**HEART DISEASE PREDICTION SYSTEM**

**A Thesis Submitted**

**In Partial Fulfilment of the Requirements**

**for the Degree of**

**MASTER OF COMPUTER APPLICATIONS**

**By**

**ADITYA AWASTHI**

**(University Roll No. 2000290140010)**

**NEHA BANSAL**

**(University Roll No. 2000290140075)**

**MOHD. BILAL**

**(University Roll No. 2000290140070)**

**Under the Supervision of**

**Dr. Arun Kumar Tripathi**

### **Professor**



**Submitted to**

**Department Of Computer Applications**

**KIET Group of Institutions, Ghaziabad**

**Uttar Pradesh-201206**

**(May 2022)**

**CERTIFICATE**

Certified that **Aditya Awasthi (2000290140010), Neha Bansal (2000290140075), Mohd. Bilal (200029014005776)** have carried out the project work having “**Heart Disease Prediction System”** for Master of Computer Applications from Dr. A.P.J. Abdul Kalam Technical University (AKTU**)** (formerly UPTU), Technical University, Lucknow under my supervision. The project report embodies original work, and studies are carried out by the student himself / herself and the contents of the project report do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Date:

**Aditya Awasthi** **(University Roll No.2000290140010)**

**NehaBansal (University No.2000290140075)**

**Mohd.Bilal(University Roll No.2000290140070)**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

**Dr. Arun Kumar Tripathi**

**Professor**

**Department of Computer Applications**

**KIET Group of Institutions, Ghaziabad**

**Signature of Internal Examiner Signature of External Examiner**

**Dr. Ajay Shrivastava**

**Head, Department of Computer Applications KIET Group of Institutions, Ghaziabad**

**Acknowledgement**

Success in life is never attained single handedly. My deepest gratitude goes to my thesis supervisor, **Dr. Arun Tripathi** for his guidance, help and encouragement throughout my research work. Their enlightening ideas, comments, and suggestions.

Words are not enough to express my gratitude to Dr. Ajay Kumar Shrivastava, Professor and Head, Department of Computer Applications, for his insightful comments and administrative help at various occasions.

Fortunately, I have many understanding friends, who have helped me a lot on many critical conditions.

Finally, my sincere thanks go to my family members and all those who have directly and indirectly provided me moral support and other kind of help. Without their support, completion of this work would not have been possible in time. They keep my life filled with enjoyment and happiness.

**TABLE OF CONTENT**

|  |  |
| --- | --- |
| Certificate | i |
| Acknowledgement | ii |
| Abstract | iii |
| List Of Chapters | iv |

|  |  |
| --- | --- |
| 1. Introduction | 7 |
| Overview | 7 |
| 1.1 Problem Definition | 8 |
| 1.2 Motivation | 9 |
| 1.3 Objectives | 9 |
| 1.3.1 Specific Objectives | 9 |
| 1.4 Justification | 10 |
| 1.5 Scope | 11 |
| 1.6 Limitations | 11 |
| 1.7 Types of cardiovascular diseases | 11-12 |
| 1.7.1 Prevalence of cardiovascular diseases | 13 |
| 1.8 Block Diagram | 13-14 |
| 1.9 Project Flow Chart | 15 |
| 1.10 Data Flow Diagram | 16 |
| 1.11 Proposed Model | 17 |
| 1. Literature Review | 18-20 |
| 1. Related Work | 21-22 |
| 3.1 Finding from related work | 22-23 |
| 1. Proposed System Architecture | 24 |
| Overview | 24 |
| 4.1 Hardware Components | 24-25 |
| 4.2 Software Components | 25 |
| 4.2.1 Operating System (Windows 8/8.1/10) | 25 |
| 4.2.2 Coding Language (Python) | 26 |
| 4.2.3 Platform Chrome v50 | 26 |
| 4.2.4 Jupyter Notebook | 26 |
| 4.3 Proposed System Architecture | 27 |
| * 1. Design Pattern | 27 |
| 1. Datasets | 29-30 |
| Overview | 29-30 |
| 1. Methodology | 31 |
| 6.1 K-Mean Algorithm | 31-32 |
| 6.2 Random Forest | 33 |
| 6.3 Support Vector Machine | 34 |
| 6.4 Decision Tree | 34-35 |
| 1. Experiments | 36 |
| 7.1 Data Preparation | 36 |
| 7.2 Exploratory Analysis | 37 |
| 7.3 Checking the Skewness of Data | 38 |
| 7.4 Checking the stats of normal distribution of data | 39 |
| 1. Research Paper | 40-50 |
| Heart Prediction System | 40-50 |

**ABSTRACT**

Heart-related diseases or cardiovascular diseases (CVDs) are the main reason for a huge number of deaths in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need fora reliable, accurate, and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data.

Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart-related diseases. Heart is the next major organ comparing to the brain which has more priority in the Human body. It pumps the blood and supplies it to all organs of the whole body. Prediction of occurrences of heart diseases in the medical field is significant work. Data analytics is useful for prediction from more information and it helps the medical centre to predict various diseases. A huge amount of patient-related data is maintained on monthly basis. The stored data can be useful for the source of predicting the occurrence of future diseases.

Some of the data mining and machine learning techniques are used to predict heart diseases, such as Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM). Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad. To reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Data mining techniques and machine learning algorithms play a very important role in this area. The researchers accelerating their research works to develop software with the help of machine learning algorithms which can help doctors to decide both prediction and diagnosing of heart disease. The main objective of this research project is to predict the heart disease of a patient using machine learning algorithms.

**CHAPTER 1**

**INTRODUCTION**

**OVERVIEW**

The heart is a vital organ in the human body. It delivers blood to every region of our bodies. If it fails to operate properly, the brain and numerous other organs will stop operating, and the person will die within a few minutes. Changes in lifestyle, work-related stress, and poor eating habits all contribute to a rise in the incidence of heart disease. Heart disease has risen to become one of the leading causes of death all over the world.

According to the World Health Organization, 12 million people die each year from heart disease around the world. Since a few years, the global burden of cardiovascular disease has been quickly increasing. Many studies have been carried out in an attempt to define the most important components in heart disease and to precisely forecast the overall risk. Heart disease is also referred to as a "silent killer" because it causes death without causing noticeable symptoms.

Early detection of cardiac disease is critical for implementing lifestyle modifications in high-risk people and, as a result, reducing consequences. This study tries to predict future heart illness by evaluating patient data and using machine-learning algorithms to classify whether they have heart disease or not.

Data on numerous health-related concerns is collected by medical institutions all over the world. These data can be used to gain meaningful insights utilising a variety of machine learning techniques. However, the amount of data collected is enormous, and the data is frequently noisy. Machine learning approaches can quickly investigate these datasets, which are too large for human minds to understand. As a result, these algorithms have recently proven to be quite beneficial in precisely predicting the existence or absence of heart-related disorders. The use of information technology in the health-care industry is growing by the day to help doctors make better decisions. It aids doctors and physicians in disease management, pharmaceutical development, and the identification of patterns and linkages among diagnosis data. Many people who would benefit from preventative care are missed by current methods for predicting cardiovascular risk, while others receive unneeded treatment. By utilising intricate connections between risk factors, machine learning provides an opportunity to increase accuracy. We wanted to see if machine learning could help with cardiovascular risk prediction.

