2d-12	[-1, 64, 56, 56]
BatchNorm2d-13	[-1, 64, 56, 56]
ReLU-14	[-1, 64, 56, 56]
Conv2d-15	[-1, 64, 56, 56]
BatchNorm2d-16	[-1, 64, 56, 56]
ReLU-17	[-1, 64, 56, 56]
BasicBlock-18	[-1, 64, 56, 56]
Conv2d-19	[-1, 128, 28, 28]
BatchNorm2d-20	[-1, 128, 28, 28]
ReLU-21	[-1, 128, 28, 28]
Conv2d-22	[-1, 128, 28, 28]
BatchNorm2d-23	[-1, 128, 28, 28]
Conv2d-24	[-1, 128, 28, 28]
BatchNorm2d-25	[-1, 128, 28, 28]
ReLU-26	[-1, 128, 28, 28]
BasicBlock-27	[-1, 128, 28, 28]
Conv2d-28	[-1, 128, 28, 28]
BatchNorm2d-29	[-1, 128, 28, 28]
ReLU-30	[-1, 128, 28, 28]
Conv2d-31	[-1, 128, 28, 28]
BatchNorm2d-32	[-1, 128, 28, 28]
ReLU-33	[-1, 128, 28, 28]
BasicBlock-34	[-1, 128, 28, 28]
Conv2d-35	[-1, 256, 14, 14]
BatchNorm2d-36	[-1, 256, 14, 14]
ReLU-37	[-1, 256, 14, 14]
Conv2d-38	[-1, 256, 14, 14]
BatchNorm2d-39	[-1, 256, 14, 14]
Conv2d-40	[-1, 256, 14, 14]
BatchNorm2d-41	[-1, 256, 14, 14]
ReLU-42	[-1, 256, 14, 14]
BasicBlock-43	[-1, 256, 14, 14]
Conv2d-44	[-1, 256, 14, 14]
BatchNorm2d-45	[-1, 256, 14, 14]
ReLU-46	[-1, 256, 14, 14]
Conv2d-47	[-1, 256, 14, 14]
BatchNorm2d-48	[-1, 256, 14, 14]
ReLU-49	[-1, 256, 14, 14]
BasicBlock-50	[-1, 256, 14, 14]
Conv2d-51	[-1, 512, 7, 7]

BatchNorm2d-52	[-1, 512, 7, 7]
ReLU-53	[-1, 512, 7, 7]
Conv2d-54	[-1, 512, 7, 7]
BatchNorm2d-55	[-1, 512, 7, 7]
Conv2d-56	[-1, 512, 7, 7]
BatchNorm2d-57	[-1, 512, 7, 7]
ReLU-58	[-1, 512, 7, 7]
BasicBlock-59	[-1, 512, 7, 7]
Conv2d-60	[-1, 512, 7, 7]
BatchNorm2d-61	[-1, 512, 7, 7]
ReLU-62	[-1, 512, 7, 7]
Conv2d-63	[-1, 512, 7, 7]
BatchNorm2d-64	[-1, 512, 7, 7]
ReLU-65	[-1, 512, 7, 7]
BasicBlock-66	[-1, 512, 7, 7]
AdaptiveAvgPool2d-67	[-1, 512, 1, 1]
Linear-68	[-1, 2]

4.2.1.2.2 LENET

The “LeNet” model is designed as a convolutional neural network using the Keras library. It includes a set of convolutional and pooling layers, which contribute to extracting features from images and processing data efficiently.

Model structure:

Convolutional layer (Conv2D): The first layer: contains 6 filters (filters) with size 5x5, with the ReLU function activated, takes the input images with dimensions (224, 224, 1) (where 1 refers to the gray channel). The third layer: contains 16 filters with the same size (5x5) with the ReLU function activated. The fifth layer: contains 120 filters with size 5x5 with the ReLU function activated.

The pooling layer (MaxPooling2D): Two layers were used for pooling, The second layer: pools the data using a pool size of 2x2, the fourth layer also pools the data using a pool size of 2x2.

Layer for Flattening: This layer converts the 2 Dimensions feature maps generated by the convolutional layers into 1 Dimensions vector, making it easier to pass the data to the fully connected layers.

Fully Connected Layers: Layer 7 Contains 84 units with ReLU function activation, where all units from the previous layer are connected, Layer 8 It is the output layer that contains 2 units for binary classification, with softmax function activation used to distribute the probabilities between classes.

Model Assembly: We constructed the model using the Adam optimizer, a widely recognized and efficient optimizer for training neural networks. We employed the categorical_crossentropy loss function, which is suitable for multi-class classification issues.

Summary of Results: We used a test set to evaluate the model's performance after training it on an image dataset. The outcomes highlight the effectiveness of the "LeNet" architecture in pattern recognition and show how well the model can classify photos. The model achieved 99.05% specificity, 98.39% sensitivity, 99% training accuracy, and 98.79% test accuracy.

4.2.1.2.3 CNN

The Keras toolkit is used by the Convolutional Neural Network (CNN) model to classify ECG data that is represented as images. Data preparation is the first step in the multi-phase model creation process, which ends with model training and evaluation.

Model Structure:

Convolutional Layers: First layer: Contains 32 filters of size 3x3 with ReLU enabled, second layer Contains 64 filters of size 3x3, third layer Contains 128 filters of size 3x3.

Pooling Layers: After each convolutional layer, a 2x2 MaxPooling2D layer was used to reduce the dimensionality of the data and reduce complexity.

Flatten Layer: Converts the output from the convolutional layers to a one-dimensional vector.

Fully Connected Layers: Fully Connected Layer with 128 units with ReLU enabled and 50% Dropout layer to reduce overfitting, The output layer includes a single unit with sigmoid function activation to suit binary classification.

Model assembly: The model was assembled using Adam optimizer, using binary_crossentropy loss function, It works well for jobs involving binary categorization.

4. Model training: The model was trained using the training data, with its performance evaluated on the test dataset, the number of epochs was set to 10.

5. Model evaluation: The model was evaluated using the test dataset, and the accuracy of the model was calculated, where the accuracy and missing value for both training and validation of the model are displayed.

Confusion Matrix was calculated to evaluate the performance of the model at the class level, and metrics such as: Accuracy Test 97.67%, Accuracy Training 98%, Sensitivity 97.01%, Specificity 98.78%, Mean Squared Error 0.0171

The training results show that the model achieved good accuracy, indicating the effectiveness of its structure in processing and successfully classifying ECG data. And the table show the Structure of CNN we used

4.2.1.2.4 ALEXNET

An image classification model was developed using Convolutional Neural Network (CNN) based on AlexNet architecture. Building the model involves several stages, starting from data processing to training and evaluation.

1. Data preparation Different libraries such as TensorFlow, NumPy, and Pandas were imported to process the data and build the model. Custom functions were used to load and process images from ECG data folders. These functions include transforming the images to a specific range (0-1) and resizing them to 224x224 pixels, using only the gray channel. A training set (70%) and a test set (30%) were created from the dataset.

2. Model building The AlexNet model was built using Keras library. The model consists of: Convolutional layers (Conv2D): Multiple convolutional layers

were used to select features from the images, where 96, 256, and 384 filters were added across the different layers, Pooling layers (MaxPooling2D): These layers were included to reduce the spatial map dimensions and reduce the computational complexity, Flatten Layers: This layer was used to transform the spatial maps into 1D vectors, allowing the data to be passed to the connected layers. Dense Layers: Two connected layers were used, each containing 4096 units, with ReLU activation function and 50% Dropout to avoid overfitting, Output Layer: It consists of two units that utilize the SoftMax activation mechanism, which facilitates binary classification.

3. Model Clustering

The Adam algorithm and categorical cross-entropy loss optimized the model, making it suitable for a variety of classification applications.

4. Model Training

The model underwent training on the training set for 10 epochs utilizing a batch size of 100. We evaluated the model's performance by analyzing accuracy and loss during each interval.

5. Model Evaluation

We estimated accuracy and loss on the test set. The confusion matrix and classification report provided more insights into the model's accuracy and F1 score. The model identified ECG images in the final evaluation, demonstrating that convolutional neural networks like AlexNet can be powerful medical image processing tools. The model was 98.79% accurate, highlighting the value of deep learning in therapy. The table shows this model's algorithm architecture.

Table 8

AlexNet 8 Structure

Output Shape	Layer (type)
(None, 54, 54, 96)	conv2d_15 (Conv2D)
(None, 26, 26, 96)	max_pooling2d_9 (MaxPooling2D)
(None, 26, 26, 256)	conv2d_16 (Conv2D)
(None, 12, 12, 256)	max_pooling2d_10 (MaxPooling2D)
(None, 12, 12, 384)	conv2d_17 (Conv2D)
(None, 12, 12, 384)	conv2d_18 (Conv2D)
(None, 12, 12, 256)	conv2d_19 (Conv2D)

(None, 5, 5, 256)	max_pooling2d_11 (MaxPooling2D)
(None, 6400)	flatten_3 (Flatten)
(None, 4096)	dense_9 (Dense)
(None, 4096)	dropout_6 (Dropout)
(None, 4096)	dense_10 (Dense)
(None, 4096)	dropout_7 (Dropout)
(None, 2)	dense_11 (Dense)

4.2.1.2.5 CNN and SVM

This model combines the strengths of a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM) for picture classification.

1-CNN

Convolutional neural networks, or CNNs, are a particular kind of neural network made specially to evaluate data with a structured format, like pictures. Convolutional Neural Networks excel in classification tasks, especially with images. Convolutional Layers: Convolutional layers employ filters to analyze patterns inside the image, such as edges, corners, and complex features. These filters enable the discernment of the essential attributes inside the image.

