DEVELOPMENT OF A PREDICTION MODEL FOR CARDIOVASULAR DISEASES

BY

ADEYANJU MOYOSORE
(BU21CSC1019)

A PROJECT SUBMITTED

TO

COMPUTER SCIENCE PROGRAMME,
COLLEGE OF COMPUTING AND COMMUNICATION STUDIES,
BOWEN UNIVERSITY, IWO, OSUN STATE, NIGERIA.

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD DEGREE OF BACHELOR OF SCIENCE (B.Sc.) IN COMPUTER SCIENCE

JULY 2025

CERTIFICATION

This is to certify that this project ‘’Development of a prediction model for cardiovascular diseases’’ was carried out by ADEYANJU MOYOSORE (BU21CSC1019) of the Computer Science programme under my supervision .

…… ……

Miss Busolami Oluwadamilare **Date**

(Supervisor)

…… ……

Dr D.O. OLANLOYE **Date**

(Head Of Programme)

DEDICATION

I dedicate this project to the Almighty God, my shield , glory and the lifter of my head, to Him be all the glory, praise, honour and adoration in Jesus name. I also dedicate it to my loving and caring parents Mr Christopher Adeyanju and Mrs Bolanle Adeyanju.

ACKNOWLEDGEMENTS

I thank the Almighty God, my maker, helper, provider and father for seeing me through the university, most especially my final year. I acknowledge my wonderful supervisor, Miss Busolami Oluwadamilare for her constructive criticism and commitment to bring out the best in me, . I also acknowledge the efforts of the Provost, Professor Oyelami , the Head of Programme, Dr. D.O. Olanloye, all members of staff in Computer Science, Dr. H.O. Aworinde, Dr. Chinecherem Umezuruike, Dr. (Mrs.) O.N. Emuoyibofarhe, Mr. Falana, Mr. Oyebade, Mr. Ayanniyi, Mrs. Olaniran, Dr. Ninan, Mrs. T.I. Awoniran, Mr. Adeniyi, for giving me a solid footing in the field of Computer Science. I say a big thank you to my parents Mr Chistopher Adeyanju and Mrs Bolanle Adeyanju for their prayers, moral and financial contributions., you are the best. I also thank my younger brother Morounfoluwa for being there for me always. I appreciate Miss Akinwe Envagelyn for helping me to proofread my project. I thank everyone who helped in my academics, Elvis, Talabi, Oyinkayinsola , all my course mates, James, Opeyemi, Praise and everyone that made my stay in the university a pleasant one. I am forever grateful.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Cardiovascular diseases are the leading factors responsible for death worldwide. While many approaches have implemented either Internet of Things or machine learning algorithms, few studies have employed both. This project aims to predict cardiovascular disease by classifying it into two or three risk levels based on the vital parameters used.

The proposed system, CardioPredictor was developed using an Arduino UNO R4 WiFi, MAX30102 pulse oximeter sensor, OLED display, dot board, AD8232 ECG module, and connecting wires. Electrocardiogram data was collected from Kaggle. Model training was performed using seven algorithms (Support Vector Machine, Kernel Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer Perceptron and Xgboost algorithm). Model evaluation was performed using accuracy, precision, recall and f1 score. The best model, XGBoost was deployed into CardioPredictor using coder in Matlab environment. The results obtained in this study were above 90 percent. However, XGBoost outperformed the other six models with an accuracy of 99.9 percent. Despite CardioPredictor's many advantages, the following changes might be made to make it even better. Real-world or other relevant ECG data should be used. Random Forest, Logistic Regression, LightGBM are additional machine learning techniques that can be used in addition to those used in the study. The wearable device can be powered by solar, electricity, or other sources.

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Cardiovascular diseases (CVDs) are the leading causes of death globally. In 2022, seven hundred and two thousand, eight hundred and eighty(702,880) people died from heart disease in the US, accounting for 1 in every 5 deaths. In 2017, 17.7 million deaths (17,700,000) were caused by CVDs. The total number of CVDs cases has nearly doubled from Two hundred and seventeen one million(271,000,000) in 1990 to Five hundred and twenty three million (523,000,000) in 2019(Roth et al., 2020). Majourity of CVDs deaths occur in middle and low income countries (World Health Organization, 2024), with one-third occurring prematurely in persons under 70 years old (World Health Organization, 2023) and a Seventy-six point one one percent (76.11%) death rate from CVDs in Nigeria (Ogah et al., 2023).

Cardiovascular diseases can be caused by various reasons, including physical inactivity, hypertension, being overweight or obese, and smoking.

Cardiovascular disease treatment often consists of a combination of lifestyle changes, drugs, and medical procedures aimed at symptom management, decreasing disease progression, and increasing quality of life.

Wearable devices, also known as wearables or wearable technologies, are devices that use motion and biometric sensors to measure a variety of physiologic parameters such as step count, activity intensity, HR, heart rhythm, blood pressure (BP), oxygen saturation, sleep, maximum oxygen uptake, and temperature (Hughes et al., 2023).

Wearable technology is becoming increasingly popular among cardiac patients, rehabilitation patients, and the elderly. To create and implement an effective plan for CVD prevention and treatment in older individuals, a better understanding of a wide

range of CVD risk factors, as well as devices that can aid in early detection and monitoring, is required.

1.2 STATEMENT OF THE PROBLEM

Cardiovascular disease is the top factor responsible for death all over the world. (WHO, 2021). Early prediction of cardiovascular disease allows for quick treatment, reduces the threat of severe complications and enhances the general health of the patient. The dataset used by (Islam et al.,2023) is nine hundred and twenty records and it causes overfitting. This study developed a device called CardioPredictor(a device that predicts CVDs by assessing its risk level) and uses a dataset of more than seven hundred thousand records.

1.3 AIM OF THE STUDY

The aim of this study is to predict cardiovascular diseases by classifying them into risk levels using CardioPredictor (a device that predicts CVDS by assessing its risk level) by using some machine learning algorithms(Support Vector Machine, Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neigbour, Multilayer perceptron and XGBoost algorithm) and Internet of Things.

1.4 OBJECTIVES OF THE STUDY

The objectives of this study are to:

1. To design a system capable of predicting CVDs.
2. To perform collection of electrocardiogram data from Kaggle
3. To perform model training and model evaluation in Matrix Laboratory environment.
4. To deploy the best model into CardioPredictor(a device that predicts CVDS by assessing its risk level) using Coder.

1.5 METHODOLOGY.

The methodology used to achieve this are:

1. Designing a system suitable for the prediction of cardiovascular diseases.
2. Collecting an electrocardiogram dataset from Kaggle, which will aid in the prediction of CVDs.
3. Performing model training in the Matrix laboratory environment using various machine learning models.
4. Evaluating models with Accuracy, Precision, F1 score, Sensitivity and Recall
5. Deploy the best model into Coder.

1.6 SIGNIFICANCE OF THE STUDY

This study leverages machine learning algorithms and Internet of Things (IoT) to enable early prediction of cardiovascular diseases. By facilitating timely diagnosis, it aims to improve patient outcomes and ultimately decrease mortality rates associated with these condition.

1.7 SCOPE OF THE STUDY

This project is focused on designing a system for the prediction of cardiovascular diseases by classifying it into risk levels of CVDs using electrocardiogram, pulse rate and blood oxygen levels. It includes designing CardioPredictor (a device that predicts CVDS by assessing its risk level), data collection, model training and evaluation and deployment of the best model into the system.

1.8 STRUCTURE OF THE PROJECT

- Chapter One: Chapter one talks about the Introduction, Background of study, Statement of problem, Aims of the study, Objectives of the study, Significance of the study, Scope of the study and Methodology of the project research.
- Chapter Two: Chapter two gives details on the Conceptual review, Theoretical Review and Empirical Review used in this project.

- Chapter Three: Chapter three describes extensively the methodology used and processes involved in model development
- Chapter Four: Chapter four describes the requirements used in the project and performs comparative analysis between the project and another study.
- Chapter Five: Chapter five discusses conclusion and how the work can be improved.

CHAPTER TWO

LITERATURE REVIEW

2.1 CONCEPTUAL FRAMEWORK

This study's conceptual framework addresses the heart, cardiovascular diseases, metrics used to monitor heart states, risk factors, machine learning and its various types with examples, wearable devices, and the Matrix laboratory.

2.2 HEART

The heart is an organ located behind the sternum at the middle of the chest. It comprises two upper chambers: right and left atria, and two lower chambers, known as the right and left ventricles which make up its four chambers. (Rehman & Rehman, 2023; Hussain & Burns, 2023; Saxton et al., 2023).

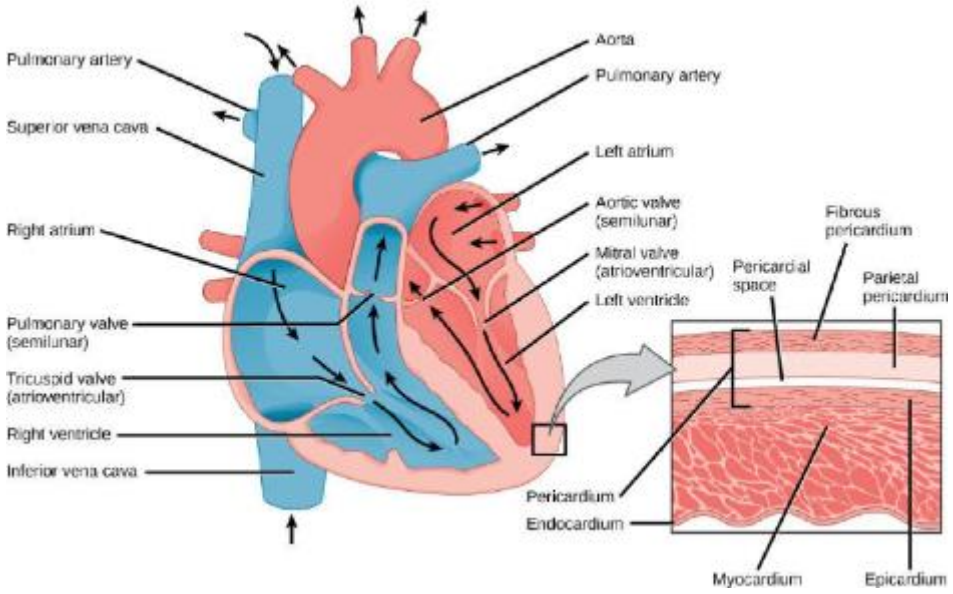


Fig 2: Picture of the Heart(Lokesh et al.,2022)

2.3 METRICS USED TO MEASURE HEART STATES

2.3.1 Pulse rate

Pulse rate is the number of times the heart beats per minute and is commonly monitored at arterial sites such as the wrist or neck. It is an important non-invasive indication of cardiovascular and autonomic nervous system health. Pulse rate is used to monitor cardiac health, detect stress and exhaustion, measure fitness and activity, analyse sleep quality, diagnose shock and dehydration, practise emergency care, and predict health issues, including sepsis, arrhythmia, and mental stress using machine learning models.