* 1. **Problem Definition**

The detection of cardiac disease is a big challenge. There are instruments that can forecast heart disease, but they are either too expensive or ineffective to quantify the risk of heart disease in humans. The mortality rate and overall consequences of heart disorders can be reduced if they are detected early. However, it is not possible to precisely monitor patients every day in all circumstances, and a doctor's 24-hour consultation is not available because it requires more intelligence, time, and skill. We can use various machine learning algorithms to analyse the data for hidden patterns because we have so much data in today's environment. In pharmaceutical data, the hidden patterns can be employed for health diagnosis.

Heart disease can be effectively managed with a combination of lifestyle changes, medication, and surgery in some circumstances. The symptoms of heart disease can be lessened and the heart's function enhanced with the correct treatment. The projected outcomes can be utilised to prevent and thereby lower the cost of surgery and other costly treatments. The ultimate goal of my research will be to accurately predict the presence of heart disease using only a few tests and features. The attributes that are taken into account are the primary foundation for testing and, for the most part, provide accurate findings. Many more input attributes might be used, but our goal is to forecast the risk of heart disease with fewer and more efficient features. Rather than the knowledge-rich data hidden in the data set and databases, decisions are frequently made purely on doctors' intuition and expertise. This approach leads to unfavourable biases, errors, and high medical expenses, all of which have an impact on the quality of care offered to patients.

* 1. **Motivation**

Machine learning techniques have been around for a long time and have been compared and used for data science analysis in a variety of ways. The main goal of this study project was to look into the feature selection methods, data preparation, and processing methods used in machine learning training models. The difficulty we confront today with first-hand models and libraries is data, where, in addition to their abundance and our cooked models, the accuracy we witness during training, testing, and actual validation has a higher variance. As a result, this project is being carried out with the goal of learning more about the models and then implementing the Logistic Regression model to train the data. Furthermore, because the goal of machine learning is to develop a computer-based system and decision support that can aid in the early detection of heart disease, we developed a model that uses logistic regression to classify whether a patient would have heart disease in ten years or not. As a result, early detection of cardiovascular disease can aid in making lifestyle adjustments in high-risk individuals, reducing consequences and potentially saving lives, which could be a major breakthrough in medicine.

* 1. **Objectives**

The main objective of this study is to create a heart prediction system. A historical heart data set can be used to find and extract hidden knowledge related with diseases.

The goal of the heart disease prediction system is to use data mining techniques on a medical data set to help forecast heart illnesses. Following are the points :

1. Using Logistic Regression, create a machine learning model to predict the future risk of heart disease.
2. To determine significant risk factors based on medical dataset which may lead to heart disease.
3. To examine feature selection methods and comprehend their principles of operation.
   * 1. **Specific Objectives**

Specific objectives are as follows:

1. Provides new approach to concealed patterns in the data.
2. Helps avoid human biasness.
3. To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
4. Reduce the cost of medical test
   1. **Justification**

Clinical choices are frequently decided on the basis of a doctor's intuition and experience rather than the knowledge-rich data contained in the dataset. This approach results in unintended biases, errors, and exorbitant medical costs, all of which have an impact on the quality of care offered to patients. The suggested Clinical decision support will be integrated with computer-based patient records in this system (Data Sets). This will reduce medical errors, improve patient safety, and lessen unwelcome variety in practise. As well as improving patient outcomes as data modelling and analysis tools, this approach looks interesting. Data mining, for example, has the ability to create a knowledge-rich environment that can aid in decision-making. Due to the large number of records in the medical data domain, it has become vital to apply data mining techniques to aid in decision support and prediction in the healthcare industry. As a result, medical data mining contributes to business intelligence, which is beneficial in disease diagnosis.

* 1. **Scope**

Integration of clinical decision support with computer-based patient records could reduce medical errors, increase patient safety, reduce undesirable practise variation, and improve patient outcomes, according to the project's scope. This approach holds promise since data modelling and analysis technologies, such as data mining, have the ability to create a knowledge-rich environment that can help to improve the quality of healthcare judgments dramatically.

* 1. **Limitations**

Medical diagnosis is seen as a significant yet complex task that must be completed accurately and efficiently. It would be really advantageous to automate the process. Clinical judgments are frequently made based on a doctor's intuition and experience rather than on the database's knowledge-rich facts. This approach results in unintended biases, errors, and exorbitant medical costs, all of which have an impact on the quality of care offered to patients. Data mining has the ability to create a knowledge-rich environment that can improve the quality of therapeutic judgments dramatically.

**1.7 Types of cardiovascular diseases**

Heart diseases or cardiovascular diseases (CVD) are a class of diseases that involve the heart and blood vessels. Cardiovascular disease includes coronary artery diseases (CAD) like angina and myocardial infarction (commonly known as a heart attack). There is another heart disease, called coronary heart disease (CHD), in which a waxy substance called plaque develops inside the coronary arteries. These are the arteries which supply oxygen-rich blood to heart muscle. When plaque begins to build up in these arteries, the condition is called atherosclerosis. The development of plaque occurs over many years. With the passage of time, this plaque can harden or rupture (break open). Hardened plaque eventually narrows the coronary arteries which in turn reduces the flow of oxygen-rich blood to the heart. If this plaque ruptures, a blood clot can form on its surface. A large blood clot can most of the time completely block blood flow through a coronary artery. Over time, the ruptured plaque also hardens and narrows the coronary arteries. If the stopped blood flow isn’t restored quickly, the section of heart muscle begins to die. Without quick treatment, a heart attack can lead to serious health problems and even death. Heart attack is a common cause of death worldwide. Some of the common symptoms of heart attack are as follows.

* Chest pain It is the most common symptom of heart attack. If someone has a blocked artery or is having a heart attack, he may feel pain, tightness or pressure in the chest.
* Nausea, Indigestion, Heartburn and Stomach Pain These are some of the often-overlooked symptoms of heart attack. Women tend to show these symptoms more than men.
* Pain in the Arms The pain often starts in the chest and then moves towards the arms, especially in the left side.
* Feeling Dizzy and Lightheaded Things that lead to the loss of balance.
* Fatigue Simple chores which begin to set a feeling of tiredness should not be ignored.
* Sweating.
* Some other cardiovascular diseases which are quite common are stroke, heart failure, hypertensive heart disease, rheumatic heart disease, Cardiomyopathy, Cardiac arrhythmia, Congenital heart disease, Valvular heart disease, Aortic aneurysms, Peripheral artery disease and Venous thrombosis. Heart diseases may Heart Disease Diagnosis and Prediction Using Machine Learning and Data… 2139 develop due to certain abnormalities in the functioning of the circulatory system or may be aggravated by certain lifestyle choices like smoking, certain eating habits, sedentary life and others. If the heart diseases are detected earlier, then it can be treated properly and kept under control. Here, early detection is the main key. Being well informed about the whys and wherefores of heart disease will help in prevention summarily.

