# DETECTION OF THYROID DISORDERS USING MACHINE LEARNING APPROCHES

*A Mini Project Report submitted to*
*JNTU Hyderabad in partial fulfillment*
*of the requirements for the award of the degree*

**BACHELOR OF TECHNOLOGY**

In

**COMPUTER SCIENCE AND ENGINEERING**

***Submitted by***

| | |
|---|---|
| **KALWAKUNTLA BHARGAVI** | **21RG1A0524** |
| **LAKAVATH SHOBHA** | **21RG1A0528** |
| **MAROJU MEGHANA** | **22RG5A0504** |
| **PAIDIPALLI DHARANI** | **22RG5A0505** |

*Under the Guidance of*

**Mr. V. SHYAM KUMAR**
*MCA,M. Tech*
*Assistant Professor*

***DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING***
**MALLA REDDY COLLEGE OF ENGINEERING FOR WOMEN**
**An UGC Autonomous Institution**
*Approved by AICTE New Delhi and Affiliated to JNTUH*
*Maisammaguda , Medchal (Dist), Hyderabad -500100,Telangana.*
*OCTOBER 2024*

*DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING*

MALLA REDDY COLLEGE OF ENGINEERING FOR WOMEN

**An UGC Autonomous Institution**

*Approved by AICTE New Delhi and Affiliated to JNTUH*

*Maisammaguda , Medchal (Dist), Hyderabad -500100, Telangana.*

*OCTOBER 2024*

## *CERTIFICATE*

This is to certify that the Mini project entitled **“DETECTION OF THYROID DISORDERS USING MACHINE LEARNING APPROACHES”** has been submitted by **KALWAKUNTLA BHARGAVI (21RG1A0524), LAKAVATH SHOBHA (21RG1A0528), MAROJU MEGHANA (22RG5A0504), PAIDIPALLI DHARANI (22RG5A0505)** in partial fulfillment of the requirements for the award of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ENGINEERING**. This record of bonafide work carried out by them under my guidance and supervision. ***The result embodied in this mini project report has not been submitted to any other University or Institute for the award of any degree.***

| | |
|---|---|
| **Mr. V. SHYAM KUMAR** | **Mrs. K. SHEETAL** |
| Assistant professor | Head of the Department Project Guide |

***External Examiner***

# ACKNOWLEDGEMENT

The Mini Project work carried out by our team in the Department of Computer Science and Engineering, Malla Reddy College of Engineering for Women, Hyderabad. ***This work is original and has not been submitted in part or full for any degree or diploma of any other university.***

We wish to acknowledge our sincere thanks to our project guide **Mr. V. Shyam Kumar,** Assistant Professor, Computer Science & Engineering for formulation of the problem, analysis, guidance and her continuous supervision during the course of work.

We acknowledge our sincere thanks to **Dr. Kanaka Durga Returi**, Principal & Professor of Department of Computer Science and Engineering and **Mrs. Sheetal Kulkarni,** Head of the Department, MRCEW and all our faculties of CSE Department for their kind cooperation in making this Mini Project work a success.

We extend our gratitude to **Sri. Ch. Malla Reddy**, Founder Chairman and **Sri. Ch. Mahender Reddy,** Secretary, **Dr. Vaka Murali Mohan,** Director for their kind cooperation in providing the infrastructure for completion of our Mini Project.

We acknowledge our special thanks to the entire teaching faculty and non-teaching staff members of the Computer Science & Engineering Department for their support in making this project work a success.

| | | |
|---|---|---|
| **KALWAKUNTLA BHARGAVI** | **HT.NO.** | **21RG1A0524________________** |
| **LAKAVATH SHOBHA** | **HT.NO.** | **21RG1A0528________________** |
| **MAROJU MEGHANA** | **HT.NO.** | **22RG5A0504________________** |
| **PAIDIPALLI DHARANI** | **HT.NO.** | **22RG5A0505________________** |

# INDEX

# ABSTRACT

Classification based Machine learning plays a major role in various medical services. In medical field, the salient and demanding task is to diagnose patient's health conditions and to provide proper care and treatment of the disease at the initial stage. Let us consider Thyroid disease as the example. The normal and traditional methods of thyroid diagnosis involve a thorough inspection and also various blood tests. The main goal is to recognize the disease at the early stages with a very high correctness. Machine learning techniques play a major role in medical field for making a correct decision, proper disease diagnosis and also saves cost and time of the patient. The purpose of this study is prediction of thyroid disease using classification Predictive Modelling followed by binary classification using Decision Tree ID3 and Naive Bayes Algorithms. The Thyroid Patient dataset with proper attributes are fetched and using the Decision Tree algorithm the presence of thyroid in the patient is tested. Further, if thyroid is present then Naïve Bayes algorithm is applied to check for the thyroid stage in the patient.

# LIST OF FIGURES

# CHAPTER1: SYSTEM ANALYSIS

## 1.1 EXISTING SYSTEM

□ In the data collection stage, small, memory-constrained and low energy-consumption sensors with a short-range communications capability are employed to collect information about the physical environment. Ethernet, WiFi, ZigBee, and wire-based technologies are combined with Transmission Control Protocol/Internet Protocol to connect the objects and users across prolonged distances during data transmission. During the data processing and utilization stage, applications process the data to obtain useful information, and may initiate control commands to act on the physical environment after making decisions based on the collected information. The coordination of diverse technologies, the heterogeneity, and the distributed nature of communications technologies proposed for the IoT by different standards development organizations [4] magnify the threat to end- to-end security in IoT applications.

### 1.1.1 Disadvantages of Existing System

1) Less accuracy
2) low Efficiency

## 1.2 PROPOSED SYSTEM

The thyroid dataset is taken from Kaggle Machine Learning website [13]. The Database mainly includes the thyroid patient records having all the necessary patient details in it. The patient record has important attributes as mentioned in the Table I. Along with this, the proposed model also takes all the records of the patient’s past clinical history showing in the Fig 1. These include whether the patient is allergic to any particular medicine, whether the patient has undergone any past thyroid surgery and also any recent thyroid test and genetic history of the patient. These also act as the major attributes since they ease the examination of the thyroid patient and reduce the thorough examination by the doctor. This saves time and eases the diagnosis process.

These attributes are stored in a dedicated cloud server which can be made private or hybrid based on the health organization’s need and interest. Among the considered attributes a train dataset is prepared and is given as the input to the classification based machine learning model. This is a supervised learning method and the designed model will generate the results based on the train dataset values. The proposed model has Decision tree and Naive Bayes algorithm to generate the results. A decision tree is a tree based algorithm which follows a top down approach build. ID3 algorithm is used to construct the decision tree. It mainly eliminates any redundant element if present and improves the accuracy of the classification.

This decision tree algorithm is applied to the thyroid patient’s records consisting of age, gender, T3, T4, TSH values. The decision tree algorithm calculates the inputted values present in the thyroid patient record. The calculation is done based on the train dataset. Therefore more the number of records in the train dataset higher the accuracy of the algorithm.

For example, if 3000 train thyroid dataset records are considered and trained to the decision tree algorithm then the generated accuracy rate will be high. Our proposed model has considered more than 3000 train dataset attributes resulting in 95% accuracy rate in the prediction results. If the patient's output value results true then further Naïve Bayes algorithm is applied to calculate which stage is the patient currently in. This adds as a major advantage to the health care staff in the easy analysis of the thyroid disease and also avoid certain lab tests if not necessary. The patient's thyroid stage here is divided into 3 stages i.e., minor, major and critical. The Naïve Bayes algorithm is applied if the Decision Tree returns thyroid true or positive value. The Naïve Bayes algorithm in machine learning is a supervised learning algorithm which is based on Bayes' Theorem.

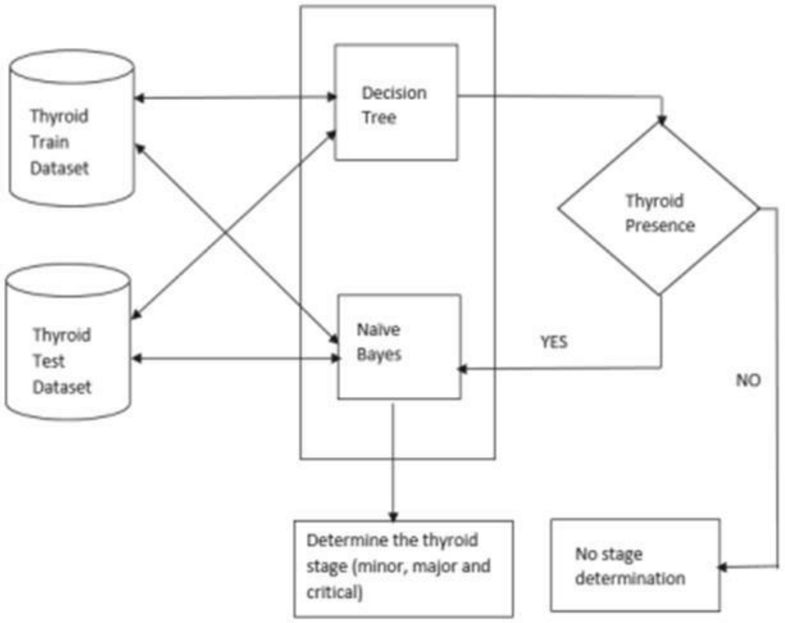


Fig. 2. Proposed machine learning classification model for thyroid disease diagnosis.

It is used for solving classification based problems. The model can be built fast and it is also cost effective one. In this way our proposed system can make a major contribution in the healthcare field and also generate positive outcomes with good accuracy with a cost and time saving method for the thyroid patients.

### 1.2.1 **Advantages of Proposed System**

1)High accuracy

2)High efficiency

## 1.3 INTRODUCTION

Thyroid disease diagnosis is not a simple task. It involves many procedures. The normal traditional way includes a proper medical examination and many blood samples for blood tests. Therefore, there is a necessity for a model which detects the thyroid disease at a very early stage of development.

In medical field machine learning plays an important role for thyroid disease diagnosis as it has various classification models based on which we can train our model with proper train dataset of the thyroid patient and can predict and give the results in an accurate manner with higher degree of correctness.

Some recent studies from Mumbai have suggested that congenital hypothyroidism is common in India. The disease occurs in 1 part of 2640 new born children, when compared to the worldwide average range of 1 in 3800 considered. Congenital hypothyroidism can lead to serious complications if not detected in early stages. Therefore, the proposed model serves the goal in early detection of thyroid disease.

Based on the obtained test values the health care staff can easily examine the condition of the patient and also skip further clinical examinations if not necessary. Hence, this approach proves to be very much beneficial to the healthcare field. A proper train dataset results into an accurate predicting model therefore reducing the overall cost of the thyroid patient treatment and also saving the time [2]. Classification algorithms are most suitable in decision making and also solving the realworld problems.

## ABOUT THYROID

The Thyroid is butterfly-shaped endocrine gland which is situated at the base of the human neck. The vital role of the thyroid gland is maintaining and balancing human metabolism and also the growth and development of the human body. The vital tasks performed by thyroid gland are blood circulation, body temperature control, muscle strength and brain functioning

[1]. Any damage or improper functioning of the gland may seriously affect the normal human body functioning [2].Therefore, proper thyroid hormone secretion results into a healthy human body. If there is either low or high secretions of the hormone it will adversely affect the human health.

The Thyroid gland mainly produces tri-iodothyronine (T3), thyroxine (T4) and Thyroid stimulating hormone (TSH). The Thyroid stimulating hormone (TSH) [3] is released by the pituitary gland which mainly stimulates the thyroid gland to produce T3 and T4 which further stimulate the metabolism of almost every tissue present in the human body. Therefore, the pituitary gland plays a vital role in controlling the production of the required amount of thyroid hormones. If the TSH production level is less then T3, T4 secretion will be more and vice versa.

The Thyroid disorder is the most common endocrine disease across the world. In a survey carried out in India, around 42 million people are suffering from this disease [1]. Thyroid disease is different from other type of endocrine diseases in terms of the mode of treatment relative attainability and the ease of predicting the disease. The high thyroid hormone secretion leads to Hyperthyroidism and low secretion leads to Hypothyroidism. Both the conditions adversely affect the human physiology and the symptoms shown for hyperthyroidism are dry skin, hair thinning, loss of weight, high blood pressure, neck enlargement and short menstrual periods. The symptoms show for hypothyroidism include the thyroid gland inflammation, weight gain, low bp, heavy menstrual periods and loss of appetite.