2.3.2 Blood oxygen levels

Blood oxygen level is the quantity of oxygen transported by red blood cells in the blood and is commonly measured as SpO_2 (peripheral capillary oxygen saturation) using pulse oximetry. This represents how effectively oxygen is distributed to tissues. Blood oxygen levels are used for respiratory health monitoring, COVID-19 severity assessment, sleep apnoea identification, critical care and ICU management, remote patient monitoring, exercise and sports medicine, and machinand machine learning models for early detection of hypoxia and other medical emergencies.

2.3.3 Electrocardiogram

Electrocardiogram (ECG) is an electrode-on-skin non invasive test to monitor the electrical activity of the heart over time. It is critical for diagnosing abnormalities in heart rhythm, rate, and conduction. It is used for arrhythmia detection, myocardial infarction diagnosis, cardiac monitoring in the intensive care unit, pre-operative assessments, telemedicine and wearable heart monitoring, machine learning-based cardiac classification, and screening for the risk of sudden cardiac death.

2.4 CARDIOVASCULAR DISEASES

Cardiovascular diseases (CVDs) are the major cause of death globally (Guarneros-Nolasco et al., 2021).They are groups of diseases of the heart and blood vessels(World Health Organisation,2021). Cardiovascular diseases have many risk factors or causes, which include physical inactivity, hypertension, overweight and obesity, and smoking (Bays et al., 2021).

Wearable devices, also known as wearables or wearable technologies, are devices that use motion and biometric sensors to measure vital signs, which include step count, activity intensity, heart rate (HR), heart rhythm, blood pressure (BP), oxygen saturation, sleep patterns, maximum oxygen uptake, and temperature (Turakhia, Desai & Harrington, 2022).

2.4.1 Types of cardiovascular diseases

Various types of cardiovascular diseases according to the World health organisation include: Coronary heart disease, Cerebrovascular disease, Peripheral arterial disease, Rheumatic heart disease, Congenital heart disease, Deep vein thrombosis, and Pulmonary embolism(World Health Organization, 2019).

Coronary heart disease: also known as coronary artery disease or ischaemic heart disease, occurs when heart arteries fo not carry sufficent blood to the heart(National Heart, Lung, and Blood Institute, 2024).

Peripheral arterial disease (PAD): is the partial or total blockage of blood flow in the body's distal arteries(Parwani et al., 2023). Risk factors include smoking, hypertension, hyperlipidemia, diabetes, obesity, and a family history of vascular disease(Mejia et al., 2021).

Rheumatic heart disease: is a chronic heart valve problem that affects millions of people worldwide and can result in acute rheumatic fever (ARF), an inflammatory

reaction and it is caused by a preventable infection from the bacterium Group A Streptococcus (Strep A)(Peters et al., 2020).

Congenital heart disease: is a structural defect of the heart and/or major vessels that exists from birth, and it can have a variety of negative short- and long-term effects(Meng et al., 2024).

Deep vein thrombosis and pulmonary embolism: happens when blood clots form in the leg veins and move to the heart and lungs(World Health Organization, 2021).

2.5 MACHINE LEARNING

Machine learning deals with how to make computers acquire knowledge without programming. Data handling through machine learning (ML) teaches machines to better process data(Mahesh, 2020).

2.5.1 TYPES OF MACHINE LEARNING

Machine learning is categorized into three primary types: Supervised learning, Unsupervised learning, and Reinforcement learning.

2.5.1.1 SUPERVISED MACHINE LEARNING ALGORITHMS

It is a kind of algorithm that uses labeled sample data to predict outcomes (Shaveta, 2023). Supervised machine learning algorithms can be classified into two groups. Classification and Regression.

2.5.1.1a CLASSIFICATION

Classification is a supervised learning process which groups data into different classes(Alnuaimi & Albaldawi, 2024). The following models were utilised in this study: Support Vector Machine, Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer Perceptron, and XGBoost algorithm, which will be detailed below:

2.5.1.1b ALGORITHMS USED FOR CLASSIFICATION

1. **Support Vector Machine**

It is a supervised type of a machine learning method that is applied in classification, as well as regression. SVMs have been found to be more efficient in small to medium sized dataset. It is applied to diagnose a disease, image and fraud.

2. **Naive Bayes**

Naive Bayes is a probabilistic classifier which makes use of Bayes theorem and independently assumes predictors. Even though this is a fundamental assumption, it does a great job in a variety of challenging real-life situations and it is useful in sentiment analysis, medical diagnosis, email spam filtering, intrusion detection, and recommenders.

3. **Decision Tree**

A decision tree is a supervised learning algorithm that divides data into branches based on feature values to form a tree-like structure of decisions, supporting both classification and regression tasks, with popular types such as CART (Classification and Regression Trees), ID3 (Iterative Dichotomiser 3), C4.5, and CHAID (Chi-square Automatic Interaction Detection). It is used in medical diagnosis, financial forecasting, customer segmentation, credit scoring, and fault detection.

4. **Linear Discriminant**

It is a supervised tool of dimensional reduction and classification, as it classifies the multidimensional data into a low- dimension Euclidean space with the purpose of maximizing the group separation between multiple classes through the linear combinations of features. It finds application in face recognition, medical diagnosis, gene expression analysis and speech recognition. Made of several varieties that

consist of Linear Discriminant, Quadratic Discriminant Analysis (QDA) and Regularise Discriminant Analysis.

5. Multilayer Perceptron

A Multilayer Perceptron (MLP) is a feed forward artificial neural network which has as its constituents an input layer, one or more hidden layers and an output layer. Individual neurone uses non-linear activation functions, and it is trained in a manner called back propagation. The MLP is considered an essentially partitioned domain of deep learning that is competent enough to approximate complicated non-linear functions, hence suitable for classification and regression issues (Zhang et al., 2022). Medical diagnosis is carried out with the help of MLPs. ECG/EEG Signal classification ECG/EEG Signal classification MLPs are used because of the speed of convergence and adaptation. Financial forecasting, Image recognition and Spam filtering. The popularity of MLPs is based on flexibility, quick convergence and the ability to model a complicated input-output relationship.

6. **XGBoost Algorithm**

It is a powerful and efficient ensemble learning method that uses gradient-boosted decision trees. It incorporates optimization such as L1/L2 regularization, sparsity handling, and parallel tree construction to provide robust performance on both classification and regression tasks. It is used in medical diagnosis, bio-informatics, neurology, environmental engineering, finance and banking, natural language processing, and industrial engineering.

2.5.1.1c REGRESSION

Regression is a statistical method used to model relationships between a dependent variable and one or more independent variables(Li, 2024).

2.5.1.2 UNSUPERVISED LEARNING ALGORITHMS

It is a type of learning which requires no guidance or supervision. This model uncovers trends. Examples include customer segmentation, referral systems and targeted marketing campaigns (Jain & Kumar, 2022) etc.

2.5.1.2a CLUSTERING

Clustering is the process of grouping items into clusters so that those that share the most similarities are grouped and those that share few or no similarities with those in other clusters (Shaveta, 2023).

2.5.1.2b ASSOCIATION RULES

It is a data mining technique employed to identify trends, relationships, and correlations within data sets (Mudumba & Md Kabir, 2024).

2.5.1.3 REINFORCEMENT LEARNING

Reinforcement learning is characterized by a trial-and-error methodology in which agents learn from continuous feedback and interaction without being given clear instructions on what actions to perform (Jia & Wang, 2020). Based on the Markov Decision Process, Reinforcement Learning (RL) employs a variety of algorithms, including Q-Learning, Monte Carlo techniques, and dynamic programming (Jia & Wang, 2020; Daoun et al., 2021).

2.5.1.3a VALUE-BASED REINFORCEMENT

Value-based reinforcement learning (RL) is a prominent approach in which agents learn to evaluate the value of states or actions to guide their decisions (Byeon, 2023; Dabney et al., 2020). This approach entails estimating a series of value functions, referred to as the "value-improvement path," as policies evolve.(Will et al., 2020).

2.5.1.3b POLICY-BASED REINFORCEMENT

Policy-based reinforcement learning (RL) is a methodology in which agents learn optimum tactics by interacting directly with their surroundings. Numerous applications have shown promise, including through-silicon optimization using arrays in high-bandwidth memory (Kim et al., 2024).

2.6 WEARABLE DEVICES

Wearables are also referred to as wearables, wearable technology. They make it possible to track vital signs continuously, which helps with early illness detection and treatment. They monitor parameters like heart rate, calorie burn, speed, and recovery time. These gadgets provide information that can improve training regimens and performance in general. However, they have drawbacks, including expensive prices, short battery life, and data privacy issues (Shang, 2024).

Fig2.1:Picture Of A Wearable Device(Patel et al., 2021)

Wearable technologies are categorized into three primary types: wearable health technologies, wearable textile technologies, and wearable consumer electronics.

2.6.1 Wearable Health Technologies

The technology of wearable health devices has become rapidly an indispensable part of health care, providing a continuous non-stop monitoring and real-time data

collection thus assisting in both diagnosis, treatment and management of various medical conditions. (Lu et al., 2020).

2.6.2 Textile Technologies Wearable

Wearable textile technologies include smart textiles capable of sensing and responding to external stimuli, applicable in health, entertainment, and sports (Kan & Lam, 2021). By using greater surface areas for sensing and energy harvesting, these e-textiles have advantages over conventional wearable accessories (Komolafe et al., 2021).

2.6.3 Wearable Electronics Technologies

Wearable electronic technologies are emerging as useful instruments for a variety of applications, including chronic pain evaluation, health monitoring, and assistive technology (Avila et al., 2021). These devices often include sensors, data processing units, and communication interfaces (E. Sazonov & W. Daoud, 2021). Recent advancements in materials and production methods have enabled the creation of self-healing, recyclable, and adaptable wearable electronics. These technologies combine rigid, flexible, and liquid materials for increased stretch-ability and functionality (Shi et al., 2020).

2.6.4 Roles of Wearable Devices in the Prediction of CVD

Health care is one of the many fields where wearable technology is advantageous. Its uses include long-term, continuous recording of physiological or functional data, which helps with accurate diagnosis and better patient outcomes; collecting vital data in multiple locations; and continuous monitoring via wearable technology, which enables more individualized medical care and treatment (Moshawra et al., 2023).