**1.7.1 Prevalence of cardiovascular diseases**

An estimated 17.5 million deaths occur due to cardiovascular diseases worldwide. More than 75% deaths due to cardiovascular diseases occur in the middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack. India too has a growing number of CVD patients added every year. Currently, the number of heart disease patients in India is more than 30 million. Over two lakh open heart surgeries are performed in India each year. A matter of growing concern is that the number of patients requiring coronary interventions has been rising at 20% to 30% for the past few year. The rest of the paper is organized as follows. Section 2 describes some of the well-known data mining algorithms used for heart disease prediction. Section 3 describes some of the popular data mining tools used for the data analysis purpose. Section 4 summarizes the methodologies and results of previous research on heart disease diagnosis and prediction. Section 5 discusses the pros and cons on literature survey. Finally, Section 6 concludes the paper along with future scope.

* 1. **Block Diagram**

In this block diagram, there is a micro controller, a heartbeat sensor (Electrodes) circuitry, a GSM MODEM and a LCD Display. The heart beat circuitry consists of a Quad Op-amp IC and three electrodes. These electrodes are placed to the patient who is suffering with high B.P as well as heart problems. The output of this circuitry is considered into logic levels and this output is given to one of the pins of the micro controller. The GSM Modem is used for sending and receiving messages from the patient to a doctor and vice versa. Whenever the heart beat rate or the B.P. exceeds the threshold value. The micro controller will automatically send the signals to the GSM Modem. Through the GSM Modem, the message will give to the concerned person or a doctor. The block diagram is as follows in figure 1.1.

**Diagram

Description automatically generated**

## 1.10 Project Flow Chart

A flowchart is a formalized graphic representation of a logic sequence, work or manufacturing process, organization chart, or similar formalized structure. The purpose of a flow chart is to provide people with a common language or reference point when dealing with a project or process.

The figure 1.2 will be the proposed flow chart that the system will look like

Diagram

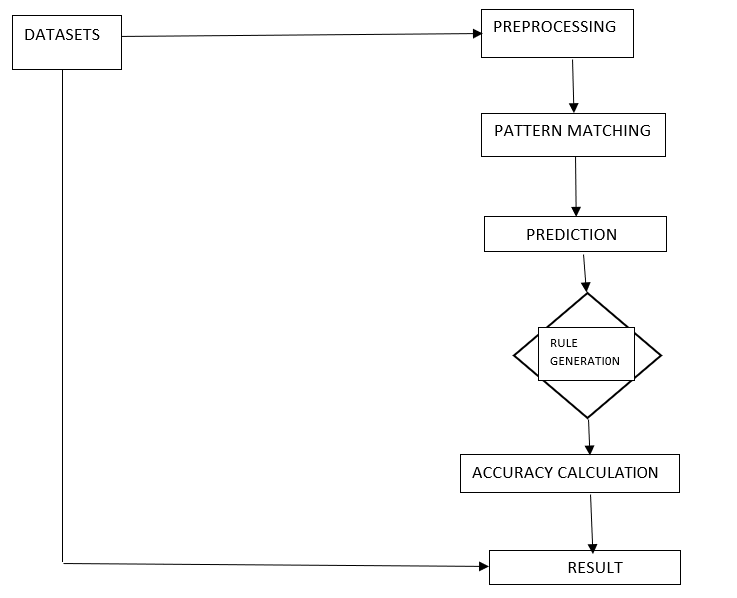
Description automatically generated

## Figure 1.2: Flowchart

## Data Flow Diagram

## 

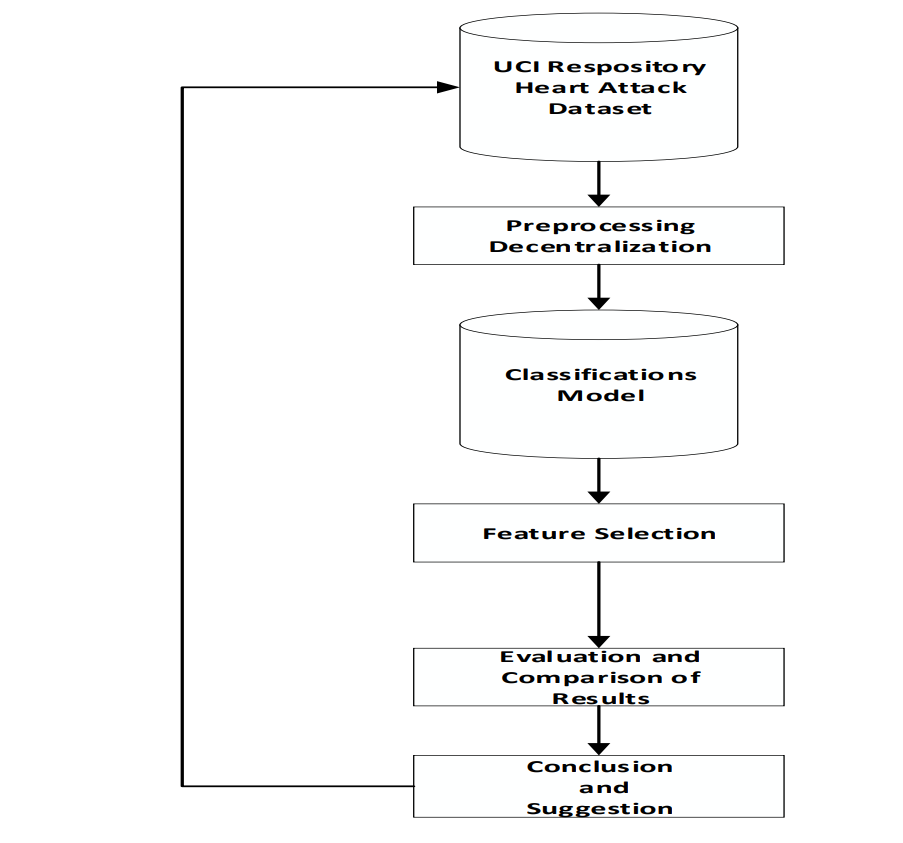
## A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. The data flow diagram is shown in figure 1.3

****

## Figure 1.3:Data flow diagram

**1.11 Proposed Model**

The proposed model is a courageous and ingenious attempt to integrate the current forms of generativist and associationism into one system

****

## Figure 1.4: Proposed Model

**CHAPTER 2**

**LITERATURE REVIEW**

The process of uncovering previously unknown patterns and trends in databases and using that information to develop predictive models is known as data mining. To discover hidden patterns and relationships from massive datasets, data mining integrates statistical analysis, machine learning, and database technology. According to the World Health Statistics 2012 report, one out of every three persons has diabetes which raised blood pressure, is a widespread problem that accounts for around half of all stroke deaths as well as heart disease.

Heart disease, commonly known as cardiovascular disease (CVD), encompasses a wide range of heart-related diseases, not just heart attacks. In many nations, including India, heart disease was the leading cause of death. In the United States, one person dies from heart disease every 34 seconds. Heart diseases include coronary heart disease, cardiomyopathy, and cardiovascular disease. The term "cardiovascular disease" refers to a variety of ailments that affect the heart and blood arteries, as well as the way blood is pumped and circulated throughout the body.