Pooling Layers: Techniques such as max pooling diminish data complexity while maintaining its essential features. This reduces data size, enhances processing speed, and diminishes the likelihood of overfitting.

Flatten Layer: This layer converts the two-dimensional matrices produced by the preceding layers into a one-dimensional matrix. This facilitates the transmission of the matrix to the fully connected layers responsible for feature classification.

Fully Connected Layers: After feature extraction, the data is relayed to dense layers for classification. The last layer employs Softmax activation to produce probability for the different classes and consists of a number of neurons corresponding to the class count.

Classifying Images Using Convolutional Neural Networks

Automated feature extraction: By learning and extracting key features from images on their own, CNNs eliminate the requirement for human feature selection. Increased image processing efficiency: Convolutional layers quickly examine incoming images to identify spatial patterns within the image.

Overfitting is less likely with CNNs since they employ shared filters and pooling layers, which lower the number of weights.

2. Using SVM to classify extracted features After extracting features from images using CNN, these features can be used to train another classification model. In this model, SVM was used to classify the extracted features for the following reasons:

High-accuracy classification: SVM is a powerful classification model, which works well with high-dimensional data, such as the features extracted from CNN. SVM finds an optimal separation between different classes by increasing the margin between data points. minimize the effect of overfitting: SVM has strong generalization capabilities by identifying the best boundary that differentiates classes, which refers to minimizing overfitting to the training dataset. Handle linearity and non-linearity features: SVM uses either a linear or nonlinear kernel (like RBF or polynomial) to classify the more complex features that can't be separable linearly.

Model operation

1. Feature extraction: CNN identifies significant characteristics from images. The flattening layer transmits the characteristics to the dense layers for classification preparation.

2. Classification using Support Vector Machines (SVM): The SVM model uses these characteristics as inputs. The SVM processes these features and categorizes them into the designated classes.

3. achieve accuracy: SVM delivers superior performance in classification, thereby enhancing the total precision of the model. The combination of CNN for feature extraction and SVM for classification shows the effectivity.

This model combines CNN and SVM to utilize CNN's capacity for feature extraction from images and SVM's expertise in precise classification of these characteristics. This method is exceptionally efficient for medical image classification tasks, as it combines the advantages of automatic feature extraction in CNNs with the enhanced classification efficacy of the SVM model. Table 4.3 shows the Structure of these combinations.

Table 9

CNN and SVM Structure

Layer (type)	Output Shape	Param
conv2d (Conv2D)	(None, 220, 220, 6)	156
max_pooling2d (MaxPooling2D)	(None, 110, 110, 6)	0
conv2d_1 (Conv2D)	(None, 106, 106, 16)	2416
max_pooling2d_1 (MaxPooling2D)	(None, 53, 53, 16)	0
flatten (Flatten)	(None, 44944)	0
dense (Dense)	(None, 120)	5,393,400
dense_1 (Dense)	(None, 84)	10,164
SVM		Classification form

4.3 Model Description

We utilize several parameters, like precision, recall, and F1-score criteria, to assess the outcomes of classifying various cardiac diseases.

Precision is the proportion of accurately anticipated positive instances to the total expected positive instances. It evaluates the precision of the model's positive predictions, indicating the proportion of accurately identified positives. Equation 5-1 presents the formula for determining precision:

$$Precision = TP/(TP + FP) \tag{4-1}$$

This statistic becomes crucial when the rate of false positives is significant, as it ensures the reliability of positive predictions.

Recall The ratio of correctly expecting positive cases to the total number of actual positive instances is measured by the recall, which is also known as sensitivity or true positive rate. It assesses how well the model can identify all relevant positive cases. Equation 5-2 presents the formula for computing recall.

$$Recall = TP/(TP + FN) \tag{4-2}$$

Recall is essential in situations where the omission of positive cases (false negatives) incurs significant costs, since it highlights the necessity of identifying as many true positives as feasible.

The F-score, or F1 score, is the harmonic mean of precision and recall, yielding a singular metric that equilibrates both measures. Equation 5-3 delineates the F-score calculation:

$$F_1 Score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \tag{4-3}$$

The F-score is useful for imbalanced datasets, as it combines both precision and recall to provide a more comprehensive assessment of a model's effectiveness. It also ensures that both precision and recall are taken into account, making it a critical metric for the overall evaluation of classification models.

4.3.1 Confusion Matrix

The confusion matrix is an important technique for evaluating the model's performance depending on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

True positives (TP): refer to the number of samples that the model accurately classifies as positive cases, or the number of cases showing the presence of myocardial in action.

true negatives (TN): The number of samples that the model correctly identified as negative cases, that is, healthy (normal) people that were correctly classified as Not Infected

False positives (FP): The number of samples that the model misclassifies as negative. This includes true negative (normal or healthy) cases that were mistakenly categorized as myocardial infarction.

False negatives (FN): represent the number of samples that were incorrectly identified as positive by the model, i.e., true positives (myocardial infarction) cases that were incorrectly classified as normal or healthy, as illustrated in Table 10.

Table 10

Confusion Matrix

Predictions	Actual Positive	Actual Negatives
Positive Predictions	TP	FP
Negative Predictions	FN	TN

4.3.2 ROC

ROC stands for Receiver Operating Characteristic, and is a curve used to evaluate the performance of classification models, especially in binary classification (i.e. classifying instances into two classes: positive and negative). The ROC curve helps visualize the relationship between sensitivity and specificity across different threshold levels.(Carrington et al., 2023)

4.4 Performance Evaluation

4.4.1 Numerical Data

4.4.1.1 Confusion Matrix

The confusion matrix is a popular technique for evaluating the model's performance based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TPs stands for the number of samples that the model correctly identified as positive cases, or more precisely, the number of people who experienced myocardial infarction in each model. TNs stand for the number of samples that the model correctly classified as negative cases or the number of people who are healthy (normal) but are classified as negative by each model. False positives (FPs) are the number of samples that the model incorrectly classifies as negative or the cases that are negative (normal or healthy) but are nonetheless classified as myocardial infarction. False negatives (FNs) are the number of samples that the model got wrongly as positive. These are the cases where the sample is positive (myocardial infarction), but the model says it is normal or healthy, as shown in figure 36, 37 and 38

Figure 36

Confusion Matrix for 80% Training and 20% Testing

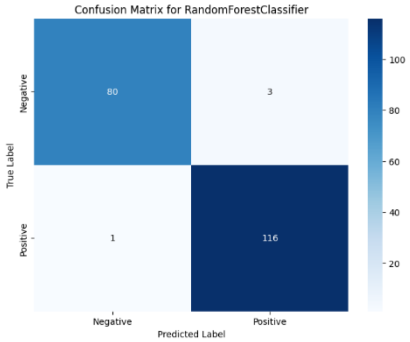
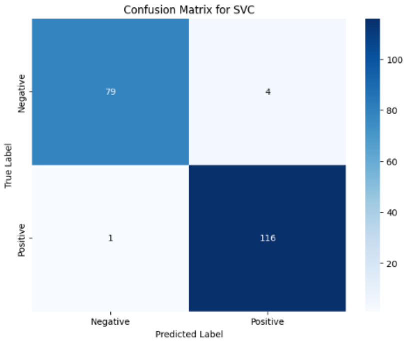
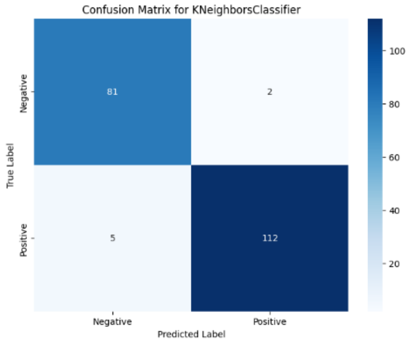
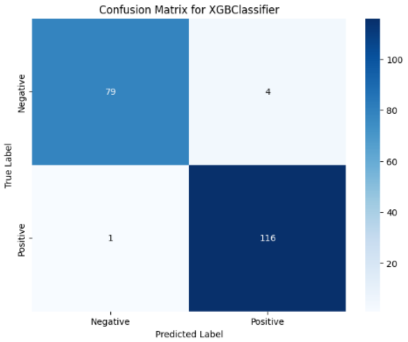
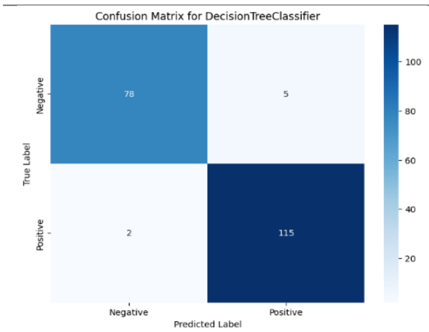
Random Forest	SVM
	