2.7 MATLAB

Matlab, which stands for "Matrix Laboratory," is a group of software programs created to solve problems in technical computing.

While serving as the Dean of the University of New Mexico's Department of Computer Science in the late 1970s, S.V. Moler developed the MATLAB system. The idea behind developing it was to make the faculty students, the users of the computer programs of Linpack and EISPACK software library without their being compelled to learn the Fortran(Moler & Little, 2020).

2.7.1 Features of Matlab

1. MatLab is a high-performance programming language developed for technical calculations. It includes programming, computation, and visualization in an intuitive setting where issues and their fixes are expressed similarly to mathematical formulas.
2. An array is the main data element of the interactive system MATLAB. It helps solve a variety of problems related to technical calculations, especially those that use matrices and vectors.
3. Simulink, a software interface, a library of mathematical functions, System MatLab, and controlled graphics are all included.
4. In MATLAB, toolboxes make it easier to study and use specialist approaches (Kurasov, 2020).

2.8 THEORETICAL REVIEW

2.8.1 Health belief model

The Health belief model(HBM) is a widely used framework which was created in the 1950s by Irwin M. Rosenstock, Godfrey M. Hochbaum, S. Stephen Kegels, and Howard Leventhal (Alyafei& Easton-Carr,2024).

According to the HBM, the coping strategies and quality of life of cardiovascular disease patients may depend on several factors, such as:

1. Their perceived susceptibility to cardiovascular disease complications and the severity of the disease
2. Their perceived benefits of following a prescribed cardiovascular disease prevention regimen and barriers to doing so
3. Their confidence in their ability to perform self-care activities and cope with cardiovascular disease
4. Their cues to action remind them to engage in cardiovascular disease prevention behaviour.

2.9 EMPIRICAL REVIEW

An empirical review reviews other empirical studies on a particular topic to put a particular research problem to rest. I will outline the articles used in my literature study below:

Islam et al. (2023) presents an IoT and machine learning-based system (Predictis) to predict risk level of cardiovascular diseases. The study addressed the problem of the lack of a user-friendly AI and IoT system for future checkups and CVD predictions. The work done involved creating a wearable gadget and designing and developing an app to assess CVD risk. The study used 10 models (K Nearest Neighbour, Naive Bayes, Random Forest, Support Vector Machine, Gradient Boosting, Stochastic Gradient Descent Classifier, XGBoost classifier, Decision Tree, AdaBoost) for two and three-level classification. The F1 score obtained was 91% for two-level classification and 80.4% for three-level classification.

Omankwu et al. (2023) presented Hybrid Deep Learning Model for Heart Disease Prediction Using Recurrent Neural Network (RNN). According to the

study, heart disease is the leading cause of death around the world. A hybrid model was created by combining recurrent neural networks and gated recurrent units, and a result of 98.6876% was achieved. Other heart disease-related data-sets should be used.

Bhatt et al. , (2023) presented Effective Heart Disease Prediction Using Machine Learning Techniques. This study developed a machine learning algorithm capable of properly predicting cardiovascular disease and lowering mortality rates associated with it. K-modes clustering, cross-validation, and grid search CV were used in conjunction with a multilayer perceptron model. This study collected real-life data on systolic and diastolic blood pressure and reached an accuracy of 87.28%. However, a single data set was utilized, a small collection of demographic and clinical characteristics was employed and the model's performance on the test data set was not examined. The authors suggested that the performance of the test data should be evaluated to establish its generalisability to new and unseen data, the effect of missing data and outliers on the accuracy of the model should be evaluated, and techniques for dealing with missing data and outliers should be developed.

Ahmad A. , Huseyin P. (2023) presented Prediction Of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm. Jellyfish Optimization was used to improve accuracy and prevent over-fitting in the Cleveland heart disease dataset. The support vector machine and the Jellyfish Optimization Algorithm produced an accuracy of 98.47%. Other datasets on heart diseases should be used.

Wu Siyang (2023) presented A Compact LSTM-SVM Fusion Model for LongDuration Cardiovascular Diseases Detection. The study used Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) algorithms, with a precision of 0.9402 for the MIT-BIH arrhythmia dataset and 0.9563 for the MIT-BIH atrial fibrillation dataset. Other datasets should be used to train the Long Short-Term

Memory and Support Vector Machine models. Other machine learning techniques should be used on the MIT-BIH arrhythmia and atrial fibrillation datasets. The authors proposed that we evaluate our model's real-time processing capability and look into the possibilities of implementing it on portable devices.

Qia et al.(2022)presented A cardiovascular Disease Prediction Model Based on routine physical examination indicators using machine learning methods: A cohort study. The researchers used machine learning algorithms to create a cardiovascular disease prediction model based on routine physical examination signs for rural populations in Xinjiang, China. Two surveys were done in Xinjiang, using CLFR, Lasso, and Random Forest approaches to find the most important physical examination indicators for cardiovascular disease. In addition, Adaboost, support vector machine, random forest, and L1-R1 algorithms were used, with the L1-R1-based prediction model outperforming all others in the Xinjiang rural population. However, the study used tiny amounts of variable information and was conducted in China.

Nayeem et al. (2022) presented Prediction Of Heart Disease Using Machine Learning Algorithms. This study used machine learning algorithms, notably K-Nearest Neighbours and Random Forest, on real-world data from the Kaggle cardiovascular dataset. To reduce irrelevant features, information gain was used, and missing values were addressed with the mean. The study accepted the limits of earlier research, handled null values, selected features, and used the Random Forest algorithm to obtain an accuracy of 95.63%. Different classification algorithms with improved feature selection approaches should have been utilized.

Nayab et al. (2021) presented Heart Disease Prediction. The study used Naive Bayes, K Nearest Neighbour, Decision Tree, Random Forest, Artificial Neural

Networks, and Naive Bayes on the UCI Machine Learning repository, with an accuracy of 88%. More heart disease datasets should be employed, various data reduction techniques should be used, more features should be included to improve algorithm implementation, and other algorithms should be used to raise classification accuracy.

Bharti et al.(2021) presented Prediction Of Heart Disease Using a combination of machine learning and deep learning. This study utilized the UCI machine learning heart disease dataset. To reduce dimensionality and extract features, this study uses a variety of machine learning and deep learning methods such as Decision Trees, Naive Bayes, Support Vector Machines, Random Forest, Artificial Neural Networks, and Principal Component Analysis. A deep learning method utilized in the study has an accuracy of 94.2%. The authors suggested using a large dataset, using different data normalisation methods and integrating heart-disease-trained machine learning and deep learning techniques.

Ali et al.(2021) presented Heart Disease Prediction using supervised machine learning algorithms: Performance analysis and comparison. This study used six machine learning techniques to detect heart disease. The Cleveland UCI repository was utilized. The K-Nearest Neighbour, Decision Tree, and Random Forest algorithms obtained 100% accuracy in the study.

Jindal et al.(2021) presented Heart disease prediction using machine learning algorithm. Logistic Regression, K-Nearest Neighbours, and a Random Forest Classifier were used in this work to analyse patients' medical histories. Combining these method resulted in an accuracy of 87.55%. Different machine learning algorithms should be employed.

Rani et al. (2021)presented A decision support system for heart disease prediction based upon machine learning. The work uses a hybrid approach that combines a genetic algorithm with recursive feature elimination to choose features from a Cleveland heart disease data set. To address data imbalance, the synthetic minority oversampling method was used, resulting in a random forest classifier with an accuracy of 86.6%. The hybrid model should be evaluated with data sets other than the dataset used. Other feature selection methods should be utilized.

Krishnan et al. (2021) presented Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction. This study addressed issues raised in previous studies using recurrent neural networks to identify heart disease on the Cleveland data sets. The study used multiple gated recurrent units (GRU), long short-term memory (LSTM), and the Adam optimizer, resulting in a 98.6876% accuracy. Other deep learning techniques could be applied alongside the generated model for better attribute selection.

Sandhya Y.(2020) presented Prediction Of Heart Disease Using Support Vector Machines. This study used support vector machines to forecast heart disease and attained an 85.97% accuracy rate. Only support vector machines were utilized. Different machine learning techniques should be used.

Khan et al, (2020)presented HDPM:An effective Heart Disease Prediction Model For A clinical decision support system. In this work, outliers were detected using DBSCAN, data balanced using SMOTE-ENN, and classified using XGBoost in a clinical decision support system. It fills a gap in previous research by applying outlier detection and data balancing approaches in heart disease prediction models more effectively. The suggested heart disease prediction model exhibited accuracy of 98.40%

and 95.90% on the datasets used, respectively. Other techniques for outlier detection and data balancing should be used

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.0 INTRODUCTION

This chapter describes the research methodology used to predict cardiovascular disease using CardioPredictor(a device that predicts CVDS by assessing its risk level) implemented with the use of Support Vector Machine, Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer perceptron and XGBoost algorithm in Matrix Laboratory.

3.1 THE EXISTING SYSTEM

Before using machine learning models to predict CVDS, traditional sphygmomanometer diagnostics, laboratory tests such as lipid profile, troponins, C-reactive protein, and ECG were used to check for electrocardiogram.

3.2 PROPOSED SYSTEM

The proposed system uses Support Vector Machine, Kernel Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer perceptron and Xgboost algorithm to predict and classify the risk level of CVDS.

3.2.1 Model development

The process of the proposed system is described below:

1. **Design of device**: The device includes an Arduino UNO R4 WIFI, a MAX30102 pulse oximetry sensor, an OLED display, a dot board, an AD8232 ECG module, and jumping wires. All components are linked to the dot board.
2. **Data collectio**n: Electrocardiogram, a vital parameter for predicting CVDS, was collected via Kaggle.
3. **Model Training**: Support Vector Machine, Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer perceptron and

XGBoost algorithm are used to classify the given data into 3 different risk levels. Data were divided into a Ninety (90)% training set and a Ten (10)% testing set.

4. **Model Evaluation**: It evaluates a machine learning model's performance, generalisability, and practical utility using statistical metrics and validation strategies.
5. **Deployment**: The best model was deployed into CardioPredictor using coder.

Some metrics for evaluation include:

a) **Accuracy:**

Accuracy refers to correctly predicted divided by all the predictions.

Accuracy=(TP+TN)/(TP+TN+FP+FN)……..Equ 1

True Positives: It refers to the value of the correctly estimated positive values.

True Negatives- the amount of negative values correctly classified.