Diagnosis is a challenging and vital activity that must be completed quickly and effectively. The diagnosis is frequently made based on the doctor's knowledge and experience. This leads to unfavourable outcomes and high medical expenses for therapies given to patients. As a result, an automatic medical diagnosing system would be extremely advantageous.

* Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods.
* (Polara Ju, Durga Prasad, & Tech Scholar, 2017) ⁠ proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing.
* (Thanh Noi, P., Kappas, M.”,2018) focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine (SVM) and Artificial Neural Network (ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.
* (Podgorelec V, Kokol P, Stiglic B, Rozman I,2002) ⁠recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms. "Big data analytics recommended Prediction and Analysis of the Occurrence of Heart Disease Using Data Mining Techniques," by Yichuan Wang, LeeAnn Kung, and Terry Anthony Byrd. The major goal is to predict the emergence of cardiac disease in order to make an early automatic diagnostic of the condition with a quick outcome. In addition, the recommended methodology is in a healthcare company with experts who have lost their knowledge and skills, this is crucial. It makes use of medical characteristics such as blood sugar and heart rate, as well as age and sex, are some of the characteristics are included to determine whether or not the person has heart disease.
* (Science & Faculty, 2009) ⁠ suggested heart disease prediction using data mining and machine learning algorithm. The goal of this study is to extract hidden patterns by applying data mining techniques.
* P. P. Chavda proposed an analysis of cardiovascular disease. This paper proposed data mining techniques to predict the disease. It is intended to provide the survey of current techniques to extract information from dataset and it will useful for healthcare practitioners. The performance can be obtained based on the time taken to build the decision tree for the system. The primary objective is to predict the disease with a smaller number of attributes
* . (A & Naik, 2016) ⁠ recommended to develop the prediction system which will diagnosis the heart disease from patient’s medical data set. 11 risk factors of input attributes have considered to build the system. After analysis of the data from the dataset, data cleaning and data integration was performed. He used k-means and naïve Bayes to predict heart disease. This paper is to build the system using historical heart database that gives diagnosis. 11 attributes have considered for building the system. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used.
* (Singh et al., 2018) Heart rate variability was for the detection of coronary artery disease. Fisher method and generalised discriminant analysis with binary classifiers were used for the detection of important features. P. Kaur, R. Kumar, M. Kumar”
* Internet of Things (IoT) is usually referred to as physical things connected to the internet world with limited storage and processing ability. It is still struggling with performance, interoperability, security, and privacy issues challenges with a huge scope for improvement soon”. Elmachtoub, A., Liang, J. C. N., McNellis, R A decision tree includes categorical data (YES/NO) as well as numeric data.
* (Arthiban and Srivatsa ,2012) Diabetes is one of the main causes of heart disease. e classifiers used are e Naive Bayes and SVM for extracting important features and classification purpose.

**CHAPTER 3**

**RELATED WORK**

Many studies are being conducted on disease prediction using artificial intelligence, machine learning algorithms, the Internet of Things, and other technologies.

This paper suggested an IoT-based health monitoring system based on a random forest algorithm. Using a random forest method, many diseases such as heart disease, diabetes, and breast cancer were predicted and attained with maximum accuracy in the dermatology dataset. The combined cardiovascular illness recommendation system employs IoT in a cloud setting to solve a multi-class classification issue that can predict eight different types of cardiovascular disorders. To increase accuracy, the model used a feature selection strategy. Lung Cancer developed an Io-based prediction system employing segmentation and cluster-based fuzzy classification in one study. The proposed system, which was created in the MATLAB environment, is designed to classify radiographic images of the lungs. He has created an online platform for clinical decision support systems that is web-based (CDSS). It is based on the most effective deep neural network (DNN). They created a cloud based CDSS system to forecast the severity of chronic kidney disease (CKD). They used Internet of Things sensors to collect clinical data from relevant patients, anticipate normal and abnormal states, and obtain the best DNN accuracy. Researchers have developed an Android-based monitoring device that can track the heart rate of heart disease patients. We built a model that can trigger an alarm if a patient's heart rate is irregular using a decision tree method. It delivers an Io-based system that uses machine learning algorithms to detect cardiac problems early.

Researchers looked at IoT security and privacy elements such as security criteria, utilization, and categories of healthcare attacks. To deal with noisy missing values, they employ a decision tree-based classifier of how people maintain social and economic bases for sustainability. Developed a mobile healthcare environment that is capable of driving.

With growing development in the field of medical science alongside machine learning various experiments and research has been carried out in these recent years releasing the relevant significant papers. The paper [1] propose heart disease prediction using KStar, J48, SMO, and Bayes Net and Multilayer perceptron using WEKA software. Based on performance from different factor SMO (89% of accuracy) and Bayes Net (87% of accuracy) achieve optimum performance than KStar, Multilayer perceptron and J48 techniques using k-fold cross validation. The accuracy performance achieved by those algorithms are still not satisfactory. So that if the performance of accuracy is improved more to give batter decision to diagnosis disease. [2]In a research conducted using Cleveland dataset for heart diseases which contains 303 instances and used 10-fold Cross Validation, considering 13 attributes, implementing 4 different algorithms, they concluded Gaussian Naïve Bayes and Random Forest gave the maximum accuracy of 91.2 percent. Using the similar dataset of Framingham, Massachusetts, the experiments were carried out using 4 models and were trained K Neighbors Classifier, Support Vector Classifier, Decision Tree Classifier and Random Forest Classifier.

**3.1 Finding from related work**

We did not find any reliable model which can accurately predict cardiovascular disease in developing nations. Most of the researchers developed the cardiovascular disease models based on single and two algorithm combinations which are not strong enough to make accurate predictions in developing nations. To overcome these research gaps, we performed the following:

I. We used three different feature selections and also the five classifiers for cardiovascular disease prediction. Till date, we have not found any research work which utilized such classifiers for feature selections and cardiovascular disease predictions. Mostly, research scholars

used only one or two algorithms in their research work.

II. We did not find any research work pertaining to cardiovascular disease prediction based upon clinical data for the area of Jammu and Kashmir (India) using machine learning techniques, therefore, this research work may be considered as the first cardiovascular disease prediction model for such developing areas.

III. We applied machine learning techniques for cardiovascular disease because it is one of the best and most innovative techniques for prediction work and is also used for other disciplines like cancer disease prediction, stock market

predictions and weather forecasting, etc.

**CHAPTER 4**

**PROPOSED SYSTEM ARCHITECTURE**

**OVERVIEW**

This section depicts the overview of the proposed system and illustrates all the components, techniques and tools are used for developing the entire system. To develop an intelligent and user-friendly heart disease prediction system, an efficient software tool is needed to train huge datasets and compare multiple machine learning algorithms. After choosing the robust algorithm with best accuracy and performance measures, it will be implemented on the development of the smartphone-based application for detecting and predicting heart disease risk level. Hardware components like Arduino/Raspberry Pi, different biomedical sensors, display monitor, buzzer etc. are needed to build the continuous patient monitoring system. Portrays the block diagram of the whole system workflow.