KNN	**XGBoost**
	
Decision Tree	
	

Figure 37

Confusion Matrix for 70% Training and 30% Testing

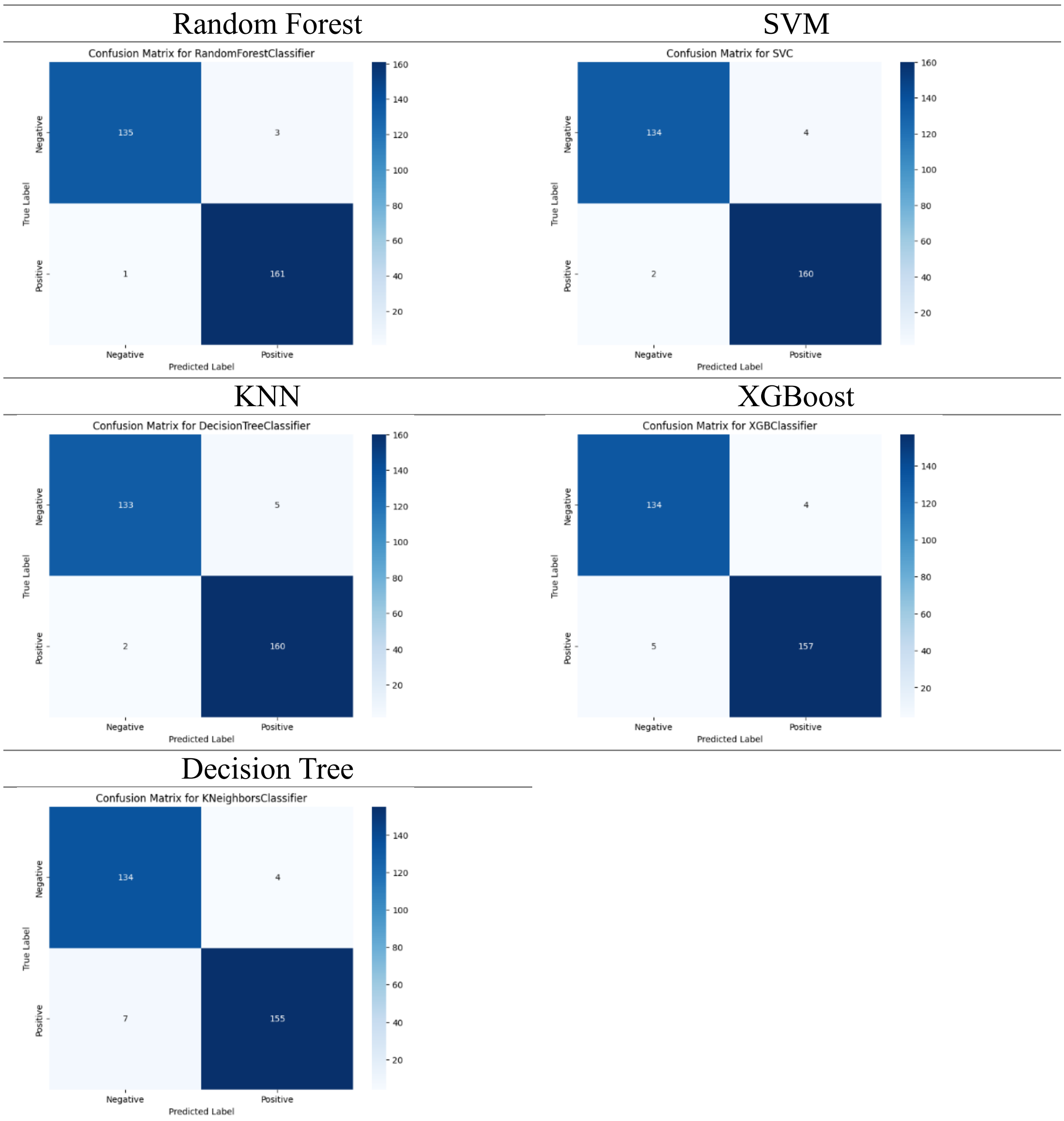
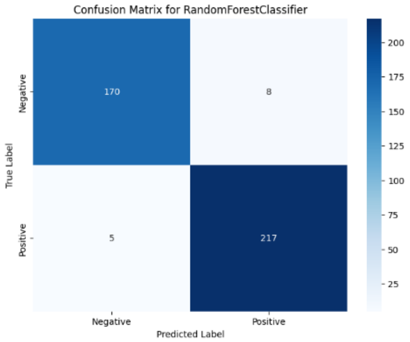


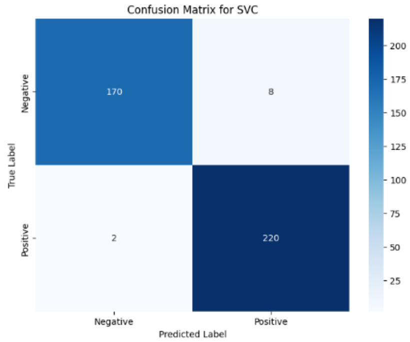
Figure 38

Confusion Matrix for 60% Training and 40% Testing

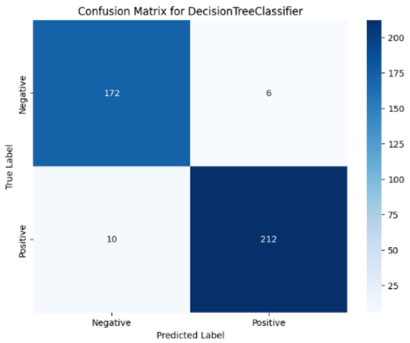
Random Forest



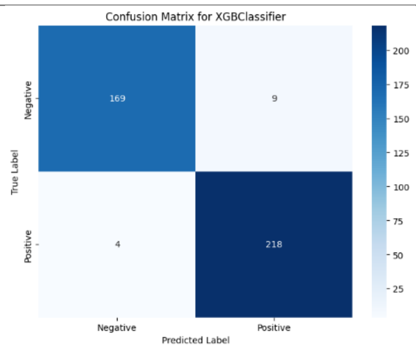
SVM



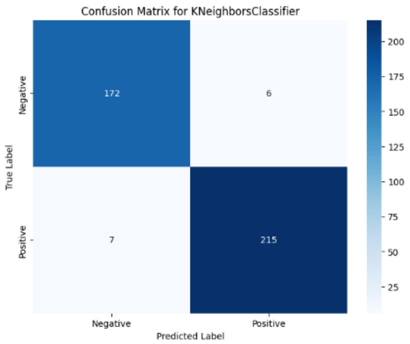
KNN



XGBoost



Decision Tree



4.4.1.2 ROC

ROC Curve is a graph used to evaluate the performance of a classification model by comparing the true positive rate with the false positive rate at different prediction thresholds. We used it to evaluate predictive models and the model's ability to distinguish between patients with MI and normal patients. And According to Figure 39,40 and 41. illustrate the ROC

Figure 39

Roc Form 20%-80$

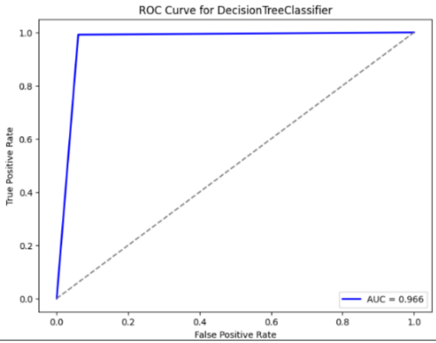
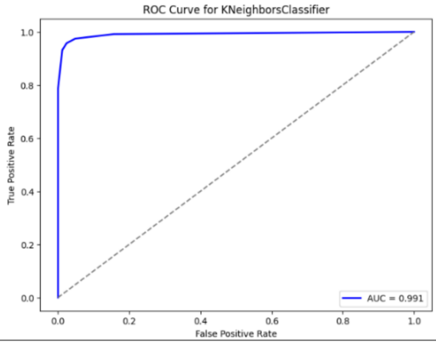
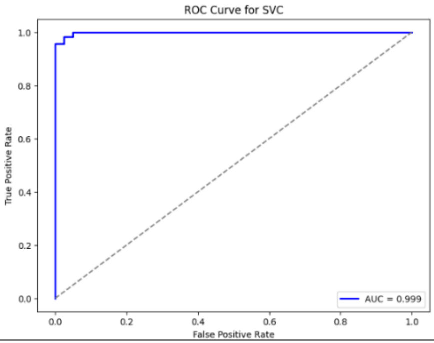
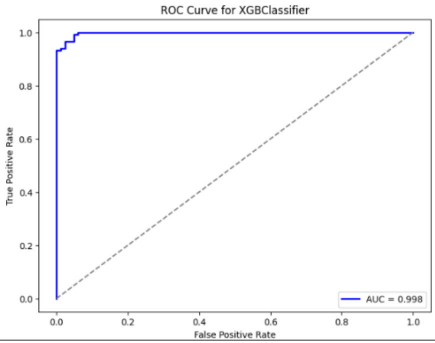
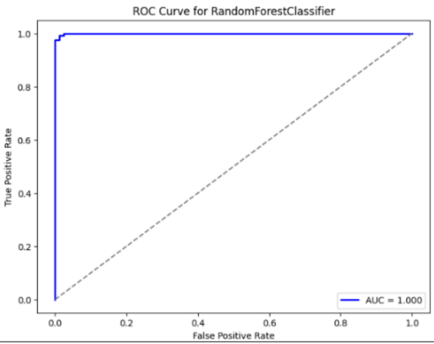
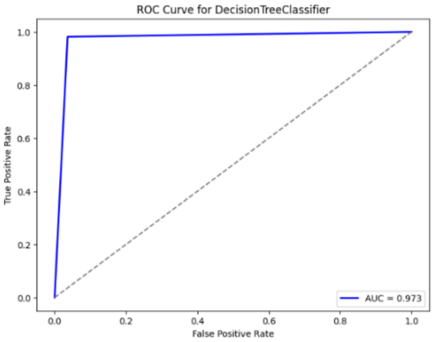
Random Forest	SVM
	