False positives: These are the cases of incorrectly forecasted positive cases.

False Negatives: These are the amount of incorrect predictions of the negative projections.

b) Precision:

Precision shows the the percentage of the correctly forecast-ed positive cases among all forecast-ed positives.

Precision = TP / (TP + FP)……..Equ 2

c) **Recall** :

Recall is the percent match between the actual positives and the number of positives that were predicted correctly

Recall = TP / (TP + FN)……..Equ 3

d) **F1-Score**:

The F1-score combines precision and recall. It is calculated as the harmonic mean between the recall and the precision.

F1 Score=2*Precision*Recall/Precision*Recall……..Equ 3

3.3 SYSTEM IMPLEMENTATION

3.3.1 Architecture of the system

The architecture of the designed model is to forecast cardiovascular disease in people and classify the risk level. The architecture consists of various components that ensure the accurate prediction of CVDS. The architecture encompasses data acquisition, model training and validation, and deployment.

Data collection

Medical data- Collecting electrocardiogram data from Kaggle (https://www.kaggle.com/datasets/bjoernjostein/ptbxl-electrocardiography-database?utm_source).

Train/Test Split

Splitting data into 90% training and 10% testing sets to evaluate how well model performed.

Model training

Selection of algorithm: Using Support Vector Machine, Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer perceptron and Xgboost algorithm to classify CVDs into risk levels.

Training: Training the developed model on the 90% training set.

Model Validation

Performance metrics: Calculating Accuracy, F1 score, Recall, Precision and Specificity on the developed model.

Deployment

Deployment: Deploying the best-trained model into the wearable device using Coder.

3.4 Flowchart of the system

Flowcharts are graphical representations of processes or algorithms. Every fundamental stage of the process is labeled with a unique symbol, with a short description of the process following it. The oval represents the beginning and end of a process. The rectangle symbolises a process. The parallelogram represents the system's input and output operations. The diamond represents decisions that are either Yes or No. The arrows tell the direction of control flow and the relationships between the various flowchart symbols.

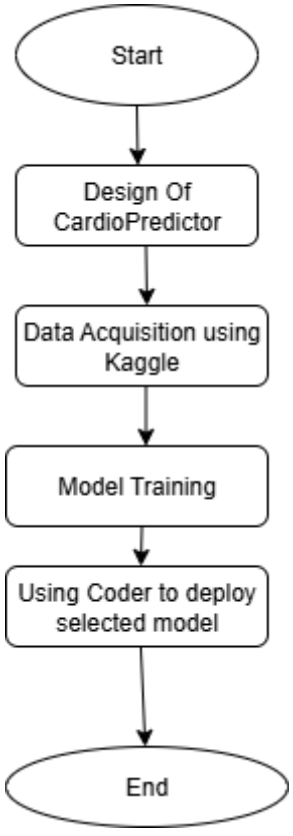


Fig 3: Flowchart of the proposed system

CHAPTER FOUR

SYSTEM IMPLEMENTATION AND RESULTS

4.1 SYSTEM OVERVIEW

The chapter details the implementation of the designed system and its outcome. The primary objective is to employ the Internet of Things (IOT), machine learning, and C++ to forecast cardiovascular disorders, and divide them into two (high or low risk) or three (high, medium, and low) risk categories.

4.2 SYSTEM REQUIREMENTS

4.2.1 HARDWARE REQUIREMENT

A few hardware requirements must be met to set up development and ensure the seamless deployment of CardioPredictor: A device designed for the early prediction of cardiovascular disorders. Here are the essential hardware requirements listed below:

1. Arduino UNO R4 WIFI: The Arduino UNO R4 WIFI is a 32-bit embedded processor that communicates with the MAX30102 sensor, ECG module, and OLED display. It was configured using the C++ programming language.
2. MAX 30102 pulse oximeter: The MAX 30102 pulse oximeter measures heart rate and oxygen saturation and connects to a wristband for continuous monitoring.
3. AD8232 ECG sensor: .It measures the heart rate and blood oxygen levels.
4. Jumping Wires: Jumping wires are simple and reusable wires which make temporary electrical connections between different parts of a circuit.
5. OLED display: It’s a stack of organic thin films positioned between an anode and a cathode that is used to display the results of crucial metrics such as blood oxygen levels, etc.

6. Dot board: The components described above are integrated for optimal placement, routing, and stable connections.

4.2.2 Hardware specifications

The MAX 30102 pulse oximeter is linked to microcontroller pins via jumper wires (metallic conductors with either male or female ends).

1. The MAX 30102 pulse oximeter voltage input is coupled to the microcontroller's 5V logic. The ground connects to the microcontroller's ground pin. The serial clock (SCL) and serial data (SDA) are connected to the microcontroller's pins (A5) and (A4), respectively.

Some connections, eg serial clock and serial data, are direct to the microcontroller, whereas voltage input and ground are via breadboard.

1. The AD8232 ECG sensor was connected to the microcontroller via jumper cables. Connections were made from 3.3 volts logic of the sensor to the 3.3 volts logic of the microcontroller. Ground of the sensor to the ground of the microcontroller. The output pin of the sensor is connected to pin (A1) of the microcontroller.

The LO negative (Leads off comparator output) is connected to pin 11 of the microcontroller. The LO positive (Lean ON comparator) is connected to pin 10 of the microcontroller. All connections are directly to the microcontroller.

The OLED display voltage input is coupled to the microcontroller's 5 volt voltage logic. Ground to ground of the microcontroller. The serial clock (SCL) and serial data (SDA) are linked to the microcontroller's serial clock and serial data pins via the breadboard.

1. The breadboard was used to connect the 5 volt logic pins and the ground pin. The serial clock (SCL) and serial data (SDA) of both the MAX 30102 pulse oximeter and the OLED display are 12C devices.

4.3 SOFTWARE REQUIREMENTS

4.3.1 Programming Arduino in C++

The Arduino UNO R4 WIFI was programmed in C++ to make it functional and able to store readings from the electrocardiogram sensors. It did this by downloading and installing the Arduino from this site: https://www.arduino.cc/en/software. I used edition Arduino 2.3.4

Here is the code used for this below:

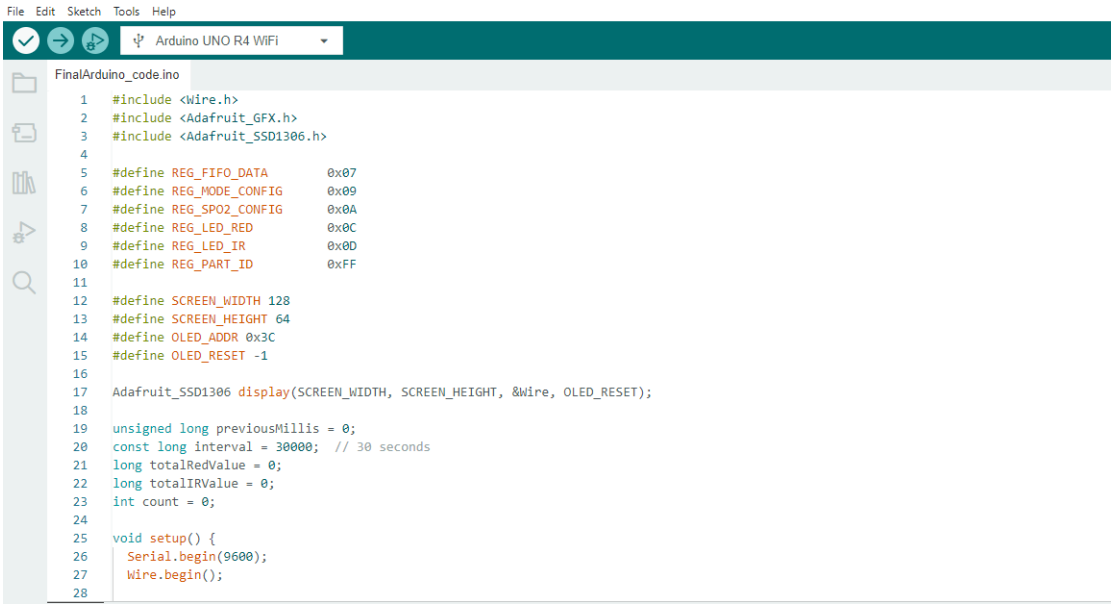


Fig 4 :Arduino Code for program

Fig 4.1 :Arduino Code for program CONTD

```
void loop() {
  int ecgValue = analogRead(A1);
  Serial.print("ECG: ");
  Serial.println(ecgValue);

  long redValue, irValue;
  read_max30102_data(redValue, irValue);

  if (redValue < 50000 || irValue < 50000) {
    Serial.println("No finger detected or values too low.");
  } else {
    totalRedValue += redValue;
    totalIRValue += irValue;
    count++;

    unsigned long currentMillis = millis();
    if (currentMillis - previousMillis >= interval) {
      float averageSpO2 = calculateSpO2(totalRedValue / count, totalIRValue / count);
      float pulseRate = random(60, 140);  // Simulated BPM (replace with real pulse if needed)

      String prediction = classifyRisk(averageSpO2, pulseRate);

      Serial.print("Avg SpO2: ");
      Serial.print(averageSpO2);
      Serial.print(" %, Pulse: ");
      Serial.print(pulseRate);
      Serial.print(" BPM, Prediction: ");
      Serial.println(prediction);
```

Fig 4.2: Arduino Code for program CONTD

```
      display.clearDisplay();
      drawHeart();
      displayHeartRate(averageSpO2, prediction, pulseRate);
      display.display();

      totalRedValue = 0;
      totalIRValue = 0;
      count = 0;
      previousMillis = currentMillis;
    }
  }

  delay(100);
}

// ==== Sensor Setup ====
void max30102_setup() {
  delay(500);
  write_max30102_register(REG_MODE_CONFIG, 0x40); // reset
  delay(100);
  write_max30102_register(REG_MODE_CONFIG, 0x03); // SpO2 mode
  write_max30102_register(REG_SPO2_CONFIG, 0x27); // 411Hz, 16-bit
  write_max30102_register(REG_LED_RED, 0xFF);     // max current
  write_max30102_register(REG_LED_IR, 0xFF);      // max current
  Serial.println("MAX30102 initialized");
}
```

