"Detection Of Obesity Using machine Learning"

Submitted in partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer Engineering

Submitted To

Dr. Babasaheb Ambedkar Technological University, Lonere (M.S.)

Submitted By:

Ms. Borase Shraddha Gokul	**PRN: T2051641191014**
Ms. Bhoi Vaishnavi Deepak	**PRN: T2051641245051**
Ms. Patil Vaishnavi Arun	**PRN: T2051641245049**
Mr. Patel Umesh Sharad	**PRN: T2051641245044**

Under the Guidance of

Prof. A. P. Khan

DEPARTMENT OF COMPUTER ENGINEERING

P.S.G.V.P. MANDAL'S
D.N.PATEL COLLEGE OF ENGINEERING
SHAHADA, DIST- NANDURBAR (M.S.)

YEAR 2023-2024

P.S.G.V.P. MANDAL'S

D. N. PATEL COLLEGE OF ENGINEERING

SHAHADA, DIST- NANDURBAR (M.S.)

CERTIFICATE

This is to certify that

Ms. Borase Shraddha Gokul	***PRN: T2051641191014***
Ms. Bhoi Vaishnavi Deepak	***PRN: T2051641245051***
Ms. Patil Vaishnavi Arun	***PRN: T2051641245049***
Mr. Patel Umesh Sharad	***PRN: T2051641245044***

Has satisfactorily completed Project Phase-II entitled

"Detection Of obesity using Machine Learning"

As prescribed by Dr. Babasaheb Ambedkar Technological University, Lonere (M.S.) as a part of syllabus for the partial fulfillment in Bachelor of Technology in Computer Engineering for Academic Year 2023-2024.

GUIDE	**H.O.D.**
Prof. A. P. Khan	Prof. V.S. Mahajan
EXAMINER	**PRINCIPAL**
____________	Prof. Dr. N.J. Patil

P.S.G.V.P. MANDAL'S

D. N. PATEL COLLEGE OF ENGINEERING

SHAHADA, DIST- NANDURBAR (M.S.)

TESTING REPORT

This is to certify that the system

"Detection Of Obesity Using Machine learning"

Developed by

Ms. Borase Shraddha Gokul	***PRN: T2051641191014***
Ms. Bhoi Vaishnavi Deepak	***PRN: T2051641245051***
Ms. Patil Vaishnavi Arun	***PRN: T2051641245049***
Mr. Patel Umesh Sharad	***PRN: T2051641245044***
Ms. Borase Shraddha Gokul	***PRN: T2051641191014***

has been successfully tested and is operating as per the specifications.

Date : ___/___/2024

Place: Shahada

GUIDE	**PROJECT INCHARGE**
Prof. A. P. khan	Prof. V.S. Mahajan/Prof. S.H. Shaikh

H.O.D.

Prof. V.S. Mahajan

TABLE OF CONTENTS

Chapter No.	Content	Page No.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENT

The Acknowledgement is just like a drop in the ocean of the deep sense of gratitude within our heart for people who helped us out of the most embarrassing part of our life when we were standing at the most difficult step towards our dream of life.

Many people have contributed to the success of this. Although a single sentence hardly suffices, we would like to thank Almighty God for blessing us with His grace. We extend our sincere and heartfelt thanks to **Prof. V.S. Mahajan**, Head of Department, Computer Engineering, for providing us with the right ambience for carrying out this work.

We are grateful and sincerely appreciate the effort of our respected in-charge **Prof. V.S. Mahajan & Prof. S.H. Shaikh** who acted as a fulcrum for us and supported us during the ups and downs of our project. We are profoundly indebted to our project guide **Prof. A. P. Khan** for innumerable acts of timely advice, encouragement and we sincerely express our gratitude to him.

We express our immense pleasure and thankfulness to Principal
Prof. Dr. N.J. Patil and all the teachers and staff of the Department of Computer Engineering for their cooperation and support.

Ms. Borase Shraddha Gokul
Ms. Bhoi Vaishnavi Deepak
Ms. Patil Vaishnavi Arun
Mr. Patel Umesh Sharad

ABSTRACT

Obesity has become a worldwide health concern as it is becoming a threat to the future. Thousands of diseases as well as risks and death are associated to it. A near prediction of a disease will help both doctors and patients to act and minimize if not total eradication of the root cause or work on preventing the disease symptom from further deterioration. Going through a patient's medical history is one of the methods of identifying a disease which is most time consuming as processing manually and it comes with error-prone analyses and expense. Therefore, there is need to scientifically develop a predicting model of the occurrence of the disease or its existence using an automated technique as it is becoming a need of the day. In this research work, we used machine learning techniques on a public clinical available dataset to predict obesity status using different machine learning algorithms.

Keywords: *Obesity, Machine Learning, BMI, Overweight, Random Forest Classifier, Decision Tree Classifier, Support Vector Machine and Algorithm, Naïve Bayes*.

1. INTRODUCTION

1.1. Introduction to Project Domain

Machine Learning Obesity detection using machine learning is an important application in the health care domain. Machine learning models can be used to predict and diagnose obesity based on various input data, including patient characteristics, medical history, and other relevant features. Obesity classification refers to the categorization of individuals based on their body weight and fat distribution. It is an important concept in healthcare and research because obesity is associated with various health risks and can inform treatment and prevention strategies. There are several methods and criteria used to classify obesity, with the most common being based on body mass index (BMI) and waist circumference.

The main concern of this paper is to analyze people for obesity and make them aware of the obesity risk factor. This paper aims to predict the obesity risk. The analysis is conducted into two parts where firstly it read the data and then checks the data if it matches the factor with obesity, and then it will show the result. For our analysis, first, we collect raw data sets for our analysis depend on some factors. In addition, we pre process those data, then we applied machine learning supervised algorithms to check the accuracy, sensitivity, specificity, precision, recall. Then we found which algorithm works more optimal and detected the actual outcome.

This work aimed at developing a machine learning model that would be an improved version of the existing model on obesity prediction using a local dataset with obesity principal components. And the objectives are:

i. To determine the best set of features for obesity prediction.
ii. To develop a model that will predict obesity of individuals using machine learning techniques with locally available dataset.
iii. To use supervised machine-learning algorithm to enhance the obesity level estimation for future prediction.
iv. Test and evaluate the performance of the model using the evaluation metrics.

1.2. Problem Definition

Obesity is a complex medical condition characterized by the excessive accumulation of body fat, which can have serious implications for an individual's health and well being. To address the growing concern of obesity and its associated health risks, there is a need to develop an effective classification system that can accurately identify and categorize individuals into different obesity classes based on their body mass index and other relevant features. Based on this perception, it is stated that the describes the word obesity as an abnormal or excess fat accumulation that may damage health. Lots of people above 16 years of age are suffering from alteration of their weight. This overweight can be caused by lots of intake of starchy foods that are high in fat. More also it is caused by lack of physical exercise. According to obesity is a general health problem around the world and it can occur in teens, children and adults.

1.3. Existing Systems

In the early years, traditional and clinical procedures are being use in predicting obesity and this can be time-consuming, as it demands the employment of trained physicians in the processes needed to obtain diagnosed result. Most times, complications are due to late intervention as manual methods are mostly carried out when symptoms has manifested in patients. In recently, various researches have been carried out in the area of obesity prediction using machine learning techniques. These techniques being employ by different researchers produced varying results of different accuracy depending on its challenges as its related to the individual. To diagnosed obesity, physical exams are performed on the patient, then tests are recommended by experts, these test and examinations generally include; taking the patient health history, generally examining the patient physically (checking height, blood pressure, heart rate and temperature), calculating the body mass index (BMI), measuring the waist circumference, lifestyle, hereditary background and checking for other problems. These processes require a high level expertise and lots of time.

1.4. Objectives of Proposed System

The proposed model will be build using the Python programming language, applying machine learning algorithm on the dataset collected from a healthcare center to model a system that will be capable of predicting obesity in patients with dataset that has been technically automated to suit the purpose of this research.

1.5. Proposed Methodology

- SVM is a supervised learning algorithm used for both classification and regression tasks.
- SVM aims to find an optimal hyper-plane that separates the data points of different classes or predicts continuous target values.
- Decision Trees are versatile supervised learning algorithms used for classification and regression tasks.
- Decision Trees create a tree-like model of decisions and their possible consequences based on the features in the input data.
- Random Forest is a popular machine learning algorithm used for classification and regression tasks due to its high accuracy.
- Data Collection: Gather a representative dataset of individuals that includes features relevant to obesity classification. This might include age, gender, weight, height, body measure men, and any other relevant factors.
- Data Preprocessing: Clean the dataset by handling missing values, removing outliers, and normalizing or standardizing features as needed
- Feature Extraction: If the dataset contains many features, consider using feature selection or feature extraction techniques to reduce dimensionality and focus on the most important variables.
- Clustering Algorithms Selection: Choose the clustering algorithms you want to compar.