KNN	**XGBoost**
	
Decision Tree	
	

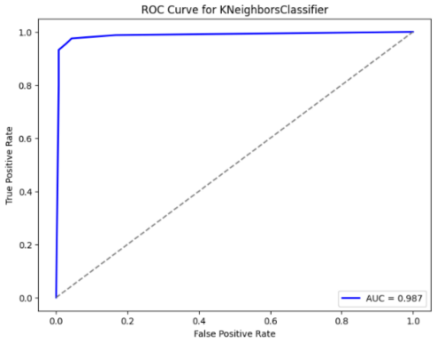
Figure 40

ROC Form 30%-70%

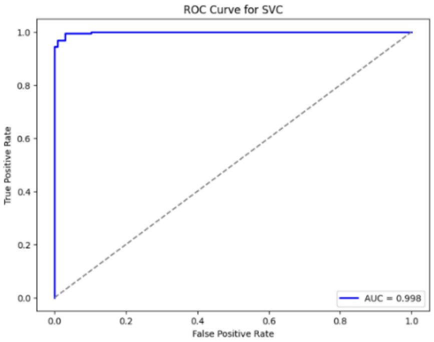
Random Forest



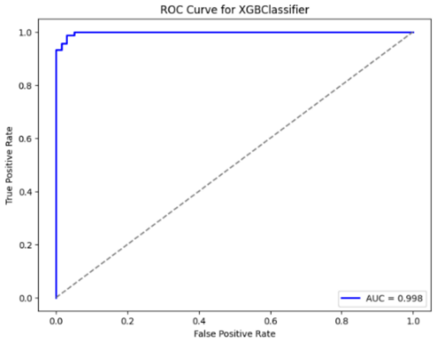
SVM



KNN



XGBoost



Decision Tree

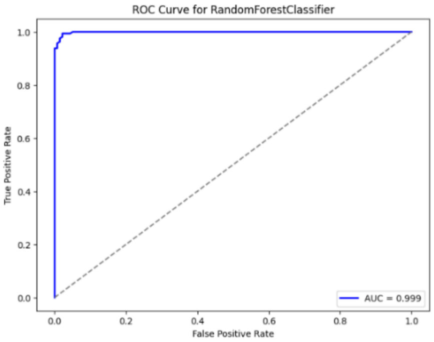
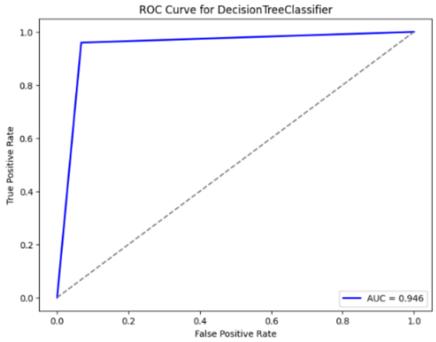


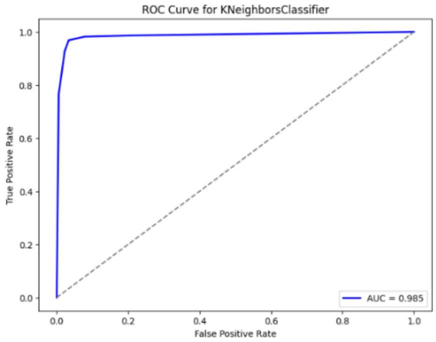
Figure 41

ROC Form 40%-60%

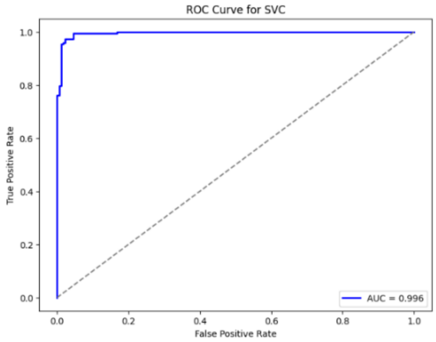
Random Forest



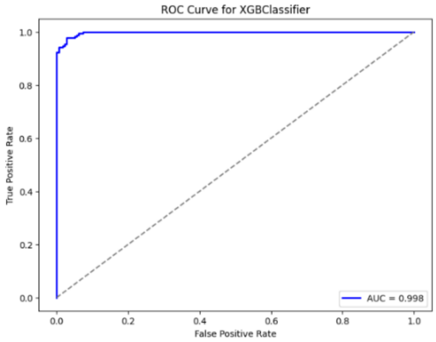
SVM



KNN



XGBoost



Decision Tree

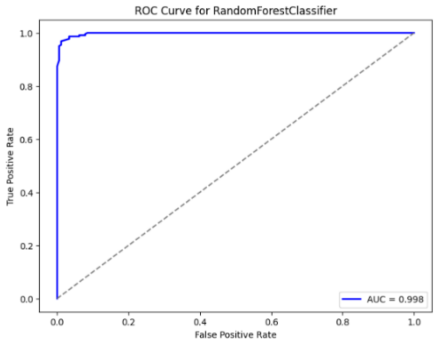


Table 11

Description of Results for 80% Training and 20% Testing

Learning Parameters	Random Fores	SVC	KNN	Decision Tree	XGBoost
Training ratio (%)	80%	80%	80%	80%	80%
Accuracy (%)	98%	98.4%	97.3%	96.1%	97.5%
Sensitivity	99.1%	99.1%	95.7%	98.3%	99.1
Specificity	96.4	95.2%	97.6%	94%	95.2
F1 score	99.2%	97.9%	97%	97.1%	97.9%

Table 12

Description of Results for 70% Training and 30% Testing

Learning Parameters	Random Fores	SVC	KNN	Decision Tree	XGBoost
Training ratio (%)	70%	70%	70%	70%	70%
Accuracy (%)	99%	98.3%	97.1%	97.7%	97%
Sensitivity	99.4%	98.8%	97.5%	98.8%	96.9%
Specificity	97.8%	97.1%	97.1%	96.4%	97.1%
F1 score	98.8%	98.2%	96.6%	97.2%	97.2%

Table 13

Description of Results for 60% Training and 40% Testing

Learning Parameters	Random Forest	SVC	KNN	Decision Tree	XGBoost
Training ratio (%)	60%	60%	60%	60%	60%
Accuracy (%)	96.8%	98.2%	97.2%	96%	96.8%
Sensitivity	97.7%	99.1%	96.8%	95.5%	98.2%
Specificity	95.5%	95.5%	96.6%	96.6%	94.9%
F1 score	97.1%	97.8%	97.1%	94.8%	97.1%

Table 14

Comparison between Related Studies and our Studies

Reference	**Model**	**Dataset**	**Preprocessing**	**Classes**	**Sensitivity**	**Specificity**	**f1-score**	**Accuracy**
(Hossen et al., 2021)	Random Forest, Decision Tree, and Logistic Regression	UCI Cleveland dataset	Data cleaning , Missing Data Analysis Standard Scalling	2(cardiac disease and non-cardiac disease)	86.96% 73.91% 92.31%	73.91% 69.57% 91.89%	82 72 92	80 72 92
(Akella & Akella, 2021)	Generalized linear model Decision tree Random forest Support-vector machine Neural network k-Nearest neighbor	UCI Cleveland dataset	Remove missing values, Normalizing, correlation matrix	2(coronary artery disease)	80% 74.47% 82.61% 79.59% 93.80% 78.72%		87.86% 79.70% 87.51% 86.62% 89.84% 84.19%	87.64% 79.78% 87.64% 86.52% 93.03% 84.27%

(Garg et al., 2021)	K-NN) Random Forest	Kaggle	analyze the data correlation matrix	2(Heart disease)	85.19% 78.57%	88.24% 84.85%	85.19 % 80%	86.88% 81.96%
(Tama et al., 2020)	Random Forest Gradient Boosting Machine XGBoost	Z-Alizadeh Sani dataset Statlog dataset Cleveland dataset Hungarian dataset	Correlation-based Feature Selection, Particle Swarm Optimization	2(Coronary Heart Disease)				in Statlog 78.90% In Hungarian 91.13% Z-Alizadeh Sani 87.65%
(Alshehri & Alharbi, n.d.)	AdaBoost SVM Decision Tree Random Forest	StatLog UCI Z-Alizadeh Sani from UCI CVD dataset	Normalization	2(CVD)	RF 82% RF 92% RF 70%	RF 82% RF,Adaboot 92% RF 70%	RF 84% RF,adaboost 91% RF 71%	STATLOG RF 84% Z-Alizadeh RF 91% CVD dataset Svm,Rf,adaboost 72%
(Sharma et al., 2020)	DNN	Cleveland Heart disease dataset	Exploratory Data Analysis Categorical Encoding	2 heart disease	81.03%	82.81%	81.03%	81.96%
(Xiao et al., 2022)	Logistic Regression (LR) Decision Tree (DT) Random Decision Forest (RDF) Naive Bayes (NB) Support Vector Machine (SVM)	Zhuzhou Central Hospital	Dealing with missing values	2(Myocardial Infarction)			56.5% 53.2$ 50.3% 57% 48% 45.3% 10.3%	72.1% 64.4% 73.3% 71.7% 74.9% 73.7% 66.3%

Adam et al. (2024)	Logistic Regression Random Forest Deep Learning	electronic health records (EHRs)	feature selection				80% 85% 90%	80% 85% 90%
(Nazir et al., 2024.)	Logistic Regression Random Forest Deep Learning Neural Network	PubMed, Scopus, Web of Science, and IEEE Xplore,		cardiovascular diseases			78% 81% 85%	80% 82% 85%
Our study	Random Fores SVC KNN Decision Tree XGBoost	Kaggle	feature selection and correlation	2(MI and normal)	Random forest in split 70-30 97.8%	SVM in split 70-30 99.4%	Random forest in split 80-20 99.2%	Random forest in split 70-30 99%

4.4.1.3 Summery

We evaluated and compared five different machine learning models, namely Random Forest, SVC, KNN, Decision Tree, and XGBoost, using three different training ratios: 80%, 70%, and 60%. At 80% training ratio:

The accuracy of the models was high, with SVC recording the highest accuracy at 98.4%, followed by Random Forest at 98%., Sensitivity showed excellent results with all models, with the best level reaching 99.1% for both Random Forest and SVC. As for the typicality, it was 96.4% for Random Forest and 95.2% for SVC.
Random Forest had the highest F1 score at 99.2%. At 70% training ratio:

The good performance continued with a D-score of 99% for Random Forest and 98.3% for SVC. Sensitivity was also highest for Random Forest at 99.4%.
While the typicality was 97.8% for Random Forest and 97.1% for SVC.
The F1 score also showed improvement, with Random Forest scoring 98.8%.

At a 60% training rate: A slight drop in performance was observed, with accuracy at 98.2% for SVC and 96.8% for both Random Forest and XGBoost.

For sensitivity, Random Forest was 97.7% and SVC was 99.1%. The result was 95.5% for both Random Forest and SVC. then for the F1 score result, it was 97.8% for SVC and 97.1% for both Random Forest and XGBoost.

The results show that the RF model performs exceptionally well at all training rates, leading other models by the test accuracy and F1. Although SVC has elevated sensitivity, it is a robust option for machine-learning applications.

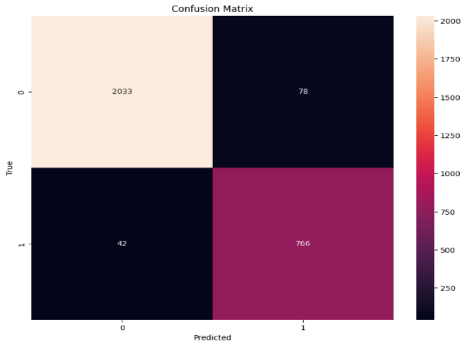
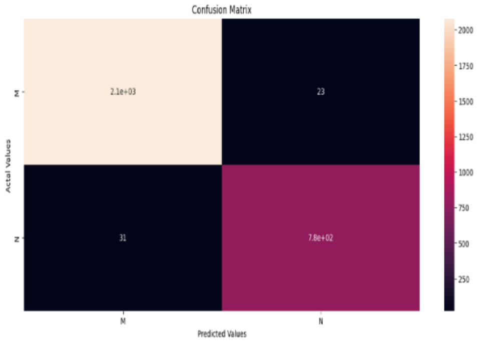
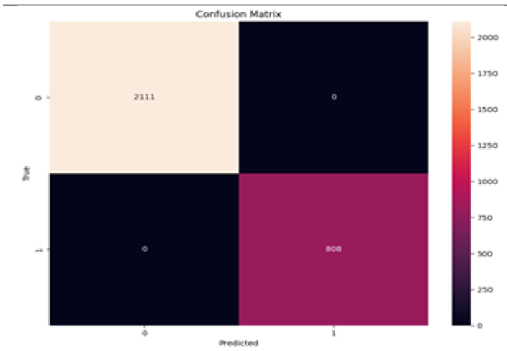
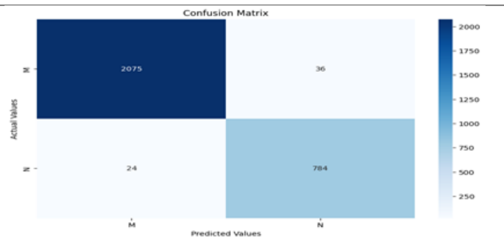
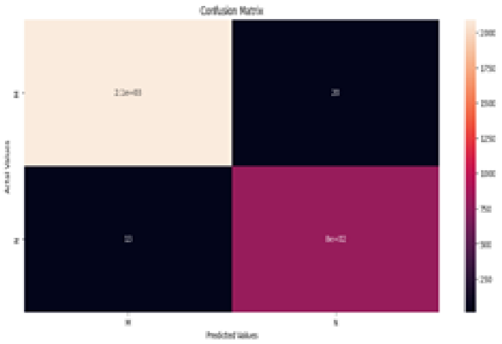
4.4.2 Image Data

4.4.2.1 Confusion Matrix

The confusion matrix is a popular technique for evaluating the model's performance based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TPs stands for the number of samples that the model correctly identified as positive cases, or more precisely, the number of people who experienced myocardial infarction in each model. TNs stand for the number of samples that the model correctly classified as negative cases or the number of people who are healthy (normal) but are classified as negative by each model. False positives (FPs) are the number of samples that the model incorrectly classifies as negative or the cases that are negative (normal or healthy) but are nonetheless classified as myocardial infarction. False negatives (FNs) are the number of samples that the model got wrongly as positive. These are the cases where the sample is positive (myocardial infarction), but the model says it is normal or healthy, as shown in figure 42.

Figure 42

Confusion Matrix for Image Data

<table>
<tr><th>ALEXNET</th><th>CNN</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><th>RESNEY18</th><th>CNN+SVM</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><th>LENET</th><th></th></tr>
<tr><td>
</td><td></td></tr>
</table>

4.4.2.2 Training Loss per Epoch

Refers to the loss value calculated after each epoch. The losses of each batch within the epoch are summed, and the average total loss for that epoch is calculated. These values are used to monitor the performance of the model during training; if the loss decreases over time, this is an indication that the model is learning correctly. If the loss continues to decrease, the model is learning. If it stops decreasing, this may indicate the need to adjust the training settings. If the loss increases or does not decrease, there may be problems such as overfitting or insufficient data According to Figure 43 and 44 illustrate the training loss per epoch

Figure 43

Training Loss per Epoch form Image Data

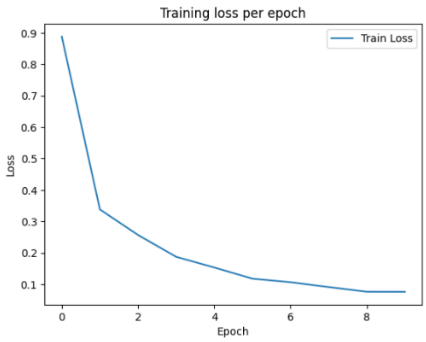
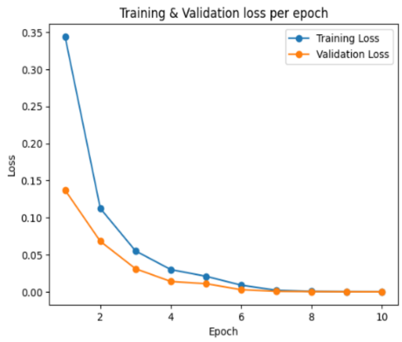
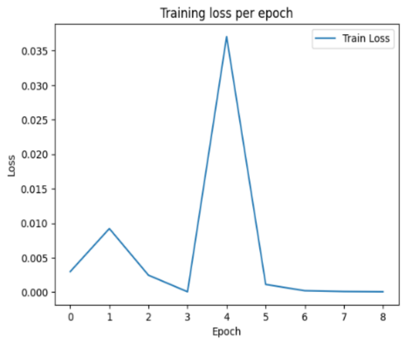
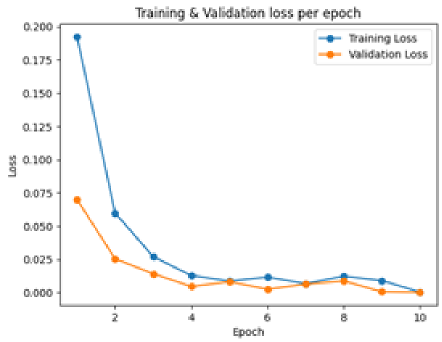
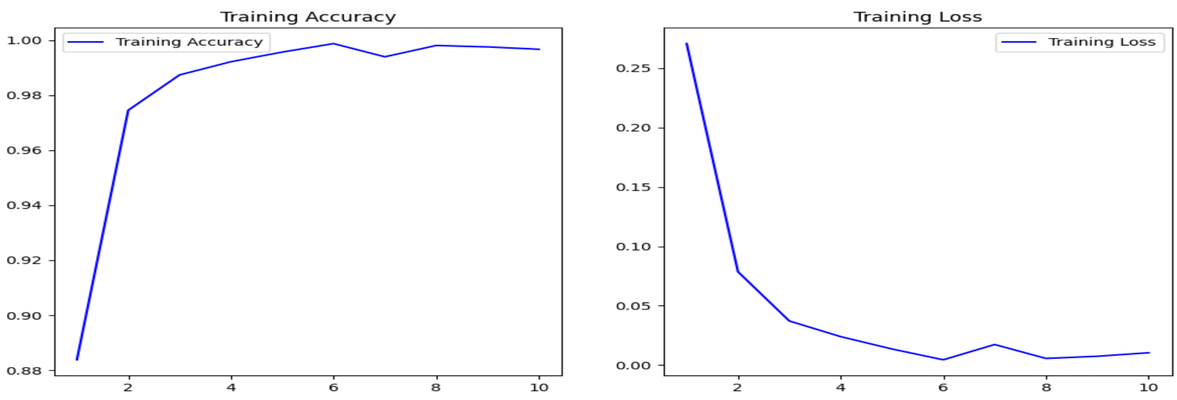
ALEXNET	CNN
	
ResNet18	**LENET**
	

Figure 44

Training Accuracy and Loss for CNN+SVM MODEL



4..4.2.3 Training Accuracy per Epoch

Refers to the accuracy value calculated after each epoch while training the model. It is calculated by comparing the predictions produced by the model to the correct values for all the data processed during that epoch.

It is used to track how much the model's performance has improved during training.