Fig 4.3: Arduino Code for program CONTD

```
      display.clearDisplay();
      drawHeart();
      displayHeartRate(averageSpO2, prediction, pulseRate);
      display.display();

      totalRedValue = 0;
      totalIRValue = 0;
      count = 0;
      previousMillis = currentMillis;
    }
  }

  delay(100);
}

// ==== Sensor Setup ====
void max30102_setup() {
  delay(500);
  write_max30102_register(REG_MODE_CONFIG, 0x40); // reset
  delay(100);
  write_max30102_register(REG_MODE_CONFIG, 0x03); // SpO2 mode
  write_max30102_register(REG_SPO2_CONFIG, 0x27); // 411Hz, 16-bit
  write_max30102_register(REG_LED_RED, 0xFF);     // max current
  write_max30102_register(REG_LED_IR, 0xFF);      // max current
  Serial.println("MAX30102 initialized");
}
```

Fig 4.4: Arduino Code for program CONTD

```
// ==== Calculations ====
float calculateSpO2(long redValue, long irValue) {
  if (irValue == 0) return 0.0;
  float SpO2 = 120 - (25 * ((float)redValue / (float)irValue));
  return constrain(SpO2, 0, 100);
}

String classifyRisk(float spO2, float pulseRate) {
  if (pulseRate > 120 || spO2 < 90) {
    return "High Risk";
  } else if (pulseRate > 100 || spO2 < 95) {
    return "Medium Risk";
  } else {
    return "Low Risk";
  }
}

// ==== Display ====
void drawHeart() {
  display.fillCircle(3, 3, 2, SSD1306_WHITE);
  display.fillCircle(7, 3, 2, SSD1306_WHITE);
  display.fillTriangle(1, 3, 9, 3, 5, 6, SSD1306_WHITE);
}
```

Fig 4.5: Arduino Code for program CONTD

```
void displayHeartRate(float spO2, String prediction, float pulseRate) {
  display.setTextSize(1);
  display.setTextColor(SSD1306_WHITE);
  display.setCursor(0, 0);
  display.print("SpO2: ");
  display.print(spO2);
  display.println(" %");

  display.setCursor(0, 20);
  display.print("Pulse: ");
  display.print(pulseRate);
  display.println(" BPM");

  display.setCursor(0, 40);
  display.print("Prediction: ");
  display.println(prediction);
}
```

Fig 4.6: Arduino Code for program CONTD

4.3.2 Model training in MATLAB environment

This project used an electrocardiogram dataset obtained from Kaggle(https://www.kaggle.com/datasets/bjoernjostein/ptbxl-electrocardiography-database?utm_source) that contained 704, 718 records, no data cleaning was done and it used 7 models to predict and classify CVDS: Support Vector Machine, Kernel Naive Bayes, Decision Tree, Linear Discriminant, K Nearest Neighbour, Multilayer Perceptron, and Xgboost algorithm to predict for the risk levels of cardiovascular diseases, the best three from the (Islam et al, 2023) study, which are K Nearest Neighbour, Multilayer Perceptron, and Xgboost algorithm.

The code used for this is written below.

```
% Step 1
data = readtable('ecg1.csv');
X = data(:, 1:end-1);
Y = categorical(data{:, end});

% Step 2
model = fitcensemble(X, Y, ...
    'Method', 'LogitBoost', ...
    'Learners', 'Tree', ...
    'NumLearningCycles', 100, ...
    'LearnRate', 0.1);

% Step 3
predictedLabels = predict(model, X);

% Step 4
confMat = confusionmat(Y, predictedLabels);
accuracy = sum(diag(confMat)) / sum(confMat(:));

disp("Confusion Matrix:");
disp(confMat);
disp("Accuracy:");
disp(accuracy);
```

Fig 4.7: Picture of Xgboost algorithm code

```
% Step 1
data = readtable('ecg1.csv');

% Step 2
predictors = data(:, 1:end-1);
prediction = data{:, end};

% Step 3
prediction = categorical(prediction);

% Step 4
data_for_app = [predictors, table(prediction)];

% Step 5
classificationLearner
```

Fig 4.8: Picture of Matlab code used for other algorithms

4.3.3 Deployment of MATLAB code

After model training was performed, the MATLAB code of K Nearest Neighbour was deployed using the following steps.

1. Export the best model to the MATLAB workspace, which is located in the Learn tab's Export section.
2. Generate a MATLAB function from the exported model, navigate to the Learn Tab, Export section, and click Generate function.
3. Refine your MATLAB algorithm to ensure MATLAB follows restrictions for code generation.
4. Open the MATLAB Coder app from the APPS tab or use the codegen command.
5. Install a C++ compiler.
6. Integrate the generated C++ Code into the firmware project.

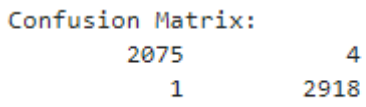


Fig 4.9 : Confusion Matrix For Xgboost algorithm

4.4 RESULT AND DISCUSSION

The metrics used in the study are compared below using 90% training and 10% testing sets. :

Models	Accuracy	Precision	Recall	F1 Score
Xboost algorithm	99.9%	99.8%	99.5%	99.4%
Support Vector Machine	98.9%	98.3%	99.0%	99.5%
Knearest Neighbour	98.7%	97.8%	99.0%	99.5%
Linear Discriminant	97.8%	97.8%	98.9%	99.5%
Multi Layer	98.4%	97.8%	98.4%	99.2%

Perception				
Decision Tree	98.0%	97.8%	97.5%	98.7%
Naive Bayes	96.3%	98.0%	93.4%	96.6%

Table 1: Comparison of Machine Learning Algorithms

Seven machine learning methods were evaluated in this study for risk classification and CVDs prediction. Xgboost was the best algorithm and it outperfomed other algorithms.