1.5.1. System Architecture

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

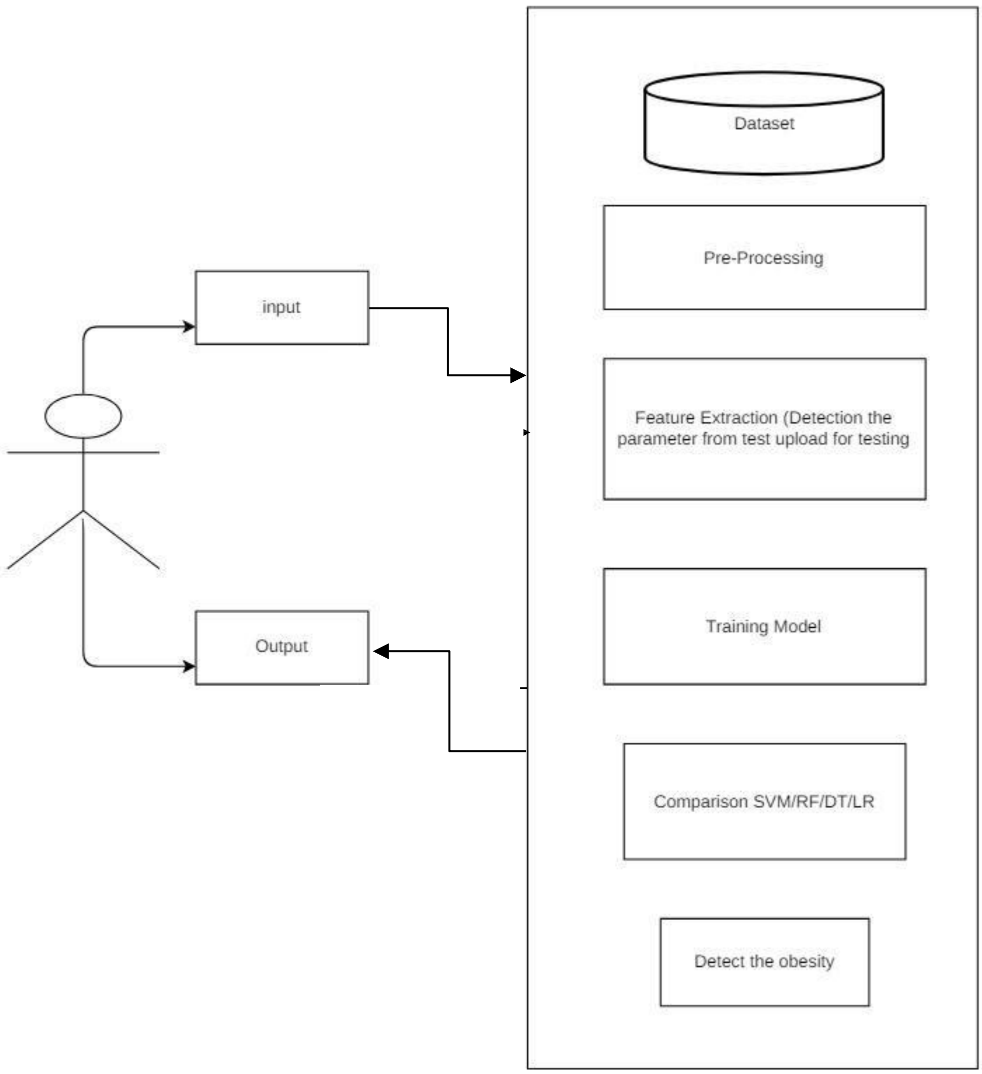

Figure 1.1. System Architecture of ODS

1.5.2. System Modules

1.5.2.1. User Information Module:

This holds the information of the patient or user which consists of details of the features being exhibited. The features details includes full name, gender, age, family history, physiological, cholesterol level, blood pressure level, body mass index and life style of the patient.

1.5.2.2. Prediction Module:

This is the model's core module; it takes the patient's details features from the user information module and analyzes it using the created model, producing a prediction result that would be sending to the result analysis module

1.5.2.3. Result Analysis Module:

At this module, the outcome of the prediction from the prediction module is interpreted into an understandable format. The patient's overall features will be analyzed and cclarifications are given to identify the level at its risk. The interpreted details are sent to the result display module.

1.5.2.4. Result Display Module:

The patient's personal information (full name, age and gender) from the user information module, the diagnosis outcome from the prediction module along with the remarks on the individual features details from the analysis module are received at the result display module to give output on the screen.

1.6. Applicability

Machine learning models can predict an individual's risk of developing obesity, allowing for proactive interventions and preventive measures. Machine learning models can analyze large healthcare datasets to support epidemiological studies, identify trends in obesity prevalence, and aid in public health research. By helping individuals manage their obesity effectively, these systems can improve the quality of life for those affected, reduce the risk of associated health problems, and enhance overall well-being.

Early detection: ML can help identify people at risk of obesity, enabling early intervention.

Improved screening: It can be used for large-scale screening programs.

Personalized medicine: ML models can consider various factors to suggest personalized prevention strategies.

2. LITERATURE SURVEY

2.1. Fundamental Concepts

This In [1] paper, An obesity patient's postoperative health status provides insight into how their surgical treatment performed. Physicians must review prior patient data at every postoperative appointment in order to review the patient's state and assess the postoperative risk of readmission. In order to facilitate this procedure, we create a technique for identifying indicators and evaluating weight fluctuations, enabling prompt identification of possible problems and clinical readmission hazards. Two methods that are based on neural networks and conventional machine learning will be compared in this research. One kind of attentive recurrent neural networks is contrasted with classical machine learning on the problem of obesity associated entity extraction. We find that when utilising neural networks to process a limited input collection.

In [2] paper, We introduce FREGEX, a regular expression-based approach for autonomously extracting characteristics from biological literature. Tokens were recovered from biomedical literature using the Smith-Waterman and Needleman- Wunsch sequence alignment techniques, and they were represented by common patterns. To assess the efficacy of the suggested approach, three manually annotated datasets containing data on obesity, obesity categories, and smoking behaviours were utilised. For comparison, characteristics derived from successive token sequences (engrams) were utilised, and the TF-IDF vector model was utilised to mathematically describe both kinds of features. The performance of the trained Support Vector Machine and Naïve Bayes classifiers was eventually employed to evaluate the efficacy of the feature extraction techniques. The findings show that regular expression-based features not only enhanced both classifiers' performance in all.

College Of Engineering 2023 In [3] paper, Accurate medical data analysis aids in early disease identification, patient care, and community services, in addition to the biomedical and healthcare sectors' tremendous rise in machine learning. On the other hand, incomplete and unexamined medical data intake reduces the analysis's accuracy of a disease. The goal of this work is to create a cutting-edge system that simplifies machine learning algorithms for the accurate prediction of obesity and related disorders, taking into account the Indian population. The development of a system that takes into account factors influencing a person's internal, mental, emotional, and physical health as well as factors that contribute to obesity was deemed necessary in order to address the issue and offer healthier options. Among the prevalent illnesses brought on by obesity.

In [4] paper, Metabolic illnesses, including type 2 diabetes mellitus, obesity, and metabolic syndrome, are highly prevalent in developed countries' populations and necessitate ongoing clinical and pharmaceutical interventions as they progress. This is stated in the [4] study. Insulin resistance is linked to obesity in more than 90% of obese patients, who are "protected" against it. Along with signal processing techniques, infrared spectroscopy has been studied as a non-invasive tool on biofluids in the search for new and predictive biomarkers. This is a presentation of a study that used infrared spectroscopy to profile saliva in a group of obese people who were metabolically aberrant but not obese compared to control patients. To provide a standardized and consistent approach for saliva profiling in various genetic areas of interest, analysis has been done.

In [5] paper, Recent research found that genetics plays an important role in obesity risk analysis besides lifestyles. Many literatures are focusing on analyzing the effect of Single Nucleotide Polymorphism (SNPs) towards obesity to facilitate personalized medication. However, SNPs data are normally large and noisy, which affects the accuracy and computational complexity on data processing and analysis. Therefore, efficient data reduction is essential to yield better analysis results and reduce computational complexity in the experimentations. In this paper, we investigated feature selection process in obesity related SPNs analysis using Forward attribute reduction based on neighborhood rough set model (FARNeM).

In [6] paper, A fuzzy medical diagnostic decision system for helping support to evaluate patients with anginal chest pain and obesity clinical condition is proposed in this paper. Such an approach is based on the Braun Wald symptomatic classification, the fuzzy set theory and fuzzy logic, and a risk obesity factor determined by a simplified Fuzzy Body Mass Index (FBMI). The fuzzy Braun Wald symptomatic classification intertwined with the fuzzy obesity risk factor overwhelm the current rapid access chest pain clinic approaches that do not discriminate the obesity comorbidity or takes into account the subjectiveness, uncertainty, imprecision, and vagueness concerning such a clinical health condition. The resulting fuzzy obesity-based Braun Wald symptomatic chest pain assessment is an alternative to support healthcare professionals in primary health care for patients with anginal chest pain worsened by the obesity clinical condition.

In [7] paper, One of the most important challenges in the analysis of high-throughput genetic data is the development of efficient computational methods to identify statistically significant Single Nucleotide Polymorphisms (SNPs). Genome-wide association studies (GWAS) use single-locus analysis where each SNP is independently tested for association with phenotypes. The limitation with this approach, however, is its inability to explain genetic variation in complex diseases. Alternative approaches are required to model the intricate relationships between SNPs. Our proposed approach extends GWAS by combining deep learning stacked autoencoders (SAEs) and association rule mining (ARM) to identify epistatic interactions between SNPs. Following traditional GWAS quality control and association analysis, the most significant SNPs are selected and used in the subsequent analysis to investigate epistasis.

In [8] paper, A Fuzzy Obesity Index for being used as an alternative in obesity treatment and bariatric surgery indication (BSI) is presented in this paper. Obesity is nowadays understood as universal epidemy and became an important source of death and comorbidities. The search for a more accurate method to evaluate obesity and to indicate a better treatment is important in the world health context. In this paper the Body Mass Index (BMI) is first modified and treated as fuzzy sets. BMI is characterized by its capacity of weight excess and is considered the main criteria for obesity treatment and BSI. Nevertheless, the fat excess related to the Body Fat (BF) is the principal harmful factor in obesity disease, that is usually neglected.

In [9] paper, Nowadays medical research plays an important role of community safeguard by finding the solutions to health-related problems. An early detection or Co- morbidity Detection of a disease can help both patients and doctors to act and eradicate the root cause or work on preventing further deterioration of the detected disease symptoms. Hence a need to detect co-morbidity or the existing disease in an automated or semi-automated fashion has become a need of the hour. In this paper we have used machine learning and deep learning techniques on publicly available i2b2 clinical datasets to detect the chronic disease status like obesity.

According to the [10] study, musculoskeletal and metabolic comorbidities are the primary causes of the increased health issues associated with overweight and obesity.This study sought to determine how subjects with diabetic neuropathic foot and varying body mass indices differed in terms of plantar pressure distribution (BMI). Twelve subjects had their peak plantar pressure measured while they were level walking. Based on their BMI values, the subjects were divided into three groups, each with four participants: nonobese, overweight, and obese. Using the Padar-X in-shoe pressure measurement system, the peak plantar pressure for the forefoot, midfoot, and hindfoot regions was ascertained. In contrast to the forefoot and hindfoot, the midfoot showed lower peak plantar pressure. In contrast to the non-obese group, the obese group displayed a noticeably higher peak plantar pressure in the midfoot.

In [11] paper, We study the box pushing problem with a set of robots. The objective is to design a distributed algorithm that lets the robots self-coordinate to move the objects.We propose a synchronous algorithm that allows the robots to self-coordinate to relocate the box. We use Timed Input/Output Automata to describe the algorithms and show that the algorithm correctly completes the task. Then we extend the algorithm to deal with obstacles and robot failures. We implement the algorithm in SyRof (a testbed built at California State University Long Beach.) The testbed consists of four robots equipped with omnidirectional wheels to simulate drones and an autopilot that provides a synchronous system. The resulting implementation allows the robots to complete the task successfully.