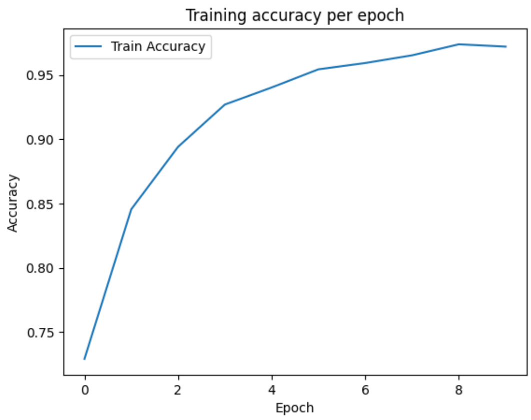
Performance is evaluated during training if the accuracy increases over time, this indicates that the model is learning well from the data.

Detecting problems if the accuracy stops improving or starts to fluctuate, this may indicate problems such as overfitting or missing data. According to Figure 45, illustrate the training acc per epoch.

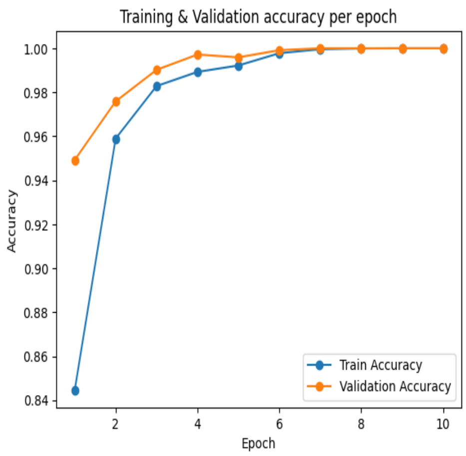
Figure 45

Training Accurse per Epoch Form Image Data

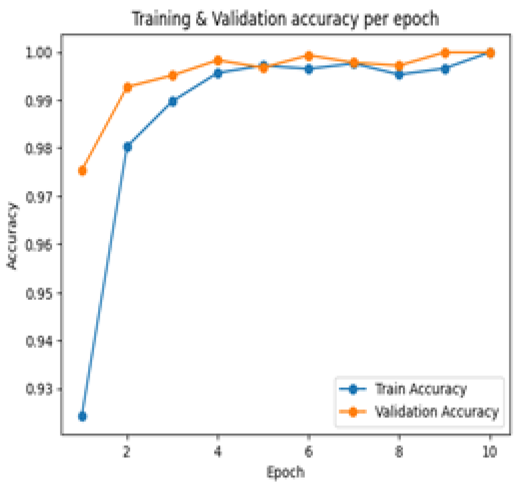
ALEXNET



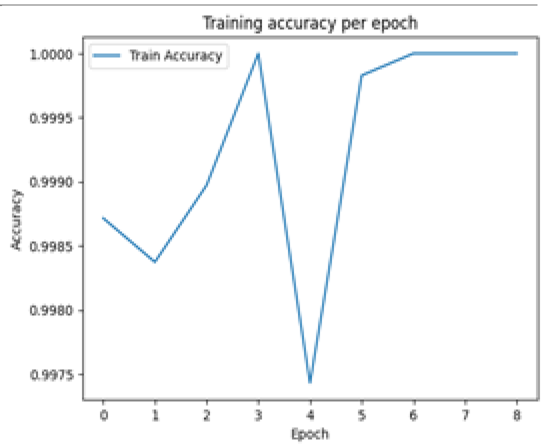
CNN



LENET



ResNet18



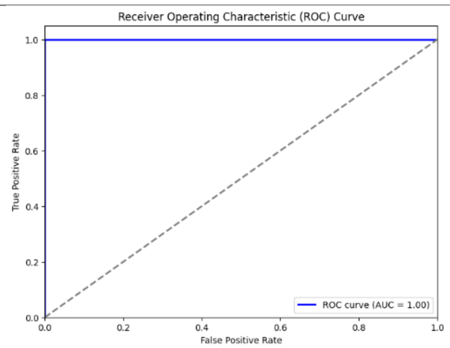
4.4.2.4 ROC

ROC Curve is a graph used to evaluate the performance of a classification model by comparing the true positive rate with the false positive rate at different prediction thresholds. We used it to evaluate predictive models and the model's ability to distinguish between patients with MI and normal patients. And According to Figure 46, illustrate the ROC.

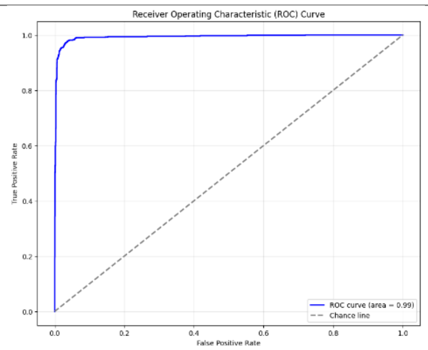
Figure 46

ROC for Image Data

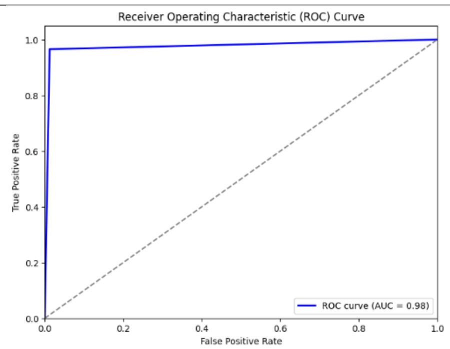
RESNET 18



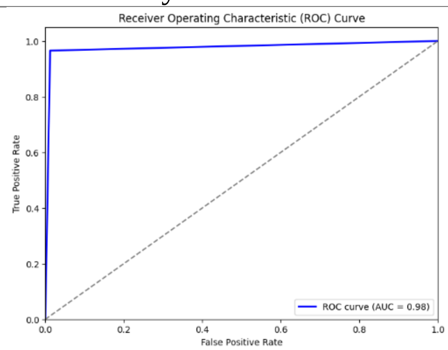
ALEXNET



CNN



Hybrid Model



LENET

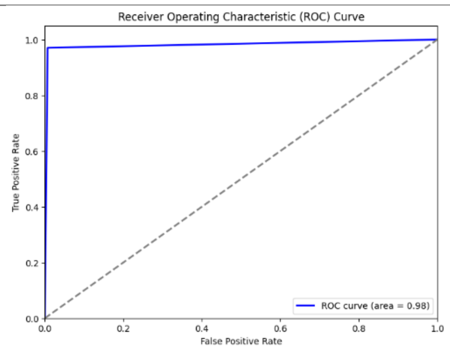


Table 15

Images Data Evaluation

Learning Parameters	RESNET 18	ALEXNET	LENET	CNN	Hybrid Model CNN+SVM
Training ratio (%)	70	70	70	70	**70**
Batch Size	32	32	32	32	32
No of epochs	20	10	100	10	10
Accuracy test (%)	**99**%	95.89	98.79	98**%**	**98%**
Accuracy train (%)	99%	96.79**%**	99%	97.67**%**	97.8**%**
Sensitivity	1	94.80	98.39	97.01**%**	98.29%
Specificity	1	96.31	99.05	98.78**%**	97.03%
Achieved mean square error	0	0.0411	0.0001278	0.0171	0.03

Table 16

Comparison Between Related Studies and our Studies

Reference	**Model**	**Dataset**	**Preprocessing**	**Classes**	**Sensitivity**	**Specificity**	**f1-score**	**Accuracy**
(Mahmoud et al., 2022)	LeNet-5 Modified Version of LeNet-5	Ch.Pervaiz Elahi Institute of Cardiology Multan	Resize images, MaxPooling Dropout	4(Myocardial Infarction Patients, abnormal heartbeat, history of (MI), Normal ECG			88% 99%	89.24 % 98.38 %
(Tippannavar et al., 2022)	CNN	MIT BIH arrhythmia database's ECG records	Convert data to 2D images	7(APC, LBBB, NOR, PAB, PVC, RBBB, VEB)			98%	98.1%

(Manolkar & Gawande, 2023)	DNN	MIT-BIH Arrhythmia Database	removing noise, filtering, normalizing	2(Chronic Heart Failure and normal)	93.96%	98.70%	95.49 %	96.9%
(Lida Kermanidis et al., 2024)	FFT,CNN–LSTM	PhysioNet dataset	FFT, ECG signal segmentation	3(ARR, CHF, NSR)	ARR: 99.09%, CHF: 99.87%, NSR: 89.01%	ARR: 99.19%, CHF: 99.88%, NSR: 99.35%	ARR: 98.89%, CHF: 99.67%, NSR: 92.02%	97.4% 98.90%
(Zhang et al., 2024)	CNN-LSTMSE,	MIMIC-III	ECG signal segmentation, Z-Score	2(CHF, normal)	99.03%	99.64%		In 12 second 99.09%
(Altaf et al., 2024)	ResNet18	ASCERTAIN dataset	STFT	5(encompassing extra-version, neuroticism, agreeableness, conscientiousness, and openness)			90%	98%
OUR STYDE	RESNET 18 ALEXNET LENET CNN HYPERMODEL CNN+SVM	Kaggle	Resizing	2(Mi and normal)	Resnet18 100%	Resnet18 100%		Resnet 18 99%

4.4.2.5 Summary

The table compares five different machine learning models: ResNet 18, AlexNet, LeNet, CNN, and Hybrid Model CNN+SVM. The same data settings were used for each model, with reference to the learning parameters and performance of each model.

Training ratio: 70% of the data was used for training in all models. Batch size: The batch size was set to 32 for all models. Number of epochs ResNet 18: 20 epochs. AlexNet: 10 epochs. LeNet: 100 epochs. CNN: 10 epochs. Hybrid Model CNN+SVM: 10 epochs.