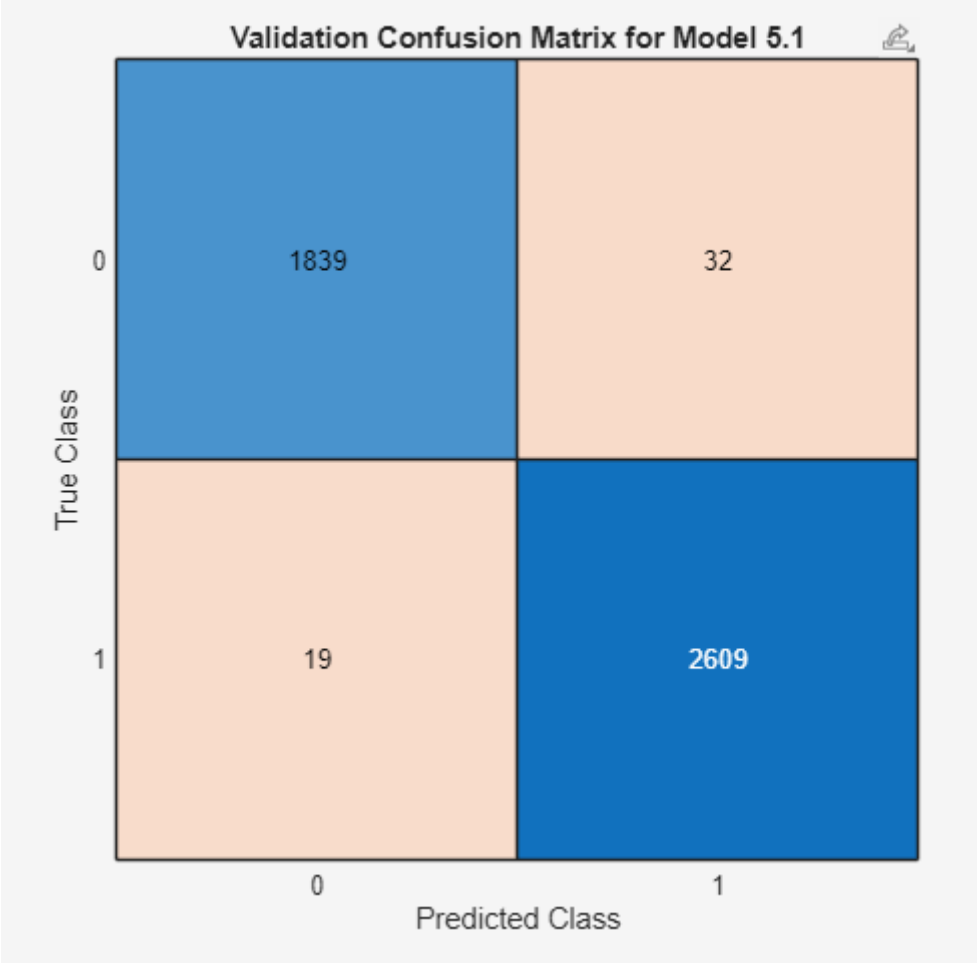


Fig 4.10 : Confusion Matrix For Support Vector Machines

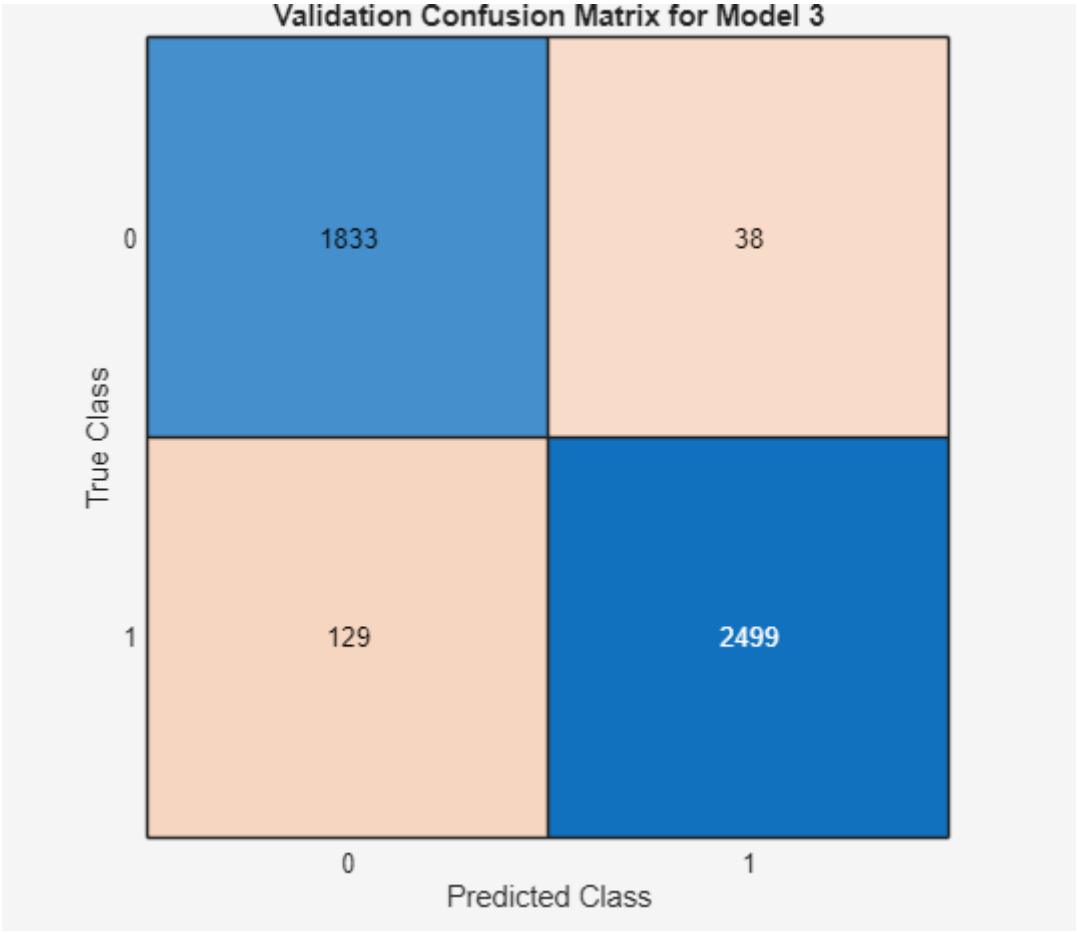


Fig 4.11 : Confusion Matrix For Naive Bayes

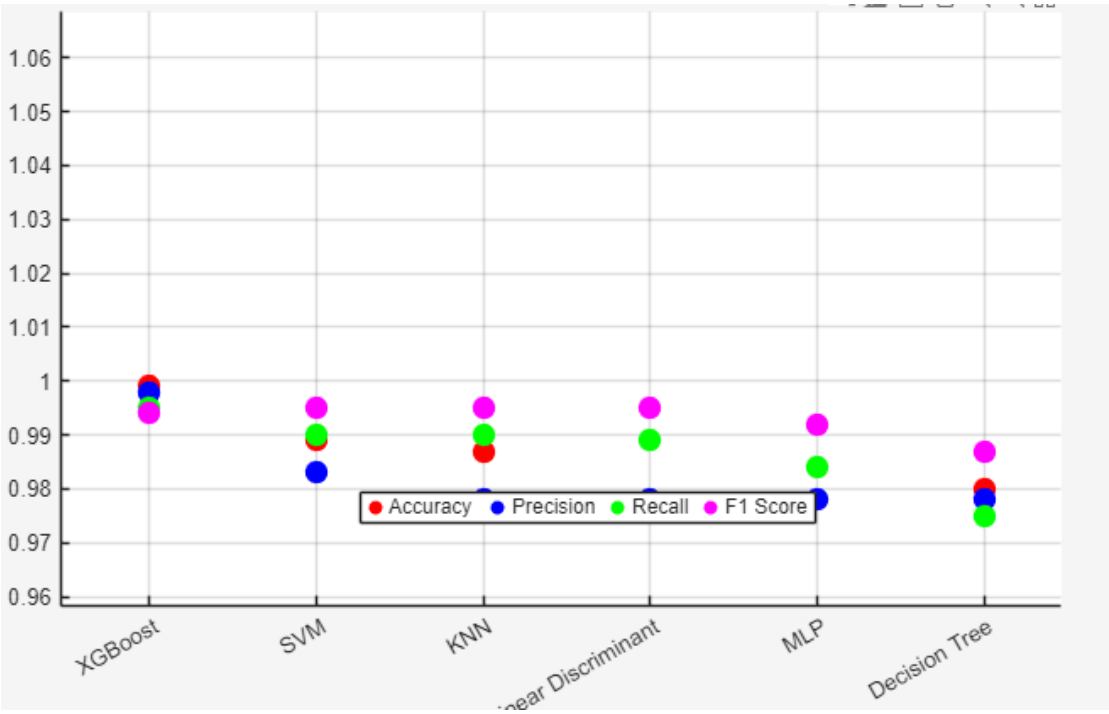


Fig 4.12: Scatter Plot of used evaluation metrics

Features	Used existing methods to compare with the result of proposed model	Model Evaluation Metrics Used	Comparison Of Result
Islam et al. (2023)	No	Uses only F1 score	The highest F1 score for two level classification is KNN of 87.8% and XGB of about 77.7%
Proposed System	Yes	Uses F1 score, Accuracy, Recall and Precsion	All models achieved an f1 score of above 90%

Table 2: Comparison with related work

With very little variation, the device's predictions closely matched those of conventional clinical instruments.

CHAPTER FIVE

SUMMARY, CONCLUSION & RECOMMENDATION

5.1 SUMMARY

CardioPredictor was proposed in this study and it is a system which predicts cardiovascular diseases and classifies them into two or three risk levels depending on the vital parameters used. This device can be used in a hospital for early prediction of CVDs and to reduce the time spent in the hospital on tests.

Machine learning and neural networks were applied for forecasting and classifying the risk levels of CVDS. The developed device measures blood oxygen levels and pulse rate, an electrocardiogram and displays results using an OLED display.

5.2 CONCLUSION

The design and development of CardioPredictor will result in the early detection of CVDs and a reduction in deaths caused by them. This device can be used in hospitals to predict CVDs early on, reducing hospitalization and test time.

5.3 RECOMMENDATIONS

Despite CardioPredictor's many advantages, the following changes might be made to make it even better. Real-world or other relevant ECG data should be used. The wearable device can be powered by solar, electricity, or other sources.

REFERENCES

Rehman, I., & Rehman, A. (2023). Anatomy, thorax, heart. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK470256 ; Hussain, A., & Burns, B. (2023). Anatomy, thorax, wall. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK535414; Saxton, A., Tariq, M. A., & Bordoni, B. (2023. Anatomy, thorax, cardiac muscle. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK535355

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., . . . Fuster, V. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019. Journal of the American College of Cardiology, Volume 76, Issue 25, pg 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010

World Health Organization.(2024). Cardiovascular diseases. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29?utm_source

World Health Organization.(2023). Cardiovascular diseases. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29?utm_source

Ogah, O. S., Orimolade, O. A., & Jinadu, T. O. (2023). Cardiovascular diseases in Nigeria: current status, threats, and opportunities. Circulation, 148(19), pg 1441–1444. https://doi.org/10.1161/circulationaha.123.063671

Bays, H. E., Taub, P. R., Epstein, E., Michos, E. D., Ferraro, R. A., Bailey, A. L., Kelli, H. M., Ferdinand, K. C., Echols, M. R., Weintraub, H., Bostrom, J., Johnson, H. M., Hoppe, K. K., Shapiro, M. D., German, C. A., Virani, S. S., Hussain, A., Ballantyne, C. M., Agha, A. M., & Toth, P. P. (2021). Ten things to know about ten cardiovascular disease risk factors. American Journal of Preventive Cardiology, 5, 100149. https://doi.org/10.1016/j.ajpc.2021.100149

Hughes, A., Shandhi, M. M. H., Master, H., Dunn, J., & Brittain, E. (2023b). Wearable devices in cardiovascular medicine. Circulation Research, Volume 132, Issue 5, pg 652–670. https://doi.org/10.1161/circresaha.122.322389

Dhingra, L. S., Aminorroaya, A., Oikonomou, E. K., Nargesi, A. A., Wilson, F. P., Krumholz, H. M., & Khera, R. (2023). Use of wearable devices in individuals with or at risk for cardiovascular disease in the US, 2019 to 2020. JAMA Network Open, Volume 6, Issue 6, e2316634. https://doi.org/10.1001/jamanetworkopen.2023.16634

Connected Devices and Wearables Category Report, 2030. (2023). https://www.grandviewresearch.com/pipeline/connected-devices-wearables-procurement-intelligence-report?

National Heart, Lung and blood(2022). Institute.https://www.nhlbi.nih.gov/sites/default/files/inline-images/NHLBI_Fig01_final02_01.jpg

Guarneros-Nolasco, L. R., Cruz-Ramos, N. A., Alor-Hernández, G., Rodríguez-Mazahua, L., & Sánchez-Cervantes, J. L. (2021). Identifying the main risk factors for cardiovascular diseases prediction using machine learning algorithms. Mathematics, Volume 9, Issue 20, pg 2537. https://doi.org/10.3390/math9202537

World Health Organization: WHO. (2021). Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29?

What is coronary heart disease? | NHLBI, NIH. (2024). NHLBI, NIH. https://www.nhlbi.nih.gov/health/coronary-heart-disease

Parwani, D., Ahmed, M. A., Mahawar, A., & Gorantla, V. R. (2023). Peripheral Arterial Disease: A Narrative Review. Cureus, 15 (6), e40267.

Peters, P., Yacoub, M. H., Rothenbühler, M., et al. (2020). Rheumatic heart disease: current status of diagnosis and therapy. Cardiovascular Diagnosis and Therapy,Volume 10, Issue 2, pg 315–327.

Mejia, E. L. P., Faxas, S. M., Taveras, N. T., Talpur, A. S., Jitesh, K., Khalid, M., Aruwani, S. K., Khalid, D., Khalid, H., & Memon, S. (2021). Peripheral artery disease as a risk factor for myocardial infarction. Cureus. https://doi.org/10.7759/cureus.15655

Meng, X., Song, M., Zhang, K., Lu, W., Li, Y., Zhang, C., & Zhang, Y. (2024). Congenital heart disease: types, pathophysiology, diagnosis, and treatment options. MedComm,Volume 5, Issue 7, e631.