**4.1 Hardware Components**

1. Arduino/Genuino Uno: Arduino/Genuino Uno is a microcontroller board where ATmega328P microcontroller is inserted. It has 14 digital input or output pins among them 6 can be used as PWM outputs and 6 analog inputs. Not only functional pins but also some other pins exist such as power pin, a 16 MHz quartz crystal oscillator, an ICSP header, a reset button, and a USB connection. Although there is a power jack, USB connection cable can also be used as a power supply. It has everything needed to provide support for the microcontroller, simply it can be connected to a computer using USB cable or powered with an AC-to-DC adapter or battery.
2. Heartbeat Sensor: The heartbeat sensor has emerged on the postulation of light modulation on blood flow through the finger in each pulse. Any change of light intensity through that organ (a vascular region) is predicted with the rate of heart pulses and since light is also absorbed by blood, those signal pulses are equivalent to the heartbeat pulses. It is constructed in such a way that gives a digital output of the heartbeat, while a finger is placed onto it. That digital output can also be connected to Arduino directly for measuring the Beats per Minute (BPM) rate.
3. Electric Buzzer: Buzzer is an electronic device used to generate alert sound. Here the buzzer is used to alert the caretaker during extreme condition. This sound indicates that the patient’s health is at immediate risk. Thus, nearest family members and caregivers are easily notified and help it to operate as a real-time patient monitoring system.
4. Temperature and Humidity Sensor (DHT11): DHT11 is a device which includes both humidity and temperature sensor. So, it can measure temperature and air humidity of a specific place whether at indoor or outdoor. Humidity and temperature are the important parameters to monitor patient’s comfort or to check the patient’s physical condition in that place as the patients feel uncomfortable due to rapid temperature and humidity changes which can also trigger abnormal cardiac responses.

**4.2 Software Component**

**4.2.1 Operating System (Windows 8/8.1/10)**

An operating system (OS) is [system software](https://en.wikipedia.org/wiki/System_software) that manages [computer hardware](https://en.wikipedia.org/wiki/Computer_hardware), [software](https://en.wikipedia.org/wiki/Software) resources, and provides common [services](https://en.wikipedia.org/wiki/Daemon_(computing)) for [computer programs](https://en.wikipedia.org/wiki/Computer_program).

[Time-sharing](https://en.wikipedia.org/wiki/Time-sharing) operating systems [schedule tasks](https://en.wikipedia.org/wiki/Scheduler_(computing)) for efficient use of the system and may also include accounting software for cost allocation of [processor time](https://en.wikipedia.org/wiki/Scheduling_(computing)), [mass storage](https://en.wikipedia.org/wiki/Mass_storage), printing, and other resources.

For hardware functions such as input and output and memory allocation, the operating system acts as an intermediary between programs and the computer hardware, although the application code is usually executed directly by the hardware and frequently makes system calls to an OS function or is interrupted by it. Operating systems are found on many devices that contain a computer – from cellular phones and video game consoles to web servers and supercomputers.

**4.2.2 Coding language(python)**

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

It is used for:

* web development (server-side),
* software development,
* mathematics,
* system scripting.

**4.2.3 Platform (Chrome v50)**

**4.2.4 The Jupyter Notebook**

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Its uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebook documents. The “notebook” term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context.

**4.3 Proposed System Architecture**

In this system we are implementing effective heart attack prediction system using Naïve Bayes algorithm. We can give the input as in CSV file or manual entry to the system. After taking input the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

**4.4 Design parameters**

Parameters or design variables are controlled factors that influence performance. They can be of various natures: geometric dimensions, properties of materials, structural choices, etc. They may be quantitative or qualitative, continuous, or discrete. The selection and the number of parameters also determine the definition of the optimization problem. There are many factors for increase the search space, but the optimization process will take longer. For example, a suitable geometrical shape, to ensure the validity of the modelling retained and its proper functioning, etc

The proposed system architecture is shown in figure 4.1

Diagram

Description automatically generated

## Figure 4.1: Proposed system architecture

**CHAPTER 5**

**DATASETS**

**OVERVIEW**

A dataset (example set) is a collection of data with a defined structure. Among the 11 features in this dataset, there is a target variable. It has 6 nominal variables and 5 numeric variables. The detailed description of all the features are as follows:

1. Age: Age of patients in years (numerical)

2. Sex: Gender of patient (Male - 1, Female - 0) (Nominal)

3. Chest Pain Type: A patient's chest pain is classified into four categories: typical, angina, non-anginal pain, and asymptomatic (Nominal).

4. Resting Bps: Number of mm/HG of blood pressure at resting (Numerical)

5. Cholesterol: Number of milligrams of cholesterol in a litre of blood

6. Fasting blood sugar: A value of 1 indicates a true value and 0 indicates a false value (Nominal) for fasting blood sugar levels >120 mg/dl

7. Resting ECG: In the electrocardiogram while at rest, three values are presented: 0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy (nominal)

8. Max Heart Rate: Result of maximum heart rate (Numeric)

9. Exercise Angina: Angina induced by exercise 0 portraying NO, 1 depicting Yes (Nominal)

10. Old peak: Compared to rest, exercise caused ST-depression (Numeric)

**Target variable**

11. HeartDisease : It is the target variable which we have to predict 1 means patient is suffering from heart risk and 0 means patient is normal.

Graphical user interface, application

Description automatically generated

The dataset is publicly available on the Kaggle Website at which is from an ongoing cardiovascular study on residents of the town of Framingham, Massachusetts. It provides patient information which includes over 919 records and 11 attributes. The attributes include age, sex, chest pain type, resting blood pressure, cholesterol, fasting, resting electrocardiographic results, maximum heart rate, exercise induced angina, ST depression induced by exercise, slope of the peak exercise, number of major vessels, and target ranging as heart disease from 0 to 1, where 0 is absence of heart disease. The data set is in csv (Comma Separated Value) format which is further prepared to data frame as supported by panda library in python.

**CHAPTER 6**

**METHODOLOGY**

**OVERVIEW**

Methodology refers to the overarching strategy and rationale of your research project. It involves studying the methods used in your field and the theories or principles behind them, in order to develop an approach that matches your objectives.

**6.1 K-means Algorithm**

K-means creates k groups from a set of given objects so that the members of a group are more similar. Other than specifying the number of clusters, k-means also “learns” the clusters on its own without any information about which cluster a particular observation should belong to. That’s why k-means can be called as semi-supervised learning method. K-means is especially effective over large datasets.

**Diagram

Description automatically generated**

## Figure 6.1: K-mean algorithm

K represents the number of the data point in the KNN algorithm that is close to the new data point. For instance, if K = 1, 2, 3 then it will choose one nearest neighbor, two nearest neighbors three nearest neighbor data points respectively. Now, it will classify the data point based on most of the voting. This idea is portrayed in Fig. 6.1. After classification, the new data point is represented by the green star point using the nearest neighbor technique, as depicted in Fig. 6.1. Here two classes A and B which is represented by a sky colour rectangle and black colour circle respectively. For K = 1, the star is close to the rectangle, hence the KNN algorithm classified it as a class A.