The performance of (Accuracy test) was evaluated: ResNet 18: 99% AlexNet: 95.89% LeNet: 98.79% CNN: 98% Hybrid Model CNN+SVM: 98% We got the training accuracy of ResNet 18: 99%, AlexNet: 96.79%, LeNet: 99%, CNN: 97.67%, Hybrid Model CNN+SVM: 97.8%,The sensitivity was, ResNet 18: 1, AlexNet: 94.80%, LeNet: 98.39%, CNN: 97.01%, Hybrid Model CNN+SVM: 98.29%,And (Specificity) ResNet 18: 1 AlexNet: 96.31% LeNet: 99.05% CNN: 98.78%,Hybrid Model CNN+SVM: 97.03%,Achieved mean square error ResNet 18: 0 AlexNet: 0.0411 LeNet: 0.0001278 CNN: 0.0171 Hybrid Model CNN+SVM: 0.03

The results show that ResNet 18 outperforms other models in testing and training accuracy, and also exhibits excellent sensitivity and typicality. While LeNet is among the models with the lowest mean square error.

CHAPTER V
CONCLUSION

Heart disease is one of the most serious diseases around the world, and myocardial infarction is one of the leading causes of death worldwide. Detecting this disease is very important due to the prevalence of this disease all over the world, and early detection of heart disease is one of the most important reasons that save people's lives. This disease is diagnosed by analyzing the heart's electricity and by estimating the patient's health status and the resulting symptoms. The main aim of our research is to utilize the machine learning methodology and neural networks to classify heart diseases (MI) based on (ECG) samples and clinical symptoms of patients. In this context, we used 10 different algorithms and two types of data, and 5 different algorithms were applied to each type of data. In the data of patient symptoms, we used deep learning algorithms (RandomFores, SVC, KNN, Decision Tree, XGBoost)

The data was divided into 3 groups: The first group 20% test-80% training, where the model achieved an accuracy of (98% 98.4% 97.3% 96.1% 97.5%), The second group 30% test-70% training, where the model achieved an accuracy of (99% 98.3% 97.1% 97.7% 97%), The third group 40% test-60% training and the accuracy was (96.8% 98.2% 97.2% 96% 96.8%)**.**

ECG classification 4 neural network models were used (RESNET 18, ALEXNET LENET, CNN) and a model that combines deep learning and neural networks (HYPERMODELCNN+SVM), where the model was trained on (70% training and 30% testing) and No of epochs were used in RESNET 18 20 and LENET 100 and the rest of the models 10 and Batch Size 32 was used in all models. Where Accuracy test (99% 95.89 98.79 98% 98%)And Accuracy train (99% 96.79% 99% 97.67% 97.8%)And for Sensitivity (1 94.80 98.39 97.01% 98.29%)And Specificity (1 96.31 99.05 98.78%) 97.03% and Achieved mean square error(0 0.0411 0.0001278 0.0171 0.03).

Our results, especially those obtained from ResNet 18, are in line with the idea of applying AI in healthcare with higher accuracy and precision. These models can now serve as a verification system for Myocardial Infarction (MI) diagnosis, thereby reducing misdiagnosis and providing an alternative to ease the significant and demanding workload of cardiologists.

One of the challenges and limitations of our research is the lack of sufficient datasets specifically for MI disease. With a large number of datasets, we can take advantage of different pre-trained architectures such as VGGNet and GoogleNet.

Future Work

While this study achieved important results, there are many opportunities for improvement, applying new techniques and expanding the scope of research By increasing the amount of data that leads to the application of complex algorithms to obtain more accurate results such as (Resnet50, Resnet101, and GoogLeNet) This confirms the need to collect larger and more diverse data in future studies. Hybrid models that combine different methods should be studied to achieve better results

Also, the application of the Internet of Things known as (IOT) by creating a website or a phone application to diagnose the disease Some advice that can be given in this area Since we have created the code in the Python language, it is possible to use the Amazon Web Services (AWS) provides servers for a variety of applications, including web hosting, data storage, machine learning, and analytics.

also taking into consideration whether the ECG images that are uploaded are taken with phones or do they require special cameras or a specialist to crop the important part of the ECG.

References

Adam, M., Documentacion, A. M.-R. E. de, & 2024, undefined. (n.d.). Smart Health Solutions: The Convergence of AI, Machine Learning, and Deep Learning for Brain and Heart Care. *Redc.Revistas-Csic.ComMA Adam, A MukhtarRevista Espanola de Documentacion Cientifica, 2024•redc.Revistas-Csic.Com*. Retrieved August 29, 2024, from https://redc.revistas-csic.com/index.php/Jorunal/article/download/215/180

Akella, A., & Akella, S. (2021). Machine Learning Algorithms for Predicting Coronary Artery Disease: Efforts Toward an Open Source Solution. *Future Science OA*, *7*(6). https://doi.org/10.2144/FSOA-2020-0206

Almulihi, A., Saleh, H., Hussien, A. M., Mostafa, S., El-Sappagh, S., Alnowaiser, K., Ali, A. A., & Refaat Hassan, M. (2022). Ensemble Learning Based on Hybrid Deep Learning Model for Heart Disease Early Prediction. *Diagnostics*, *12*(12), 3215. https://doi.org/10.3390/DIAGNOSTICS12123215/S1

Alshehri, G. A., & Alharbi, H. M. (n.d.). Prediction of Heart Disease using an Ensemble Learning Approach. *IJACSA) International Journal of Advanced Computer Science and Applications*, *14*(8), 2023. Retrieved August 13, 2024, from www.ijacsa.thesai.org

Altaf, M. M., Khan, S. U., Majid, M., & Anwar, S. M. (2024). Personality Trait Recognition using ECG Spectrograms and Deep Learning. *Computer Science*. https://arxiv.org/abs/2402.04326v1

Arif Ali, Z., H. Abduljabbar, Z., A. Tahir, H., Bibo Sallow, A., & Almufti, S. M. (2023). eXtreme Gradient Boosting Algorithm with Machine Learning: a Review. *Academic Journal of Nawroz University*, *12*(2), 320–334. https://doi.org/10.25007/AJNU.V12N2A1612

Armoundas, A. A., Narayan, S. M., Arnett, D. K., Spector-Bagdady, K., Bennett, D. A., Anthony Celi, L., Friedman, P. A., Gollob, M. H., Hall, J. L., Kwitek, A. E., Lett, E., Menon, B. K., Sheehan, K. A., & Al-Zaiti, S. S. (2024). Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement From the American Heart Association. *Circulation*, *149*(14), E1028–E1050.

https://doi.org/10.1161/CIR.0000000000001201/SUPPL_FILE/ARMOUNDAS

Brown, J., Gharineiat, Z., & Raj, N. (2023). CNN Based Image Classification of Malicious UAVs. *Applied Sciences (Switzerland)*, *13*(1). https://doi.org/10.3390/APP13010240

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., Mcinnes, M., & Holzinger, A. (2023). Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 329–341. https://doi.org/10.1109/TPAMI.2022.3145392

Congenital Heart Disease Types, Symptoms, and Treatment | Dr. Raghu. (n.d.). Retrieved October 15, 2024, from https://drraghu.com/resources/congenital-heart-disease/

Coronary Artery Disease - Lathrup Village, MI: Heart and Vein Center. (n.d.). Retrieved October 15, 2024, from https://www.heartandveincenter.com/contents/cardiological-services/conditions-and-diagnoses/coronary-artery-disease-condition

Damayunita, A., Fuadi, R. S., & Juliane, C. (2022). Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients. *Jurnal Online Informatika*, *7*(2), 219–225. https://doi.org/10.15575/JOIN.V7I2.919

D.B, R., & F.A., Z. (2022a). DETERMINE THE VALUE OF RISK FACTORS FOR MYOCARDIAL INFARCTION. *FAN, TA'LIM, MADANIYAT VA INNOVATSIYA JURNALI | JOURNAL OF SCIENCE, EDUCATION, CULTURE AND INNOVATION*, *1*(4), 23–28. https://mudarrisziyo.uz/index.php/innovatsiya/article/view/96

D.B, R., & F.A., Z. (2022b). DETERMINE THE VALUE OF RISK FACTORS FOR MYOCARDIAL INFARCTION. *FAN, TA'LIM, MADANIYAT VA INNOVATSIYA JURNALI | JOURNAL OF SCIENCE, EDUCATION, CULTURE AND INNOVATION*, *1*(4), 23–28. https://mudarrisziyo.uz/index.php/innovatsiya/article/view/96 *ecg_image_data*. (n.d.). Retrieved December 1, 2024, from https://www.kaggle.com/datasets/erhmrai/ecg-image-data/data

G, A., Ganesh, B., Ganesh, A., Srinivas, C., Dhanraj, & Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. *Global Transitions Proceedings*, *3*(1), 127–130. https://doi.org/10.1016/J.GLTP.2022.04.008

Garg, A., Sharma, B., & Khan, R. (2021). Heart disease prediction using machine learning techniques. *IOP Conference Series: Materials Science and Engineering*, *1022*(1). https://doi.org/10.1088/1757-899X/1022/1/012046

Heart Disease Risk Factors In Young Adults | Sprint Medical. (n.d.). Retrieved October 15, 2024, from https://sprintmedical.in/blog/heart-disease-risk-factors-in-young-adults

Hossen, M. D. A., Tazin, T., Khan, S., Alam, E., Sojib, H. A., Monirujjaman Khan, M., & Alsufyani, A. (2021). Supervised Machine Learning-Based Cardiovascular Disease Analysis and Prediction. *Mathematical Problems in Engineering*, *2021*(1), 1792201. https://doi.org/10.1155/2021/1792201

Key Deep Learning Architectures: LeNet-5 | by Max Pechyonkin | Medium. (n.d.). Retrieved October 16, 2024, from https://medium.com/@pechyonkin/key-deep-learning-architectures-lenet-5-6fc3c59e6f4

Li, L., & Liu, Z. P. (2022). A connected network-regularized logistic regression model for feature selection. *Applied Intelligence*, *52*(10), 11672–11702. https://doi.org/10.1007/S10489-021-02877-3/FIGURES/8

Lida Kermanidis, K., Maragoudakis, M., Eleyan, A., & Alboghbaish, E. (2024). Electrocardiogram Signals Classification Using Deep-Learning-Based Incorporated Convolutional Neural Network and Long Short-Term Memory Framework. *Computers 2024, Vol. 13, Page 55*, *13*(2), 55. https://doi.org/10.3390/COMPUTERS13020055

Lindahl, B., & Mills, N. L. (2023). A new clinical classification of acute myocardial infarction. *Nature Medicine 2023 29:9*, *29*(9), 2200–2205. https://doi.org/10.1038/s41591-023-02513-2

Machine Learning Applications and Examples - IABAC. (n.d.). Retrieved October 15, 2024, from https://iabac.org/blog/machine-learning-applications-and-examples

Mahmoud, S., Gaber, M., Farouk, G., & Keshk, A. (2022). Heart Disease Prediction Using Modified Version of LeNet-5 Model. *International Journal of Intelligent Systems and Applications*, *14*(6), 1–12. https://doi.org/10.5815/IJISA.2022.06.01

Manolkar, O., & Gawande, N. (2023). ECG Analysis for Chronic Heart Failure Detection using Deep Learning. *International Journal for Research in Applied Science and Engineering Technology*, *11*(6), 1592–1597. https://doi.org/10.22214/IJRASET.2023.53733

Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review 2023 56:9*, *56*(9), 10031–10066. https://doi.org/10.1007/S10462-023-10424-4

ML | Primeiros passos com AlexNet – Acervo Lima. (n.d.). Retrieved October 16, 2024, from https://acervolima.com/ml-primeiros-passos-com-alexnet/

Myocardial infarction complications. (n.d.). Retrieved December 1, 2024, from https://www.kaggle.com/datasets/rafatashrafjoy/myocardial-infarction-complications

Nazir, M., Documentacion, I. H.-R. E. de, & 2024, undefined. (n.d.). Charting New Frontiers: AI, Machine Learning, and Deep Learning in Brain and Heart Health. *Redc.Revistas-Csic.ComMB Nazir, I HussainRevista Espanola de Documentacion Cientifica, 2024•redc.Revistas-Csic.Com*. Retrieved August 29, 2024, from https://redc.revistas-csic.com/index.php/Jorunal/article/download/214/179

Norton, C. (2017). Acute coronary syndrome: assessment and management. *Nursing Standard (Royal College of Nursing (Great Britain) : 1987)*, *31*(29), 61–71. https://doi.org/10.7748/NS.2017.E10754

Optimizers in Deep Learning - Scaler Topics. (n.d.). Retrieved October 15, 2024, from https://www.scaler.com/topics/deep-learning/optimizers-in-deep-learning/

Salari, N., Morddarvanjoghi, F., Abdolmaleki, A., Rasoulpoor, S., Khaleghi, A. A., Hezarkhani, L. A., Shohaimi, S., & Mohammadi, M. (2023). The global prevalence of myocardial infarction: a systematic review and metaanalysis. *BMC Cardiovascular Disorders*, *23*(1), 1–12. https://doi.org/10.1186/S12872-023-03231-W/FIGURES/2

Sharma, V., Rasool, A., & Hajela, G. (2020). Prediction of Heart disease using DNN. *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*, 554–562. https://doi.org/10.1109/ICIRCA48905.2020.9182991

Soydaner, D. (n.d.). *A Comparison of Optimization Algorithms for Deep Learning*.

Supervised Machine Learning, Unsupervised Machine Learning, and Deep Learning - CFA, FRM, and Actuarial Exams Study Notes. (n.d.). Retrieved October 15, 2024, from https://analystprep.com/study-notes/cfa-level-2/quantitative-method/supervised-machine-learning-unsupervised-machine-learning-deep-learning/

Tama, B. A., Im, S., & Lee, S. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. *BioMed Research International*, *2020*(1), 9816142. https://doi.org/10.1155/2020/9816142

Tippannavar, S. S., Harshith, R., Shashidhar, R., Sweekar, S. C., & Jain, S. (2022). ECG based Heart Disease Classification and Validation using 2D CNN. *Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022*, 1182–1186. https://doi.org/10.1109/IC3I56241.2022.10073094

Wereski, R., Kimenai, D. M., Bularga, A., Taggart, C., Lowe, D. J., Mills, N. L., & Chapman, A. R. (2022). Risk factors for type 1 and type 2 myocardial infarction. *European Heart Journal*, *43*, 127–135. https://doi.org/10.1093/eurheartj/ehab581

What Is Artificial Intelligence? | The Motley Fool. (n.d.). Retrieved October 15, 2024, from https://www.fool.com/terms/a/artificial-intelligence/

What is Heart Failure? – PB Cardiovascular. (n.d.). Retrieved October 15, 2024, from https://pbcardiovascular.com/site/what-is-heart-failure/

Why Arrhythmia Occurs？Top 5 Things You Should Know About Arrhythmia – Wellue. (n.d.). Retrieved October 15, 2024, from https://getwellue.com/blogs/cardiovascular/why-arrhythmia-occurs-top-5-things-you-should-know-about-arrhythmia?currency=EUR

Xiao, C., Guo, Y., Zhao, K., Liu, S., He, N., He, Y., Guo, S., & Chen, Z. (2022). Prognostic Value of Machine Learning in Patients with Acute Myocardial Infarction. *Journal of Cardiovascular Development and Disease*, *9*(2). https://doi.org/10.3390/jcdd9020056

Zhang, C. J., Yuan-Lu, Tang, F. Q., Cai, H. P., Qian, Y. F., & Chao-Wang. (2024). Heart failure classification using deep learning to extract spatiotemporal features from ECG. *BMC Medical Informatics and Decision Making*, *24*(1), 1–17. https://doi.org/10.1186/S12911-024-02415-4/FIGURES/10

Appendix A

Similarity Report

Yakın Doğu Üniversitesi

QUICK SUBMIT | NOW VIEWING: ALL PAPERS ▾

Submit

AUTHOR	TITLE	SIMILARITY	FILE	PAPER ID
Mustafa Jamal Ahmed	Abstract	0%		2535909157
Mustafa Jamal Ahmed	Conclusion	0%		2535909172
Mustafa Jamal Ahmed	Chapter 3	4%		2535909167
Mustafa Jamal Ahmed	Chapter 1	7%		2535909160
Mustafa Jamal Ahmed	Chapter 2	11%		2535909164
Mustafa Jamal Ahmed	Chapter 4	11%		2535909169
Mustafa Jamal Ahmed	All Chapters	19%		2535909173

Name-Surname	Signature
Supervisor: Assoc. Prof. Dr. Abdullahi Umar Ibrahim	
Co-supervisor: Assoc. Prof. Dr. Süleyman AŞIR	
Student name: Mustafa Jamal Ahmed	

CURRICULUM VITAE

PERSONAL INFORMATION

Full Name	: MUSTAFA JAMAL AHMED AHMED
Nationality	: Iraqi
Date and Place of Birth	: 20 August1998, Baghdad.
Marital Status	: Single

EDUCATION

a. Schools Attended with Dates

- ❖ Near East University, Nicosia 2022-2024
- ❖ Al-Esraa University, Baghdad 2017-2021
- ❖ Maikop State Technological University,Maikop 2016-2017
- ❖ Central High School for Boys, Diyala 2013-2016
- ❖ Martyrs of Islam Middle School, Diyala 2010-2013
- ❖ Ghassaneh Elementary School, Diyala 2004-2010

b. Qualifications Obtained with Dates

- ❖ Masters in Biomedical Engineering Nov, 2024
- ❖ Bachelor Degree in Medical Instrumentation Engineering July, 2022

c. Research Project/Thesis

- ❖ Master Thesis "INTEGRATION OF DIAGNOSTIC TECHNOLOGY FOR HEART DISEASE (MYOCARDIAL INFARCTION) USING ARTIFICIAL INTELLIGENCE TO IMPROVE CARDIOVASCULAR HEALTHCARE"
- ❖ Bachelor Thesis "SMART STICK FOR THE BLIND USING ARDUINO"

d. Membership of Professional Bodies/Affiliations

- ❖ Iraqi Association of Engineers

EMPLOYMENT HISTORY

Laboratory Assistant at University of Bilad Al-Rafidain, Department of Computer Engineering Technology 2021-

INTERESTS

My interests lie in the fields of technology and electronics. I have gained extensive experience in programming through my work in the Computer Engineering Department for three years, as well as through my master's studies and working on my thesis project in artificial intelligence.

REFEREES

(Assoc.Prof.Dr.) Abdullahi Umar Ibrahim
Biomedical Engineering
Near East University
Mobile: +90 548 831 43 46
Email: abdullahi.umaribrahim@neu.edu.tr

(Assist.Prof.Dr.)Mohammed Q. Mohammed
Medical Instrumentation Engineering
Al-Esraa Universiry
Email: dr.mohammed@esraa.edu.iq