Thiyagarajan, M., & P, J. (2022). DEEP VEIN THROMBOSIS – A SYSTEMATIC REVIEW. Global Journal for Research Analysis, 4–8. https://doi.org/10.36106/gjra/4700229

Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction. Health & Social Care in the Community, 2023, 1–10. https://doi.org/10.1155/2023/1406060

Almuhanna, M. A., Alanazi, M. H., Ghamdi, R. N. A., Alwayli, N. S., Alghamdi, I. S. G., Qari, A. A., Alzahid, A. A., Alharbi, F. F., Alwagdani, N. M. A., & Alharthi, S. A. (2022). Tachycardia Evaluation and its Management Approach, Literature Review. World Journal of Environmental Biosciences, 11(1), 4–8. https://doi.org/10.51847/7mah6swjqy ; Hobohm, L., Becattini, C., Ebner, M., Lerchbaumer, M. H., Casazza, F., Hasenfuß, G., Konstantinides, S. V., & Lankeit, M. (2020). Definition of tachycardia for risk stratification of pulmonary embolism. European Journal of Internal Medicine, 82, pg 76–82. https://doi.org/10.1016/j.ejim.2020.08.009

Sidhu, S., & Marine, J. E. (2023). Evaluating and managing bradycardia. Trends in Cardiovascular Medicine, 30(5), pg 265–272. https://doi.org/10.1016/j.tcm.2019.07.001

Whitaker, J., Wright, M. J., & Tedrow, U. (2023). Diagnosis and management of ventricular tachycardia. Clinical Medicine, Volume 23, Issue 5, pg 442–448. https://doi.org/10.7861/clinmed.2023-23.5.cardio3

Mahesh, B. (2020). Machine Learning Algorithms - a review. International Journal of Science and Research (IJSR), Volume 9, Issue 1, pg 381–386. https://doi.org/10.21275/art20203995

Shaveta, N. (2023d). A review on machine learning. International Journal of Science and Research Archive, Volume 9, Issue 1, pg 281–285. https://doi.org/10.30574/ijsra.2023.9.1.0410

Alnuaimi, A. F., & Albaldawi, T. H. (2024b). Concepts of statistical learning and classification in machine learning: An overview. BIO Web of Conferences, 97, 00129. https://doi.org/10.1051/bioconf/20249700129

Li, Z. (2024). Regression fundamentals. Geographic Information Science & Technology Body of Knowledge, 2024(1). https://doi.org/10.22224/gistbok/2024.1.11

Jain, N., & Kumar, R. (2022). A Review on Machine Learning & It's Algorithms. International Journal of Soft Computing and Engineering, Volume 12, Issue 5, pg 1–5. https://doi.org/10.35940/ijsce.e3583.1112522

Mudumba, B., & Kabir, M. F. (2024b). Mine-first association rule mining: An integration of independent frequent patterns in distributed environments. Decision Analytics Journal, 10, 100434. https://doi.org/10.1016/j.dajour.2024.100434

Jia, J., & Wang, W. (2020). Review of Reinforcement Learning Research. In *2020 35th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE. https://doi.org/10.1109/YAC51587.2020.187; Daoun, A., Bouzid, S., Haj-Said, H., & Sharaf, A. M. (2021). Reinforcement Learning in Wireless Sensor Networks: A Survey. *IEEE Communications Surveys & Tutorials*, Volume 23, Issue 2, xxxx–xxxx. https://doi.org/10.1109/COMST.2020.xxxxx

Byeon, H. (2023). Advances in value-based, policy-based, and deep learning-based reinforcement learning. International Journal of Advanced Computer Science and Applications, Volume 14, Issue 8.https://doi.org/10.14569/ijacsa.2023.0140838 ; Dabney, W., Barreto, A., Rowland, M., Dadashi, R., Quan, J., Bellemare, M. G., & Silver, D. (2020). The Value-Improvement Path: Towards Better representations for Reinforcement learning. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2006.02243

Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, Volume *11*, *Issue* 1. https://doi.org/10.1186/s40537-024-00973-y

Franjić, S., & Galić, D. (2021). In Shortly about Neural Networks. Advances in Computer and Communication, Volume 2, Issue 1,pg 15–19. https://doi.org/10.26855/acc.2021.08.001

Khan, A., Hayat, S., Zhong, Y., Arif, A., Zada, L., & Fang, M. (2022). Computational and topological properties of neural networks by means of graph-theoretic parameters. Alexandria Engineering Journal, 66, pg 957–977. https://doi.org/10.1016/j.aej.2022.11.001

Lu, L., Zhang, J., Xie, Y., Gao, F., Xu, S., Wu, X., & Ye, Z. (2020). Wearable Health Devices in Health Care: Narrative Systematic Review. JMIR Mhealth and Uhealth, Volume 8, Issue 11, e18907. https://doi.org/10.2196/18907

Kan, C., & Lam, Y. (2021). Future trend in wearable electronics in the textile industry. Applied Sciences, Volume 11, Issue 9, pg 3914. https://doi.org/10.3390/app11093914

Komolafe, A., Zaghari, B., Torah, R., Weddell, A. S., Khanbareh, H., Tsikriteas, Z. M., Vousden, M., Wagih, M., Jurado, U. T., Shi, J., Yong, S., Arumugam, S., Li, Y.,

Yang, K., Savelli, G., White, N. M., & Beeby, S. (2021b). E-Textile Technology Review–From Materials to Application. IEEE Access, 9, pg 97152–97179. https://doi.org/10.1109/access.2021.3094303

Shi, C., Zou, Z., Lei, Z., Zhu, P., Zhang, W., & Xiao, J. (2020b). Heterogeneous integration of rigid, soft, and liquid materials for self-healable, recyclable, and reconfigurable wearable electronics. Science Advances, Volume 6, Issue 45. https://doi.org/10.1126/sciadv.abd0202

Sazonov, E., & Daoud, W. A. (2021b). Grand challenges in wearable electronics. Frontiers in Electronics, 2. https://doi.org/10.3389/felec.2021.668619

Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H., & Raad, A. (2023). Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature review. Sensors, Volume 23, Issue 2, pg 828. https://doi.org/10.3390/s23020828

Kurasov, D. (2020). Mathematical modeling system MatLab. *Journal of Physics: Conference Series, 1691*, Article 012123.

Srinvias, N. T. a. S., Donald, N. a. D., Sameena, N. M., Rekha, N. K., & Srihith, N. I. D. (2023). Unlocking the Power of Matlab: A comprehensive survey. International Journal of Advanced Research in Science Communication and Technology, pg 20–31. https://doi.org/10.48175/ijarsct-9005

Literature Review and Steps to Writing Empirical Review. (2022). https://punchng.com/project-chapter-two-literature-review-and-steps-to-writing-empirical-review/.

The health belief model of behavior change. (2025). PubMed. https://pubmed.ncbi.nlm.nih.gov/39163427/

Green, E. C., Murphy, E. M., & Gryboski, K. (2020). The Health Belief Model. In K. Sweeny, M. L. Robbins, & L. M. Cohen (Eds.), The Wiley Encyclopedia of Health Psychology (gp. 211–214). John Wiley & Sons. https://doi.org/10.1002/9781119057840.ch68

Islam, M. N., Raiyan, K. R., Mitra, S., Mannan, M. M. R., Tasnim, T., Putul, A. O., & Mandol, A. B. (2023). Predictis: an IoT and machine learning-based system to predict risk level of cardio-vascular diseases. BMC Health Services Research, Volume 23, Issue 1. https://doi.org/10.1186/s12913-023-09104-4

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, Volume 16, Issue 2, pg 88. https://doi.org/10.3390/a16020088

Ahmad, A. A., & Polat, H. (2023). Prediction of heart disease based on machine learning using Jellyfish optimization algorithm. Diagnostics, Volume 13, Issue 14, pg 2392. https://doi.org/10.3390/diagnostics13142392

Wu, S. (2023). A compact LSTM-SVM fusion model for Long-Duration Cardiovascular Diseases Detection. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2312.09442

Qian, X., Li, Y., Zhang, X., Guo, H., He, J., Wang, X., Yan, Y., Ma, J., Ma, R., & Guo, S. (2022b). A cardiovascular disease prediction model based on routine physical examination indicators using machine learning methods: a cohort study. Frontiers in Cardiovascular Medicine, 9. https://doi.org/10.3389/fcvm.2022.854287

Nayeem, M. J. N., Rana, S., & Islam, M. R. (2022). Prediction of heart disease using machine learning algorithms. European Journal of Artificial Intelligence and Machine Learning, Volume , Issue 3, pg 22–26. https://doi.org/10.24018/ejai.2022.1.3.13

Akhtar, N. (2021). Heart Disease Prediction. Heart Disease Prediction. https://www.researchgate.net/publication/349140147_Heart_Disease_Prediction

Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. Computational Intelligence and Neuroscience, Volume 2021, Issue 1. https://doi.org/10.1155/2021/8387680

Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Computers in Biology and Medicine, 136, 104672. https://doi.org/10.1016/j.compbiomed.2021.104672

Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. IOP Conference Series Materials Science and Engineering, 1022(1), 012072. https://doi.org/10.1088/1757-899x/1022/1/012072

Rani, P., Kumar, R., Ahmed, N. M. O. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, Volume 7, Issue 3, pg 263–275. https://doi.org/10.1007/s40860-021-00133-6

Krishnan, S., Magalingam, P., & Ibrahim, R. (2021). Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction. International Journal of Power Electronics and Drive Systems/International Journal of

Electrical and Computer Engineering, Volume 11, Issue 6, pg 5467. https://doi.org/10.11591/ijece.v11i6.pp5467-5476

Sandhy, Y. A. (2020). Prediction of Heart Diseases using Support Vector Machine. International Journal for Research in Applied Science and Engineering Technology, Volume 8, Issue 2 , pg 126–135. https://doi.org/10.22214/ijraset.2020.2021

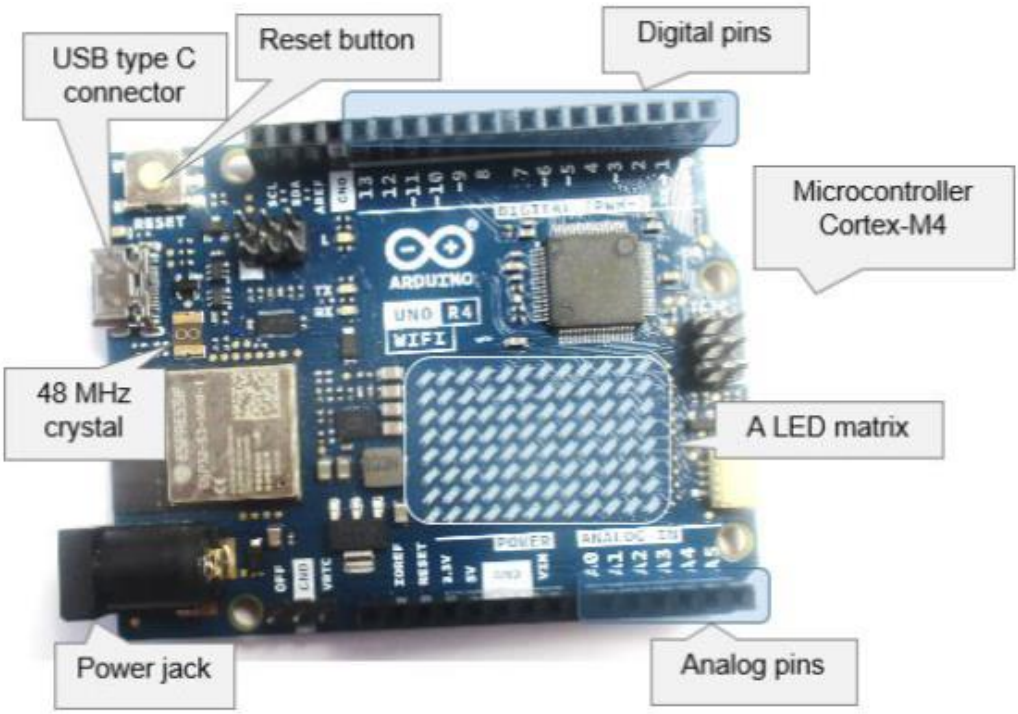
Khan, M. A. (2020). An IoT framework for heart disease prediction based on MDCNN classifier. IEEE Access, 8, pg 34717–34727. https://doi.org/10.1109/access.2020.2974687