**6.2 Random Forest**

Random forest is a classification algorithm, and it constructs multiple decision trees during the training phase. The random forest algorithm takes a final decision to choose the trees based on most of the voting. It reduces the risk of overfitting of the model based on the utilization of multiple trees. It works efficiently on a

large database and produces highly accurate results.

Steps of Random Forest algorithm:

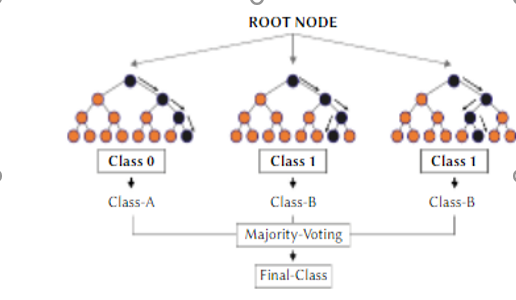
Step 1: Pick random R data objects from the training dataset.

Step 2: Create a DT (Decision Tree) for the K data point.

Step 3: Picks the n-tree subset from the newly generated trees and execute

step 1 and step2

Step 4: Take the decision or result based on most voters portrays the working of the Random Forest algorithm. We made three different Decision trees from the dataset, but the final classification algorithm takes decisions based upon most voters.



## Figure 6.2: Random Forest algorithm

**6.3 Support Vector Machine (SVM)**

SVM is a classifier or a classification algorithm that uses the hyperplane to classify the data. The primary objective of SVM must find the best and optimal hyper-plane in n-dimension space because it requires such type of hyper-plane which has a maximummargin. Hence, the thumb rule is choosing the hyper-plane, which separates the two classes better. The circle and rectangle are two classes and separated

from each other by hyper-plane. But the queries raised here that how many hyper-planes are possible, and which one has to choose out of possible hyper-planes. It will be chosen based upon the maximum margin between the classes. For the non-linear data points, SVM uses the kernel trick to draw the hyper-plane between them. Hence, it can

say that SVM also works for non-linear data points.

**6.4 A Decision Tree**

A decision tree can be used for both regression and classification problems. It is a non-parametric supervised learning method. It is a tree-structured classifier where the characteristic of a dataset is defined by internal nodes and branches of the tree represent the decision rules and the outcome is defined by each leaf node.

Diagram

Description automatically generated

## Figure 6.3: Decision tree

There are two nodes in the Decision tree, which are the leaf node and the decision node. Decision nodes have several branches and are used to make some decisions while leaf nodes are the outcome of such decisions and do not have any further branches. A test or decision is carried out based on features of the dataset. A decision tree asks a question, and it further partitions the tree into sub-trees based on the appropriate response (Yes / No), as portrayed. A decision tree includes categorical data (YES/NO) as well as numeric data.

**CHAPTER 7**

**EXPERIMENTS**

**OVERVIEW**

An experiment is a procedure carried out to support or refute a hypothesis, or determine the efficacy or likelihood of something previously untried. Experiments provide insight into cause-and-effect by demonstrating what outcome occurs when a particular factor is manipulated.

**7.1 Data Preparation**

Since the dataset consists of 4240 observations with 388 missing data and 644 observations to be risked for heart disease, two different experiments were performed for data preparation. First, we checked by dropping the missing data, leaving with only 3751 data and only 572 observations risked for heart disease.Chart, bar chart

Description automatically generated

## Figure 7.1

## Where 0 defines normal person and 1 defines person with disease.

**7.2 Exploratory Analysis:**

**Chart

Description automatically generated**

Correlation Matrix visualization Before Feature Selection Correlation Matrix Visualization It shows that there is no single feature that has a very high correlation with our target value. Also, some of the features have a negative correlation with the target value and some have positive. The data was also visualized through plots and bar graphs.

**7.3 Checking the Skewness of the Data**

For checking the attribute values and determining the skewness of the data (the asymmetry of a distribution), many distribution plots are plotted so that some interpretation of the data can be seen. Different plots are shown, so an overview of the data could be analyzed. e distribution of age and sex, the distribution of chest pain and trestbps, the distribution of cholesterol and fasting blood, the distribution of ecg resting electrode and thalach, the distribution of exang and oldpeak, the distribution of slope and ca, and the distribution of thal and target all are analysed and the conclusion is drawn as shown in Figures 2 and 3 of figure 7.3. By analysing the distribution plots, it is visible that thal and fasting blood sugar is not uniformly distributed and they needed to be handled; otherwise, it will result in overfitting or underfitting of the data .

Chart, histogram

Description automatically generated

## Figure 7.3

**7.4 Checking Stats of the Normal Distribution of Data.**

Checking the features which are important for heart disease and not important for heart disease is shown in Figures 4 and 5 of figure 7.4, respectively. Here the important factors show a different variation which means it is important. e conclusion which can be drawn from these statistical figures is that we can see a Gaussian distribution which is important for heart disease and no Gaussian distribution which is playing that much important role in heart disease.

Chart, histogram

Description automatically generated

## Figure 7.4

**CHAPTER 8**

**Research Paper**

**Heart Disease Prediction System**

The main objective of this study is to create a heart prediction system. A historical heart data set can be used to find and extract hidden knowledge related with diseases.

**ABSTRACT**

In addition to information, communications, and technology, there are new areas such as machine learning, the Internet of Things, and cloud computing. These technologies can save millions of lives in the healthcare world and can be used in healthcare systems that lack medical expertise. In the healthcare field, these technologies have the potential to save millions of lives and can be employed in healthcare systems that lack medical competence. Fast food consumption has risen in recent decades, contributing to high cholesterol, diabetes, and other health issues affecting the heart and other organs. Another factor that contributes to health issues, such as cardiovascular disease, is lifestyle changes. Cardiovascular is also a top cause of the deaths, according to the World Health Organization. The purpose of this study was to analyse existing cardiovascular disease data to predict heart disease early and prevent its onset. Records of heart disease patients are extracted from India and stored via the cloud. The proposed task predicts the likelihood of heart disease and ranks patient risk levels by performing various data mining techniques such as naive arrays, decision trees, logistic regression, and random forest. In this study we analysing the performance of different machine learning algorithm by comparison.

**Introduction-**

Heart attacks, arrhythmias, pericardial illnesses, and fatalities have all increased in the last several decades over the world (CVDs). In the India, CVD kills one person every minute. Many researchers have experimented with machine learning classification algorithms in order to diagnose cardiovascular illness and assist medical practitioners throughout the world in improving local health systems. According to the World Health Organization (WHO), cardiovascular disease causes about a quarter of all fatalities worldwide, with more than three-quarters of those deaths happening in low- and middle-income nations. In India, 25 percent of the population aged 2569 dies from cardiovascular disease [1]. The Internet of Things (IoT) is also known as the physical Internet since it is linked to the Internet but has limited storage and processing capabilities. We're still dealing with challenges like performance, interoperability, security, and privacy, and there's a lot of room for progress [2]. [3] [4].

Smart sensors are utilized for continuous monitoring of specific patient ailments, and IoT has shown to be the gold standard for medical systems. Biomedical sensors are examples of smart sensors that collect health-related data and send it to doctors via the cloud/edge for further diagnosis. As a result, regardless of region, IoT can assist in bridging the gap between patients and doctors [5]. Cloud computing provides nearly limitless processing and storage capacity. Data mining is a clever technology used to analyse large numbers of documents and extract new information, despite being a more advanced solution for solving IoT technological challenges. [6]. Various machine learning algorithms can be utilized to produce specific judgements, estimations, and predictions. The majority of medical data is now collected via computer systems, yet it is not used for analysis everywhere in the world. It accumulates as an old handwritten trace in the database and is no longer useful. Cancer, cardiovascular disease, diabetes, and dengue fever can all be predicted using this information [7]. As a result, we suggest a novel information technology (IT) model. The IoT ML Cloud model combines machine learning with IoT and cloud computing, resulting in three integrated technologies that work together to eliminate hurdles in existing and future global systems. future. Medical attention. Medical assistive technology and services are inextricably linked to public health and improved medical facilities. In the prediction of chronic diseases, the integration of cloud computing and IoT for healthcare-oriented modern technology applications is critical. High security, enhanced efficiency, virtualization, dependability, scalability, resource sharing, cost reduction, medical monitoring, management, and administration systems are all advantages of the growth of public cloud (cloud computing) in hospitals. The process is highly efficient and precise.

**Related Work-**

Many studies are being conducted on disease prediction using artificial intelligence, machine learning algorithms, the Internet of Things, and other technologies.

This paper suggested an IoT-based health monitoring system based on a random forest algorithm [8]. Using a random forest method, many diseases such as heart disease, diabetes, and breast cancer were predicted and attained with maximum accuracy in the dermatology dataset. [9] The combined cardiovascular illness recommendation system [10] [11] employs IoT in a cloud setting to solve a multi-class classification issue that can predict eight different types of cardiovascular disorders. To increase accuracy, the model used a feature selection strategy. Lung Cancer developed an Io-based prediction system employing segmentation and cluster-based fuzzy classification in one study. The proposed system, which was created in the MATLAB environment, is designed to classify radiographic images of the lungs. He has created an online platform for clinical decision support systems that is web-based (CDSS). It is based on the most effective deep neural network (DNN). They created a cloud-based CDSS system to forecast the severity of chronic kidney disease (CKD). They used Internet of Things sensors to collect clinical data from relevant patients, anticipate normal and abnormal states, and obtain the best DNN accuracy. [12] Researchers have developed an Android-based monitoring device that can track the heart rate of heart disease patients. We built a model that can trigger an alarm if a patient's heart rate is irregular using a decision tree method. It delivers an Io-based system that uses machine learning algorithms to detect cardiac problems early.

[13] Researchers looked at IoT security and privacy elements such as security criteria, utilization, and categories of healthcare attacks. To deal with noisy missing values, they employ a decision tree-based classifier of how people maintain social and economic bases for sustainability. Developed a mobile healthcare environment that is capable of driving.

**Objective-**

The goal of this study is to examine the current data on cardiovascular illnesses in order to predict heart disease at an earlier stage and so prevent it from occurring. The heart disease patient data was imported from India and stored in the cloud. For the prediction of cardiac illnesses, the stored data is pre-processed and further analysed using machine learning techniques.

The benefits of developing this study article include increased security, enhanced efficiency, virtualization, dependability, and scalability, which can promote resource sharing, cost savings, and a highly efficient and accurate medical monitoring, management, and administration system.

Additionally, this study aims to identify patients with cardiac disease.

**Dataset description-**

Among the 11 features in this dataset, there is a target variable. It has 6 nominal variables and 5 numeric variables. The detailed description of all the features are as follows:

1. Age: Age of patients in years (numerical)

2. Sex: Gender of patient (Male - 1, Female - 0) (Nominal)

3. Chest Pain Type: A patient's chest pain is classified into four categories: typical, angina, non-anginal pain, and asymptomatic (Nominal).

4. Resting Bps: Number of mm/HG of blood pressure at resting (Numerical)

5. Cholesterol: Number of milligrams of cholesterol in a liter of blood

6. Fasting blood sugar: A value of 1 indicates a true value and 0 indicates a false value (Nominal) for fasting blood sugar levels >120 mg/dl

7. Resting ECG: In the electrocardiogram while at rest, three values are presented: 0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy (nominal)

8. Max Heart Rate: Result of maximum heart rate (Numeric)

9. Exercise Angina: Angina induced by exercise 0 portraying NO, 1 depicting Yes (Nominal)

10. Oldpeak: Compared to rest, exercise caused ST-depression (Numeric)

**Target variable**

11. HeartDisease : It is the target variable which we have to predict 1 means patient is suffering from heart risk and 0 means patient is normal.

**Machine Learning Algorithms**

Five machine learning techniques are employed to develop an illness prediction model in this research.K Nearest Neighbor (KNN), Decision Tree Classifier (DTC), Support Vector Machine (SVM), Random Forest (RF), and Nave Bayes are all used in this model (NB).As the dataset has output class labels, supervised algorithms are suitable for handling class label problems.

**Big data in healthcare information processing-**

In the healthcare system, big data is utilized to anticipate high-risk diseases that can save a person's life, predict the patient's state from two chronic diseases at the same time, and save money on medical services [18]. The management and treatment of diseases information is maintained in a big data database sequentially in many nations. The Conventional Electrocardiogram (ECG) collects a large amount of data in a short amount of time. The healthcare gadget uses cloud computing and an IoT structure. The more expensive home use devices are ECG based. For this reason, an ECG tracking device is designed, which works in conjunction with the IoT cloud.

Diagram

Description automatically generated

The five representations that differentiate the enormous facts are volume, velocity, diversity, truth, and value. The Size of the dataset is represented by the Volume in huge facts, and its length ranges from (10) 12 bytes to (10) 21 bytes[19]. Velocity represents the movement of the facts as well as the arrival of fresh facts. The numerous sorts of facts and the issues that come from them are represented by variety. The obligation of the facts is indicated by Veracity due to the lack of confidence in fact versions. Figure 1 depicts the data processing within the healthcare device. Value refers to the knowledge derived from the significance of a fact. However, there are several limitations to traditional data analytics within the large data's fitness informatics [20]. As a result, advancements in big data in healthcare devices increased quickly in the form of smart phones, implantable and wearable devices, and real-time sensors.

**Cloud computing with IoT architecture in healthcare-**

In a fitness care gadget, a large amount of data from the patient's frame sensor needs to be saved and regulated properly. Information from sensors is currently saved and controlled for processing via cloud computing with IoT via wearable or implanted sensors [23]. Previously, a completely wearable sensor based on the Wireless Body Area Network (WBAN) was employed, which included the data gathering, communication networking, and provider layer.

The patient's body temperature and blood pressure, as well as other physiological markers, are detected by this wearable sensor. Furthermore, IoT with cloud computing necessitates energy consumption in sensing devices for low-energy communication methods. In Zigbee, IEEE 802.15. four is commonly used in LR WPANs (low price WPANs) to improve the records replacing between Personal Operating Space (POS) at low energy environments of around 10m. IPv6 with Less Power Wireless Network (6LoWPAN) was developed to connect electricity-limited WPAN devices to the internet, further complicating the IoT concept [21]. It does, however, assist fragmentation systems in repairing IPv6 datagrams. Figure 10 shows this ..

Figure 10: Cloud computing with IoT architecture in healthcare**-**

Diagram

Description automatically generated

The Conventional Electrocardiogram (ECG) collects a large amount of data in a short amount of time. The healthcare gadget uses cloud computing and an IoT structure. The more expensive home use devices are ECG based. For this reason, an ECG tracking device is designed, which works in conjunction with the IoT cloud.

**Predictive analytics system for Cardiovascular data-**

According to World Health Organization (WHO) data, cardiovascular illnesses are one of the worst infectious diseases in the world, affecting roughly 30% of the global population [23]. The goal of the pattern or data mining technique is to categorize Cardiovascular connected to determine illness severity. The severity is calculated using a parameter value that is depending on the diagnostic stage. Because a large amount of data is available in the network, some of it is unbalanced, filtering the dataset is critical [24]. For data categorization approaches, modern standard schemes such as the Naive Bayes (NB) classifier, j48, CART, plain logistic, decision tree, and multilayer perception are employed [25]. Furthermore, all of these classifiers were created by researchers for various body diseases.

**CONCLUSION**

Heart disease is becoming more widespread in people everywhere, including in our own country (India). As a result, anticipating disease before infection lowers the risk of death. Lot of press has been gotten from this forecast. Our findings are part of a bigger study on the identification and prediction of heart disease. With SVM and KNN, a genuine Algerian dataset with excellent results, a neural network obtained 93 percent accuracy. Our research is strong because we examined the algorithm's stability on datasets of varied sizes. In the end, neural networks were the most successful. To determine the connections between the attributes, we conducted feature selection experiments or employed correlation matrices. There are certain methods to magnify these approaches. B. Using deep learning techniques Increasing the size of the dataset, employing various approaches for attribute selection, and applying algorithms.

**REFERENCES**

[1] V. Fuster, B. B. Kelly, eds., “Promoting cardiovascular health in the developing world: a critical challenge to achieve global health,” National Academies Press, 2010, pp 465.

[2] P. Kaur, R. Kumar, M. Kumar, “A healthcare monitoring system using random forest and internet of things (IoT)”, 2019, pp. 19905-1916.

[3] I. Cvitić, D. Peraković, M. Periša, M. Botica, “Novel approach for detection of IoT generated DDoS traffic,” Wireless Networks, 2019, pp. 1-14, doi: 10.1007/s11276-019-02043-1.

[4] A. Tewari, B.B. Gupta, “Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework,” Future generation computer systems”, 2002, pp 909-920.

[5] I. Cvitic, D. Perakovic, M. Perisa,M. Botica, “Definition of the IoT device classes based on network traffic low features”, in 4th EAI International Conference on Management of Manufacturing Systems. Springer, Cham, 2020, pp 1-17.

[6] P.K. Senyo, E. Addae, R. Boateng, “Cloud computing research: “A review of research themes, frameworks, methods and future research directions”, International Journal of Information Management, vol. 38, no. 1, pp 128-139, 2018.

[7] Y. Deepthi, Kalyan K.P., Vyas M., Radhika K., Babu D.K., Krishna Rao N.V. “Disease Prediction Based on Symptoms Using Machine Learning”. In: Sikander A., Acharjee D., Chanda C., Mondal P., Verma P. (eds) Energy Systems, Drives and Automations. Lecture Notes in Electrical Engineering, vol. 664, Springer, 2020

[8] F. Jabeen, M. Maqsood, M. A. Ghazanfar, F. Aadil, S. Khan, K. Kim, “An IoT based efficient hybrid recommender system for cardiovascular disease,” Peer-to-Peer Networking and Applications, vol.12, no. 5, pp 1263-1276, 2019.

[9] D. Palani, K. Venkatalakshmi, “An IoT Based Predictive Modelling for Predicting Lung Cancer Using Fuzzy Cluster Based Segmentation and Classification,” Journal of medical systems,vol. 43, no. 2, pp. 1-12.

[10] A. Benjemmaa, H. Lti, M. Ben Ayed, “Design of Remote Heart Monitoring System for Cardiac Patients”, In International Conference on Advanced Information Networking and Applications. Springer, Cham, 2019. pp.963-976.

[11] Hashi, E. K., Zaman, M. S. U., Hasan, M. R. “An expert clinical decision support system to predict disease using classification techniques”. International Conference on Electrical, Computer and Communication Engineering (ECCE),2017. pp. 396-400

[12] K. Saravananathan, T. Velmurugan, “Analyzing Diabetic Data using Classification Algorithms in Data Mining,” Indian Journal of Science and Technology, vol. 9, no. 43, pp 1-6, 2016.

[13] D. Dziak, B. Jachimczyk, W. Kulesza, “IoT-Based Information System for Healthcare Application: Design Methodology Approach,” Applied Sciences, vol. 7, no. 6, pp. 596, 2017.

[14] A.M. Koli, M. Ahmed, “Machine Learning Based Parametric Estimation Approach for Poll Prediction”, Recent Advances in Computer Science and Communications, vol. 14, no. 4, pp. 1287 - 1299, doi: 10.2174/2666255813666191204112601.

[15] P. P. Chavda, “Early Detection of Cardiac Disease Using Machine Learning,” In 2nd International Conference on Advances in Science & Technology (ICAST) 2019.http://dx.doi.org/10.2139/ssrn.3370813.

[16] Thanh Noi, P., Kappas, M.” Comparison of random forest, k-nearest neighbour, and support vector machine classifiers for land cover classification using Sentinel-2 imagery.Sensors”, vol. 18, no. 1, pp. 18, 2018.

[17] Podgorelec V, Kokol P, Stiglic B, Rozman I. “Decision trees: an overview and their use in medicine”, Journal of medical systems, vol. 26, pp 445-463, 2002.

[18] Avati, Anand, et al. "Improving palliative care with deep learning." BMC medical informatics and decision making 18.4 (2018): 122.

[19] Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." International journal of information management 35.2 (2015): 137-144.

[20] Wang, Yichuan, LeeAnn Kung, and Terry Anthony Byrd. "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations." Technological Forecasting and Social Change 126 (2018): 3-13.

[21] Liu, Ruiqing, et al. "Profile of Consecutive Fecal Calprotectin Levels in the Perioperative Period and Its Predictive Capacity for Early Endoscopic Recurrence in Crohn’s Disease." Diseases of the Colon & Rectum 62.3 (2019): 318-326.

[22] Manogaran, Gunasekaran, et al. "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system." Future Generation Computer Systems 82 (2018): 375-387.

[23] World Health Organization. "Compendium of good practices in the implementation of the Tuberculosis Action Plan for the WHO European Region 2016–2020." (2019).

[24] Alonso-Betanzos, Amparo, et al. "Preprocessing in High Dimensional Datasets." Advances in Biomedical Informatics. Springer, Cham, 2018. 247-271.

[25] Chen, Wei, et al. "A Comparative Study of Functional Data Analysis and Generalized Linear Model Data-Mining Techniques for Landslide Spatial Modeling." Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier, 2019. 467-484.