# CHAPTER-2: LITERATURE SURVEY

1. Bibi Amina Begum et al. [1] have proposed different Thyroid prediction techniques using data mining approaches. They have considered different dataset attributes for prediction and have explained the classification techniques in data mining like Decision Tree, Backpropagation Neural Network, SVM and density based clustering. They have analyzed the correlation of T3, T4 and TSH with hyperthyroidism and hypothyroidism.
2. Ankita Tyagi et al. [2] have studied various classification based machine learning algorithms. They have considered train data set from UCI Machine Learning repository and compared and analyzed the performance metric of decision tree, support vector machine and K-nearest neighbor.
3. Aswathi A K et al. [3] have proposed a training model consisting of 21 thyroid causing attributes. They have proposed partial swarm optimization to optimize the support vector machine parameters.
4. M. Deepika et al. [4] have performed a general empirical study on various disease diagnosis like Diabetes, Breast Cancer, Heart disease, Thyroid prediction and have compared the accuracy rate by applying SVM, Decision tree and Artificial Neural Networks.
5. Sumathi A et al. [5] have considered Thyroid data preprocessing mainly by applying the decision tree algorithm. They have first calculated the mean values of T3, T4 and TSH and considered as the preprocessing stage. Later on they have applied machine learning based feature selection and feature construction. Further they have applied classification based J48 algorithm which is a continuation of ID3 algorithm and calculated the results.
6. I Md. Dendi Maysanjaya et al. [6] have analyzed a comparison on various classification methods used to diagnose thyroid disease. They have compared by using Artificial Neural Networks, Radial Based Function, Learning Vector Quantization, Back Propagation Algorithm and Artificial Immune recognition system and concluded the comparison results. Among that they found out that Multilayer Perceptron has the highest accuracy of 96.74%.
7. Ammulu K et al. [7] have proposed a Thyroid Prediction System based on data mining classification algorithm. They have used random forest approach to predict the results using Weka open source tool used for data mining. Using this tool they have applied random forest algorithm with 25 thyroid data attributes and predicted the results accordingly.

8. Roshan Banu D et al. [8] have conducted a study on different data mining techniques to detect thyroid disease. They have done study on Linear Discriminant analysis, Kfold cross validation, and Decision tree. They have analyzed various splitting rules for the attributes of Decision tree. They have also compared the obtained values.
9. Dr.B.Srinivasan et al. [9] have conducted a study on diagnosis of the thyroid disease using different data mining approaches. They have explained the major cause of the thyroid disease and have also given description about Decision Tree, Naïve Bayes classification and SVM.
10. Sunila Godara et al. [10] have performed a Prediction on Thyroid Disease using various machine learning techniques. They have considered Logistic Regression and Support Vector Machine as the main Thyroid detection models. They have concluded that these two proposed classifier methods are the best when the number of classes increases in the thyroid prediction model.
11. [11] Ankita Tyagi and Rikitha Mehra "Interactive Thyroid Disease Prediction System Using Machine Learning Techniques" 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018), 20-22 Dec, 2018, Solan, India.
12. Sunila Godara. [12] They have used Logistics Regression and SVM machine learning Technique to analyze Thyroid Dataset. Comparison was made between these two algorithms based on Precision, Recall, F measure, ROC, RMS error. Logistic Regression turned out has best classifier.
13. YongFeng Wang. [13] Thyroid Nodule is diagnosed for benign or malignant type using Ultrasound images of thyroid by image analysis - radiomics and deep learning based approaches. Comparison is made between these two approaches.
14. Hitesh Garg. [14] Feed Forward Neural Network is used for feature extraction and segmentation from Ultrasound images to predict the tumors. The accuracy and other factors were measured and all the average values were above 86%.
15. Gou and Du proposed a system [15] that consists of a Generalized Discriminant Analysis and Wavelet Support Vector Machine System (GDA_WSVM) approach for the analysis of thyroid illnesses which incorporates three phases.
16. Yang et al.'s targets are diagnosing thyroid illnesses with a professional system [16]. These are feature extraction – feature reduction phase, classification phase, and test of GDA_WSVM for correct diagnosis of thyroid diseases phase.

17. Poudel et al. [17] proposed that information benefit primarily based on a synthetic immune popularity system (IG-AIRS) might help diagnose thyroid characteristics primarily based totally on laboratory tests and might open the manner to numerous unwell diagnoses aided by the use of the latest scientific exam data.
18. Shankar et al. [18] evaluated the TUSP automated detection technique to predict thyroid disease by removing the long ultrasound imaging process.
19. Aswathi and Antony [19] used unlabelled data to perform unsupervised learning to improve thyroid classification problems and optimize them.
20. Banu [20] has targeted growing an AIS-based device gaining knowledge of a classifier for clinical analysis and investigating the functionality of the proposed classifier. The proposed classifier efficiently advanced the identity manner of thyroid gland disease.

# CHAPTER-3: SYSTEM DESIGN

## 3.1 System Modules

1) **Upload Thyroid Dataset:** using this module we will upload dataset details and then application will read and display dataset values and then find and plot graph of normal and thyroid patients count.
2) **Preprocess Dataset:** using this module we will preprocess dataset and then remove missing values and then convert all non-numeric data into numeric data and then shuffle and then split dataset into train and test where application using 80% dataset for training and 20% for testing.
3) **Run Naive Bayes Algorithm:** using this module we will input 80% dataset to Naïve Bayes to train a model and 20% test data will be applied on trained model to calculate prediction accuracy.
4) **Run SVM Algorithm:** using this module we will input 80% dataset to SVM to train a model and 20% test data will be applied on trained model to calculate prediction accuracy.
5) **Run Random Forest Algorithm:** using this module we will input 80% dataset to Random Forest to train a model and 20% test data will be applied on trained model to calculate prediction accuracy.
6) **Comparison Graph:** using this module we will plot accuracy comparison graph between all algorithms.
7) **Predict Disease from Test Data:** using this module we will upload test data and then random forest will predict weather test data is normal or contains thyroid disease.

## 3.2 System Architecture

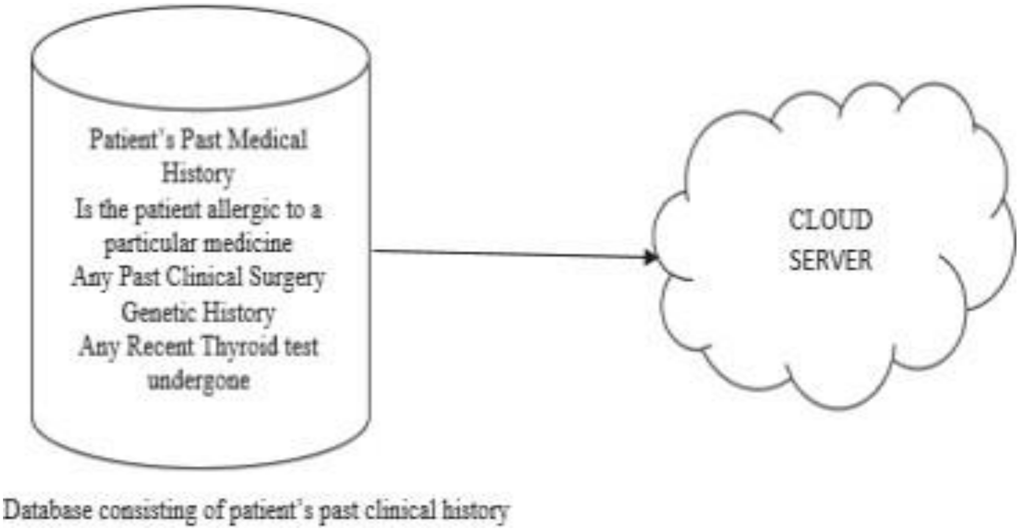

Patient's Past Medical History
Is the patient allergic to a particular medicine
Any Past Clinical Surgery
Genetic History
Any Recent Thyroid test undergone
CLOUD SERVER
Database consisting of patient's past clinical history

**Fig.3.1: System Architecture**

## 3.3 System Requirements

### 3.3.1 Hardware Requirements

| | | |
|---|---|---|
| System | : | i3 |
| Ram | : | 4 GB |
| Hard Disk | : | 40 GB |

### 3.3.2 Software Requirements:

| | | |
|---|---|---|
| Operating system | : | Windows8 |
| Coding Language | : | python |

## 3.4 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.

### 3.4.1 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.

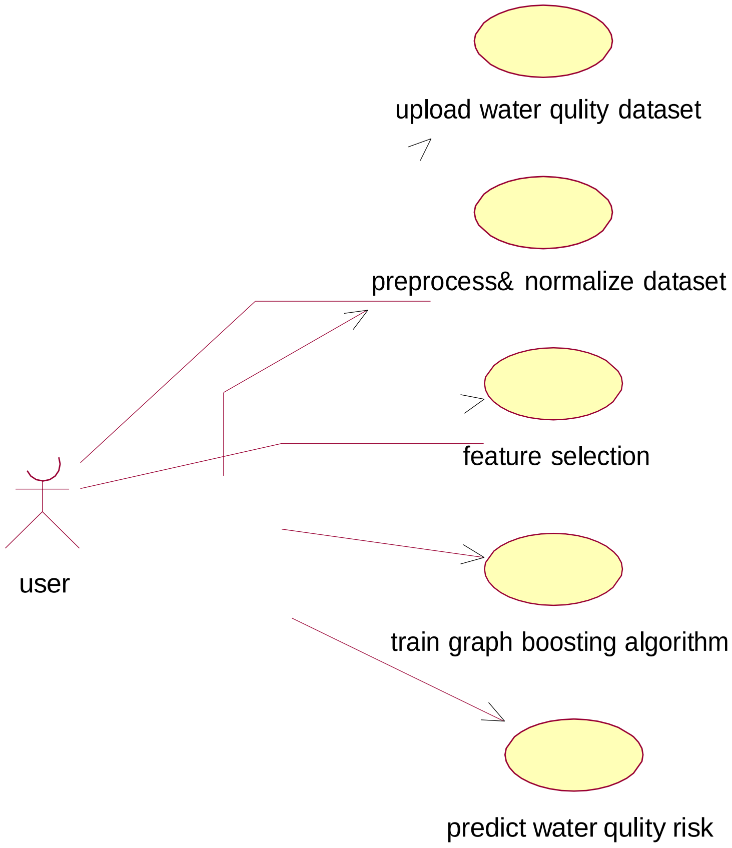

upload water qulity dataset
preprocess& normalize dataset
feature selection
user
train graph boosting algorithm
predict water qulity risk

**Fig. 3.2 : Use Case Diagrams**

### 3.4.2 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.

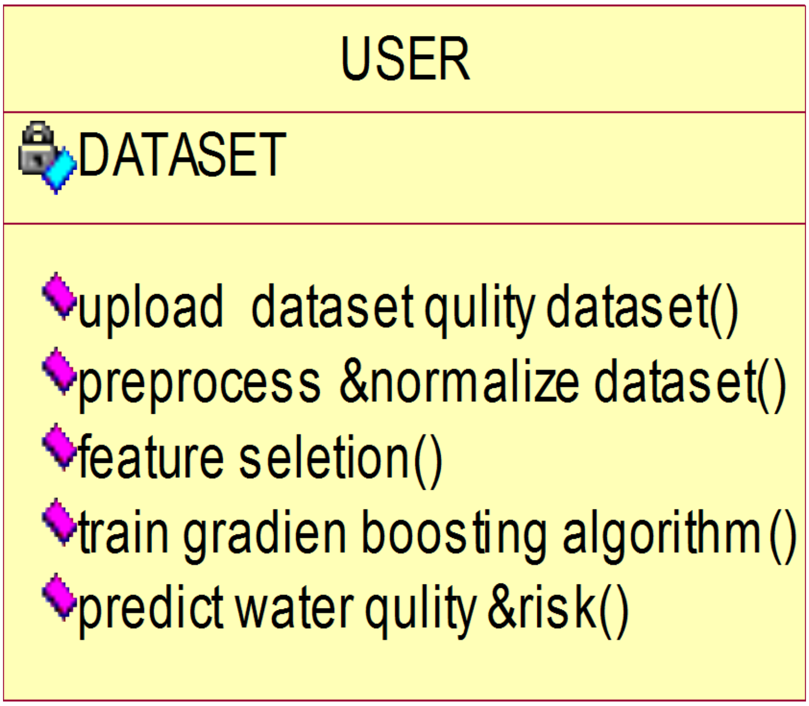

USER
DATASET
upload dataset qulity dataset()
preprocess &normalize dataset()
feature seletion()
train gradien boosting algorithm()
predict water qulity &risk()

**Fig. 3.3: Class Diagram**

### 3.4.3 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.

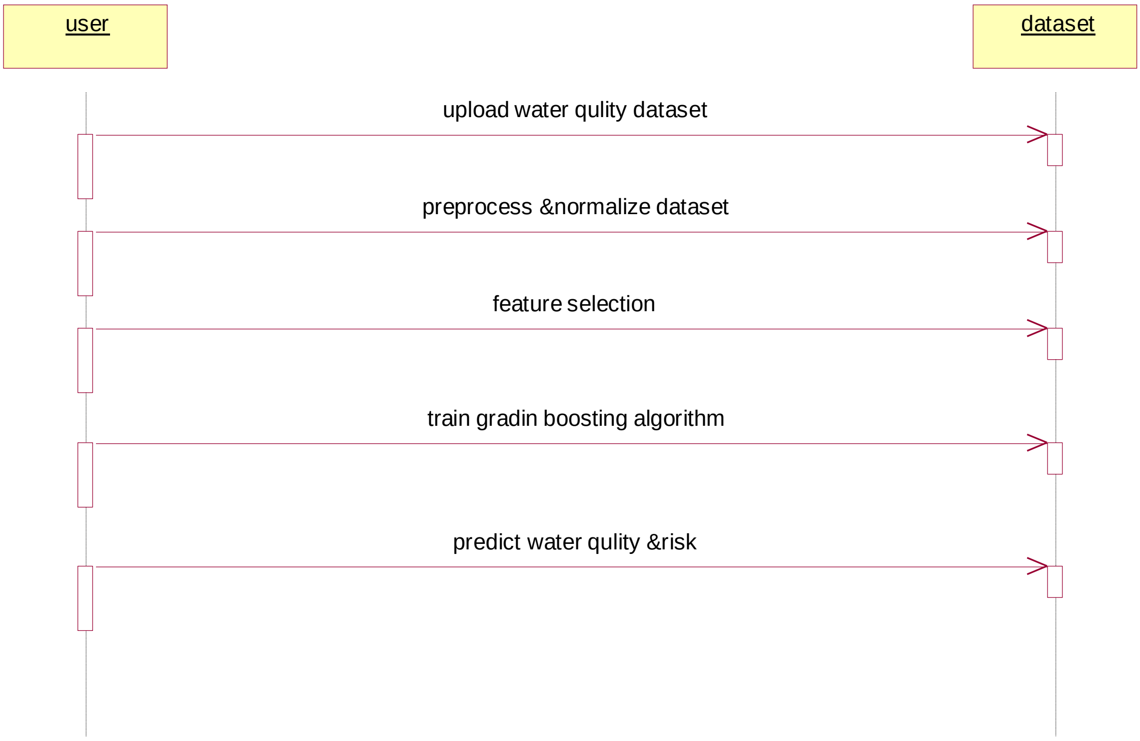

user
dataset
upload water qulity dataset
preprocess &normalize dataset
feature selection
train gradin boosting algorithm
predict water qulity &risk

Fig. 3.4 : Sequence Diagram

### 3.4.4 Collaboration 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.

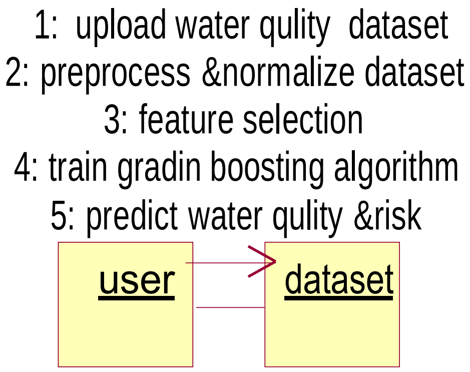

1: upload water qulity dataset
2: preprocess &normalize dataset
3: feature selection
4: train gradin boosting algorithm
5: predict water qulity &risk
user
dataset

**Fig. 3.5 : Collaboration Diagram**

# CHAPTER-4: INPUT AND OUTPUT DESIGN

## 4.1 INPUT DESIGN

Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations.

This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user-friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases.

## 4.2 OUTPUT DESIGN

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the users id depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user, rests with the administrator only.

The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

# CHAPTER-5: SOFTWARE ENVIRONMENT

## 5.1 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

### 5.1.1 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.

## 5.1.2 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.

## 5.1.3 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 existenceof 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 Data Base 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.

## 5.2 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 Venners[1], 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."

## 5.3 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.

## 5.4 Categories Of Machine Leaning :-

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.

### 5.4.1 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 programing 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.

### 5.4.2 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.

### 5.4.3 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

## 5.5 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!!!.

### 5.5.1 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. But if you want to focus on R&D in Machine Learning, then mastery of Linear Algebra and Multivariate Calculus is very important as you will have to implement many ML algorithms from scratch.

### (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 colour, 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 labelled data. Using labelled 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.

## 5.5.2 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.

## 5.5.3 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.

## 5.6 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 7.3:

- 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.

## 5.7 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. Arbitrary data-types can be defined using Numpy which allows Numpy to seamlessly and speedily integrate with a wide variety of databases.

## 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.

## 5.8 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 :

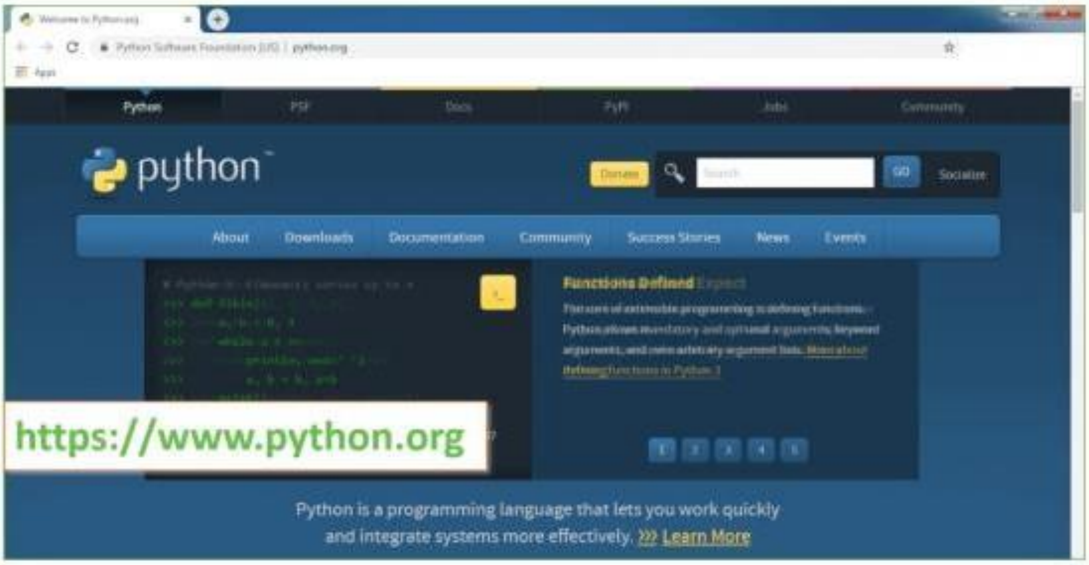
There have been several updates in the Python version over the years. The question is how to install Python? It might be confusing for the beginner who is willing to start learning Python but this tutorial will solve your query. The latest or the newest version of Python is version 3.7.4 or in other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

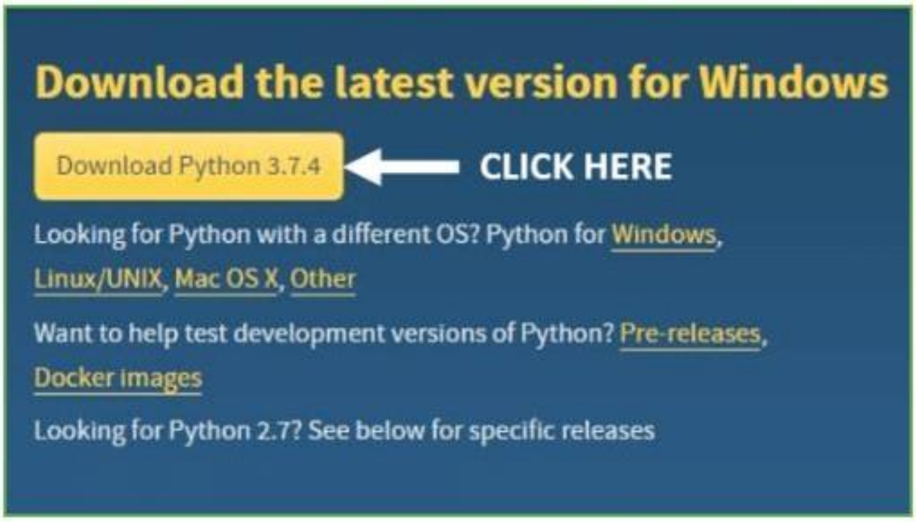


https://www.python.org

Python is a programming language that lets you work quickly and integrate systems more effectively. >>> Learn More

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

**Step2:** Click on the Download Tab.



Download the latest version for Windows

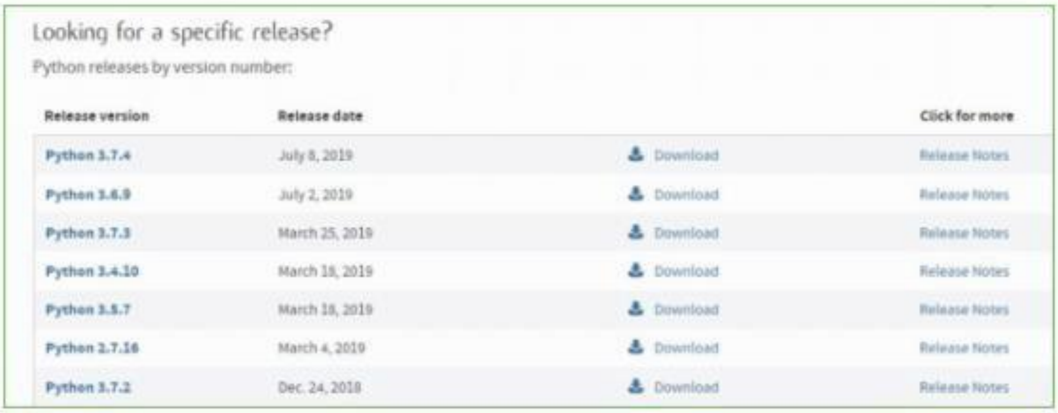
Download Python 3.7.4 ← CLICK HERE

Looking for Python with a different OS? Python for Windows, Linux/UNIX, Mac OS X, Other

Want to help test development versions of Python? Pre-releases, Docker images

Looking for Python 2.7? See below for specific releases

**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.



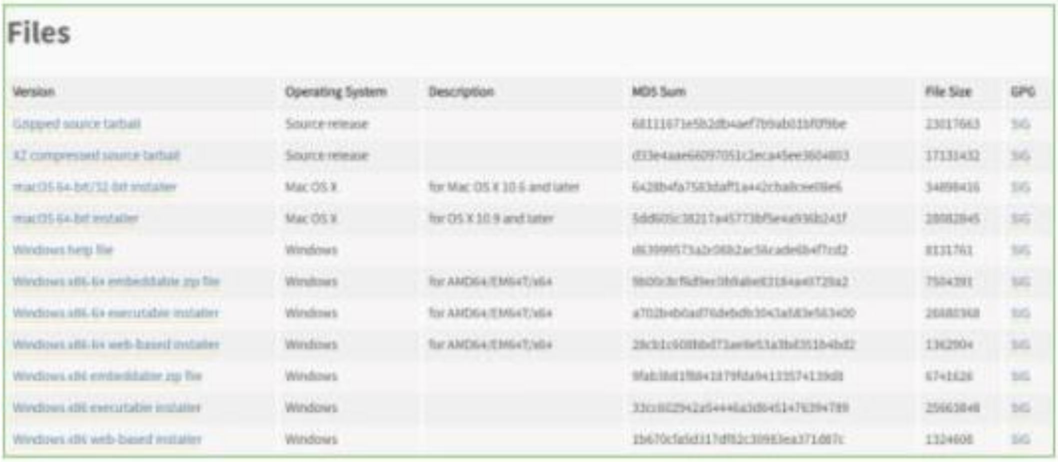
Looking for a specific release?

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 | | 68111671e5b2db4aef7b9ab01bf0f9be | 23017663 | SIG |
| XZ compressed source tarball | Source release | | d33e4aae66097051c2eca45ee3604803 | 17131432 | SIG |
| macOS 64-bit/32-bit installer | Mac OS X | for Mac OS X 10.6 and later | 6428b4fa7583daff1a442cba8cee08e6 | 34898416 | SIG |
| macOS 64-bit installer | Mac OS X | for OS X 10.9 and later | 5dd605c38217a45773bf5e4a936b241f | 28082845 | SIG |
| Windows help file | Windows | | d63999573a2c06b2ac56cade6b4f7cd2 | 8131761 | SIG |
| Windows x86-64 embeddable zip file | Windows | for AMD64/EM64T/x64 | 9b00c8cf6d9ec0b9abe83184a40729a2 | 7504391 | SIG |
| Windows x86-64 executable installer | Windows | for AMD64/EM64T/x64 | a702b4b0ad76debdb3043a583e563400 | 26680368 | SIG |
| Windows x86-64 web-based installer | Windows | for AMD64/EM64T/x64 | 28cb1c608bbd73ae8e53a3bd351b4bd2 | 1362904 | SIG |
| Windows x86 embeddable zip file | Windows | | 9fab3b81f8841879fda94133574139d8 | 6741626 | SIG |
| Windows x86 executable installer | Windows | | 33cc602942a54446a3d6451476394789 | 25663848 | SIG |
| Windows x86 web-based installer | Windows | | 15670cfa5d317df82c30983ea371d87c | 1324608 | SIG |

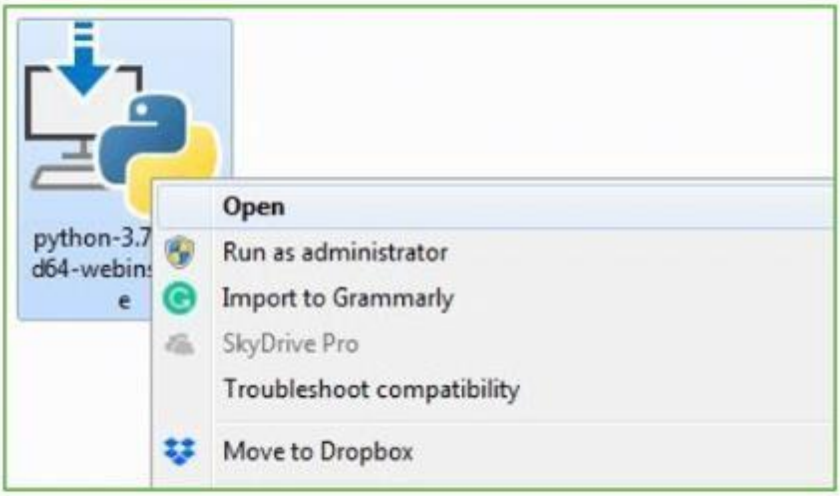
- To download Windows 32-bit python, you can select any one from the three options: Windows x86 embeddable zip file, Windows x86 executable installer or Windows x86 web-based installer.
- To download Windows 64-bit python, you can select any one from the three options: Windows x86-64 embeddable zip file, 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.

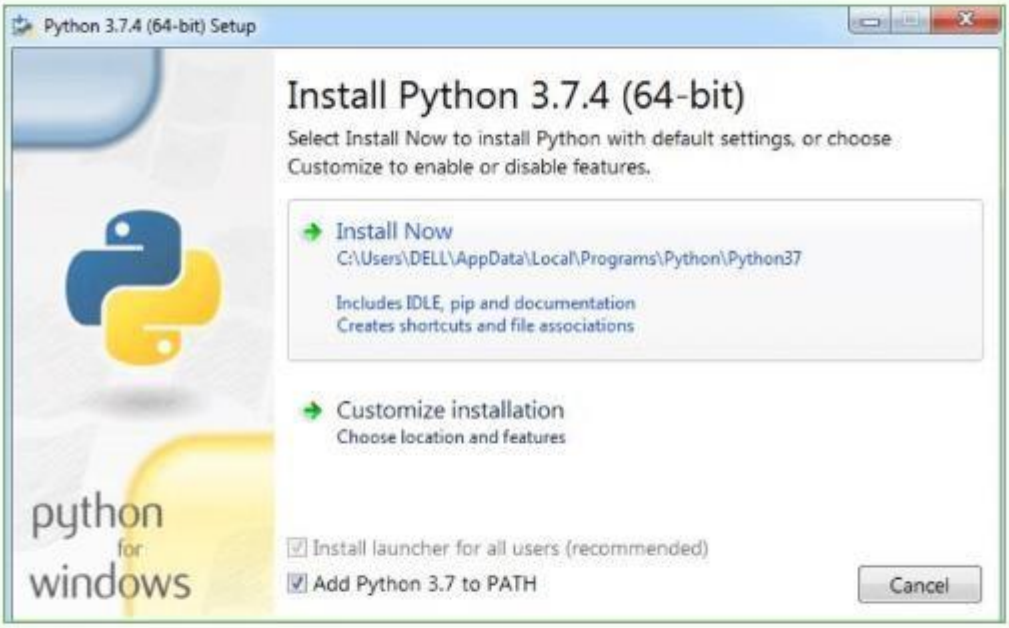
**Note:** To know the changes or updates that are made in the version you can click on the Release Note Option.

## 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.



Python 3.7.4 (64-bit) Setup

Install Python 3.7.4 (64-bit)

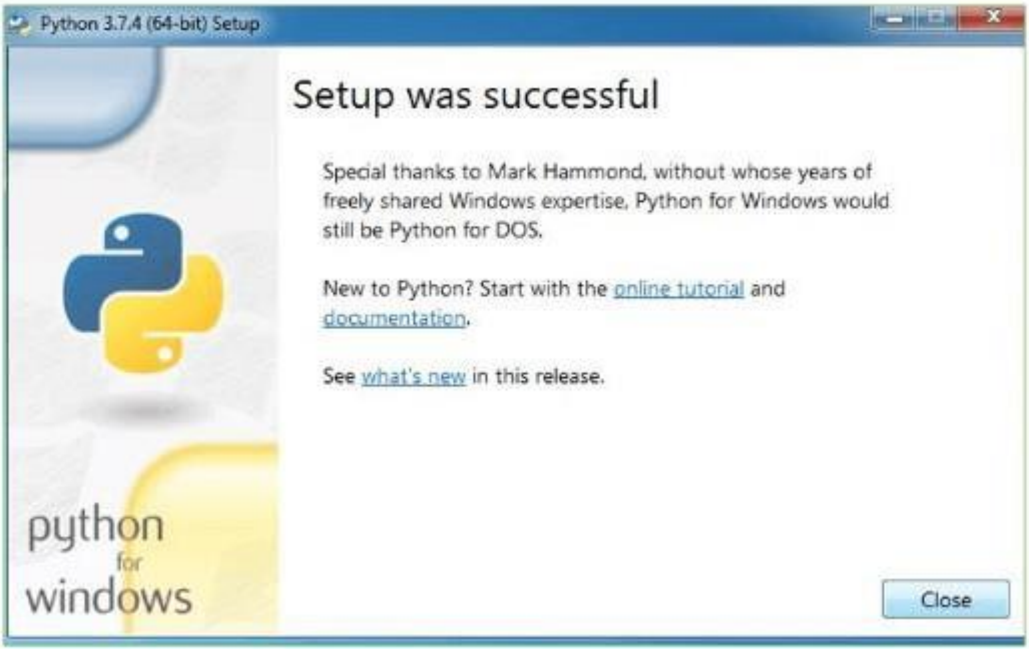
Select Install Now to install Python with default settings, or choose Customize to enable or disable features.

Install Now
C:\Users\DELL\AppData\Local\Programs\Python\Python37
Includes IDLE, pip and documentation
Creates shortcuts and file associations
Customize installation
Choose location and features
python
for
windows
Cancel

☑ Install launcher for all users (recommended)

☑ Add Python 3.7 to PATH

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



Setup was successful

Special thanks to Mark Hammond, without whose years of freely shared Windows expertise, Python for Windows would still be Python for DOS.

New to Python? Start with the online tutorial and documentation.

See what's new in this release.

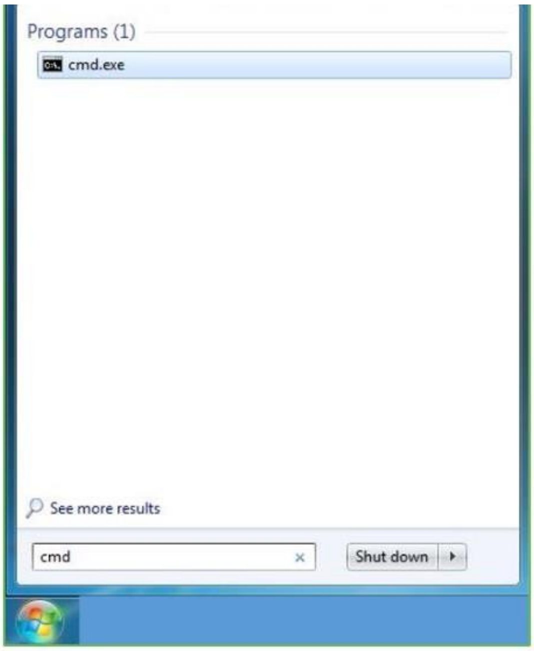
With these above three steps on python installation, you have successfully and correctly installed Python. Now is the time to verify the installation.

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

## Verify the Python Installation

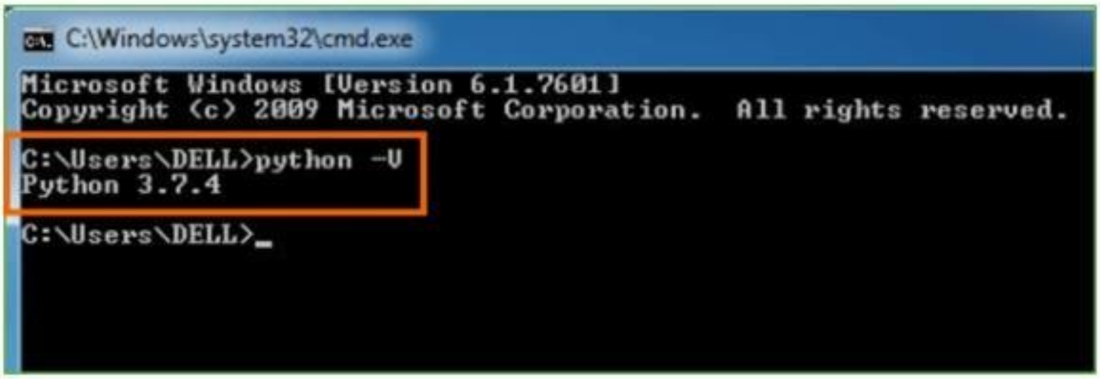
**Step 1:** Click on Start.

**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.



```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

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

C:\Users\DELL>_
```

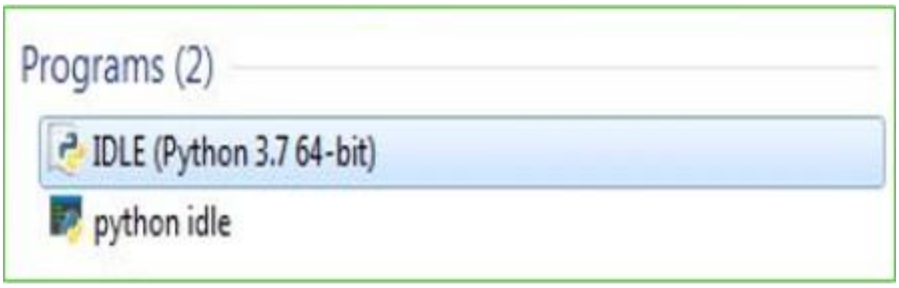
**Step 5:** You will get the answer as 3.7.4.

**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”.



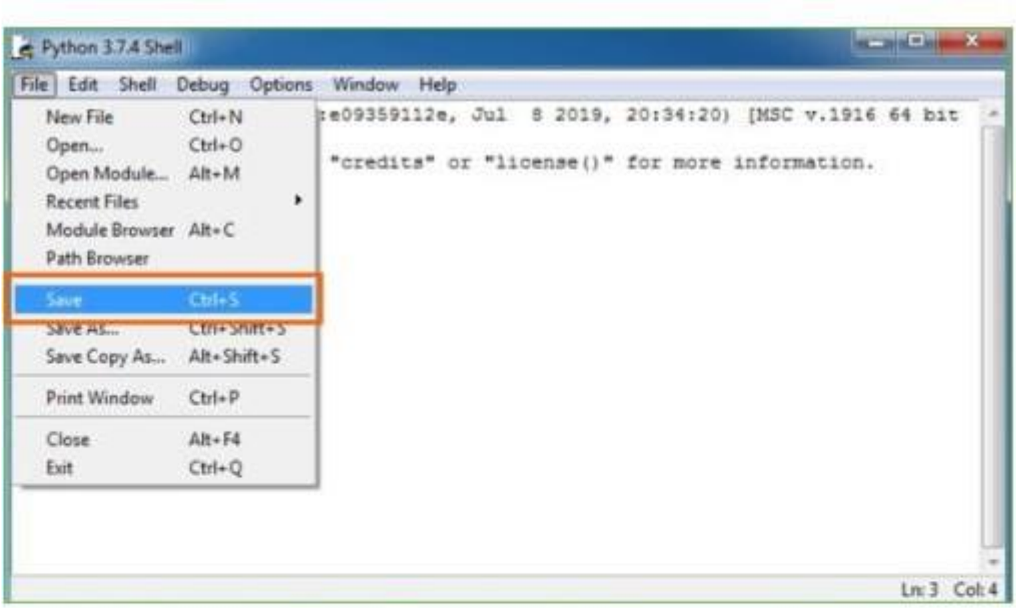
Programs (2)

IDLE (Python 3.7 64-bit)

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.**



```
:e09359112e, Jul  8 2019, 20:34:20) [MSC v.1916 64 bit

 "credits" or "license()" for more information.
```

**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**.

# CHAPTER-6: SYSTEM STUDY

## FEASIBILITY STUDY

A feasibility study assesses the practicality of using machine learning (ML) for detecting thyroid disorders by evaluating economic, technical, and social factors. Here’s an analysis of the feasibility in this context:

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

### 6.1 ECONOMICAL FEASIBILITY

Developing and deploying ML models for thyroid disorder detection requires initial investment in data collection, software development, and hardware or cloud computing resources. Once trained, ML models can offer faster and more cost-effective diagnostics compared to traditional methods, potentially reducing the need for expensive lab tests and hospital visits.

### 6.2 TECHNICAL FEASIBILITY

Machine learning models require large, high-quality datasets, including patient history, lab results and other relevant biomarkers.ML models like neural networks, support vector machines (SVMs), and decision trees have shown promise in detecting thyroid conditions. These models can be trained to identify patterns in medical data that signal the presence of thyroid disorders.

### 6.3 SOCIAL FEASIBILITY

Medical professionals must trust the accuracy and reliability of ML systems. This will require thorough validation of models and education on how to interpret ML-based results. Public perception of AI in healthcare is a significant factor. Building trust through transparent communication, accuracy, and reliability is critical.

# CHAPTER-7: SYSTEM TESTING

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.

## 7.1 TYPES OF TESTS

### 7.1.1 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.

### 7.1.2 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 satisfaction, as shown by successfully 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.

### 7.1.3 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.

### 7.1.4 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.

### 7.1.5 White Box Testing

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

### 7.1.6 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, 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.

# CHAPTER-8: RESULTS

To run project double click on ‘**run.bat**’ file to get below screen.

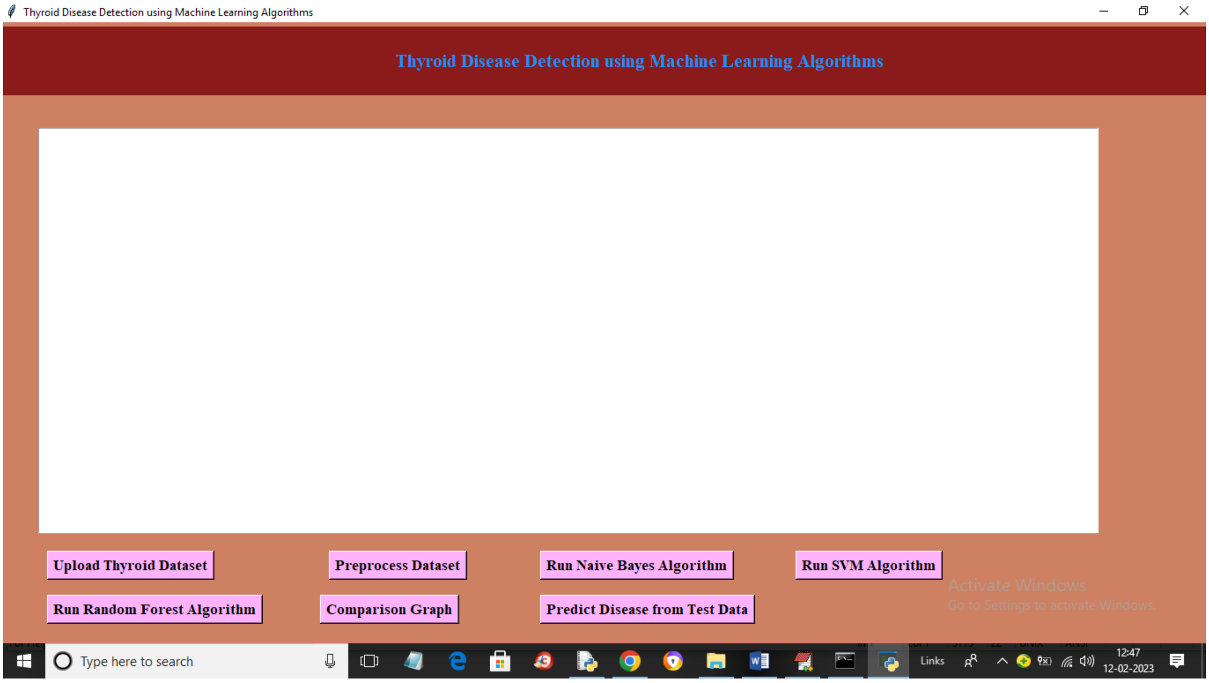
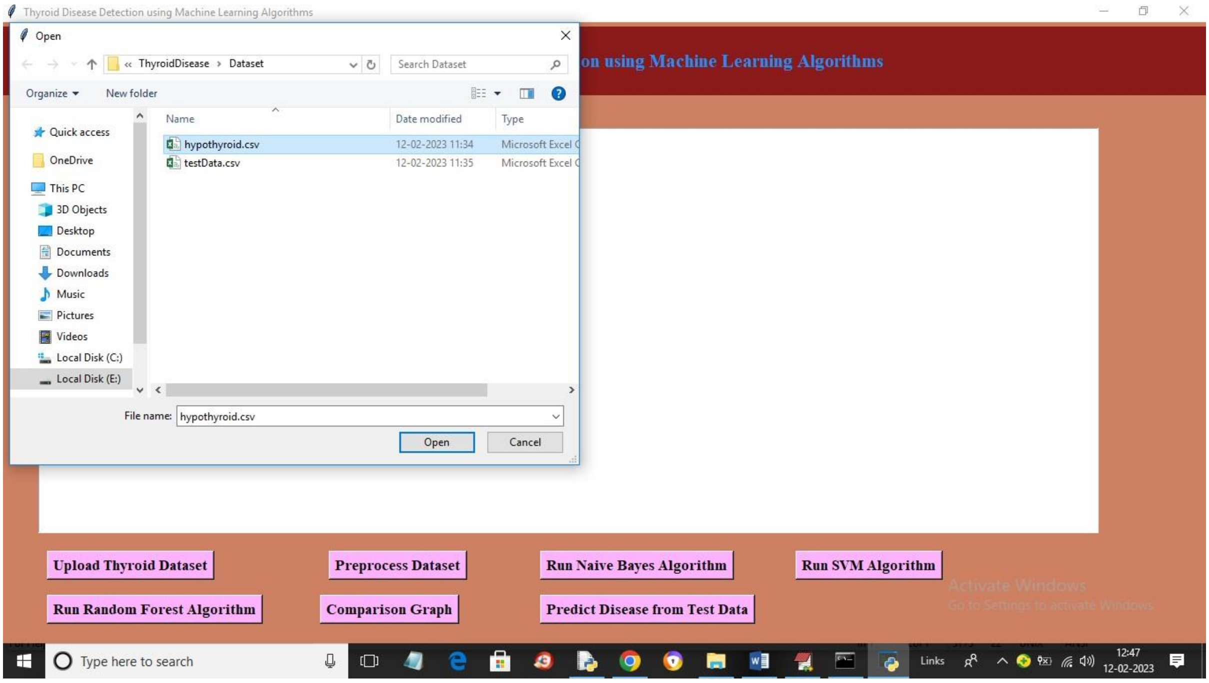

Thyroid Disease Detection using Machine Learning Algorithms
Thyroid Disease Detection using Machine Learning Algorithms
Upload Thyroid Dataset
Preprocess Dataset
Run Naive Bayes Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Comparison Graph
Predict Disease from Test Data
Type here to search
12:47
12-02-2023

Fig.8.1: Screenshot of Home Page

In above screen click on ‘**Upload Thyroid Dataset**’ button to load dataset and get below output.

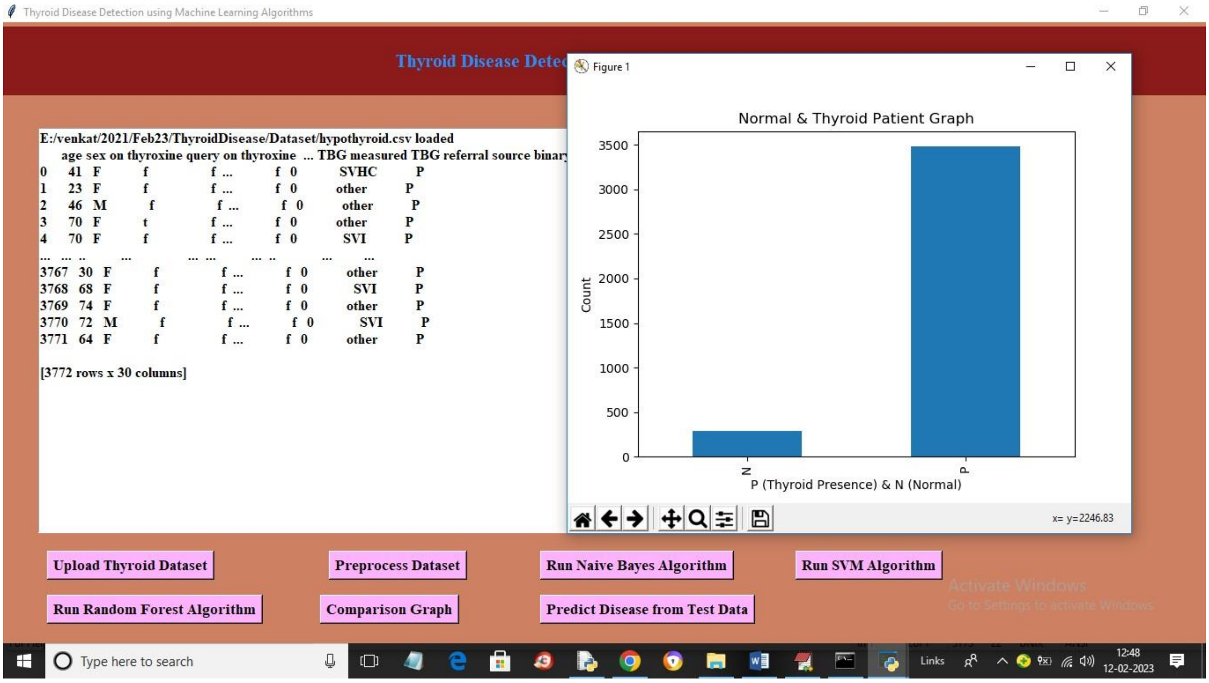


| Name | Date modified | Type |
|---|---|---|
| hypothyroid.csv | 12-02-2023 11:34 | Microsoft Excel C |
| testData.csv | 12-02-2023 11:35 | Microsoft Excel C |

File name: hypothyroid.csv

Fig.8.2: Screenshot of Upload Dataset

In above screen selecting and uploading thyroid dataset and then click on ‘**Open**’ button to load dataset and get below output.

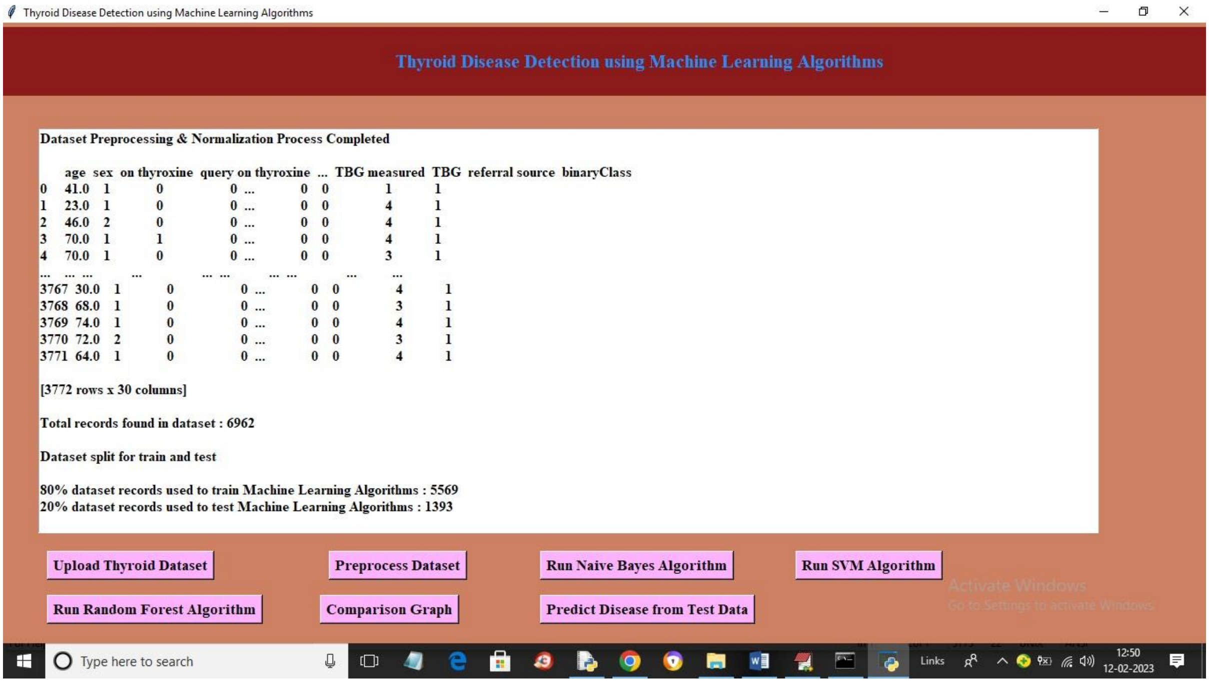


```
E:/venkat/2021/Feb23/ThyroidDisease/Dataset/hypothyroid.csv loaded
     age sex on thyroxine query on thyroxine  ... TBG measured TBG referral source binar
0     41  F          f              f ...        f  0        SVHC        P
1     23  F          f              f ...        f  0       other        P
2     46  M          f              f ...        f  0       other        P
3     70  F          t              f ...        f  0       other        P
4     70  F          f              f ...        f  0         SVI        P
...   ... ..       ...            ... ...      ... ..         ...      ...
3767  30  F          f              f ...        f  0       other        P
3768  68  F          f              f ...        f  0         SVI        P
3769  74  F          f              f ...        f  0       other        P
3770  72  M          f              f ...        f  0         SVI        P
3771  64  F          f              f ...        f  0       other        P

[3772 rows x 30 columns]
```

Thyroid Disease Detection using Machine Learning Algorithms
Thyroid Disease Detec
Figure 1
Normal & Thyroid Patient Graph
Count
3500
3000
2500
2000
1500
1000
500
0
N
P
P (Thyroid Presence) & N (Normal)
x= y=2246.83
Upload Thyroid Dataset
Preprocess Dataset
Run Naive Bayes Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Comparison Graph
Predict Disease from Test Data
Type here to search
Links
12:48
12-02-2023

**Fig.8.3: Screenshot of Loading Dataset**

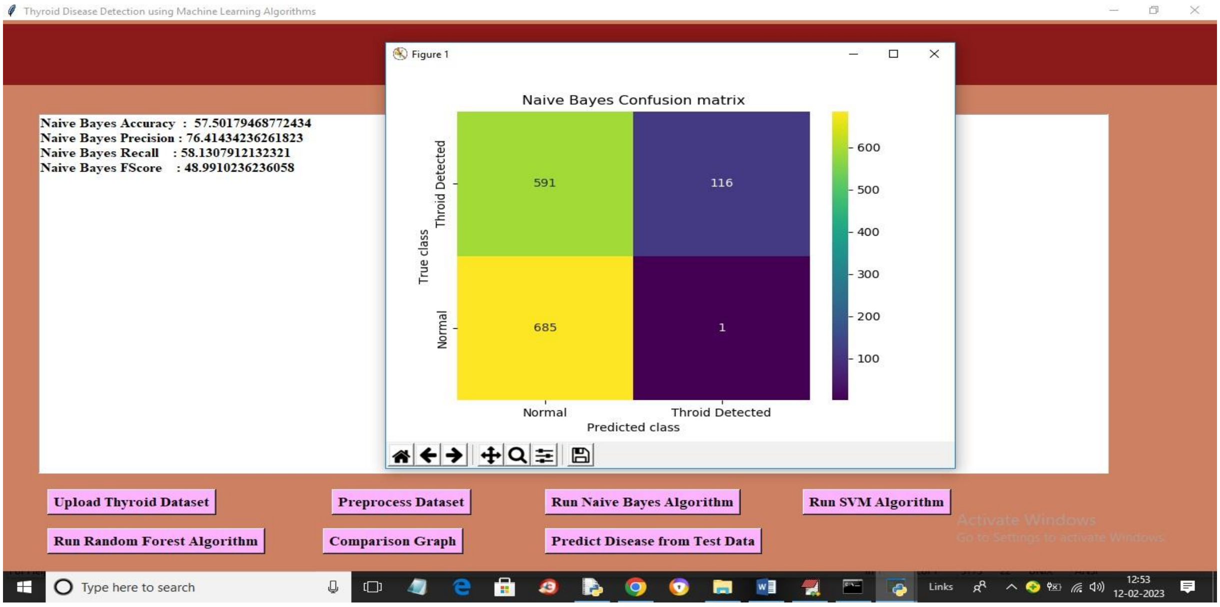
In above screen dataset loaded and in graph x-axis represents N (normal) and P (thyroid presence) and y-axis represents number of records and in above dataset values we can see some are non-numeric and some are numeric and machine learning algorithms accept only numeric values so we need to process dataset to encode non-numeric values to numeric values so click on ‘**Preprocess Dataset**’ button to get below output.


Thyroid Disease Detection using Machine Learning Algorithms
Thyroid Disease Detection using Machine Learning Algorithms
Upload Thyroid Dataset
Preprocess Dataset
Run Naive Bayes Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Comparison Graph
Predict Disease from Test Data
Type here to search
Links
12:50
12-02-2023

```
Dataset Preprocessing & Normalization Process Completed

     age  sex  on thyroxine  query on thyroxine  ...  TBG measured  TBG  referral source  binaryClass
0    41.0   1             0                   0  ...             0    0                1            1
1    23.0   1             0                   0  ...             0    0                4            1
2    46.0   2             0                   0  ...             0    0                4            1
3    70.0   1             1                   0  ...             0    0                4            1
4    70.0   1             0                   0  ...             0    0                3            1
...   ...  ...          ...                 ...  ...           ...  ...              ...          ...
3767  30.0   1             0                   0  ...             0    0                4            1
3768  68.0   1             0                   0  ...             0    0                3            1
3769  74.0   1             0                   0  ...             0    0                4            1
3770  72.0   2             0                   0  ...             0    0                3            1
3771  64.0   1             0                   0  ...             0    0                4            1

[3772 rows x 30 columns]

Total records found in dataset : 6962

Dataset split for train and test

80% dataset records used to train Machine Learning Algorithms : 5569
20% dataset records used to test Machine Learning Algorithms : 1393
```

## Fig.8.4: Screenshot of Preprocess Dataset

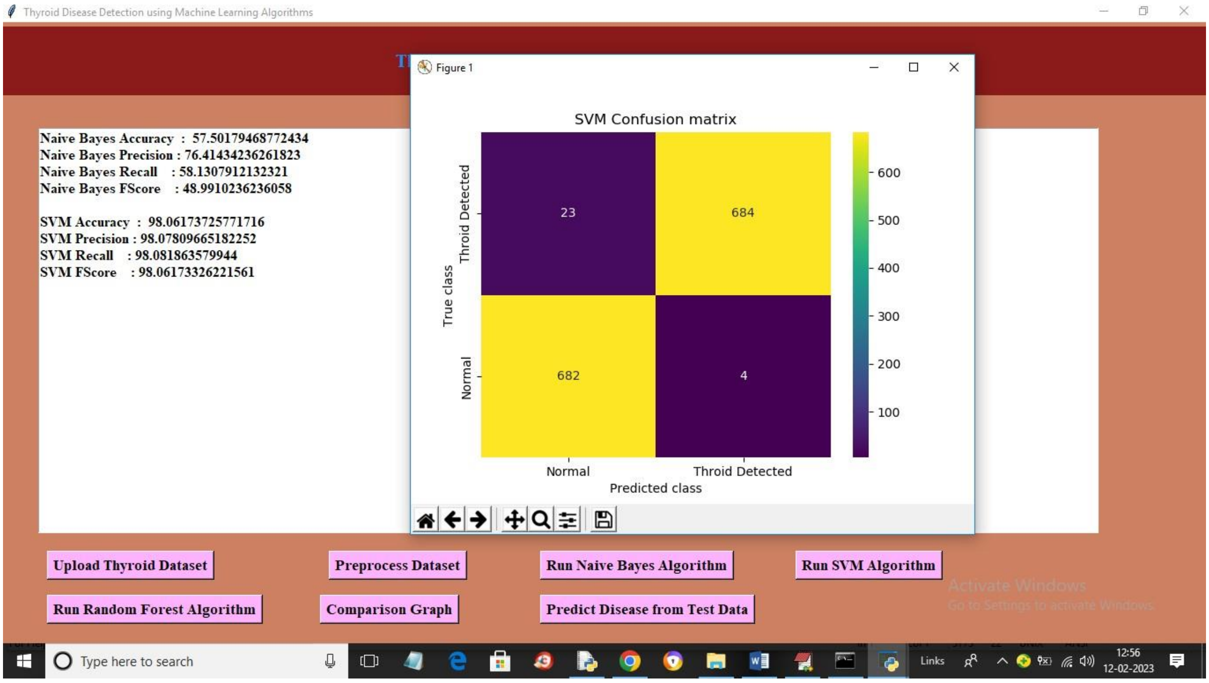
In above screen we can see all values are converted to numeric format and then we can see dataset contains 6962 records where application using 80% records (5569) for training and 1393 (20%) records for testing. Now click on '**Run Naïve Bayes Algorithm**' button to train Naïve Bayes on 80% dataset and test on 20% data to get below prediction accuracy.



```
Naive Bayes Accuracy  :  57.50179468772434
Naive Bayes Precision : 76.41434236261823
Naive Bayes Recall    : 58.1307912132321
Naive Bayes FScore    : 48.9910236236058
```

## Fig.8.5: Screenshot of Run Navie Bayes Algorithm

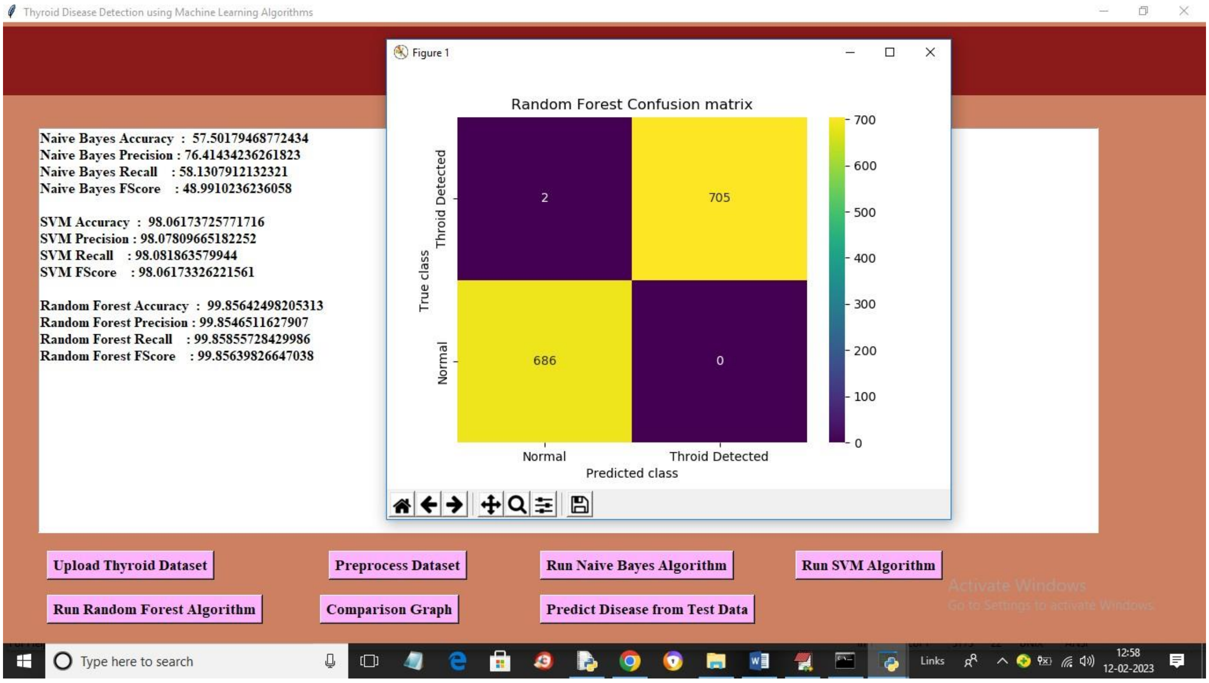
In above screen with Naïve Bayes we got 57% accuracy and we can see other metrics like precision, recall and FSCORE. In confusion matric graph x-axis represents Predicted Labels and y-axis represents True Labels and yellow and light blue color in diagonal represents correct prediction and dark blue and green box contains incorrect prediction count. Now close above graph and then click on ‘**Run SVM Algorithm**’ button to get below output.



```
Naive Bayes Accuracy  :  57.50179468772434
Naive Bayes Precision : 76.41434236261823
Naive Bayes Recall    : 58.1307912132321
Naive Bayes FScore    : 48.9910236236058

SVM Accuracy  :  98.06173725771716
SVM Precision : 98.07809665182252
SVM Recall    : 98.081863579944
SVM FScore    : 98.06173326221561
```

**Fig.8.6: Screenshot of Run SVM Algorithm**

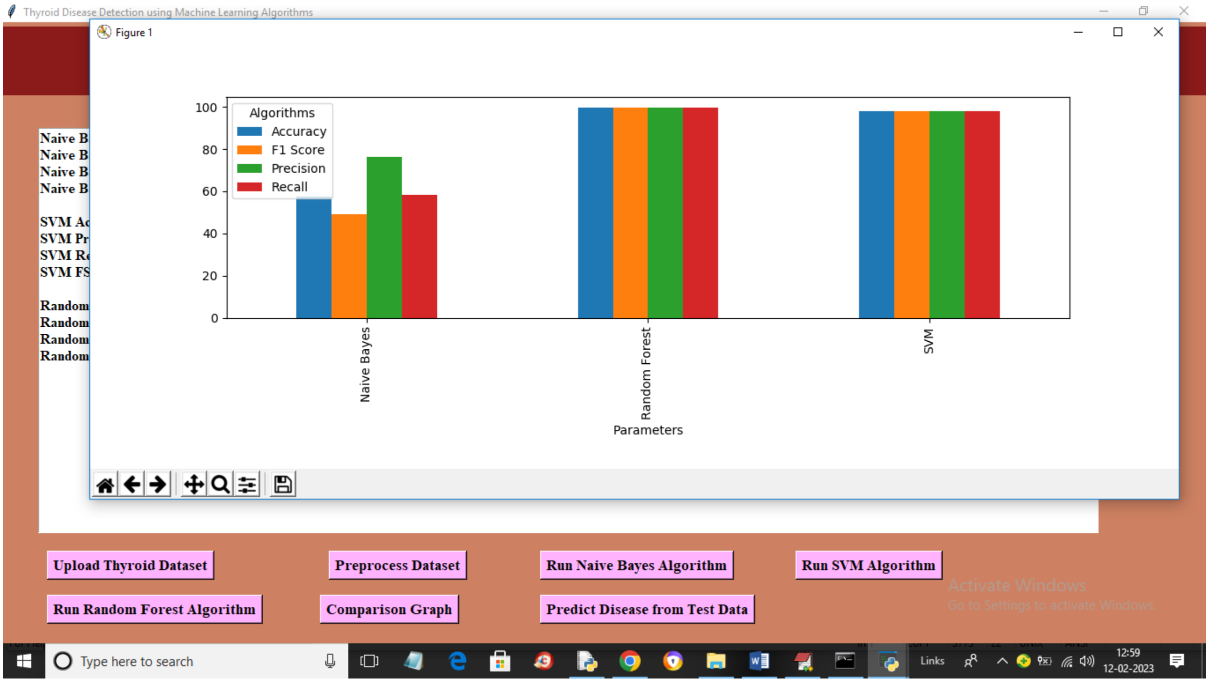
In above screen with SVM we got 98% accuracy and in confusion matrix graph yellow boxes contains correct prediction count and blue boxes contains incorrect prediction count. Now close above graph and then click on ‘**Run Random Forest Algorithm**’ button to get below output.



```
Naive Bayes Accuracy  :  57.50179468772434
Naive Bayes Precision : 76.41434236261823
Naive Bayes Recall    : 58.1307912132321
Naive Bayes FScore    : 48.9910236236058

SVM Accuracy  :  98.06173725771716
SVM Precision : 98.07809665182252
SVM Recall    : 98.081863579944
SVM FScore    : 98.06173326221561

Random Forest Accuracy  :  99.85642498205313
Random Forest Precision : 99.8546511627907
Random Forest Recall    : 99.85855728429986
Random Forest FScore    : 99.85639826647038
```

## Fig.8.7: Screenshot of Run Random Forest Algorithm

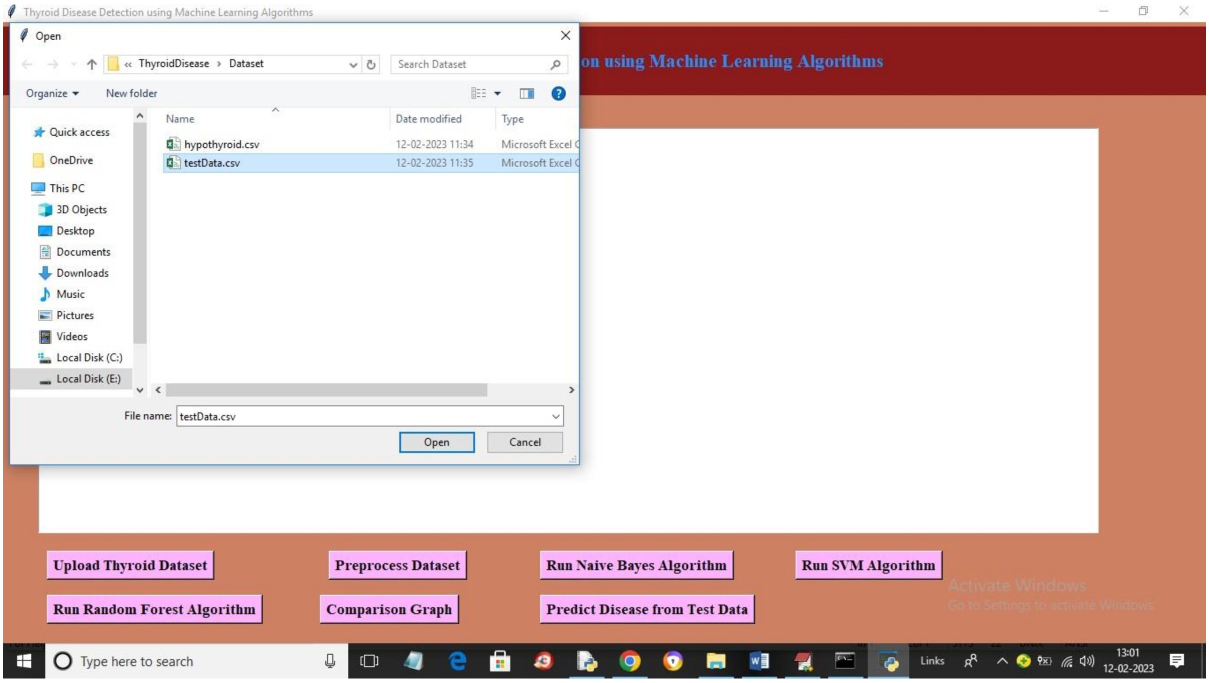
In above screen with Random Forest we got 99% accuracy and now click on ‘**Comparison Graph**’ button to get below graph.



```
Naive B
Naive B
Naive B
Naive B

SVM Ac
SVM Pr
SVM Re
SVM FS

Random
Random
Random
Random
```

## Fig.8.8: Screenshot of Comparison Graph

In above graph x-axis represents algorithm names and y-axis represents metric values like accuracy, precision, recall in different bar color and in all algorithms Random Forest got high accuracy. Now close above graph and then click on ‘**Predict Disease from Test Data**’ to upload test data and get prediction output.

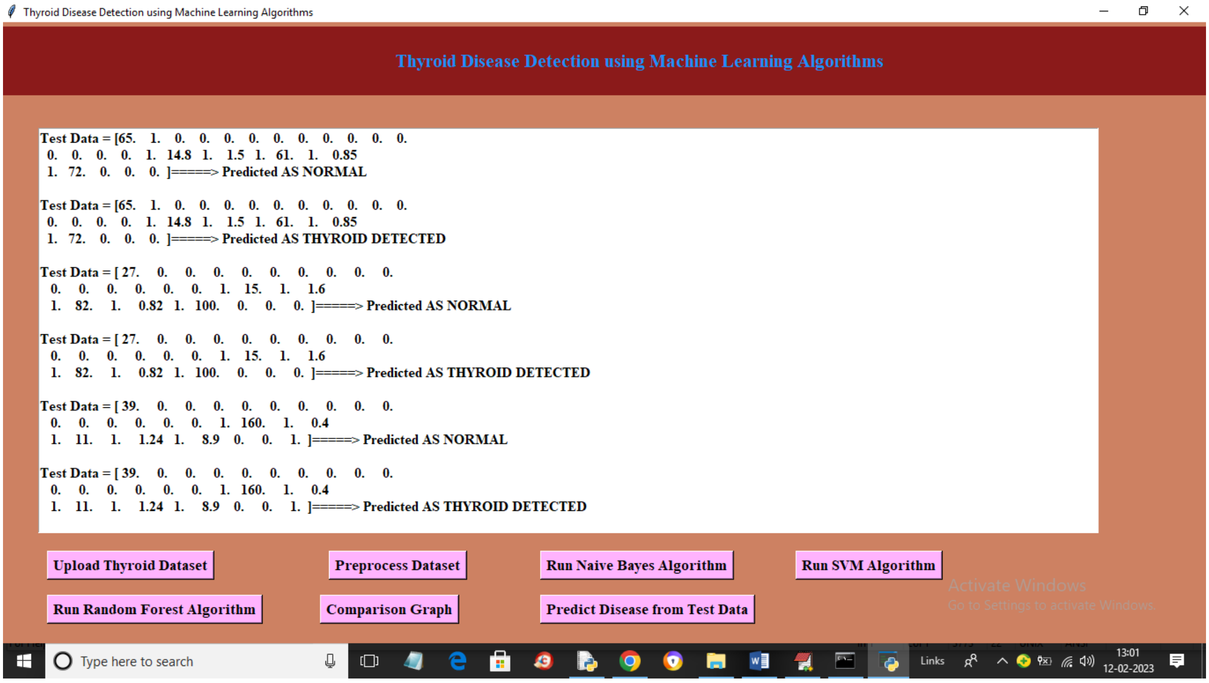


```
hypothyroid.csv    12-02-2023 11:34    Microsoft Excel C
testData.csv       12-02-2023 11:35    Microsoft Excel C
```

File name: testData.csv

## Fig.8.9: Screenshot of Predict Disease from Test Data

In above screen selecting and uploading testData.csv file and then click on ‘**Open**’ button to get below output.



Thyroid Disease Detection using Machine Learning Algorithms
Thyroid Disease Detection using Machine Learning Algorithms
Upload Thyroid Dataset
Preprocess Dataset
Run Naive Bayes Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Comparison Graph
Predict Disease from Test Data
Activate Windows
Go to Settings to activate Windows.
Type here to search
Links
13:01
12-02-2023

```
Test Data = [65.   1.   0.   0.   0.   0.   0.   0.   0.   0.   0.   0.
  0.   0.   0.   0.   1.  14.8  1.   1.5  1.  61.   1.   0.85
  1.  72.   0.   0.   0.  ]=====> Predicted AS NORMAL

Test Data = [65.   1.   0.   0.   0.   0.   0.   0.   0.   0.   0.   0.
  0.   0.   0.   0.   1.  14.8  1.   1.5  1.  61.   1.   0.85
  1.  72.   0.   0.   0.  ]=====> Predicted AS THYROID DETECTED

Test Data = [ 27.    0.    0.    0.    0.    0.    0.    0.    0.    0.
   0.    0.    0.    0.    0.    0.    1.   15.    1.    1.6
   1.   82.    1.    0.82  1.  100.    0.    0.    0.  ]=====> Predicted AS NORMAL

Test Data = [ 27.    0.    0.    0.    0.    0.    0.    0.    0.    0.
   0.    0.    0.    0.    0.    0.    1.   15.    1.    1.6
   1.   82.    1.    0.82  1.  100.    0.    0.    0.  ]=====> Predicted AS THYROID DETECTED

Test Data = [ 39.    0.    0.    0.    0.    0.    0.    0.    0.    0.
   0.    0.    0.    0.    0.    0.    1.  160.    1.    0.4
   1.   11.    1.    1.24  1.    8.9   0.    0.    1.  ]=====> Predicted AS NORMAL

Test Data = [ 39.    0.    0.    0.    0.    0.    0.    0.    0.    0.
   0.    0.    0.    0.    0.    0.    1.  160.    1.    0.4
   1.   11.    1.    1.24  1.    8.9   0.    0.    1.  ]=====> Predicted AS THYROID DETECTED
```

## Fig.8.10: Screenshot of Predicted Data

In above screen in square bracket we can see test data and after arrow symbol we can see predicted output

# CHAPTER 9: CONCLUSION & FUTURE ENHANCEMENT

Thus the proposed work will be very much useful to identify the thyroid disease in a patient at an early stage using classification based machine learning techniques. These algorithms give various levels of precision and accuracy. These methods also aid in decreasing the unwanted redundant data from the patient’s database. The algorithms used in the proposed model are cost effective and also have good output performance and speed. These classification methods make the treatment of the thyroid patient simple by reducing further complex procedures with an affordable price.

## FUTUREENHANCEMENT:

Various activity patterns are studied and analyzed to detect anomaly behaviors. Anomaly detection from the travel log or trajectory data is a more promising process nowadays. This is very challenging due to several reasons such as mobility patterns are unique in every user; the trajectories are dynamic and need frequent updating. Detecting outlier/ anomaly from those dynamic and updatable datasets need more concentration. The techniques should be developed carefully. Surveyed various applications related to the public transit records such as identifying pickpocket suspects, daily activity pattern detection, bus route planning, traffic abnormality detection, etc. From this summary; a new technique canbe developed and integrated into the recent applications.

# CHAPTER-10: BIBLIOGRAPHY

## 1. Research Papers:

"Machine Learning Approaches for Diagnosis of Thyroid Disorders" – Explore various ML algorithms and their effectiveness in thyroid disease detection.

"A Comparative Study of Machine Learning Techniques for Predicting Thyroid Disease" – An analysis of different models (e.g., SVM, Random Forest) applied to thyroid datasets.

## 2. Datasets:

UCI Machine Learning Repository: Look for the "Thyroid Disease" dataset which provides clinical features for analysis.

Kaggle Datasets: Search for thyroid-related datasets, often containing patient data with various features and outcomes.

## 3. Books:

"Pattern Recognition and Machine Learning" by Christopher Bishop – A foundational text that covers machine learning techniques applicable to medical diagnosis.

1. "Machine Learning for Healthcare" by R. S. D. P. K. M. D. M. and H. P. P.

Offers insights into machine learning techniques specifically tailored for healthcare applications.

2. "Deep Learning for Medical Image Analysis" by Geert Litjens et al.

While focused on medical imaging, it provides a comprehensive understanding of deep learning approaches that can be applied to diagnostic tasks, including thyroid conditions.

3. "Applied Machine Learning for Healthcare" by A. M. K. and A. P.

Focuses on practical implementations of machine learning in healthcare settings, including various disease predictions.

4. "Artificial Intelligence in Healthcare: Anticipating Challenges to Ethics, Privacy, and Bias" by Adam Bohr and Kaveh Memarzadeh.

## 4. Review Articles:

"A Review of Machine Learning Techniques for Thyroid Disease Detection" – Summarizes the state of research in this field and highlights effective methods and challenges.

"Predictive Modeling of Thyroid Disease Using Machine Learning Algorithms" – A study that compares various ML algorithms in predicting thyroid conditions.

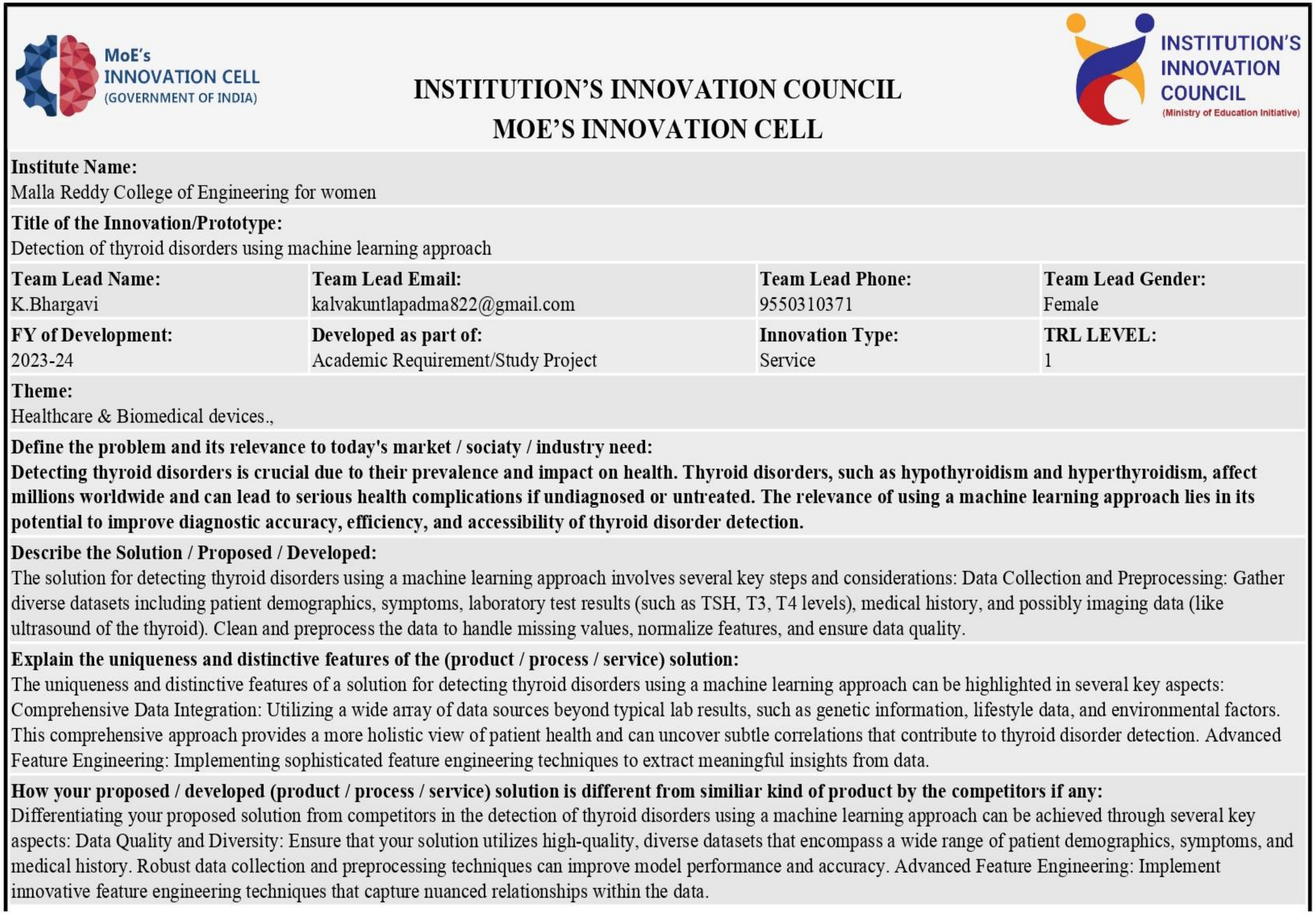
"Deep Learning in Thyroid Disease: A Review" – Reviews the application of deep learning techniques in diagnosing thyroid disorders.

## 5. Journals:

Journals like "IEEE Transactions on Biomedical Engineering" or "Journal of Medical Systems" often publish relevant articles on ML applications in healthcare.

## 6. Online Resources:

Courses or tutorials on platforms like Coursera or edX that focus on machine learning in healthcare.



## INSTITUTION'S INNOVATION COUNCIL
## MOE'S INNOVATION CELL

**Institute Name:**
Malla Reddy College of Engineering for women

**Title of the Innovation/Prototype:**
Detection of thyroid disorders using machine learning approach

| | | | |
|---|---|---|---|
| **Team Lead Name:**<br>K.Bhargavi | **Team Lead Email:**<br>kalvakuntlapadma822@gmail.com | **Team Lead Phone:**<br>9550310371 | **Team Lead Gender:**<br>Female |
| **FY of Development:**<br>2023-24 | **Developed as part of:**<br>Academic Requirement/Study Project | **Innovation Type:**<br>Service | **TRL LEVEL:**<br>1 |

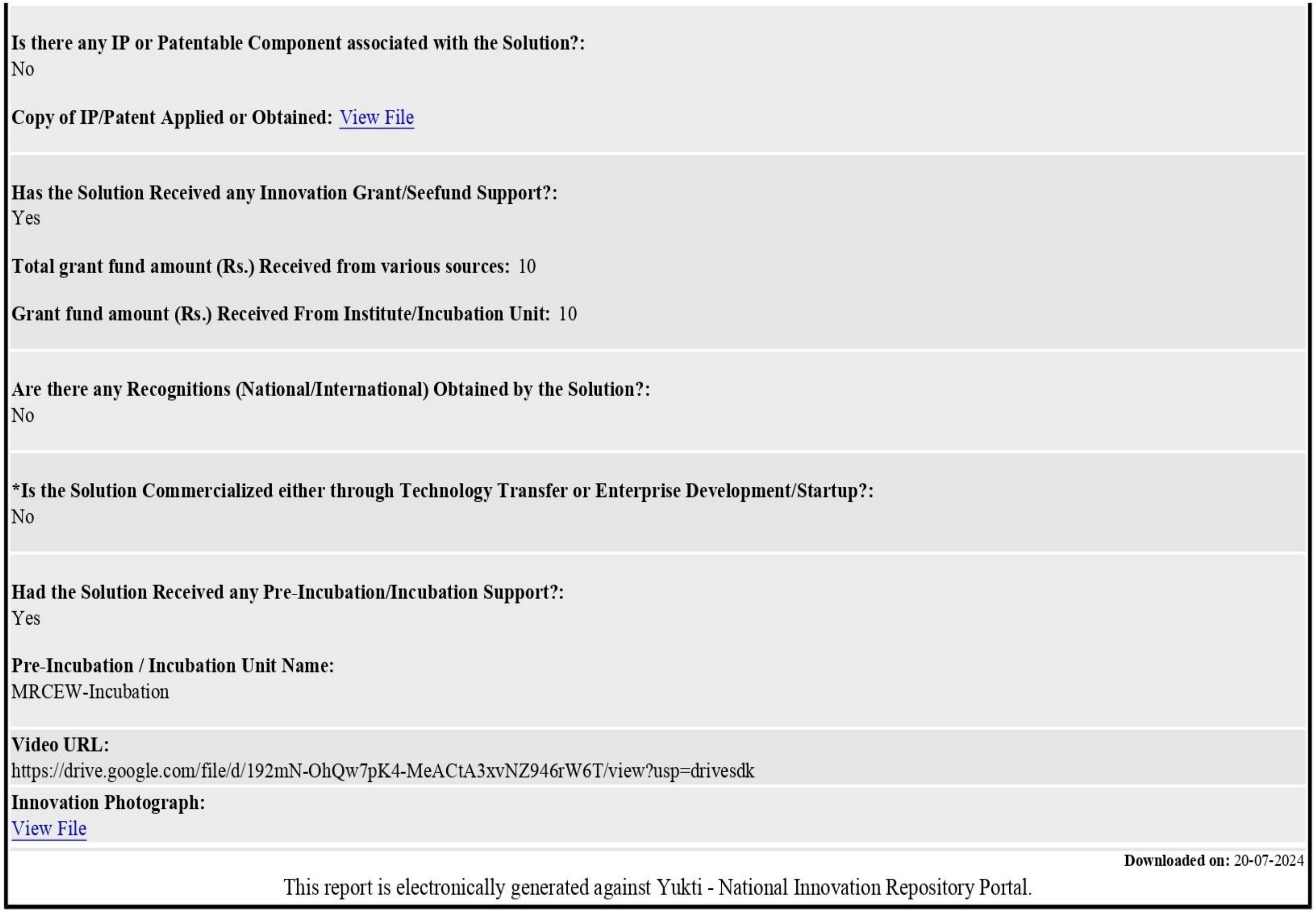
**Theme:**
Healthcare & Biomedical devices.,

**Define the problem and its relevance to today's market / sociaty / industry need:**
**Detecting thyroid disorders is crucial due to their prevalence and impact on health. Thyroid disorders, such as hypothyroidism and hyperthyroidism, affect millions worldwide and can lead to serious health complications if undiagnosed or untreated. The relevance of using a machine learning approach lies in its potential to improve diagnostic accuracy, efficiency, and accessibility of thyroid disorder detection.**

**Describe the Solution / Proposed / Developed:**
The solution for detecting thyroid disorders using a machine learning approach involves several key steps and considerations: Data Collection and Preprocessing: Gather diverse datasets including patient demographics, symptoms, laboratory test results (such as TSH, T3, T4 levels), medical history, and possibly imaging data (like ultrasound of the thyroid). Clean and preprocess the data to handle missing values, normalize features, and ensure data quality.

**Explain the uniqueness and distinctive features of the (product / process / service) solution:**
The uniqueness and distinctive features of a solution for detecting thyroid disorders using a machine learning approach can be highlighted in several key aspects: Comprehensive Data Integration: Utilizing a wide array of data sources beyond typical lab results, such as genetic information, lifestyle data, and environmental factors. This comprehensive approach provides a more holistic view of patient health and can uncover subtle correlations that contribute to thyroid disorder detection. Advanced Feature Engineering: Implementing sophisticated feature engineering techniques to extract meaningful insights from data.

**How your proposed / developed (product / process / service) solution is different from similiar kind of product by the competitors if any:**
Differentiating your proposed solution from competitors in the detection of thyroid disorders using a machine learning approach can be achieved through several key aspects: Data Quality and Diversity: Ensure that your solution utilizes high-quality, diverse datasets that encompass a wide range of patient demographics, symptoms, and medical history. Robust data collection and preprocessing techniques can improve model performance and accuracy. Advanced Feature Engineering: Implement innovative feature engineering techniques that capture nuanced relationships within the data.



**Is there any IP or Patentable Component associated with the Solution?:**
No

**Copy of IP/Patent Applied or Obtained:** View File

**Has the Solution Received any Innovation Grant/Seefund Support?:**
Yes

**Total grant fund amount (Rs.) Received from various sources:** 10

**Grant fund amount (Rs.) Received From Institute/Incubation Unit:** 10

**Are there any Recognitions (National/International) Obtained by the Solution?:**
No

***Is the Solution Commercialized either through Technology Transfer or Enterprise Development/Startup?:**
No

**Had the Solution Received any Pre-Incubation/Incubation Support?:**
Yes

**Pre-Incubation / Incubation Unit Name:**
MRCEW-Incubation

**Video URL:**
https://drive.google.com/file/d/192mN-OhQw7pK4-MeACtA3xvNZ946rW6T/view?usp=drivesdk

**Innovation Photograph:**
View File

**Downloaded on:** 20-07-2024

This report is electronically generated against Yukti - National Innovation Repository Portal.