Omankwu, Chinecherem.O. &. Ubah I. (2023). Hybrid Deep learning model for heart disease prediction using Recurrent Neural Network (RNN). Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.801

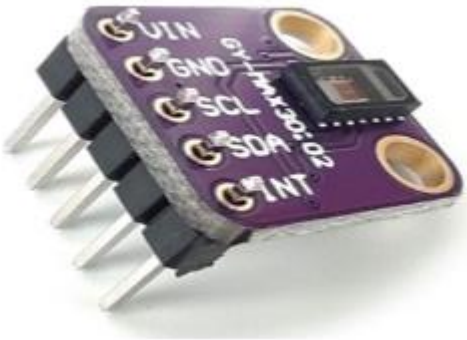
Moler, C., & Little, J. (2020). A history of MATLAB. Proceedings of the ACM on Programming Languages, 4(HOPL), 1–67. https://doi.org/10.1145/3386331

APPENDICES

Appendix A:

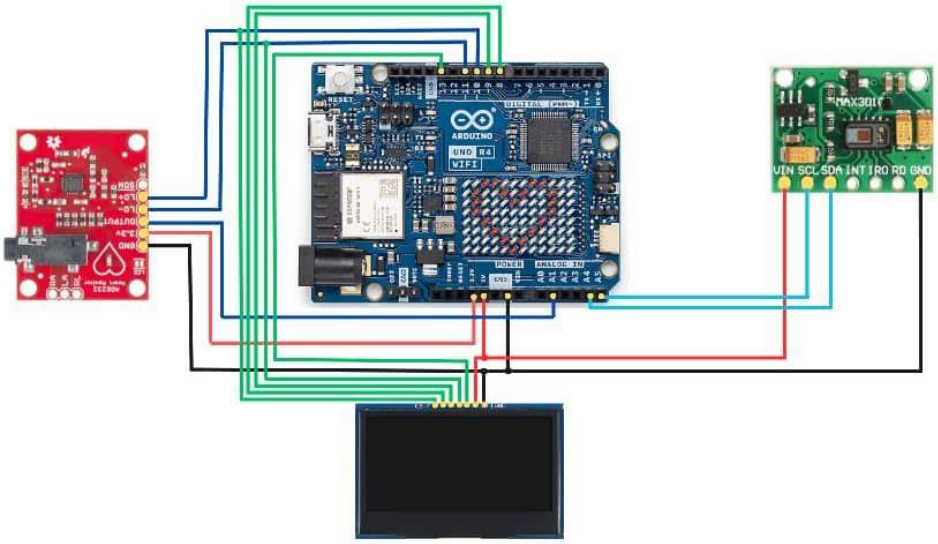


Appendix B:



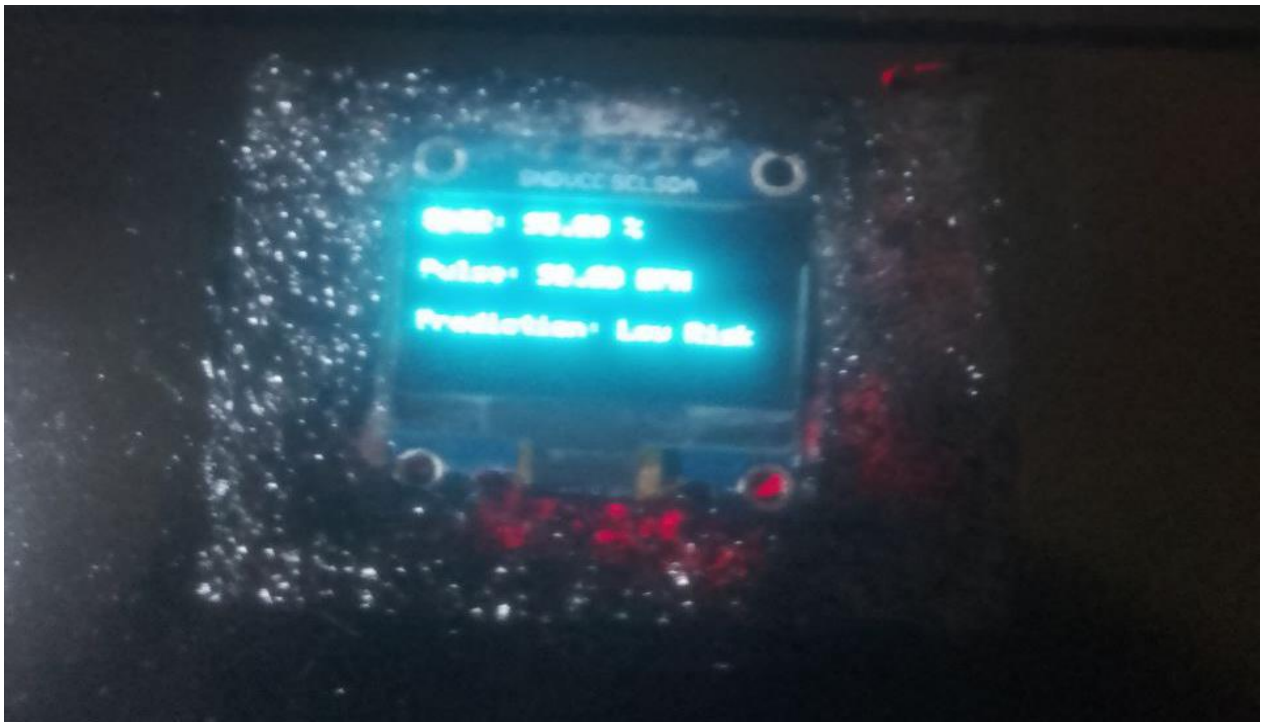
Appendix C:

Appendix D:

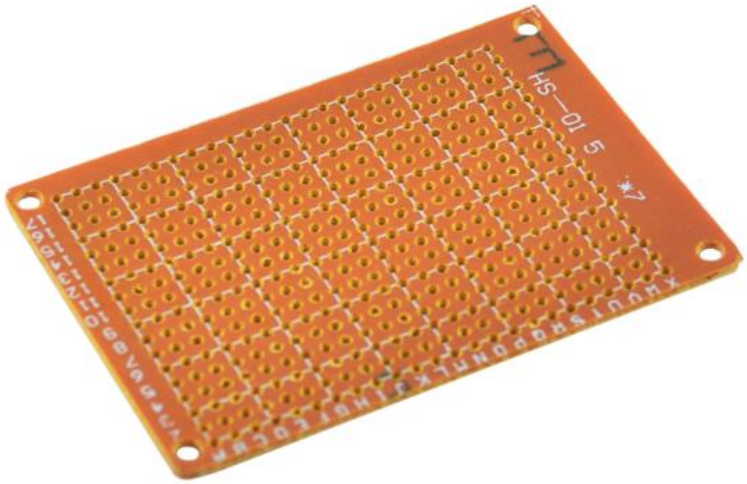


Appendix E:

Appendix F:

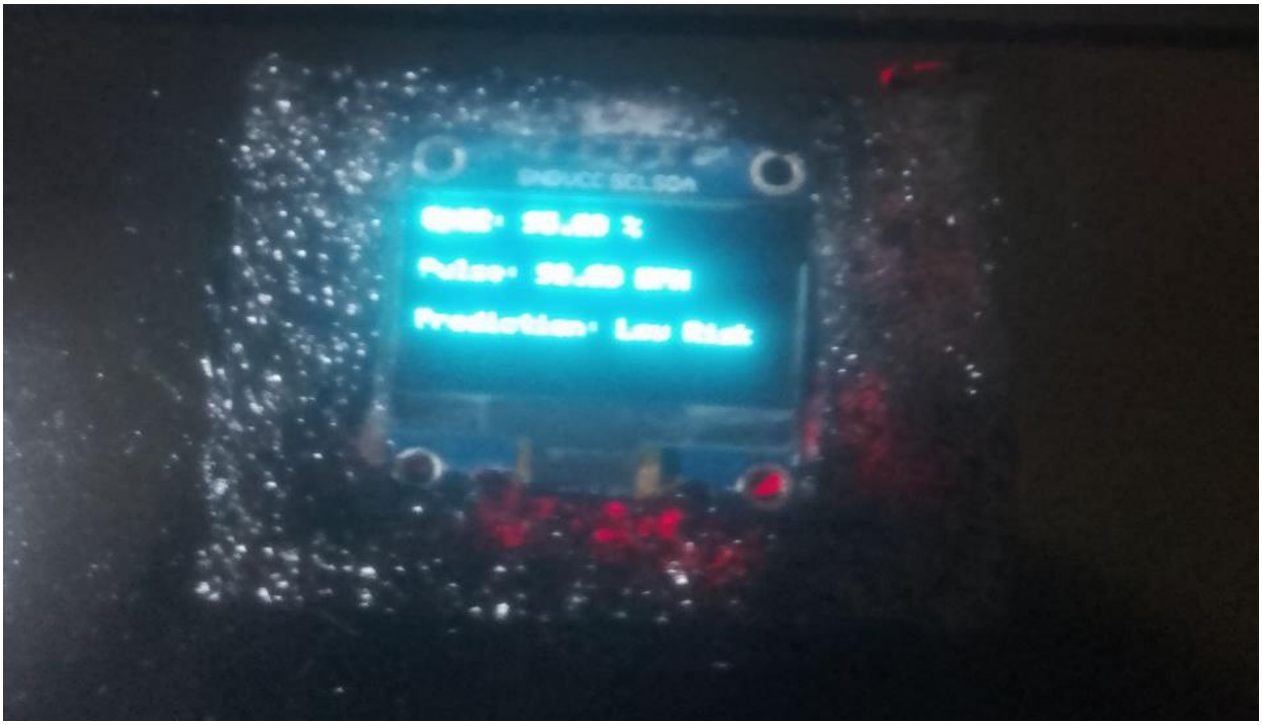


Appendix G:



Appendix H:

Appendix I:



Appendix J: