

A Project Report
On
MACHINE LEARNING FOR REAL TIME
HEART DISEASE PREDICTION

**Submitted in partial fulfillment of the requirements for the award of the
degree of**

BACHELOR OF TECHNOLOGY
IN
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

By
D.RUPIKA

(218H1A4311)

SK.MOHAMMED FAROOQ
(218H1A4350)

G.BHARATH
(218H1A4319)

K.SAI SREENIVAS
(218H1A4364)

CH.VENKATA
KRISHNA
(218H1A4309)

Under the Esteemed Guidance of

Mr. N.VENKATESWARA RAO –Assist. professor
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MVR College of Engineering & Technology
(AUTONOMOUS)

Paritala, NTR Dist., PIN 521180 A.P. India.



MVR College of Engineering & Technology

Department of Computer Science & Engineering

Paritala, Krishna Dist., PIN 521180 A.P.India.

2024 – 2025

CERTIFICATE

This is to certify that the project report entitled “**MACHINE LEARNING FOR REAL -TIME HEART DISEASE PREDICTION**” being submitted for the partial fulfilment of the requirements for the award of the degree in Bachelor of Technology in Computer Science & Engineering (Artificial Intelligence) from MVR College of Engineering and Technology, is a bonafide work done by all the team members under my guidance during the academic year 2024-2025 and it has been found suitable for acceptance according to the requirement of the university. This work is not submitted to any other institution or university for the award of any degree.

Mr.N.VENKATESWARA RAO

Project Guide

Mr.Ch.SREENU BABU

Head of the Department

ACKNOWLEDGEMENT

I express my deep sense of gratitude and indebtedness to **Dr.U.YEDUKONDALU**, Principal of our college, who gave very good support to me during this work and others who assisted me in this project work.

I also thank **Mr.Ch.SREENUBABU**, Head of the Department of Computer Science & Engineering, who gave very good support to me during this work.

I also thank my guide, **Mr.VENKATESWARA RAO**, Asst. Professor of CSE Department, MVR College of Engineering & Technology, for his valuable guidance during the course of this project work. I am much indebted to his for suggesting a challenging and interactive project and his valuable advice at every stage of this work. I am very much thankful to his for his coordination in this regard.

Similarly, I am grateful to my friends I have worked with and consider myself extremely privileged and fortunate to work with such honest, hardworking and loyal friends.

I wish my warm and grateful thanks to the staff of our department of MVR College of Engineering & Technology for the assistance for their continuous encouragement in completing my work successfully.

Finally, I thank one and all who directly or indirectly helped me in completing project successfully.

DECLARATION

I hereby declare that this project entitled “**MACHINE LEARNING FOR REAL - TIME HEART DISEASE PREDICTION**” has been undertaken by me and this work has been submitted to MVR college of Engineering & Technology affiliated to JNTUK, Kakinada, in partial fulfilment of the requirements for award of the degree of Bachelor of technology in COMPUTER SCIENCE AND ENGINEERING(ARTIFICIAL INTELLIGENCE).

I further declare that this project work has not been submitted in full or part for the award of any degree of this in any other educational institutions.

DEVARAPALLI RUPIKA
218H1A4311

GOLLAPALLI BHARATH
218H1A4317

SHAIK MOHAMMED FAROOQ
218H1A4350

KAVURI SAI SRINIVAS
218H1A4364

CHITLA VENKAT KRISHNA
218H1A4309

LIST OF CONTENTS

ABSTRACT

CHAPTER I: INTRODUCTION

CHAPTER 2 : LITERATURE SURVEY

CHAPTER 3: SYSTEM ANALYSIS

3.1 Existing System

3.2 Proposed System

3.3 Modules

3.4 Process Model Used With Justification

3.5 Software Requirement Specification

3.5.1. Overall Description

3.5.2. External Interface Requirements

CHAPTER 4: SYSTEM DESIGN

CHAPTER 6: TESTING

CHAPTER 7: SCREEN SHOTS

CHAPTER 8: CONCLUSION

REFERENCES

Abstract:

In today's era deaths due to heart disease has become a major issue approximately one person dies per minute due to heart disease. This is considering both male and female category and this ratio may vary according to the region also this ratio is considered for the people of age group 25-69. This does not indicate that the people with other age group will not be affected by heart diseases. This problem may start in early age group also and predict the cause and disease is a major challenge nowadays. Here in this paper, we have discussed various algorithms and tools used for prediction of heart diseases

Introduction

The contents of this paper mainly focus on various data mining practices that are valuable in heart disease forecast with the assistance of dissimilar data mining tools that are accessible. If the heart doesn't function properly, this will distress the other parts of the human body such as brain, kidney etc. Heart disease is a kind of disease which effects the functioning of the heart. In today's era heart disease is the primary reason for deaths. WHO-World Health Organization has anticipated that 12 million people die every year because of heart diseases. Some heart diseases are cardiovascular, heart attack, coronary and knock. Knock is a sort of heart disease that occurs due to strengthening, blocking or lessening of blood vessels which drive through the brain or it can also be initiated by high blood pressure.

The major challenge that the Healthcare industry faces now-a-days is superiority of facility. Diagnosing the disease correctly & providing effective treatment to patients will define the quality of service. Poor diagnosis causes disastrous consequences that are not accepted

Records or data of medical history is very large, but these are from many dissimilar foundations. The interpretations that are done by physicians are essential components of these data. The data in real world might be noisy, incomplete and inconsistent, so data preprocessing will be required in directive to fill the omitted values in the database.

Even if cardiovascular diseases is found as the important source of death in world in ancient years, these have been announced as the most avoidable and manageable diseases. The whole and accurate management of a disease rest on on the well-timed judgment of that disease. An correct and methodical tool for recognizing high-risk patients and mining data for timely analysis of heart infection looks a serious want.

Different person body can show different symptoms of heart disease which may vary accordingly. Though, they frequently include back pain, jaw pain, neck pain, stomach disorders, and tininess of breath, chest pain, arms and shoulders pains. There are a variety of different heart diseases which includes heart failure and stroke and coronary artery disease [3].

Even though heart disease is acknowledged as the supreme chronic sort of disease in the world, it can be most avoidable one also at the same time. A healthy way of life (main prevention) and timely analysis (inferior prevention) are the two major origins of heart disease director. Conducting steady check-ups (inferior prevention) shows outstanding role in the judgment and early prevention of heart disease difficulties. Several tests comprising of angiography, chest X-rays, echocardiography and exercise tolerance test support to this significant issue. Nevertheless, these tests are expensive and involve availability of accurate medical equipment.

Heart expert's create a good and huge record of patient's database and store them. It also delivers a great prospect for mining a valued knowledge from such sort of datasets.

There is huge research going on to determine heart disease risk factors in different patients, different researchers are using various statistical approaches and numerous programs of data mining approaches. Statistical analysis have acknowledged the count of risk factors for heart diseases counting smoking, age, blood pressure, diabetes, total cholesterol, and hypertension, heart disease training in family, obesity and lack of exercise. For prevention and healthcare of patients who are about to have addicted of heart disease it is very important to have awareness of heart diseases.

Researchers make use of several data mining techniques that are accessible to help the specialists or physicians identify the heart disease. Commonly used procedures used are decision tree, k-nearest and Naïve Bayes. Other different classification based techniques used are bagging algorithm, kernel density, sequential minimal optimization and neural networks, straight Kernel self-organizing map and SVM (Support Vector Machine). The next section clearly provides details of techniques that were used in the study.

LITERATURE SURVEY

In year 2000, research conducted by ShusakuTsumoto [5] says that as we human beings are unable to arrange data if it is huge in size we should use the data mining techniques that are available for finding different patterns from the available huge database and can be used again for clinical research and perform various operations on it.

Y. Alp Aslandogan, et. al. (2004), worked on three different classifiers called K-nearest Neighbour (KNN), Decision Tree, Naïve Bayesian and used Dempsters' rule for this three viewpoint to appear as one concluding decision. This classification based on the combined idea show increased accuracy [6].

Carlos Ordonez (2004), Assessed the problematic to recognize and forecast the rule of relationship for the heart disease. A dataset involving medical history of the patients having heart disease with the aspects of risk factors was accessed by him, measurements of narrowed artery and heart perfusion. All these restrictions were announced to shrink the digit of designs, these are as follows:

- 1) The features should seem on a single side of the rule.
- 2) The rule should distinct various features into the different groups.
- 3) The count of features available from the rule is organized by medical history of people having heart disease only. The occurrence or the nonappearance of heart disease was predicted by the author in four heart veins with the two clusters of rules [7].

Franck Le Duff (2004), worked on creating Decision tree quickly with clinical data of the physician or service. He suggested few data mining techniques which can help cardiologists in the predication survival of patients. The main drawback of the system was that the user needs to model [8].

Boleslaw Szymanski, et. al. (2006), operated on a novel experiential to check the aptitude of calculation of scarce kernel in SUPANOVA. The author used this technique on a standard boston housing market dataset for discovering heart diseases, measurement of heart activities and prediction of heart diseases were found 83.7% correct which were measured with the help of support vector machine and kernel equivalent to it. A quality result is gained by spline kernel with the help of standard boston housing market database [9].

Kiyong Noh, et. al. (2006) made use of a classification technique for removal of multi-parametric structures by accessing HRV and ECG signals. Kiyong used the FP-growth algorithm as the foundation of this technique that is associative. A rule consistency degree was gained which allows a robust press on trimming designs in the method of producing designs [10].

HeonGyu Lee, et. al. (2007), operated for the operation systems of Arithmetical and cataloguing for the addition chief of the multi-parametric feature through direct and nonlinear features of Heart Rate Variability (HRV). The dissimilar classifiers existing are cataloguing grounded on Decision Tree (C4.5), Multiple Association Rules (CMAR) and Bayesian classifiers, and Support Vector Machine (SVM) that are investigated for the valuation of the linear and nonlinear features of the HRV tables [11].

Niti Guru, et. al. (2007), functioned for forecasting of heart disease, Blood Stress and Sugar by the aid of neural systems. Hearings were accepted out on example best ever of patients. The neural system is verified with 13 types, as blood pressure, period, angiography etc. [12].

Controlled network was used for analysis of heart diseases. Training was accepted out with the support of a back-propagation technique. The secretive data was nourished at certain times by the doctor; the acknowledged technique applied on the unidentified data since the judgments with trained data and caused a grade of possible ailments that the patient is

inclining to heart disease.

Hai Wang, et al. (2008), deliberated the part of medicinal experts in medical data mining also on obtaining a model for medical awareness achievement using data mining [13].

Sellappan Palaniappan, et. al. (2008), industrialized IHDPS-Intelligent Heart Disease Prediction System by means of data mining algorithm, i.e. Naïve Bayes, Decision Trees and Neural Network. Each process has its own authority to advance right results. The unknown designs and association amongst them have been used to paradigm this method. The IHDPS is web-based, user-friendly, mountable, trustworthy and stretchy and justifiable [14].

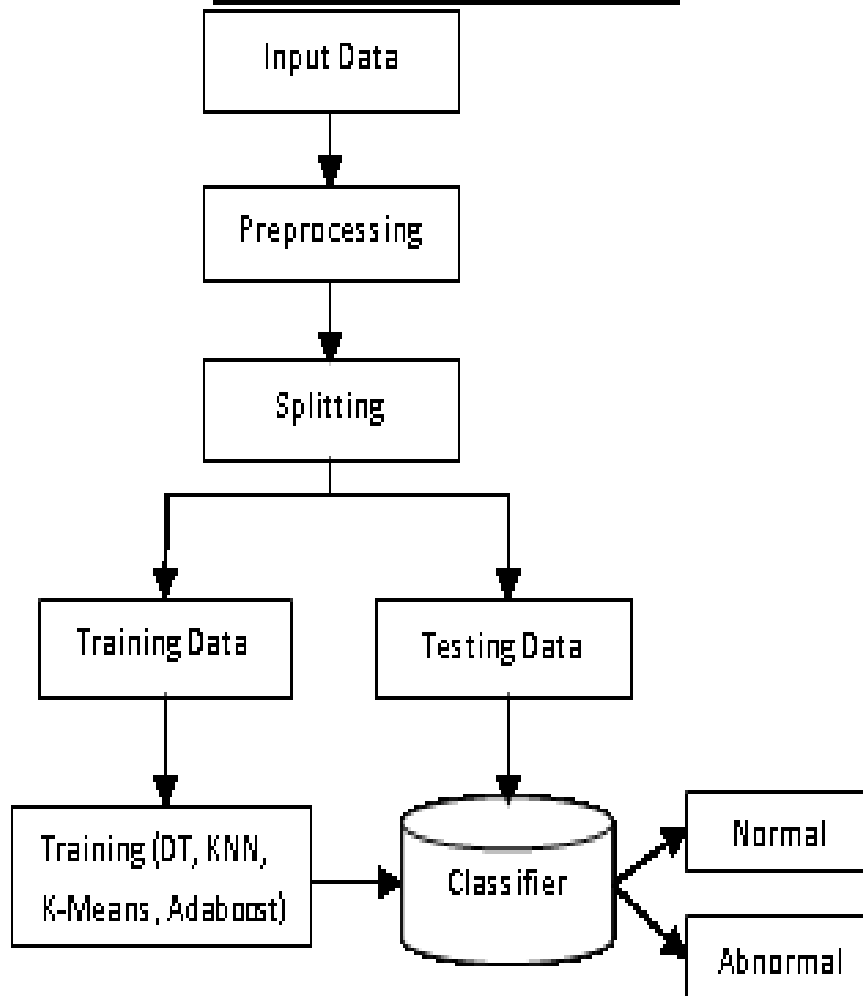
Latha Parthiban, et. al. (2008), operated on the foundation of CANFIS (co-active neuro-fuzzy implication method) for identification of heart disease. CANFIS model established the disease by integrating the neural network and fuzzy logic methods and later combined with the genetic algorithm. On the grounds of the training presentations and classification correctness found, the performance of the CANFIS model were estimated. The CANFIS prototypical is exposed as the possible for estimation of heart disease [15].

Chaitrali S. D., (2012), investigated a computation structures for heart syndrome with the help of full amount of input characteristics. A few terms related to medical like blood pressure, sex, cholesterol and 13 more attributes like this were recycled to predict the heart disease to a particular person or patient. He also made use of two different attributes like smoking and obesity. Unlike data mining performances were used like Decision trees, neural networks and naïve bayes's for analyzing the heart disease database. The concert of these practices depends on the accuracy provided by the system. The accuracy provided by decision tree is 99.62%, neural network is 100% and naïve bayes is 90.74% respectively [16].

S. Vijayarani, et. al. in (2013), made use of experimental results carried out using dissimilar classification methods for heart disease dataset. The different classification systems which were used and tested by him are Decision Stump, Random Forest and LMT tree algorithm. WEKA tool was used for comparison [17].

Harsh Vazirani, et. al. (2010), deliberates many belongings connected to the analysis of the heart disease. The main emphasis is on two kinds of the analysis methods that were used one is manual and the other is programmed analysis which contains of analysis of diseases with the help of segmental neural network and intellectual expert structure that were used for analyzing heart diseases [18].

System Architecture



Existing System

Heart related infections or Cardiovascular Diseases (CVDs) are the primary justification a colossal number of death on the planet in the course of the most recent couple of many years and has arisen as the most perilous illness, in India as well as in the entire world. Along these lines, there is a need of dependable, precise and attainable framework to analyze such sicknesses on schedule for legitimate therapy. AI calculations and methods have been applied to different clinical datasets to robotize the investigation of enormous and complex information. Numerous scientists, as of late, have been utilizing a few AI procedures to help the medical care industry and the experts in the analysis of heart related illnesses.

PROPOSED SYSTEM

We smooth out AI calculations for powerful expectation of constant illness episode in disease-continuous networks. We try the adjusted expectation models over genuine medical clinic data collected from focal China in 2013–2015. To conquer the trouble of deficient information, we utilize a latent factor model to remake the missing information. We investigate a territorial persistent infection of cerebral infarction. We propose another convolutional neural organization (CNN)- based multimodal infection hazard prediction algorithm utilizing organized and unstructured information from medical clinic. As far as we could possibly know, none of the existing work zeroed in on both information types nearby clinical huge information investigation.

3.3 SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System : MINIMUM i3.
- Hard Disk : 40 GB.
- Ram : 4 GB.

SOFTWARE REQUIREMENTS:

- **Operating System:** Windows 8
- **Coding Language:** Python 3.7

3.4 SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ◆ ECONOMICAL FEASIBILITY
- ◆ TECHNICAL FEASIBILITY
- ◆ SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

4.SYSTEM DESIGN

4.1 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

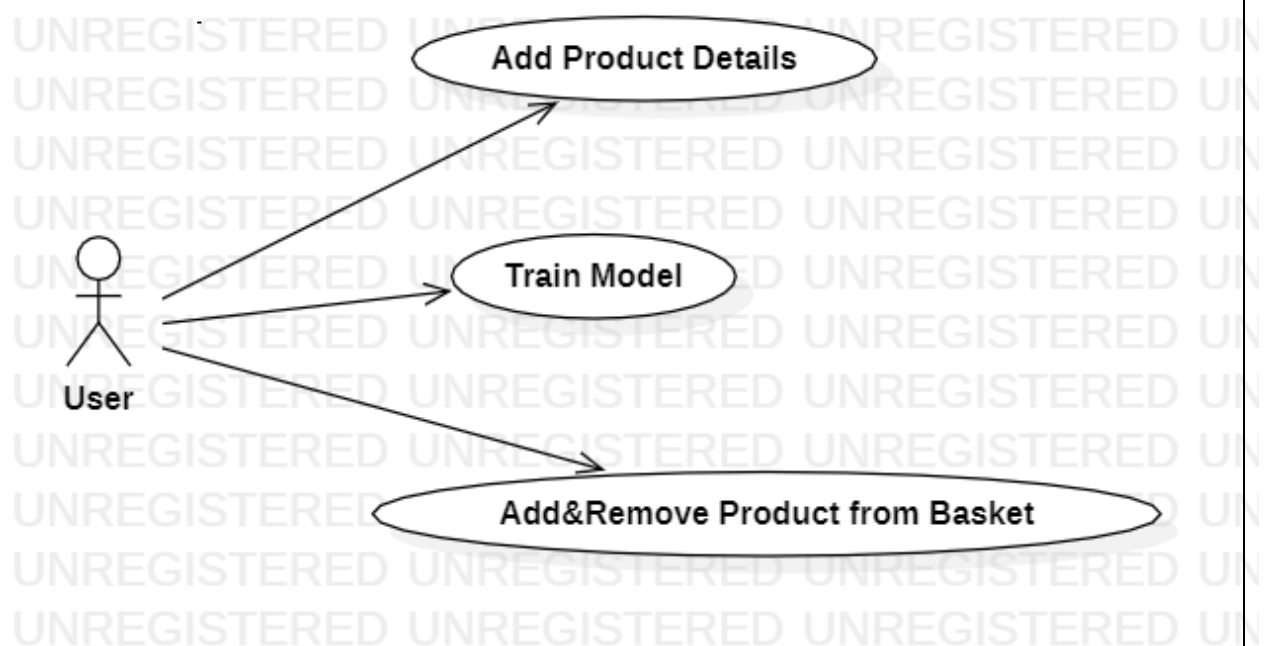
The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.

6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

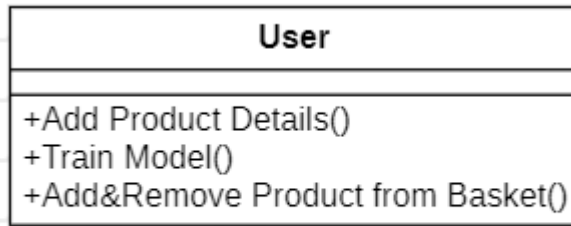
USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



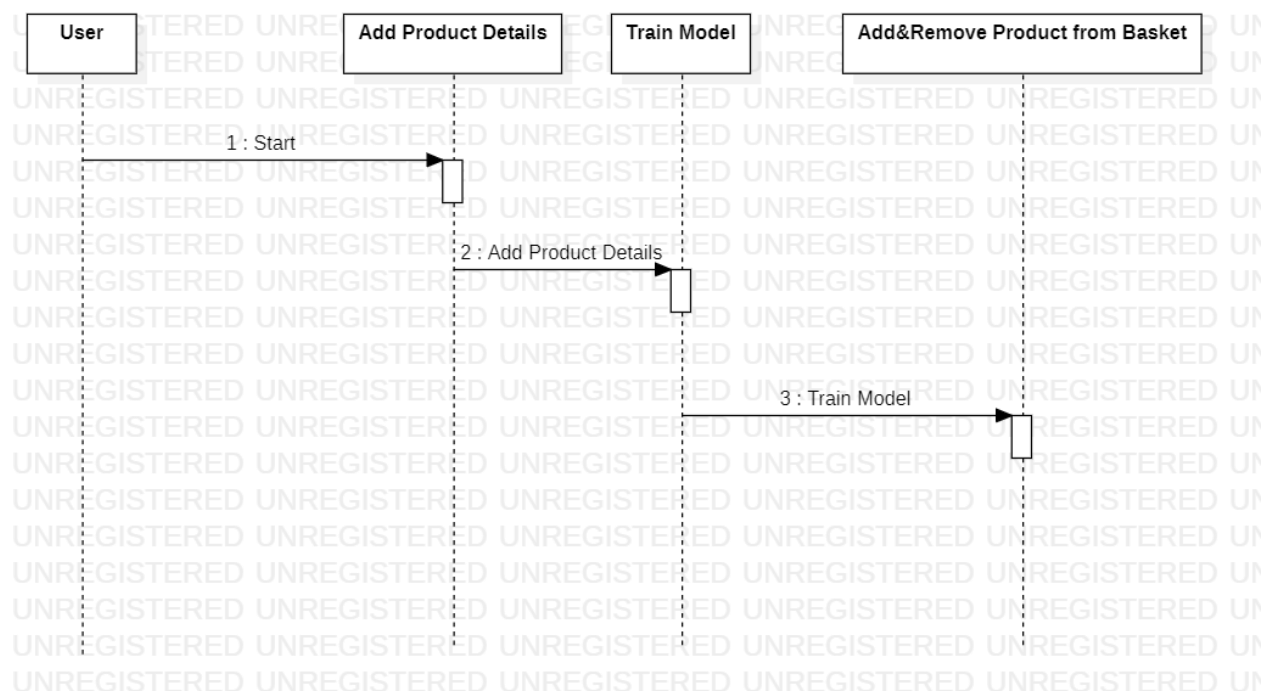
CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



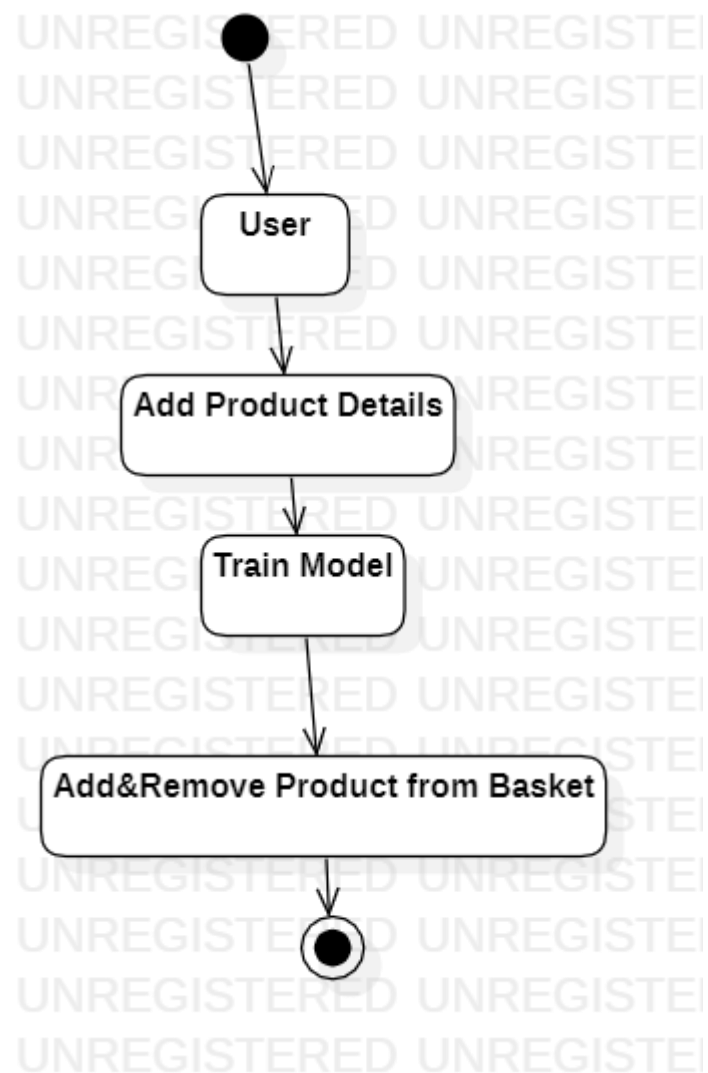
SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



4.3 IMPLEMENTATION:

MODULES:

1.Add Product Details

To build project I used some sample products image to train product identification models

2.Train Model

In this Module screen train model generated with 100% accuracy and now show product to web cam.

3. Add/Remove Product from basket

To allow application to identify product image and then show in text area and if we again show same product then application will remove from text area

5.SOFTWARE ENVIRONMENT

What is Python :-

Below are some facts about Python.

Python is currently the most widely used multi-purpose, high-level programming language.

Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java.

Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber... etc.

The biggest strength of Python is huge collection of standard library which can be used for the following –

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQt etc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like Opencv, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia

Advantages of Python :-

Let's see how Python dominates over other languages.

1. Extensive Libraries

Python downloads with an extensive library and it *contain code for various purposes like regular expressions, documentation-generation, unit-testing, web browsers, threading, databases, CGI, email, image manipulation, and more.* So, we don't have to write the complete code for that manually.

2. Extensible

As we have seen earlier, Python can be **extended to other languages**. You can write some of your code in languages like C++ or C. This comes in handy, especially in projects.

3. Embeddable

Complimentary to extensibility, Python is embeddable as well. You can put your Python code in your source code of a different language, like C++. This lets us add **scripting capabilities** to our code in the other language.

4. Improved Productivity

The language's simplicity and extensive libraries render programmers **more productive** than languages like Java and C++ do. Also, the fact that you need to write less and get more things done.

5. IOT Opportunities

Since Python forms the basis of new platforms like Raspberry Pi, it finds the future bright for the Internet Of Things. This is a way to connect the language with the real world.

6. Simple and Easy

When working with Java, you may have to create a class to print '**Hello World**'. But in Python, just a print statement will do. It is also quite **easy to learn, understand, and code**. This is why when people pick up Python, they have a hard time adjusting to other more verbose languages like Java.

7. Readable

Because it is not such a verbose language, reading Python is much like reading English. This is the reason why it is so easy to learn, understand, and code. It also does not need curly braces to define blocks, and indentation is mandatory. This

further aids the readability of the code.

8. Object-Oriented

This language supports both the **procedural and object-oriented** programming paradigms. While functions help us with code reusability, classes and objects let us model the real world. A class allows the **encapsulation of data** and functions into one.

9. Free and Open-Source

Like we said earlier, Python is **freely available**. But not only can you **download Python** for free, but you can also download its source code, make changes to it, and even distribute it. It downloads with an extensive collection of libraries to help you with your tasks.

10. Portable

When you code your project in a language like C++, you may need to make some changes to it if you want to run it on another platform. But it isn't the same with Python. Here, you need to **code only once**, and you can run it anywhere. This is called **Write Once Run Anywhere (WORA)**. However, you need to be careful enough not to include any system-dependent features.

11. Interpreted

Lastly, we will say that it is an interpreted language. Since statements are executed one by one, **debugging is easier** than in compiled languages.

Any doubts till now in the advantages of Python? Mention in the comment section.

Advantages of Python Over Other Languages

1. Less Coding

Almost all of the tasks done in Python requires less coding when the same task is done in other languages. Python also has an awesome standard library support, so you don't have to search for any third-party libraries to get your job done. This is the reason that many people suggest learning Python to beginners.

2. Affordable

Python is free therefore individuals, small companies or big organizations can leverage the free available resources to build applications. Python is popular and widely used so it gives you better community support.

The 2019 Github annual survey showed us that Python has overtaken Java in the most popular programming language category.

3. Python is for Everyone

Python code can run on any machine whether it is Linux, Mac or Windows. Programmers need to learn different languages for different jobs but with Python, you can professionally build web apps, perform data analysis and **machine learning**, automate things, do web scraping and also build games and powerful visualizations. It is an all-rounder programming language.

Disadvantages of Python

So far, we've seen why Python is a great choice for your project. But if you choose it, you should be aware of its consequences as well. Let's now see the downsides of choosing Python over another language.

1. Speed Limitations

We have seen that Python code is executed line by line. But since Python is interpreted, it often results in **slow execution**. This, however, isn't a problem unless speed is a focal point for the project. In other words, unless high speed is a requirement, the benefits offered by Python are enough to distract us from its speed limitations.

2. Weak in Mobile Computing and Browsers

While it serves as an excellent server-side language, Python is much rarely seen on the **client-side**. Besides that, it is rarely ever used to implement smartphone-based applications. One such application is called **Carbonnelle**.

The reason it is not so famous despite the existence of Brython is that it isn't that secure.

3. Design Restrictions

As you know, Python is **dynamically-typed**. This means that you don't need to declare the type of variable while writing the code. It uses **duck-typing**. But wait, what's that?

Well, it just means that if it looks like a duck, it must be a duck. While this is easy on the programmers during coding, it can **raise run-time errors**.

4. Underdeveloped Database Access Layers

Compared to more widely used technologies like **JDBC (Java DataBase Connectivity)** and **ODBC (Open DataBase Connectivity)**, Python's database access layers are a bit underdeveloped. Consequently, it is less often applied in huge enterprises.

5. Simple

No, we're not kidding. Python's simplicity can indeed be a problem. Take my example. I don't do Java, I'm more of a Python person. To me, its syntax is so simple that the verbosity of Java code seems unnecessary.

This was all about the Advantages and Disadvantages of Python Programming Language.

History of Python : -

What do the alphabet and the programming language Python have in common? Right, both start with ABC. If we are talking about ABC in the Python context, it's clear that the programming language ABC is meant. ABC is a general-purpose programming language and programming environment, which had been developed in the Netherlands, Amsterdam, at the CWI (Centrum Wiskunde & Informatica). The greatest achievement of ABC was to influence the design of Python. Python was conceptualized in the late 1980s. Guido van Rossum worked that time in a project at the CWI, called Amoeba, a distributed operating system. In an interview with Bill Venners¹, Guido van Rossum said: "In the early 1980s, I worked as an implementer on a team building a language called ABC at Centrum voor Wiskunde en Informatica (CWI). I don't know how well people know ABC's influence on Python. I try to mention ABC's influence because I'm indebted to everything I learned during that project and to the people who worked on it." Later on in the same Interview, Guido van Rossum continued: "I remembered all my experience and some of my frustration with ABC. I decided to try to design a simple scripting language that possessed some of ABC's better properties, but without its problems. So I started typing. I created a simple virtual machine, a simple parser, and a simple runtime. I made my own version of the various ABC parts that I liked. I created a basic syntax, used indentation for statement grouping instead of curly braces or begin-end blocks, and developed a small number of powerful data types: a hash table (or dictionary, as we call it), a list, strings, and numbers."

What is Machine Learning : -

Before we take a look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of *building models of data*.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models *tunable parameters* that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

Categories Of Machine Learning :-

At the most fundamental level, machine learning can be categorized into two main types: supervised learning and unsupervised learning.

Supervised learning involves somehow modeling the relationship between measured features of data and some label associated with the data; once this model is determined, it can be used to apply labels to new, unknown data. This is further subdivided into *classification* tasks and *regression* tasks: in classification, the labels are discrete categories, while in regression, the labels are continuous quantities. We will see examples of both types of supervised learning in the following section.

Unsupervised learning involves modeling the features of a dataset without reference to any label, and is often described as "letting the dataset speak for itself." These models include tasks such as *clustering* and *dimensionality reduction*. Clustering algorithms identify distinct groups of data, while dimensionality reduction algorithms search for

more succinct representations of the data. We will see examples of both types of unsupervised learning in the following section.

Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven't surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, "to make decisions, based on data, with efficiency and scale".

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programming logic, in the problems that cannot be programmed inherently. The fact is that we can't do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

Challenges in Machines Learning :-

While Machine Learning is rapidly evolving, making significant strides with cybersecurity and autonomous cars, this segment of AI as whole still has a long way to go. The reason behind is that ML has not been able to overcome number of challenges. The challenges that ML is facing currently are –

Quality of data – Having good-quality data for ML algorithms is one of the biggest challenges. Use of low-quality data leads to the problems related to data preprocessing and feature extraction.

Time-Consuming task – Another challenge faced by ML models is the consumption of time especially for data acquisition, feature extraction and retrieval.

Lack of specialist persons – As ML technology is still in its infancy stage, availability of expert resources is a tough job.

No clear objective for formulating business problems – Having no clear objective and well-defined goal for business problems is another key challenge for ML because this technology is not that mature yet.

Issue of overfitting & underfitting – If the model is overfitting or underfitting, it cannot be represented well for the problem.

Curse of dimensionality – Another challenge ML model faces is too many features of data points. This can be a real hindrance.

Difficulty in deployment – Complexity of the ML model makes it quite difficult to be deployed in real life.

Applications of Machines Learning :-

Machine Learning is the most rapidly growing technology and according to researchers we are in the golden year of AI and ML. It is used to solve many real-world complex problems which cannot be solved with traditional approach. Following are some real-world applications of ML –

- Emotion analysis
- Sentiment analysis
- Error detection and prevention
- Weather forecasting and prediction
- Stock market analysis and forecasting
- Speech synthesis
- Speech recognition
- Customer segmentation
- Object recognition
- Fraud detection
- Fraud prevention
- Recommendation of products to customer in online shopping

How to Start Learning Machine Learning?

Arthur Samuel coined the term “**Machine Learning**” in 1959 and defined it as a “**Field of study that gives computers the capability to learn without being explicitly programmed**”.

And that was the beginning of Machine Learning! In modern times, Machine Learning is one of the most popular (if not the most!) career choices. According to Indeed, Machine Learning Engineer Is The Best Job of 2019 with a 344% growth and an average base salary of **\$146,085** per year.

But there is still a lot of doubt about what exactly is Machine Learning and how to start learning it? So this article deals with the Basics of Machine Learning and also the path you can follow to eventually become a full-fledged Machine Learning Engineer. Now let's get started!!!

How to start learning ML?

This is a rough roadmap you can follow on your way to becoming an insanely talented Machine Learning Engineer. Of course, you can always modify the steps according to your needs to reach your desired end-goal!

Step 1 – Understand the Prerequisites

In case you are a genius, you could start ML directly but normally, there are some prerequisites that you need to know which include Linear Algebra, Multivariate Calculus, Statistics, and Python. And if you don't know these, never fear! You don't need a Ph.D. degree in these topics to get started but you do need a basic understanding.

(a) Learn Linear Algebra and Multivariate Calculus

Both Linear Algebra and Multivariate Calculus are important in Machine Learning. However, the extent to which you need them depends on your role as a data scientist. If you are more focused on application heavy machine learning, then you will not be that heavily focused on maths as there are many common libraries available.

(b) Learn Statistics

Data plays a huge role in Machine Learning. In fact, around 80% of your time as an ML expert will be spent collecting and cleaning data. And statistics is a field that handles the collection, analysis, and presentation of data. So it is no surprise that you need to learn it!!!

Some of the key concepts in statistics that are important are Statistical Significance, Probability Distributions, Hypothesis Testing, Regression, etc. Also, Bayesian Thinking is also a very important part of ML which deals with various concepts like Conditional Probability, Priors, and Posteriors, Maximum Likelihood, etc.

(c) Learn Python

Some people prefer to skip Linear Algebra, Multivariate Calculus and Statistics and learn them as they go along with trial and error. But the one thing that you absolutely cannot skip is Python! While there are other languages you can use for Machine Learning like R, Scala, etc. Python is currently the most popular language for ML. In fact, there are many Python libraries that are specifically useful for Artificial Intelligence and Machine Learning such as Keras, TensorFlow, Scikit-learn, etc.

So if you want to learn ML, it's best if you learn Python! You can do that using various online resources and courses such as **Fork Python** available Free on GeeksforGeeks.

Step 2 – Learn Various ML Concepts

Now that you are done with the prerequisites, you can move on to actually learning ML (Which is the fun part!!!) It's best to start with the basics and then move on to the more complicated stuff. Some of the basic concepts in ML are:

(a) Terminologies of Machine Learning

- **Model** – A model is a specific representation learned from data by applying some machine learning algorithm. A model is also called a hypothesis.
- **Feature** – A feature is an individual measurable property of the data. A set of numeric features can be conveniently described by a feature vector. Feature vectors are fed as input to the model. For example, in order to predict a fruit, there may be features like color, smell, taste, etc.
- **Target (Label)** – A target variable or label is the value to be predicted by our model. For the fruit example discussed in the feature section, the label with each set of input would be the name of the fruit like apple, orange, banana, etc.
- **Training** – The idea is to give a set of inputs(features) and it's expected outputs(labels), so after training, we will have a model (hypothesis) that will then map new data to one of the categories trained on.
- **Prediction** – Once our model is ready, it can be fed a set of inputs to which it will provide a predicted output(label).

(b) Types of Machine Learning

- **Supervised Learning** – This involves learning from a training dataset with labeled data using classification and regression models. This learning process continues until the required level of performance is achieved.
- **Unsupervised Learning** – This involves using unlabelled data and then finding the underlying structure in the data in order to learn more and more about the data itself using factor and cluster analysis models.
- **Semi-supervised Learning** – This involves using unlabelled data like Unsupervised Learning with a small amount of labeled data. Using labeled data vastly increases the learning accuracy and is also more cost-effective than Supervised Learning.
- **Reinforcement Learning** – This involves learning optimal actions through trial and error. So the next action is decided by learning behaviors that are based on the current state and that will maximize the reward in the future.

Advantages of Machine learning :-

1. Easily identifies trends and patterns -

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

2. No human intervention needed (automation)

With ML, you don't need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

3. Continuous Improvement

As **ML algorithms** gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

4. Handling multi-dimensional and multi-variety data

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

5. Wide Applications

You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

Disadvantages of Machine Learning :-

1. Data Acquisition

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

2. Time and Resources

ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.

3. Interpretation of Results

Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the algorithms for your purpose.

4. High error-susceptibility

Machine Learning is autonomous but highly susceptible to errors. Suppose you train an algorithm with data sets small enough to not be inclusive. You end up with biased predictions coming from a biased training set. This leads to irrelevant advertisements being displayed to customers. In the case of ML, such blunders can set off a chain of errors that can go undetected for long periods of time. And when they do get noticed, it takes quite some time to recognize the source of the issue, and even longer to correct it.

Python Development Steps : -

Guido Van Rossum published the first version of Python code (version 0.9.0) at alt.sources in February 1991. This release included already exception handling, functions, and the core data types of list, dict, str and others. It was also object oriented and had a module system. Python version 1.0 was released in January 1994. The major new features included in this release were the functional programming tools lambda, map, filter and reduce, which Guido Van Rossum never liked. Six and a half years later in October 2000, Python 2.0 was introduced. This release included list comprehensions, a full garbage collector and it was

supporting unicode. Python flourished for another 8 years in the versions 2.x before the next major release as Python 3.0 (also known as "Python 3000" and "Py3K") was released. Python 3 is not backwards compatible with Python 2.x. The emphasis in Python 3 had been on the removal of duplicate programming constructs and modules, thus fulfilling or coming close to fulfilling the 13th law of the Zen of Python: "There should be one -- and preferably only one -- obvious way to do it." Some changes in Python 3:

- Print is now a function
- Views and iterators instead of lists
- The rules for ordering comparisons have been simplified. E.g. a heterogeneous list cannot be sorted, because all the elements of a list must be comparable to each other.
- There is only one integer type left, i.e. int. long is int as well.
- The division of two integers returns a float instead of an integer. "//" can be used to have the "old" behaviour.
- Text Vs. Data Instead Of Unicode Vs. 8-bit

Purpose :-

We demonstrated that our approach enables successful segmentation of intra-retinal layers—even with low-quality images containing speckle noise, low contrast, and different intensity ranges throughout—with the assistance of the ANIS feature.

Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

- Python is Interpreted – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive – you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

Modules Used in Project :-

Tensorflow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

Numpy

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as an efficient multi-dimensional container of generic data.

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. **Python**

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

- Python is Interpreted – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive – you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

Install Python Step-by-Step in Windows and Mac :

Python a versatile programming language doesn't come pre-installed on your computer devices. Python was first released in the year 1991 and until today it is a very popular high-level programming language. Its style philosophy emphasizes code readability with its notable use of great whitespace.

The object-oriented approach and language construct provided by Python enables programmers to write both clear and logical code for projects. This software does not come pre-packaged with Windows.

How to Install Python on Windows and Mac :

There have been several updates in the Python version over the years. The question is how to install Python? It might be confusing for the beginner who is willing to start learning Python but this tutorial will solve your query. The latest or the newest version of Python is version 3.7.4 or in other words, it is Python 3.

Note: The python version 3.7.4 cannot be used on Windows XP or earlier devices.

Before you start with the installation process of Python. First, you need to know about your **System Requirements**. Based on your system type i.e. operating system and based processor, you must download the python version. My system type is a **Windows 64-bit operating system**. So the steps below are to install python version 3.7.4 on Windows 7 device or to install Python 3. [Download the Python Cheatsheet here](#). The steps on how to install Python on Windows 10, 8 and 7 are **divided into 4 parts** to help understand better.

Download the Correct version into the system

Step 1: Go to the official site to download and install python using Google Chrome or any other web browser. OR Click on the following link: <https://www.python.org>



Now, check for the latest and the correct version for your operating system.

Step 2: Click on the Download Tab.



Step 3: You can either select the Download Python for windows 3.7.4 button in Yellow Color or you can scroll further down and click on download with respective to their version. Here, we are downloading the most recent python version for windows 3.7.4

Looking for a specific release?

Python releases by version number:

Release version	Release date		Click for more
Python 3.7.4	July 8, 2019	Download	Release Notes
Python 3.6.9	July 2, 2019	Download	Release Notes
Python 3.7.3	March 25, 2019	Download	Release Notes
Python 3.4.10	March 18, 2019	Download	Release Notes
Python 3.5.7	March 18, 2019	Download	Release Notes
Python 2.7.16	March 4, 2019	Download	Release Notes
Python 3.7.2	Dec. 24, 2018	Download	Release Notes

Step 4: Scroll down the page until you find the Files option.

Step 5: Here you see a different version of python along with the operating system.

Files					
Version	Operating System	Description	MD5 Sum	File Size	GPG
Gzipped source tarball	Source release		68111671e5b2db4aef7b9ab01b709be	23017663	SG
XZ compressed source tarball	Source release		d33e4aa66097051c2eca45ee3604803	17131432	SG
macOS 64-bit/32-bit installer	Mac OS X	for Mac OS X 10.6 and later	6428b4fa7583daf1a442cbalcee08e6	34896416	SG
macOS 64-bit installer	Mac OS X	for OS X 10.9 and later	5dd605c38217a45773b9e4a936b241f	28882845	SG
Windows help file	Windows		d63999573a2c96b2ac56cade6b47cd2	8131761	SG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64/x64	9b00c3cf6d9ec0b9abec3184a40728a2	7504391	SG
Windows x86-64 executable installer	Windows	for AMD64/EM64/x64	a702b4b0aef76d4b5d63043a583e543400	2688368	SG
Windows x86-64 web-based installer	Windows	for AMD64/EM64/x64	28c51c08b6d73a8e53a3b4351b4bd2	1362904	SG
Windows x86 embeddable zip file	Windows		9fab1b818841879fda94133574139d8	6741626	SG
Windows x86 executable installer	Windows		33cc802942a54446a3d9451a76394789	25663848	SG
Windows x86 web-based installer	Windows		1b670cfa5d317df82c30983ea371d87c	1324606	SG

- To download Windows 32-bit python, you can select any one from the three options:

Windows x86 embeddable zip file, Windows x86 executable installer or Windows x86 web-based installer.

- To download Windows 64-bit python, you can select any one from the three options:

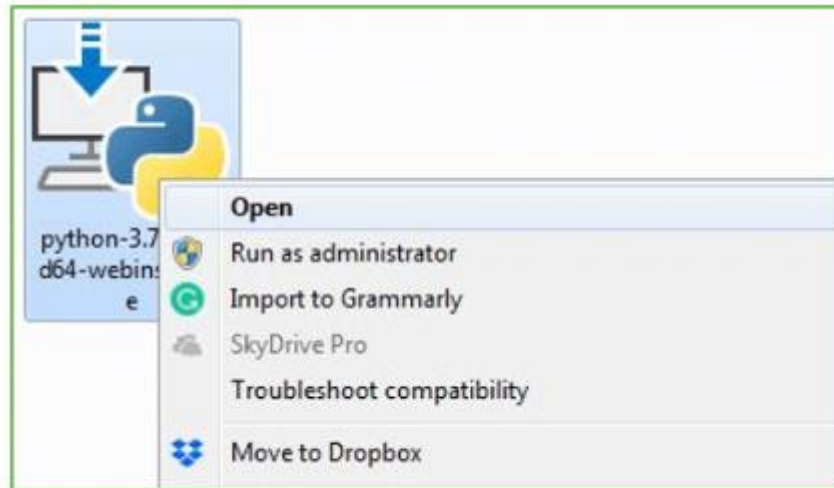
Windows x86-64 embeddable zip file, Windows x86-64 executable installer or Windows x86-64 web-based installer.

Here we will install Windows x86-64 web-based installer. Here your first part regarding which version of python is to be downloaded is completed. Now we move ahead with the second part in installing python i.e. Installation

Note: To know the changes or updates that are made in the version you can click on the Release Note Option.

Installation of Python

Step 1: Go to Download and Open the downloaded python version to carry out the installation process.



Step 2: Before you click on Install Now, Make sure to put a tick on Add Python 3.7 to PATH.



Step 3: Click on Install NOW After the installation is successful. Click on Close.



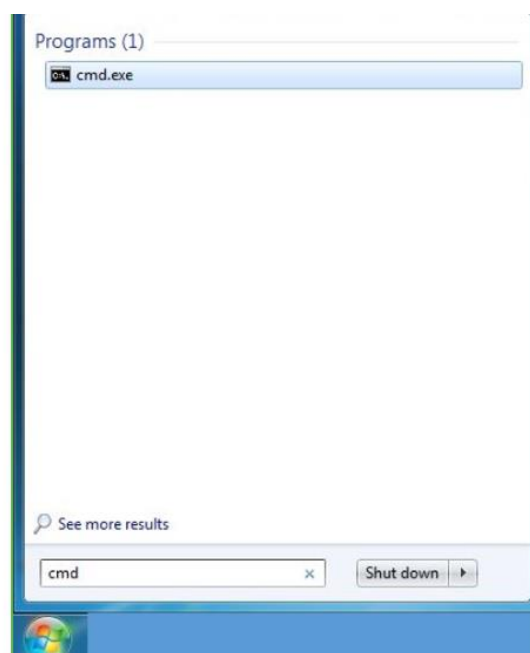
With these above three steps on python installation, you have successfully and correctly installed Python. Now is the time to verify the installation.

Note: The installation process might take a couple of minutes.

Verify the Python Installation

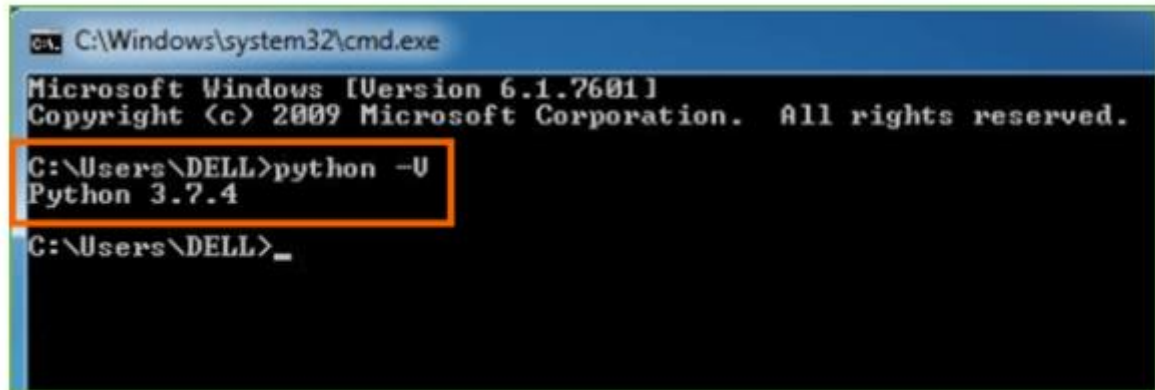
Step 1: Click on Start

Step 2: In the Windows Run Command, type “cmd”.



Step 3: Open the Command prompt option.

Step 4: Let us test whether the python is correctly installed. Type **python -V** and press Enter.



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\DELL>python -U
Python 3.7.4

C:\Users\DELL>_
```

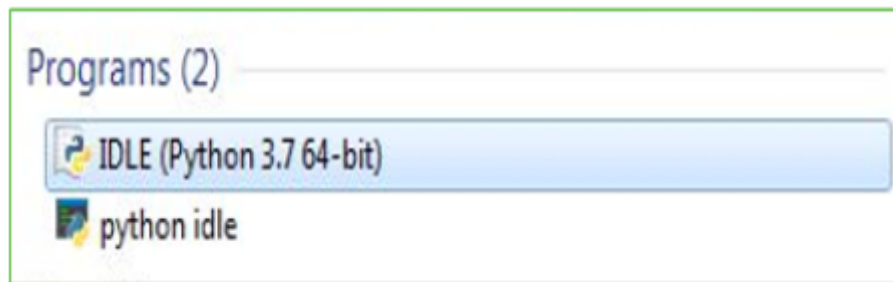
Step 5: You will get the answer as 3.7.4

Note: If you have any of the earlier versions of Python already installed. You must first uninstall the earlier version and then install the new one.

Check how the Python IDLE works

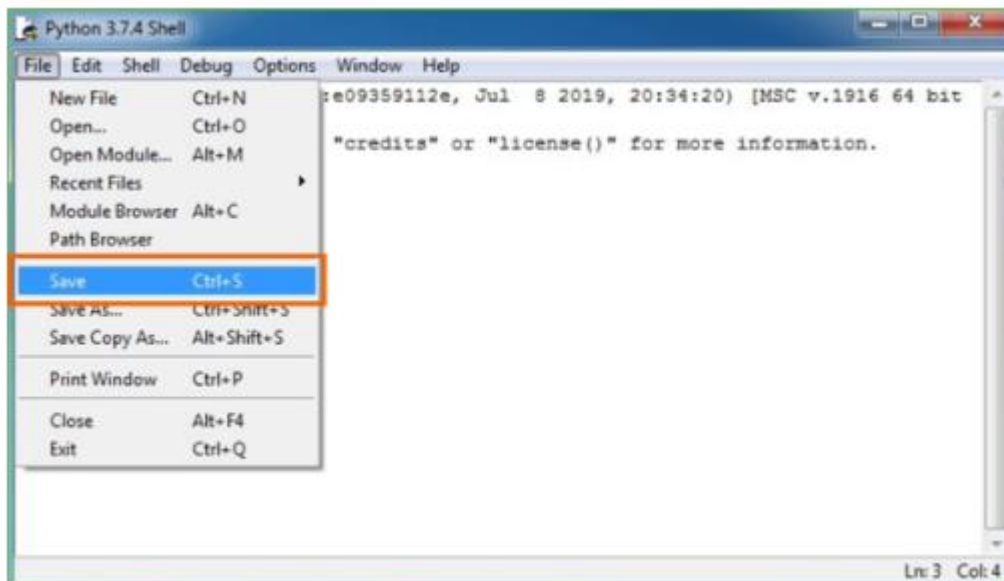
Step 1: Click on Start

Step 2: In the Windows Run command, type “python idle”.



Step 3: Click on IDLE (Python 3.7 64-bit) and launch the program

Step 4: To go ahead with working in IDLE you must first save the file. **Click on File > Click on Save**



Step 5: Name the file and save as type should be Python files. Click on SAVE. Here I have named the files as Hey World.

Step 6: Now for e.g. **enter print**

6.SYSTEM TEST

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successful unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- | | |
|---------------|---------------------------------------------------------|
| Valid Input | : identified classes of valid input must be accepted. |
| Invalid Input | : identified classes of invalid input must be rejected. |

Functions : identified functions must be exercised.
Output : identified classes of application outputs must be exercised.
Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.

- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

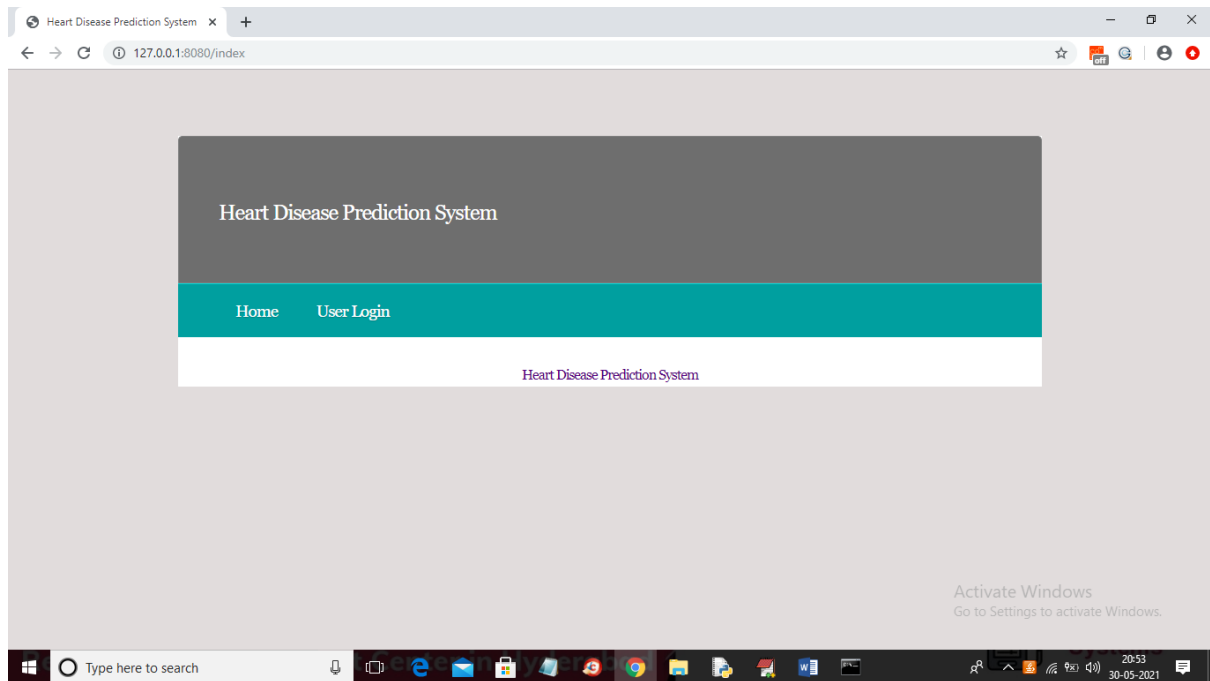
Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

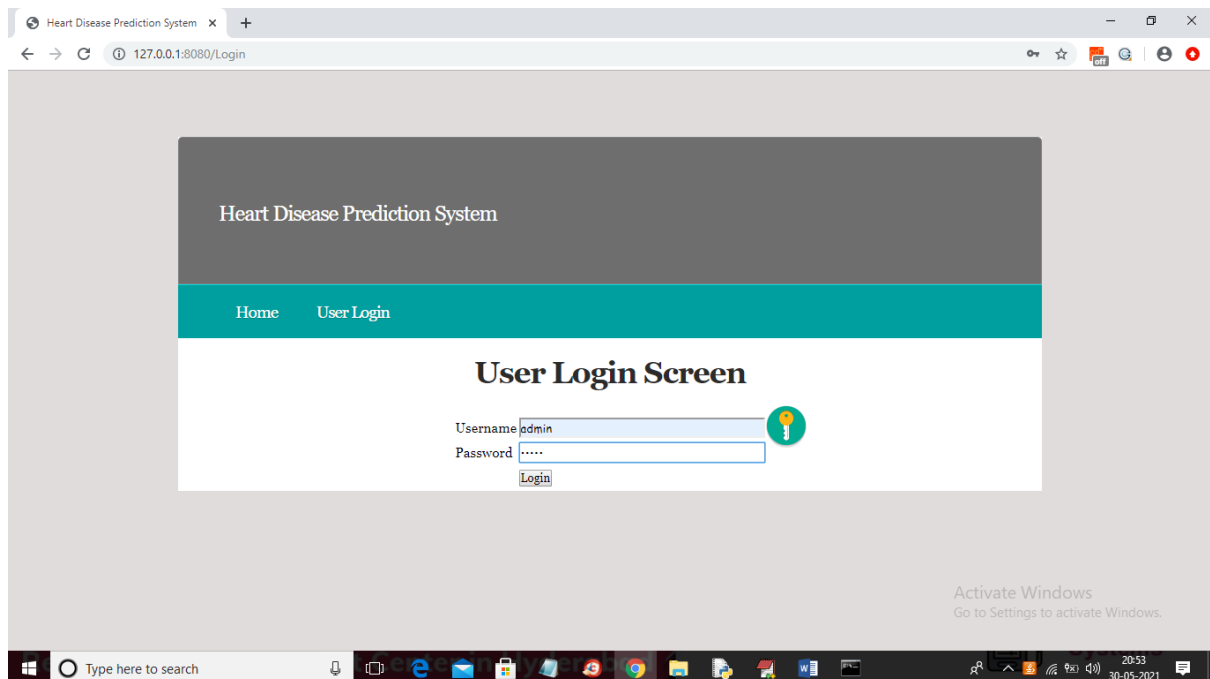
User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

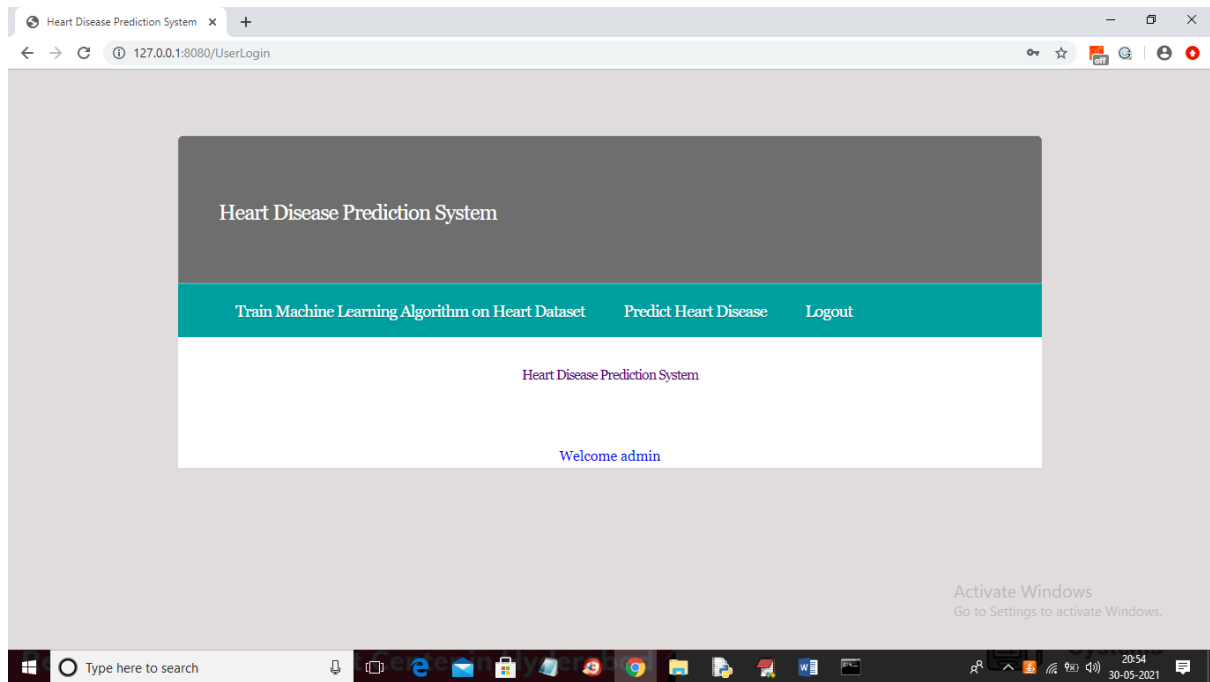
7.SCREENSHOTS



In above screen click on 'User Login' link to get below screen



In above screen enter Username as 'admin' and password as 'admin' and then click on 'Login' button to get below screen



In above screen click on ‘Train Machine Learning Algorithm on Heart Dataset’ link to train all machine learning algorithms

Heart Disease Prediction System

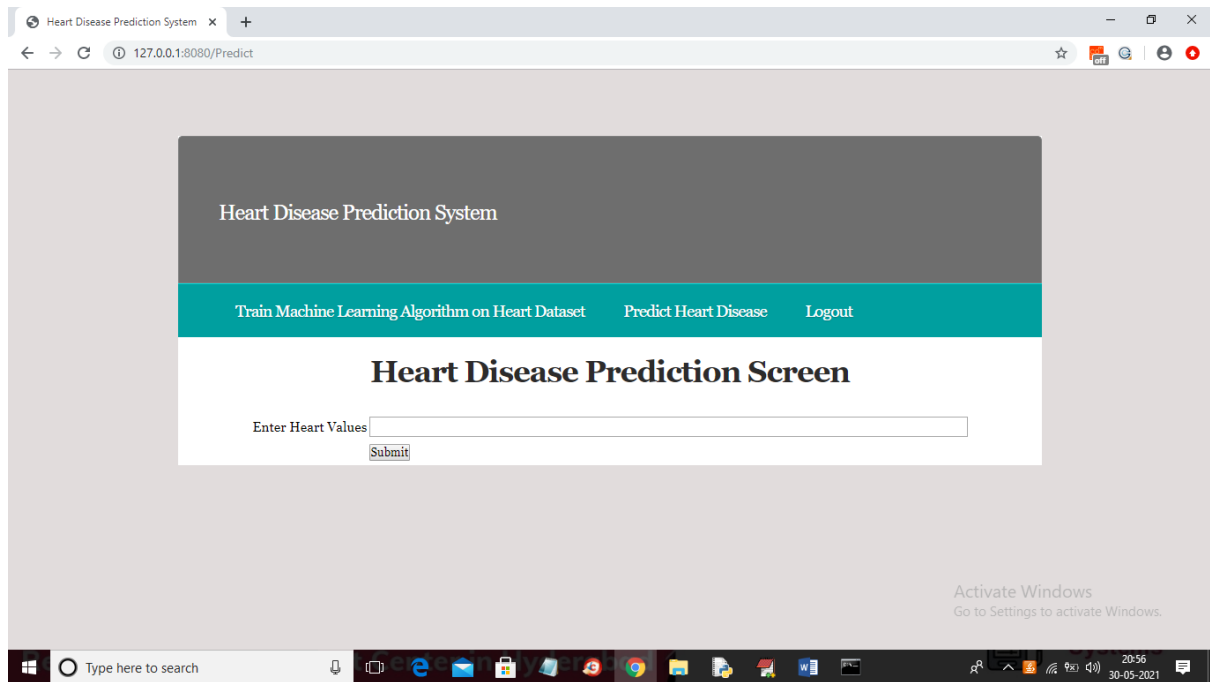
Train Machine Learning Algorithm on Heart Dataset Predict Heart Disease Logout

Heart Disease Prediction System

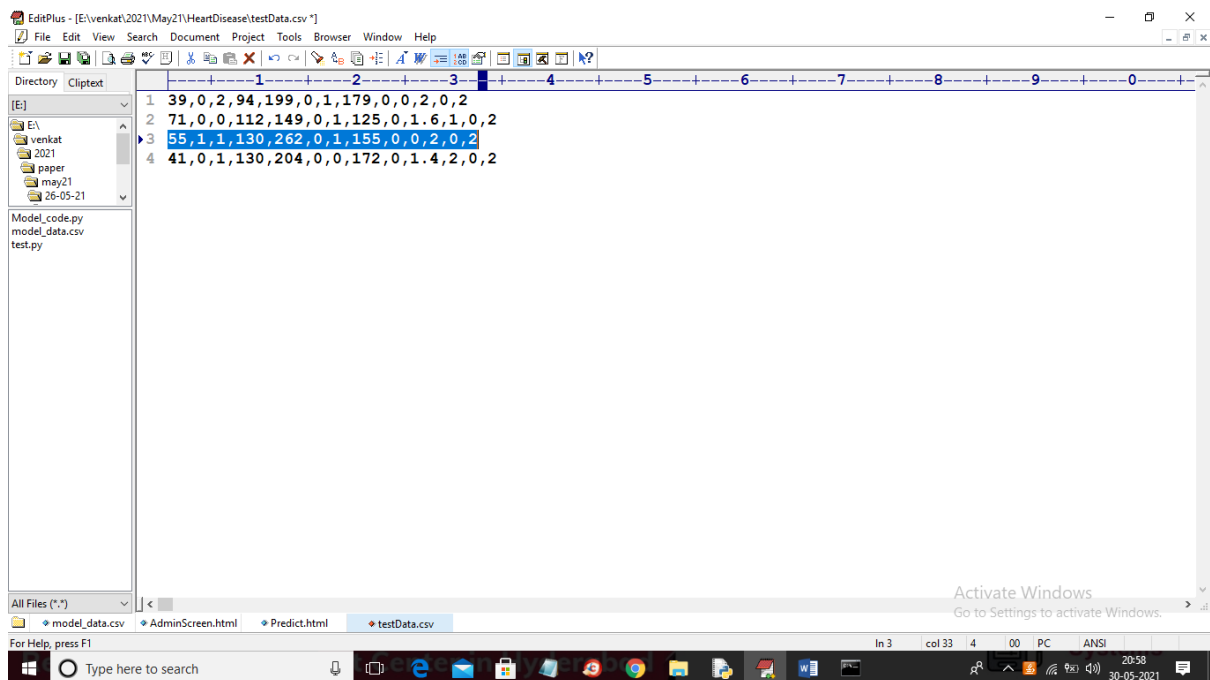
Algorithm Name	Dataset Size	Training Dataset Size	Testing Dataset Size	Accuracy
Random Forest Algorithm	303	242	61	0.8524590163934426
XGBoost Algorithm	303	242	61	0.7704918032786885
KNN Algorithm	303	242	61	0.5901639344262295
LightGBM Algorithm	303	242	61	0.8032786885245902

Activate Windows
Go to Settings to activate Windows.

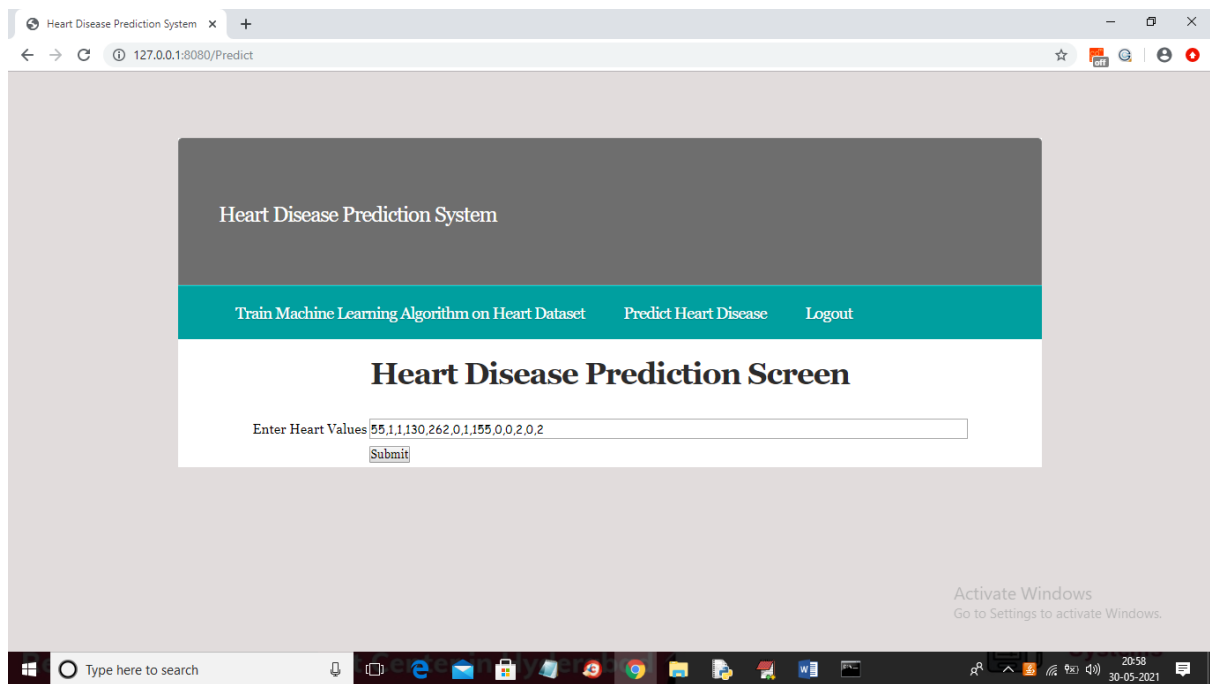
In above screen all algorithms are trained with heart dataset and we can see dataset size and the dataset size used to train and test all algorithms accuracy and in last column we can see accuracy of each algorithm. Now ML models are ready and now click on ‘Predict Heart Disease’ link to get below screen



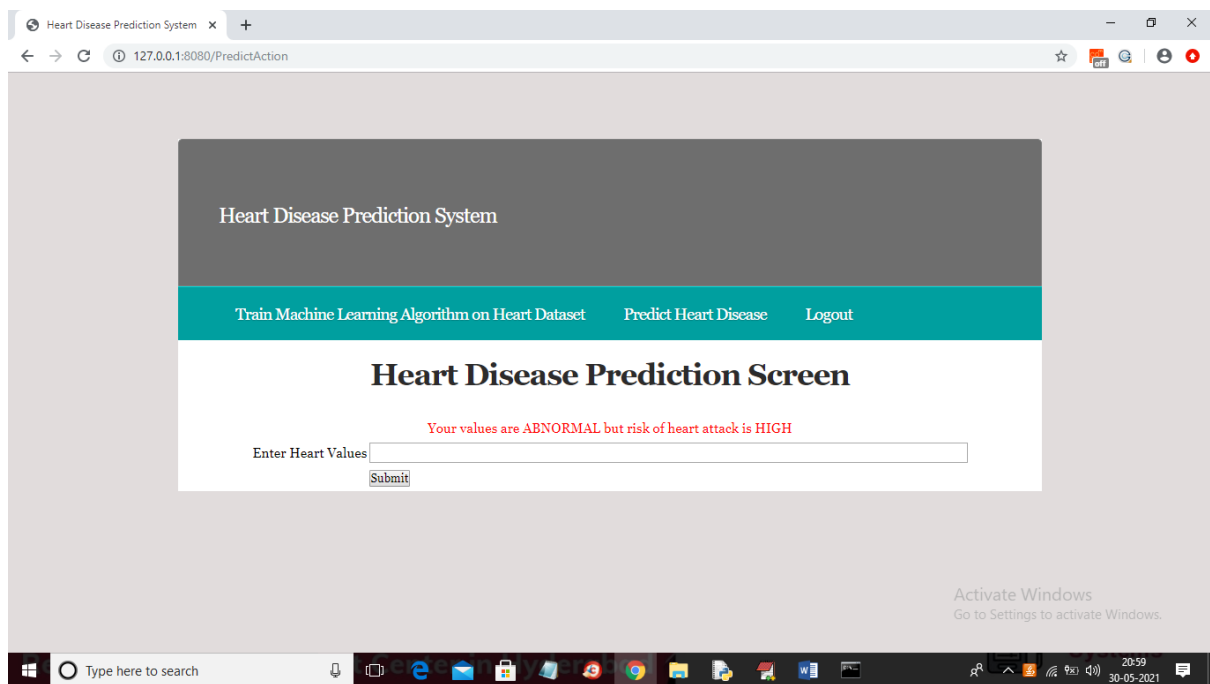
In above screen copy test data and then paste in text field



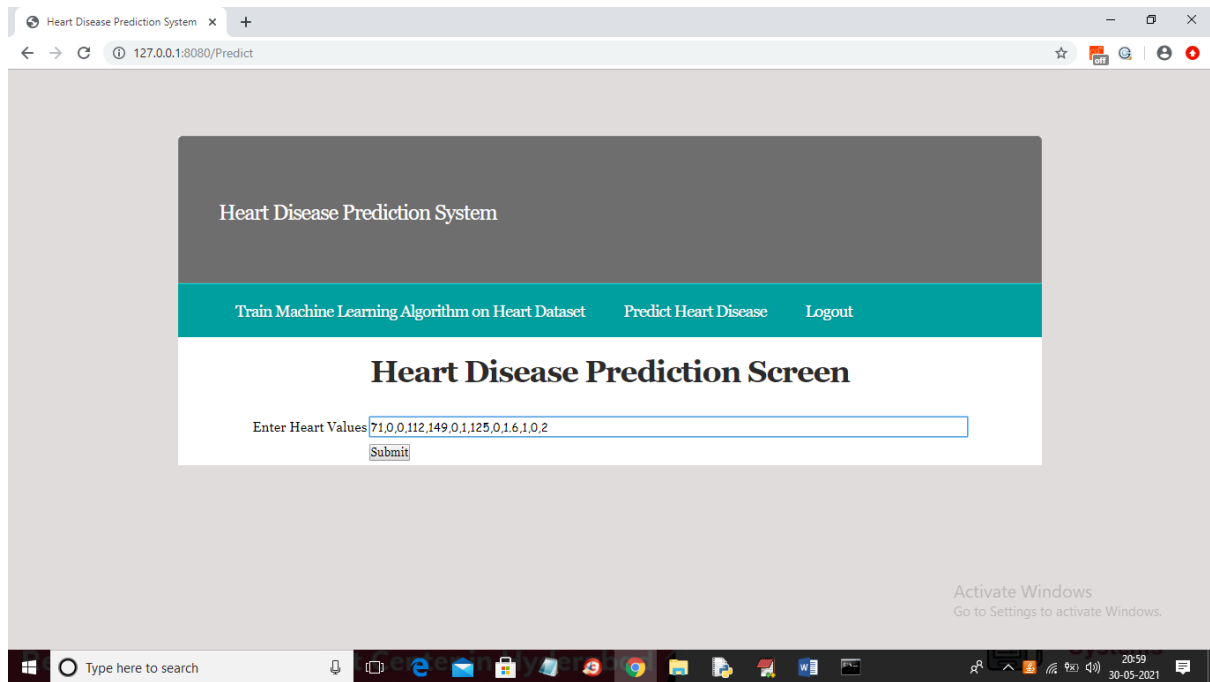
In above screen I am selecting and copy 3rd line and then paste in below screen



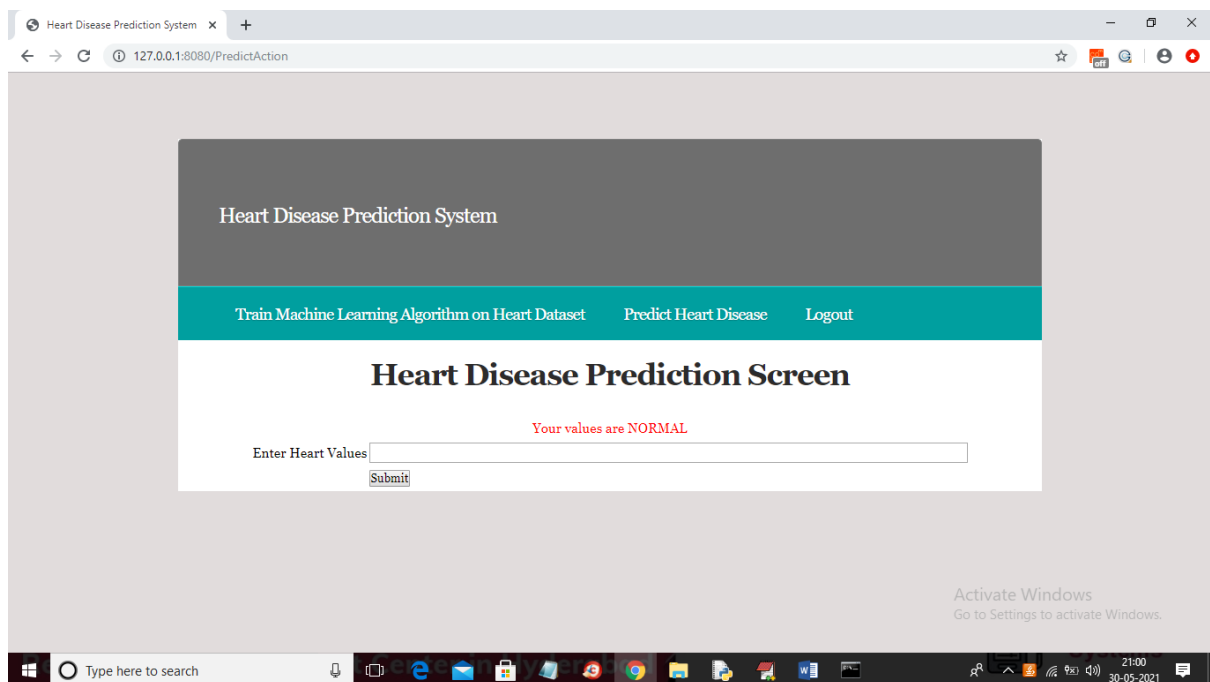
Now in above screen paste copied values and then press submit button to get below screen



In above screen heart disease detected with HIGH risk and now try with other record



Press submit button to get below result.



In above screen result is predicted as NORMAL. Similarly you can enter any own value to get prediction result

CONCLUSION

In this paper, we propose a new convolutional neural networkbased multimodal disease risk prediction (CNN-MDRP)algorithm using structured and unstructured data from hos-pital. To the best of our knowledge, none of the existingwork focused on both data types in the area of medical bigdata analytics. Compared to several typical prediction algo-rithms, the prediction accuracy of our proposed algorithmreaches 94.8% with a convergence speed which is fasterthan that of the CNN-based unimodal disease risk prediction(CNN-UDRP) algorithm

Referneces:

- [1] P. Groves, B. Kayyali, D. Knott, and S. van Kuiken, The`Big Data'Revolution in Healthcare: Accelerating Value and Innovation. USA: Center for US Health System Reform Business Technology Ofce, 2016.
- [2] M. Chen, S. Mao, and Y. Liu, ``Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171209, Apr. 2014.
- [3] P. B. Jensen, L. J. Jensen, and S. Brunak, ``Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genet.*, vol. 13, no. 6, pp. 395405, 2012.
- [4] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, ``A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 30333049, Dec. 2015.
- [5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, ``Wearable 2.0: Enable human-cloud integration in next generation healthcare system," *IEEE Commun.*, vol. 55, no. 1, pp. 5461, Jan. 2017.
- [6] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, ``Smart clothing: Con- necting human with clouds and big data for sustainable health monitor- ing," *ACM/Springer Mobile Netw. Appl.*, vol. 21, no. 5, pp. 825845, 2016.
- [7] M. Chen, P. Zhou, and G. Fortino, ``Emotion communi- cation system," *IEEE Access*, vol. 5, pp. 326337, 2017, doi: 10.1109/ACCESS.2016.2641480.
- [8] M. Qiu and E. H.-M. Sha, ``Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems," *ACM Trans. Design Autom. Electron. Syst.*, vol. 14, no. 2, p. 25, 2009.
- [9] J. Wang, M. Qiu, and B. Guo, ``Enabling real-time information service on telehealth system over cloud-based big data platform," *J. Syst. Archit.*, vol. 72, pp. 6979, Jan. 2017.
- [10] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, ``Big data in health care: Using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 11231131, 2014.
- [11] L. Qiu, K. Gai, and M. Qiu, ``Optimal big data sharing approach for telehealth in cloud computing," in *Proc. IEEE Int. Conf. Smart Cloud (Smart-Cloud)*, Nov. 2016, pp. 184189.
- [12] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, ``Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Syst. J.*, vol. 11, no. 1, pp. 8895, Mar. 2017.
- [13] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, ``Localization based on social big data analysis in the vehicular networks," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2016.2641467.

- [14] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced fingerprinting and trajectory prediction for iot localization in smart buildings," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 12941307, Jul. 2016.
- [15] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, "Risk factors and risk assessment tools for falls in hospital in-patients: A systematic review," *Age Ageing*, vol. 33, no. 2, pp. 122130, 2004.
- [16] S. Maroon, A. M. Chang, B. Lee, R. Salhi, and J. E. Hollander, "Heart score to further risk stratify patients with low TIMI scores," *Critical Pathways Cardiol.*, vol. 12, no. 1, pp. 15, 2013.
- [17] S. Bandyopadhyay et al., "Data mining for censored time-to-event data: A Bayesian network model for predicting cardiovascular risk from electronic health record data," *Data Mining Knowl. Discovery*, vol. 29, no. 4, pp. 10331069, 2015.
- [18] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," *Data Mining Knowl. Discovery*, vol. 29, no. 4, pp. 10701093, 2015.
- [19] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, "Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration," *J. Biomed. Inform.*, vol. 53, pp. 220228, Feb. 2015.
- [20] J. Wan et al., "A manufacturing big data solution for active preventive maintenance," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2017.2670505.