In [12] paper, This paper describes an efficient way of detecting the behavior and monitoring the safety of the two-wheeler driver. The Extreme Learning Machine Algorithm based Driver Condition Recognition (ELMA-DCR) is proposed to detect three parameters such as alcohol taken by the driver, accident detection, and overweight detection (more than two persons).The ELMA-DCR is evaluated by comparing the performance with other state-of-art algorithms. The results of the proposed method show that it is an accurate and efficient technique helpful for two-wheeler drivers. An experimental setup is constructed and tested under different test conditions.

In[13] paper, In recent years, Information and Communication Technologies (ICT) are becoming a promising method as an alternative to monitor and control overweight and obesity, in addition to reducing economic costs and reducing the time between visits to doctors and nutritionists. This article presents a proposal for a mobile application to keep track of and monitor the problems of overweight and obesity.The prototype provides timely information for both the patient and the nutritionist and doctor, establishing communication channels between the parties to obtain regular updates on the progress of patients, personalized and precise suggestions from nutritionists, for sustained behavior change, helping to combat the phenomenon of obesity. The proposed application is based on the approach of new mobile technologies, which allows the rapid development of apps.

In[14] paper, The epidemiological and nutritional transition process is overweightevident in the coexistence of diseases resulting from nutritional deficiencies such as anemia and malnutrition, accompanied by an increase in the prevalence of and obesity.The propouse of this work was to analyze spatial behaviour of nutritional disease and food determinants such as malnutrition, overweight and obesity, and anemia, in Cordoba city, Argentina. The 2D-variogram algorithm for anisotropic data was programmed using IDL language to study malnutrition, overweight and obesity, and anemia variability. Our results must be a spatial pattern in the distribution of the prevalence of overweight-obesity and anemia. This study provides an important basis for further research to understand behavioral, demographic, socio-economic and environmental determinants of nutrition epidemiology.

In [15] paper, The objective of this study was to use non-invasive elastography (NIVE) to detect early changes in vascular biomechanics associated with obesity in children. The NIVE algorithm also measured the intimamedia thickness (IMT) for comparison. NIVE was applied in 120 children, 60 with elevated body mass index (BMI) (≥ 85th percentile for age and sex) and 60 non-overweight.Participants were randomly selected from a longitudinal cohort, evaluating consequences of obesity in healthy children with one obese parent. The carotid wall was automatically segmented and elastograms were computed to measure the cumulated axial strain (CAS), cumulated axial translation (CAT), and maximal shear strain IMT was also computed from segmented contours.

In this paper[16], using a previously developed model, that described the complex links between respiration, heart rate and arterial blood pressure (Psa) during cardiovascular regulation, we investigate the variation of heart rate as response of vagal and sympathetic activities. For this reason, the amplitude and frequency of breathing are varied. The correlation between breathing frequency and variations of the heart rate, is studied. The implementation of the model was done using Matlab Simulink.

In this[17] paper we present a comparative study on data base management engines for supporting big data analytics for the identification of obese subjects based on Electronic Health Records. We compared relational and non-relational approaches to address scalability and performance in a tertiary hospital. The experiments have evaluated data from five different hospital services on a data-mart containing 20,706,947 records from the University Hospital La Fe of Valencia (Spain). Experiments were based on data load and query with different configurations and restrictions.

In this[18]study, to facilitate early detection of childhood obesity, we present our methodology for effective data integration that allows modelers to import new attributes from auxiliary datasets using geospatial proximity, alongside the associated data uncertainty for each data point that is caused by the data aggregation process while estimating that attribute. We have used the data uncertainty estimate as input to various machine learning algorithms, to improve obesity prediction.

In This [19] Paper,analyses accuracy of a disease is reduced when the intake and the quality of medical data is unexplored and incomplete. There felt a need to develop a system that consider parameters affecting an individual physically, internally, mentally, psychologically and emotionally that contribute to the occurrence of Obesity and suggest healthier alternatives to curb this problem. Some of the common diseases due to Obesity included in the system are Diabetes, Heart Attack, Hypertension, Osteoarthritis and Varicose Veins. The developed system will undeniably be beneficial for predicting obesity, its related diseases and for the future betterment of an individual.

The [20] paper presents an aim to understand the effect of obesity on ANS (autonomic nervous system) using HRV (heart rate variability) parameters. The statistical results of the study indicate the sympathovagal imbalance due to reduced parasympathetic activity. The statistical results were validated by incorporating the machine learning technique into the study. Machine Learning (ML) algorithm helps to identify the most important predictor that can clearly differentiate control and obesity subjects. The statistical and ML algorithm result shows changes in the sympathovagal balance due to decreased parasympathetic activity.

[21] As Wi-Fi signals are closely integrated with the human workplace, the signals are used to observe human activity in the vicinity of the Wi-Fi routers, that is, the transmitter and the receiver. The activity once correctly identified can be used to store the changes in the activity and how often one performs such activity or is there a sudden change. using these metrics and quantifying the changes can help calculate and predict the heath of the individual and provide better statistics to health professionals to identify any ailment that may have caused a change in the individuals behaviors' making them seek medical attention.

In This [22] Paper study of Ontology-based Obesity Tracking System for Children and Adolescents.The detection of obesity in children and adolescents at early stages is important and it is crucial to start individual based treatments. This research study especially emphases obesity management in childhood and adolescence stages with the contribution of Semantic Web technology. The system is an ontology-based obesity tracking system for children and adolescents which has its own Obesity Tracking Ontology and medical semantic rule knowledge base with an inference engine.

This[23]paper proposes a state evaluation method based on rejection rate of production shift. This paper uses a linear return equation to fit continuous overweight rejection number in time increment and identify the optimal slope range. By comparing the slope range, it can determine whether the status of overweight rejection of cigarette equipment is abnormal or not, and the relevant factors within the time range of abnormal occurrence are automatically correlated to support the user to carry out anomaly analysis.

This[24] paper surveys the growing body of recent literature on machine and deep learning models for obesity prediction by providing a coherent view of the limitations of the existing systems. The taxonomy of the existing literature on obesity prediction into methods used, predicted outcome, factors used, type of datasets, and the associated purpose, is discussed for analysis of the state-of-the-art. Further,computer vision-based methods for obesity prediction and interpret-able techniques for understanding the out come of the models are discussed as well. In addition, we have also identified novel research directions.The overall aim is to advance the state-of-the-art and improve the quality of discourse in this field.

This [25] paper describes the implementation of a comprehensive clinical decision support system (CDSS) for the risk factors prediction of comorbidities related to obesity and for the characterization of indirect connections between such comorbidities and non-communicable diseases. In particular, the direct correlation between obesity, diabetes, cardiovascular, and heart disease is analyzed by using machine learning (ML) predictive models, while the connection of the co-occurring disorders to the numerous additional non-communicable diseases is analyzed via a graph-based user interface. ML predictive models based on publicly available datasets, explainable artificial intelligence local and global model interpretation, and graph-based representation of noncommunicable disease connections.

This Paper [26] proposed solution created a positive impression and satisfaction from the technology acceptance perspective. Methods: Continuous co-creation process has been applied in the frame of the Design Thinking Methodology, involving children, educators and healthcare professional in the whole process. Such consider rations were used to derive the user needs and the technical requirements needed for the conception of the Internet of Things platform based on micro services. Conclusions: Main findings confirm that this ecosystem can assess behavior of children, motivating and guiding them towards achieving personal goals.

In this paper [27] the study takes a data set relating to the main causes of obesity, based on the aim to reference high caloric intake, a decrease of energy expenditure due to the lack of physical activity, alimentary disorders, genetics, socioeconomic factors, and anxiety and depression. Obesity is a worldwide disease that affects people of all ages and gender; in consequence, researchers have made great efforts to identify factors that cause it early. In this study, an intelligent method is created, based on supervised and unsupervised techniques of data mining detect obesity levels and help people and health professionals to have a healthier lifestyle against this global epidemic.

In This [28] Paper the study of overweight and obesity are one of the main lifestyle illnesses that leads to further health concerns and contributes to numerous chronic diseases, including cancers, diabetes, metabolic syndrome, and cardiovascular and machine learning techniques for the prevention and treatment of obesity. This study initially recognized the significant potential factors that influence and cause adult obesity.

It is[29] proposed that obesity, generally defined by an excess of body fat causing prejudice to health, can no longer be evaluated solely by the body mass index because it represents a heterogeneous entity excessive amounts of visceral adipose tissue and of ectopic fat largely define the cardiovascular disease risk of overweight and moderate obesity . Because of the difficulties of normalizing body fat content in patients with severe obesity, more aggressive treatments have been studied in this subgroup of individuals such as obesity surgery, also referred to as metabolic surgery. Because of the difficulties of normalizing body fat content in patients with severe obesity, more aggressive treatments have been studied in this subgroup of individuals such as obesity surgery, also referred to as metabolic surgery.

The [30] primary purpose of this study this paper is not to present a risk prediction model but to provide a review of various machine learning (ML) methods and their execution using available sample health data in a public repository related to lifestyle diseases, such as obesity, and diabetes type II. In this study, we targeted people, both male and female, in the age group of >20 and <60 excluding pregnancy and genetic factors. This paper qualifies as a tutorial article on how to use different ML methods to identify potential risk factors of obesity.

3. ANALYSIS

Analysis is concerned with understanding and modeling the application and domain within which it operates. The initial input to the analysis phase is problem statement, which describes the problem to be solved, and provides a conceptual view of the proposed system. Subsequent dialog with the customer and real-world background knowledge are additional inputs to analysis. The output from analysis is a formal model that captures the three essential aspects of the system: the objects and their relationships, the dynamic flow of control, and the functional transformation of data subject to constraints.

Requirement analysis bridges the gap between system engineering and software analysis design. Software requirement analysis involves requirement collection, classification, structuring, prioritizing and validation. Requirement analysis consists of user requirements Analysis is concerned with understanding and modeling the application and domain within which it operates. The initial input to the analysis phase is problem statement, which describes the problem to be solved, and provides a conceptual view of the proposed system.

3.1. System Development Requirements

The system development requirements detail the resources required by the system in the development phase. The resources are categorized as hardware resources and software resources.

3.1.1. Software Requirements

Table 3.1. Minimum Software Requirements

Software Type	Minimum Requirement
Platform (OS)	Windows 10
Front End	Python
Back End	SQLite , Tkinter .
Design Tool	RSA (Rational Software Architect), Edraw
Development Tool (IDE)	Spyder
Testing Tool	Manual testing.
Documentation Tool	Microsoft Office.

3.2. System Deployment Requirements

The system deployment requirements detail the resources required by the system to deploy/install/execute it on user machine. The resources are also categorized as hardware resources and software resources.

3.2.1. Hardware Requirements

Table 3.2. Minimum Hardware Requirements (Server Side)

Hardware Type	Minimum Requirement
Processor	2.66 GHz x86/x64 based
Primary Memory	4 GB RAM
Secondary Memory	5 GB HDD/ROM Space
Other Hardware	Camera 5 MP

Table 3.3. Minimum Hardware Requirements (Client Side)

Hardware Type	Minimum Requirement
Processor	2.2 GHz x86/x64 based
Primary Memory	2 GB RAM
Secondary Memory	5 GB HDD/ROM Space
Other Hardware	Camera 5 MP

3.2.2. Software Requirements

Table 3.4. Minimum Software Requirements (Server Side)

Software Type	Minimum Requirement
Platform (OS)	Windows 10
Hosting Server	Apache Tomcat, Microsoft IIS
Database Server	SQLite
Execution Environment	Spyder
Web Browser	Google Chrome

Table 3.5. Minimum Software Requirements (Client Side)

Software Type	Minimum Requirement
Platform (OS)	Windows 10
Execution Environment	Spyder
Web Browser	Google Chrome

3.3. Functional Requirements

These are the requirements that the end user specifically demands as basic facilities that the system should offer. All these functionalities need to be necessarily incorporated into the system as a part of the contract. These are represented or stated in the form of input to be given to the system, the operation performed and the output expected. They are basically the requirements stated by the user which one can see directly in the final product, unlike the non-functional requirements.

1. The system should be able to register new users and manage login credentials.
2. The system should be able to store and manage user profiles with personal and medical history details.
3. The system should be able to train, test, and periodically update the machine learning model using historical and new data.
4. The system should be able to predict obesity levels.
5. The system should be able to offer personalized health and lifestyle recommendations.

3.4. Non-Functional Requirements

These are basically the quality constraints that the system must satisfy according to the project contract. The priority or extent to which these factors are implemented varies from one project to other. They are also called non-behavioral requirements. They basically deal with issues like:

1. The system should be available most of the time and not go down frequently.
2. It should be able to handle more users and data without slowing down.
3. The system must keep user data safe by following security standards.
4. It should be easy to use on different devices like computers and phones.
5. The system must follow rules to keep patient information private.
6. The code should be well-organized and easy to update.
7. It should be easy to add new features or connect with other tools.
8. The system should work quickly and use computer resources effectively.

3.5. Functional Modeling: Data Flow Diagram

Data flow diagram (DFD), also referred to as 'Bubble chart' is a graphical technique, which is used to represent information flow, and transformers are applied when data moves from input to output.

I. Data Flow Diagram (DFD): Level 0

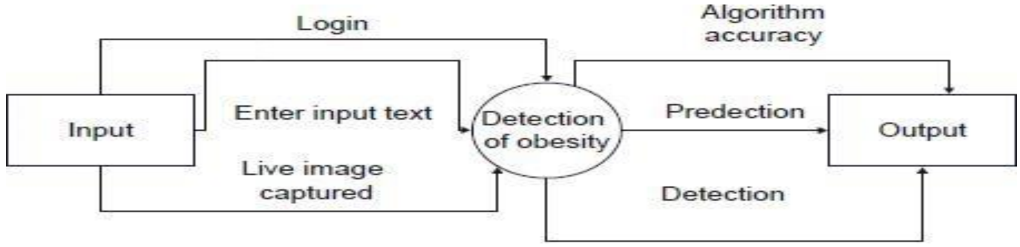

Figure 3.1. Data Flow Diagram (DFD): Level 0

II. Data Flow Diagram (DFD): Level 1

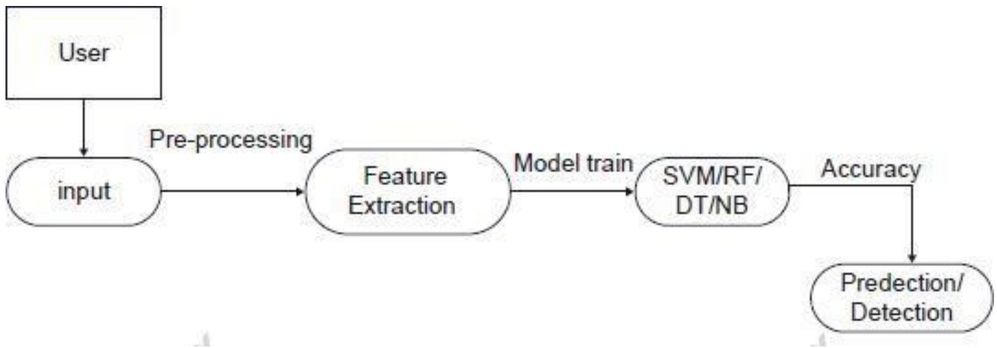

Figure 3.2. Data Flow Diagram (DFD): Level

4.1 Software Process Model

4.1.1 Description of Selected Model

The software development cycle is a combination of different phases such as designing, implementing and deploying the project. These different phases of the software development model are described in this section. The SDLC model for the project development can be understood using the following figure The chosen SDLC model is the waterfall model which is easy to follow and fits bests for the implementation of this project.

Requirements Analysis: At this stage, the business requirements, definitions of use cases are studied, and respective documentation is generated. Design: In this stage, the designs of the data models will be defined, and different data preparation and analysis will be carried out.

Implementation: The actual development of the model will be carried out at this stage. Based on the data model designs and requirements from previous stages, appropriate algorithms, mathematical models and design patterns will be used to develop the agent's back-end and front-end components.

Testing: The developed model based on the previous stages will be tested in this stage. Various validation tests will be carried out over the trained model.

Deployment: After the model is validated for its accuracy scores its ready to be deployed or used in simulated scenarios.

Maintenance: During the use of the developed solution various inputs/scenarios will be countered by the model which might affect the models overall accuracy. Or with the passing of time the model might not fit the new business requirements. Thus, the model must be maintained often to keep its desired state of operation.

4.1.2 Reason for selection

Budget and time constraints can impact the choice of SDLC model. Agile and iterative models are more flexible and can adapt to changing timelines and budgets. Waterfall may be suitable when fixed deadlines and budgets are critical. If the project involves long-term maintenance and updates, models like Agile may be preferable, as they support ongoing iterations and improvements based on user feedback and changing needs.

4.2 Cost Estimation Using Basic Cocomo Model

4.2.1 Empirical data from Existing Similar Project

Cost Model (COCOMO) is a widely used technique for estimating the cost of software development projects. There are three different versions of COCOMO: Basic COCOMO, Intermediate COCOMO, and Detailed COCOMO. I will provide you with an overview of how to perform cost estimation using the Basic COCOMO model.

Basic COCOMO divides software projects into three categories based on their complexity: Organic, Semi-detached, and Embedded. You select the category that best fits your project, and then estimate the project size in lines of code (KLOC) and apply a set of cost drivers. Identify the category of your project: Organic, Semi-detached, or Embedded.

4.2.2 Estimation Size (in KLOC)

The selection is based on factors such as project size, complexity, and the level of interaction with other software or systems. Determine the size of the software in thousands of lines of code (KLOC). This can be based on historical data, expert judgment, or similar projects. Cost Estimate The project cost can be found using any one of the models.

4.2.3 Person Required

With these values and your estimated project size, you can calculate the required staffing. The staffing estimation will provide you with the number of people needed to complete the project within the estimated time, assuming average productivity levels and other factors as per the Basic COCOMO model.

4.2.4 Estimation Technique Used

COCOMO (Constructive Cost Estimation Model) was proposed by Boehm [1981]. COCOMO predicts the efforts and schedule of a software product based on size of the software. COCOMO stands for "Constructive Cost Model". According to Boehm, software cost estimation should be done through three stages: Basic COCOMO, Intermediate COCOMO and Complete COCOMO. We are going to used Basic COCOMO that categorized projects into three types:

i. organic: Suitable for organization that has considerable experience and requirements.

ii. Semidetached: Examples of this type are developing new database management system.

iii. Embedded: Organization has little experience and stringent requirements.

The Basic COCOMO formula takes the form:

E=ab(KLoC)(b b) [person-months]

D=cb(E)(d b) [months]

P= E / D [persons]

Where E is the effort applied in person-months, KLoC is the estimated number of thousands of delivered lines of code for the project, D is total time duration to develop the system in months, and P is number of persons required to develop that system. The coefficient a_b ,c_b and the exponent b_b ,d_b are given in the next table.

Table 4.1. Coefficient values for Basic COCOMO

Software project	a_b	b_b	C_b	d_b
Organic	2.4	1.05	2.5	0.38
Semi-detached	3.0	1.12	2.5	0.35
Embedded	3.6	1.20	2.5	0.32

This project will fall in the Semi-detached category.

4.2.5 Size Estimations

Table 4.2. Size Estimation of Current System.

Software Module	LOC
Login	300
Registration	400
GUI Main	350
Check1	300
Human Skeleton	250
High	300
Low	350
Medium	350
Next	200
Coordinate_in_CSV	400
Sample_GUI	500
Extreme	350
GUI_Master	400
Training	450
GUI_ Master_Old	450
Total Estimated Lines Of Code (LOC)	5350

Line of code approximately will be 5350

Effort Estimation

The value ab and bb according to Semi-detached system is

a_b=3.0and b_b=1.12

The system falls in the Semi-detached category.

Total LOC (approx) of project is:5350LOC= 5.350 KLOC

Effort (E) =a_b(KLoC)(b_b) [Person -Month]

$E=3.0*(5.350)^{1.12}$

E= PM (Calculate)

Person-Month=20PM (approx)

Duration Estimation

Duration (D) = c_b(E) (d_b) [months]

$= 2.5*(19.62)^{0.35}$

= 2.5 *2.83

= 6.07Duration ≈6[months]

Person Required

Person Required = Effort Applied (E) / Development Time (D) [count]

=19.62/6

= 3.72[count]

Person Required =4[Persons]

Cost Estimation

We take the assumption each person charges 3000 rupees per month.

Total Estimation= 3,000*4= Rs.12,000/- +Two Mobile phone Cost

Total Estimation=Rs.12,000+15000=27,000/-

Total Estimation=Rs.27,000 /-

4.2.6 Estimation Summary

Table 4.3. Summary of different estimation.

Estimation	Value
Size of the Project	5350 LoC
Effort Required	20 Person Months
Duration Required	6 Months
Person Required	4
Cost Required	₹27,000/-

4.3 Team Structure

Team structure addresses the issue of organization of the individual project teams. As per the estimation, the project team will consist of **Four** members. The effort assignment, duties, and details of each member is given below:

Table 4.4. Team structure.

Sr. No.	Name of Team Member	Phase - I Role	Phase - II Role	E-mail ID
1.	Shraddha Gokul Borase	Leader	Leader	shraddhaborase0000@gmail.com
2.	Vaishnavi Deepak Bhoi	Member	Member	vaishnavibhoi120@gmail.com
3.	Vaishnavi Arun Patil	Member	Member	patilvaishnavi2510@gmail.com
4.	Umesh Sharad Patel	Member	Member	up41174@gmail.com

4.4. Team Scheduling

Project scheduling involves plotting project activities against a time frame. The aim of the process is to ensure that various project tasks are well coordinated, and they meet the various project objectives including timely completion of the project. The Project Table is a popular way to perform project scheduling.

Table 4.5. Project Table

Activity Name	Start Date	Actual Start Date	End Date	Actual End Date	Effort Assignment
Introduction	09/09/2023	11/09/2023	14/09/2023	15/09/2023	Ms. Vaishnavi Bhoi
Literature Survey	15/09/2023	16/09/2023	20/09/2023	21/09/2023	Mr. Umesh Patel
Analysis	21/10/2023	25/10/2023	26/10/2023	30/10/2023	Ms. Shraddha Borase
Planning	01/11/2023	02/11/2023	07/10/2023	08/11/2023	Ms. Vaishnavi Patil
Design	08/11/2023	10/11/2023	18/11/2023	19/11/2023	Ms. Shraddha Borase
Phase-I Documentation	19/11/2023	20/11/2023	27/11/2023	28/11/2023	Ms. Shraddha Borase Ms. Vaishnavi Patil Ms. Vaishnavi Bhoi Mr. Umesh Patel
Form Design	15/01/2024	17/01/2023	24/01/2023	25/01/2023	Ms. Shraddha Borase Ms. Vaishnavi Bhoi
Implementation	26/02/2024	04/03/2023	07/03/2023	11/03/2023	Ms. Shraddha Borase

Testing	14/03/2024	16/03/2023	20/03/2023	22/03/2023	Ms. Shraddha Borase Mr. Umesh Patel
Project Effort and Cost	23/04/24	24/04/2023	29/04/2023	02/05/2023	Ms. Vaishnavi Patil Ms. Vaishnavi Bhoi
Project-II Documentation	15/05/24	15/05/2023	22/05/2023	24/05/2023	Ms. Shraddha Borase Ms. Vaishnavi Patil Ms. Vaishnavi Bhoi Mr. Umesh Patel

5. DESIGN

Design uses a combination of text and diagrammatic forms to depict the requirements for data, function and behavior in a way that is relatively easy to understand and more importantly, straightforward to review for correctness, completeness, and consistency.

A diagram is the graphical presentation of a set of elements most often rendered as a connected graph of vertices (things) and arcs (relationship). These diagrams are drawn to visualize a system from different perspectives so a diagram into a system.

5.1. UML Modeling

The unified modeling language (UML) is a Graphical Language for visualization, Specifying, construction and documenting the artifacts of a software intensive system. The UML gives a standard was to write system's blueprints, covering conceptual thing, such as Business Processes & system functions, as well as concrete things, such as classes written in a specific programming language, database schemas, and reusable software components.

5.1.1. Use Case Diagram

A use case defines behavioral features of a system. Each use case is named using a verb phase expresses a goal of the system. A use case diagram shows a set of use cases and actors &their relationships. Use case diagrams address the static use case view of a system. These diagrams are especially important in organizing and modeling the behaviors of a system. It shows the graphical overview of functionality provided by the system intents actor.

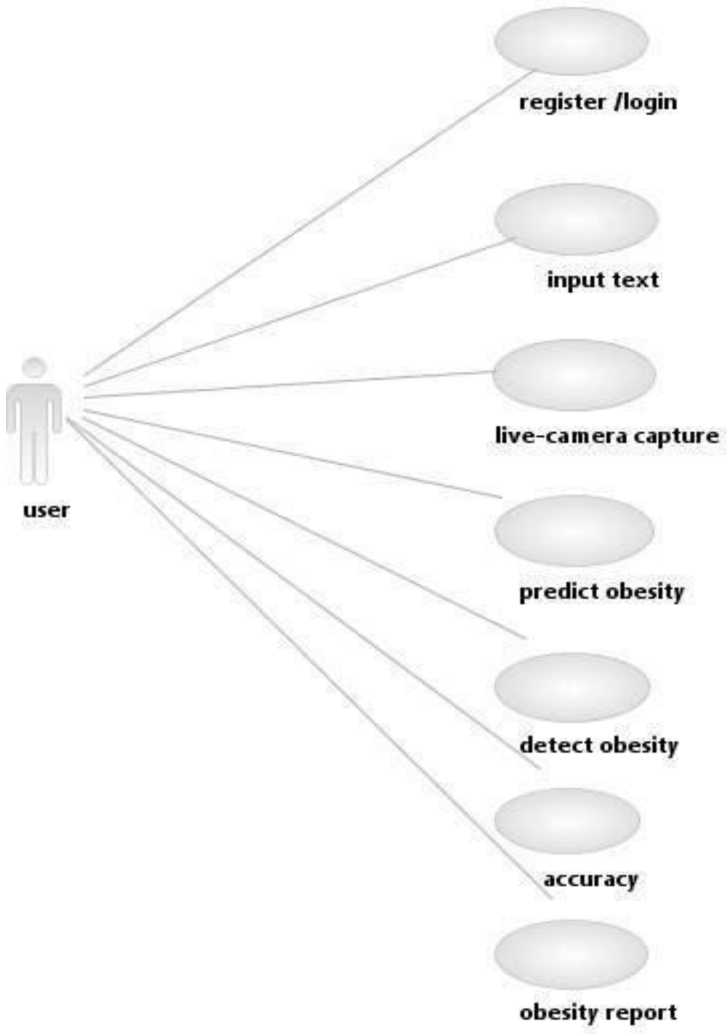

Figure 5.1. Use case Diagram for User

Table 5.1. Use case Description for Use

Use case	Description
Login	The user can login to start beginning his work.
Create Profile	User can create his own profile .
Access Application	User can access various application like keyboard,Mouse etc.
Speak command	User can give speak command to access specific application.
Give Speech Test	User can give speech test to verify command.
Logout	User can logout to exit .

5.1.2. Activity Diagram

An activity diagram of a special kind of state chart diagram that shows the flow from activity within a system. An activity addresses the dynamic view of a system. The activity diagram is often seen as part of the functional view of a system because it describes logical processes, or functions. Each process describes a sequence of tasks and the decisions that govern when and when they are performed. The flow in an activity diagram is driven by the completion of an action.

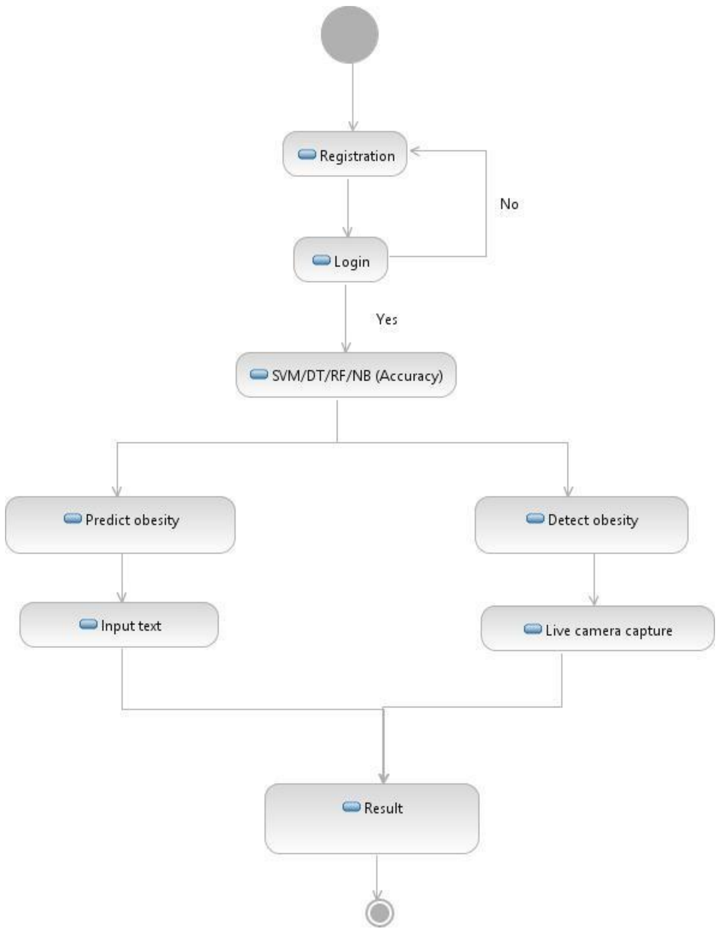

Figure 5.2. Activity Diagram for Keyboard Control

5.1.3. Sequence Diagram

Sequence diagram is a kind of interaction diagram. It shows an interaction, consisting of a set of objects and their relationships, including the message that may be dispatched among them. A sequence diagram emphasizes the time ordering of messages. As shown in the figure we can form a sequence diagram by first placing the objects that participate in the interaction at the top of our diagram. The object that initiates the interaction at the left and increasingly more subordinate objects to the right. The messages that these objects send and receive along the Y-axis, in order of increasing time from top to bottom. This gives the reader a clear visual cue to the flow of control over time.

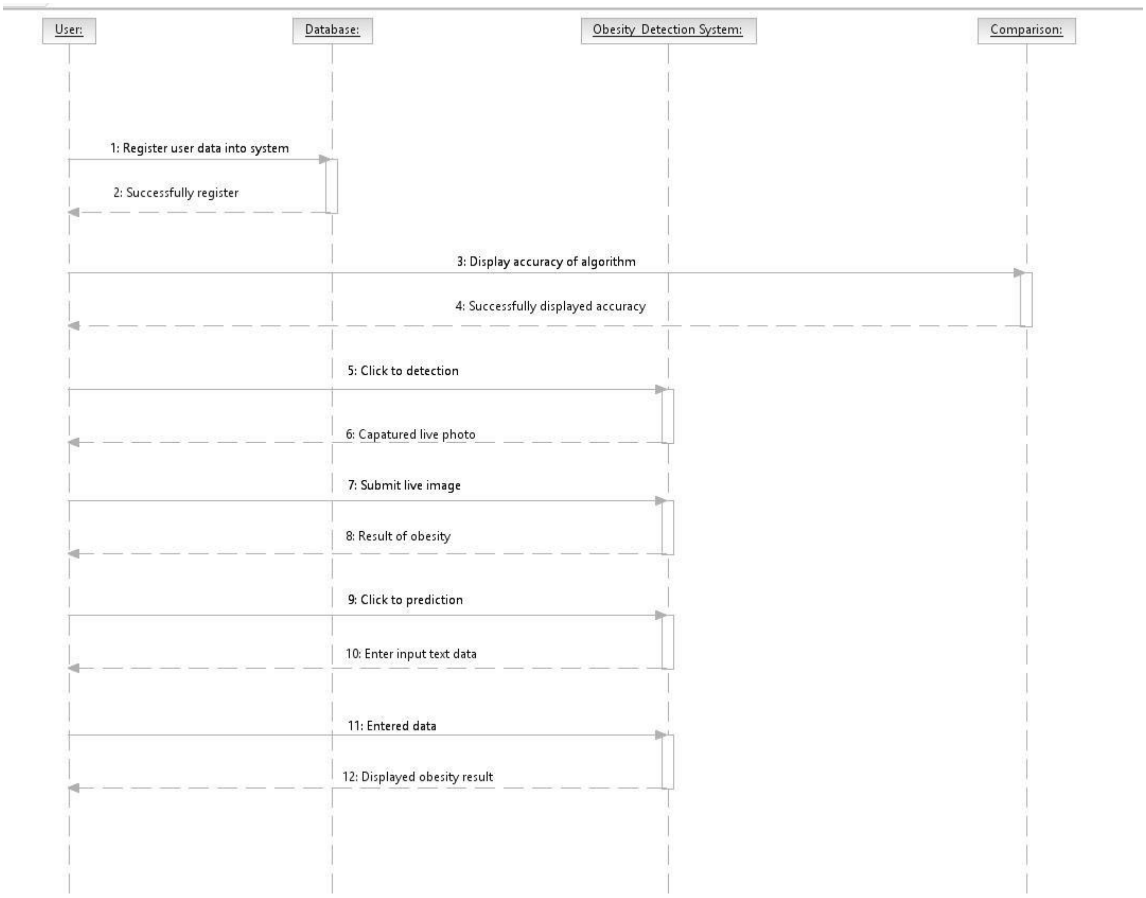

Figure 5.3. Sequence Diagram for Keyboard Control

5.1.4. Class Diagram

A class diagram shows a set of classes, interfaces and collaborations and their relationship. These diagrams are the most common diagram found in modeling object-oriented systems. Class diagram addressed the static design view of a system.

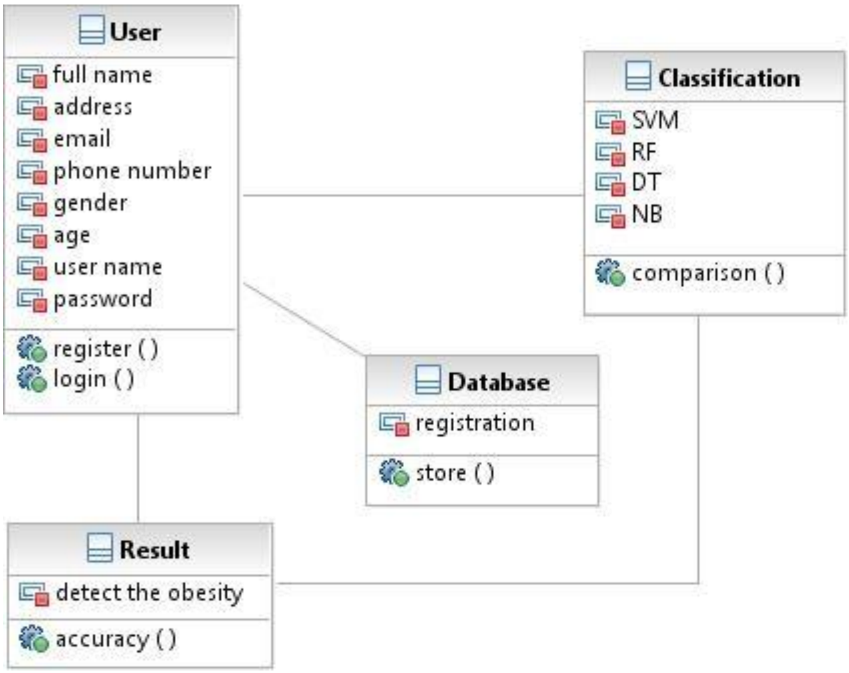

Figure 5.4. Class Diagram for Speech Enabled O.S. Control

Table 5.2. Class Diagram Description for System

Class	Description
Application GUI	Application GUI gives access to various applications such as mouse control, keyboard control etc.
User	Users can access application control through voice command.
SAPI	Speech API can convert speech to text when user gives command and gives text to speech output.
Database	Database can manage user profiles, keywords, preferences etc.

5.1.5. Component Diagram

A component diagram shows the organization and dependencies among a set of components. Component diagrams address the static implementation view of a system. Component diagrams are one of the two kinds of diagrams found in modeling the physical aspects of object-oriented systems. A component diagram shows the organization and dependencies among set of components. You can use component diagrams to model the static implementation view of a system.

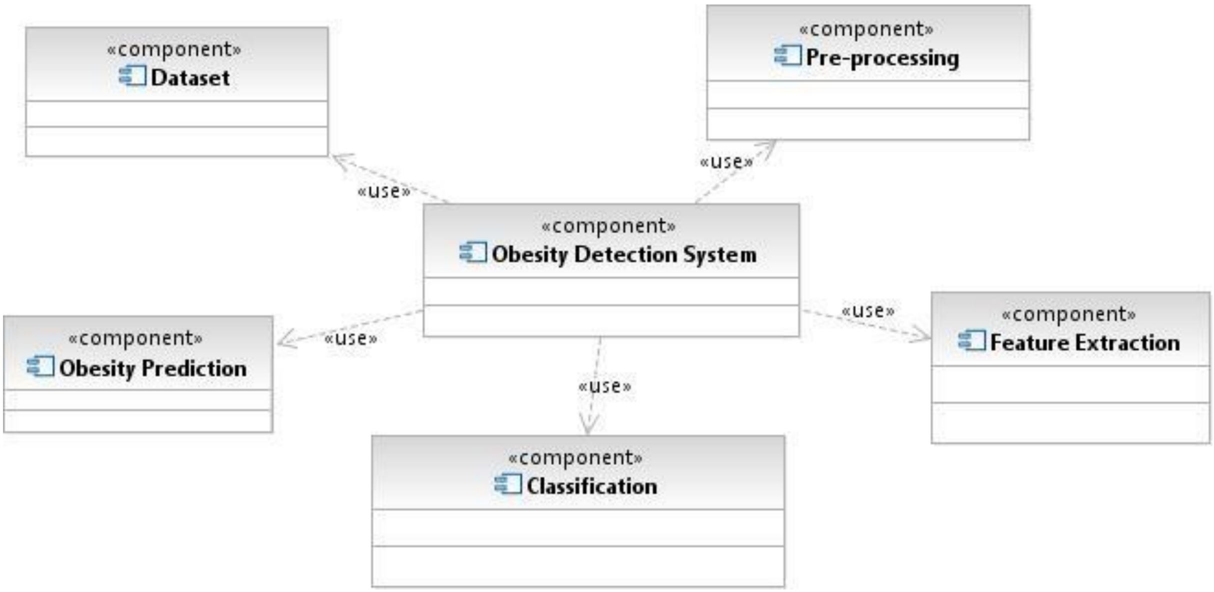

Figure 5.5. Component Diagram for Speech Enabled O.S. Control

5.1.6. Deployment Diagram

Deployment diagram shows the configuration of run time processing nodes and components that live on them. Deployment diagram addresses the static deployment view of architecture. A deployment diagram shows the configuration of run-time processing nodes and the components that live on them. Deployment diagrams address the static view of architecture. They are related to components diagram in that a node typically encloses one or more components.

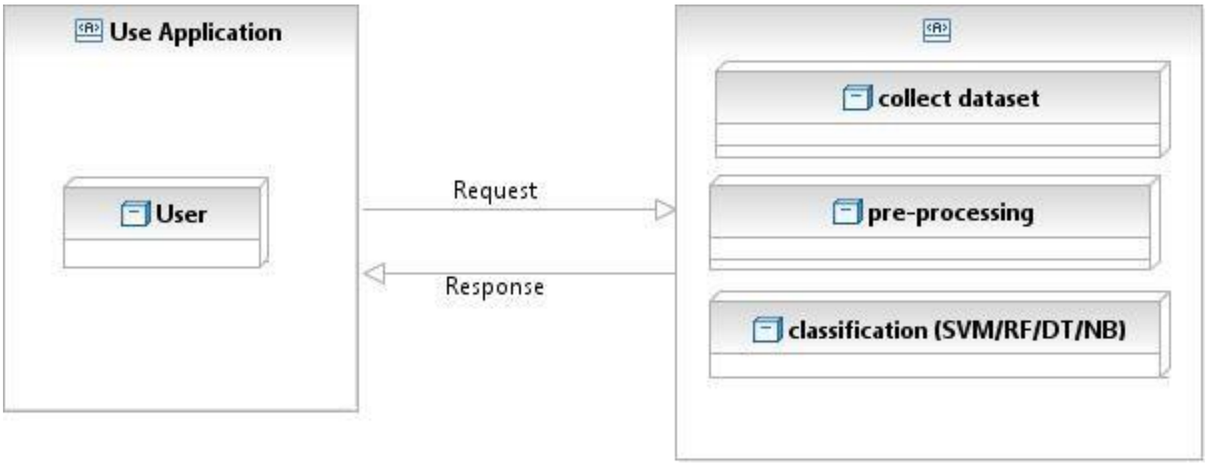

Figure 5.6. Deployment Diagram for Speech Enabled O.S. Control

5.2. Data Modelling

Data modeling shows the logical structure of a database, including the relationships and constraints that determine how data can be stored and accessed. Individual database models are designed based on the rules and concepts of whichever broader data model the designers adopt. Most data models can be represented by an accompanying database diagram.

5.2.1. Database Schema

A database schema is the way a given database is organized or structured. Not all types of databases make use of a rigid schema (e.g., NoSQL databases such as MongoDB). However, the ones that do will require you to have a structure in place before you start adding data to your database.

Table 5.3. Database Schema for Employee

Table Name		
Column Name	**Data Type (Size)**	**Constraints**
E_id	Int (05)	Primary, Not Null
E_name	Varchar (25)	Not Null
E_address	Varchar (50)	Not Null
E_contact	Int (10)	Not Null
E_email	Varchar (50)	Null

5.2.2. E-R Diagram

Entity-Relationship diagram (ERD) is a graphical technique, which is used to represent entities present in the system, and relationship those are applied between these entities. The entity-relationship (E-R) data model is based on a perception of a real world that consists of collection of basic objects, called entities, and of relationships among these objects. An entity is a “thing” or “object” in the real world that is distinguishable from other objects.

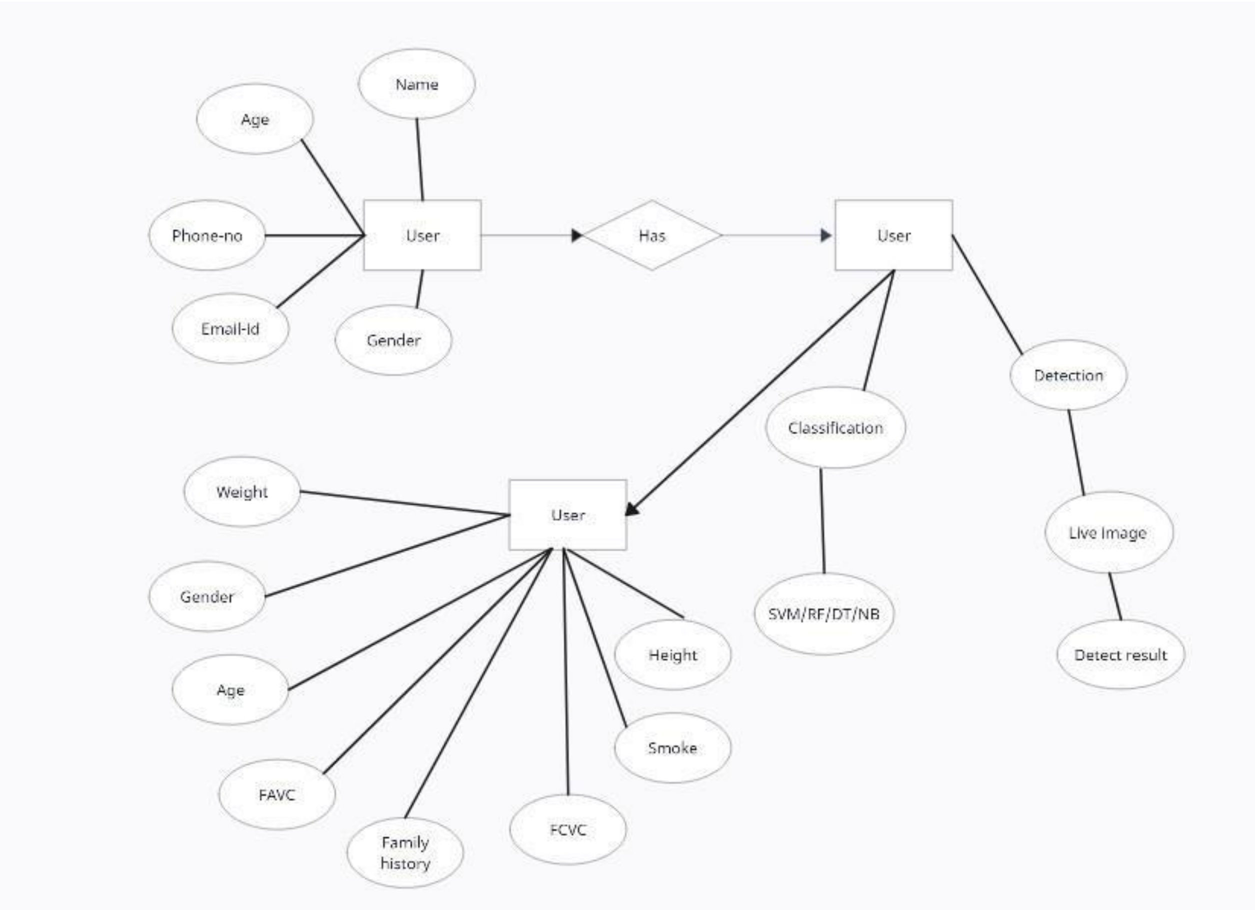

Figure 5.7. E-R Diagram for Company Database

6. IMPLEMENTATION

6.1. Implementation Language: Python

6.1.1. Features

Th Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whites pace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed, and garbage collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a" batteries included" language due to its comprehensive standard library. Python was created in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system with reference counting. Python 3.0, released in 2008, was a major revision of the language that is not completely backward compatible, and much Python 2 code does not run unmodified on Python 3. The Python 2 language was officially discontinued in 2020 (first planned for 2015), and" Python 2.7.18 is the last Python 2.7 release and therefore the last Python 2 release." No more security patches or other improvements will be released for it. With Python 2's end-of-life, only Python 3.6.x and later are supported.

6.1.2. Reason for Selection

We have selected Python to implement the system because of following reasons:

- Python programming language resembles the everyday English language, and that makes the process of learning easier. Its simple syntax allows you to comfortably work with complex systems, ensuring clear relations between the system elements.

- It offers an option to choose either to use Oops or scripting.

- There's also no need to recompile the source code, Python developers can implement any changes and quickly see the results.

- Programmers can combine Python and other languages to reach their goals.

- The next advantage of python for AI and ML development is platform independence. Python is not only comfortable to use and easy to learn but also very versatile. What we mean is that Python for machine learning development can run on any platform including Windows, MacOS, Linux, Unix, and twenty-one others.

6.1.3. Comparison with Language

This section compares the selected programming language with one of the popular programming languages in similar domain based on certain criteria.

Table 6.1. Comparison of Python with R language

Criteria	Python	R Language
Purpose	General-purpose language for web development, data analysis, and AI.	Specialized for statistical analysis and data visualization.
Syntax	Emphasizes readability and simplicity.	Optimized for statistical analysis.
Community and Libraries	Diverse community with extensive libraries.	Strong community focused on statistics.
Flexibility	More versatile for various tasks beyond statistics.	Specialized for statistical computing.
Integration	Easily integrates with other tools and languages.	Primarily used for standalone statistical analysis.
Performance	Generally faster for general-purpose tasks.	May perform better for certain statistical computations.
Industry Adoption	Widely adopted across industries.	Commonly used in academia and specific industries for statistics.

6.2. Database: SQLite

6.2.1. Features

SQLite is a popular and lightweight, server less, self-contained, and embedded relational database management system (RDBMS). While SQLite allows multiple processes to access the database simultaneously, it has some limitations in high-concurrency scenarios. It uses a file-level locking mechanism for write operations, which can impact concurrent writes. Unlike traditional RDBMS like MySQL or PostgreSQL, SQLite is serverless. It doesn’t require a dedicated server process to manage the database.

6.2.2. Reason for Selection

We have selected SQLite to store the system data because of following reasons:

- The updates your content continuously so, little or no work is lost in a case of power failure or crash.
- SQLite is less bugs prone rather than custom written file I/O codes.
- SQLite queries are smaller than equivalent procedural codes so, chances of bugs are minimal.
- SQLite is a very light weighted database so, it is easy to use it as an embedded software with devices like televisions, Mobile phones, cameras, home electronic devices, etc.
- It reduces application cost because content can be accessed and updated using concise SQL queries instead of lengthy and error-prone procedural queries.
- SQLite can be easily extended in in future releases just by adding new tables and/or columns. It also preserve the backwards compatibility.
- SQLite is portable across all 32-bit and 64-bit operating systems and big- and little-ending architectures.
- Multiple processes can be attached with same application file and can read and write without interfering each other.
- It can be used with all programming languages without any compatibility issue.

6.2.3. Comparison with SQLite

This section compares the selected database with one of the popular database systems in similar domain based on certain criteria.

Table 6.2. Comparison of SQLite with MYSQL

Criteria	SQLite	MYSQL
Deployment	Local/embedded applications, serverless.	Client-server applications, separate server process.
Concurrency	Single-user, limited concurrency.	Multi-user, high concurrency support.
Performance	Suitable for small-scale applications.	Optimized for larger-scale applications.
Features	Basic relational database features.	Rich feature set including stored procedures, triggers, etc.
Scalability	Limited scalability.	Designed for horizontal and vertical scaling.
Ease of Use	Simple setup and management.	More configuration options, potentially more complex setup.

6.4. Implementation Tool(s): Anaconda (Spyder)

6.4.1. Features

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.)that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS.It is developed and maintained by Anaconda, Inc As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition.

Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for other things than Python.

There is also a small, bootstrap version of Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages. Anaconda distribution comes with over 250 packages automatically installed, and over 7,500 additional open-source packages can be installed from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator, as a graphical alternative to the command line interface (CLI). The big difference between conda and the pip package manager is in how package dependencies are managed, which is a significant challenge for Python data science and the reason conda exists. When pip installs a package, it automatically installs any dependent Python packages without checking if these conflict with previously installed packages [citation needed]. It will install a package and any of its dependencies regardless of the state of the existing installation [citation needed]. Because of this, a user with a working installation of, for example, Google TensorFlow, can find that it stops working having used pip to install a different package that requires a different version of the dependent NumPy library than the one used by TensorFlow. In some cases, the package may appear to work but produce different results in detail.

6.4.2. Reason for Selection

We have selected Spyder to implement the system because of following reasons:

• **Interactive console:** Harness the power of as many IPython consoles as you like with full workspace and debugging support, all within the flexibility of a full GUI interface. Instantly run your code by line, cell, or file, and render plots right in line with the output or in interactive windows.

• **Editor:** Work efficiently in a multi-language editor with a function/class browser, real time code analysis tools (pyflakes, pylint, and pycodestyle), automatic code completion horizontal/vertical splitting, and go-to-definition. • Documentation viewer: Render documentation in real-time with Sphinx for any class or function, whether external or user-created, from either the Editor or a Console.

• **Variable explorer:** Inspect any variables, functions or objects created during your session. Editing and interaction is supported with many common types, including numeric/strings/bools, Python lists/tuples/dictionaries, dates/time deltas, NumPy arrays, Panda’s index/series/datagrams, PIL/Pillow images, and more.

• **Development tools:** Examine your code with the static analyzer, trace its execution with the interactive debugger, and unleash its performance with the profiler. Keep things organized with project support and a built-in file explorer, and use find in files to search across entire projects with full regex support.

6.4.3. Comparison with VS Code

This section compares the selected IDE with one of the popular IDEs available for similar purpose based on certain criteria.

Table 6.3. Comparison of Anaconda (Spyder) with VS Code

Criteria	Anaconda	Vs code
Target	Data scientists, analysts.	Developers across languages.
Focus	Scientific computing, data analysis.	Versatile code editor with extensive customization.
IDE	Complete IDE with built-in data science tools.	Lightweight editor with powerful extensions.
Included Packages	Bundled with Anaconda distribution.	Not bundled, but customizable.
User Interface	MATLAB-like layout.	Clean, customizable, and extensible.
Features	IPython console, variable explorer, plotting.	Extensions for Git, IntelliSense, debugging, task automation.

6.5 Form Design

6.5.1 Home Page

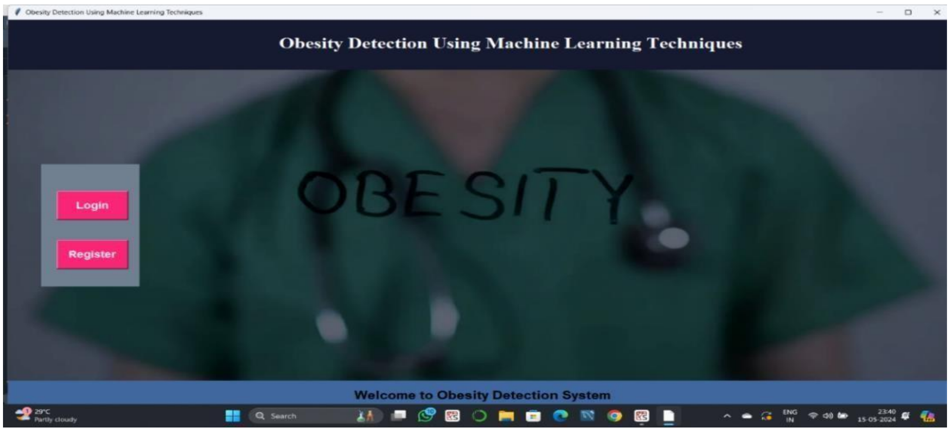

Figure 6.5.1. Home page ODS

6.4.1. Login Page

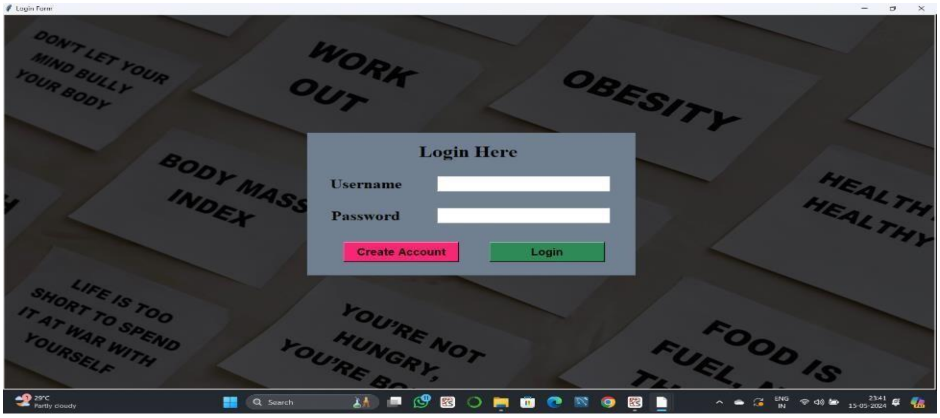

Figure 6.5.2. Login Page ODS

6.4.2. Registration Page

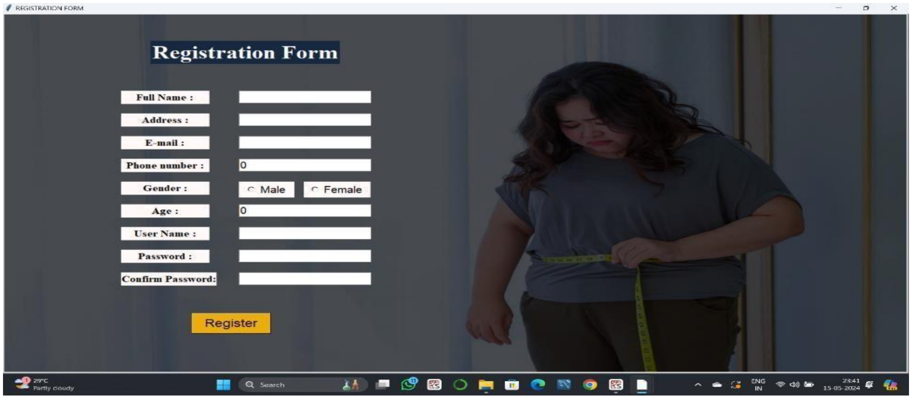

Figure 6.5.3. Registration Page ODS

6.4.3. Algorithm Page

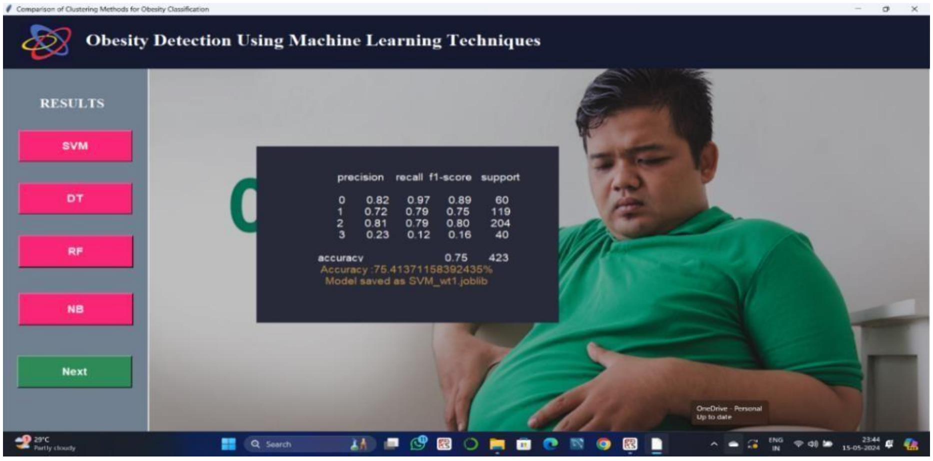

Figure 6.5.4. Algorithm page ODS

6.4.4. Check Page

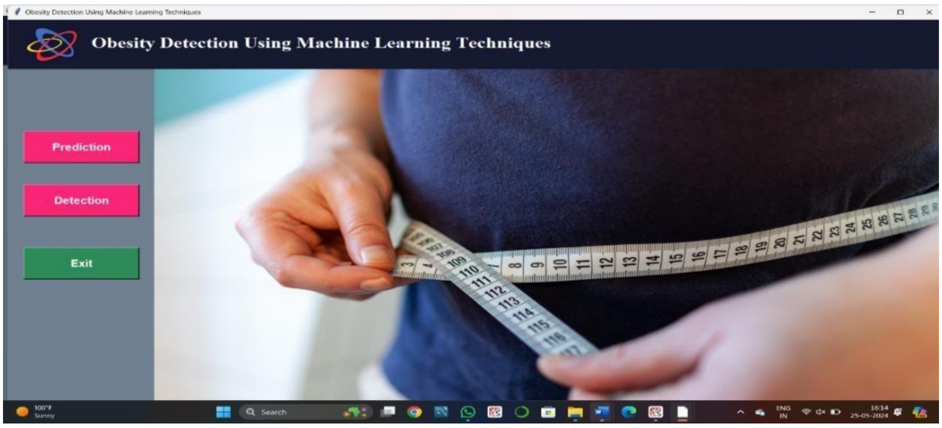

Figure 6.5.5. Check Page ODS

6.4.5. Prediction Page

Figure 6.5.6. Prediction Page ODS

6.4.6. Prediction Result (LOW)

Figure 6.5.7. Prediction Result (LOW) ODS

6.4.7. Prediction Result (Normal)

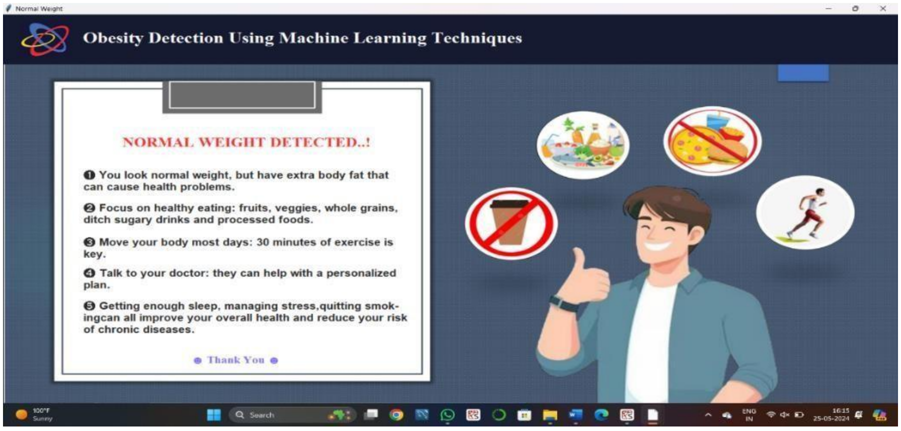

Figure 6.5.8. Prediction Result (NORMAL) ODS

6.4.8. Prediction Result (HIGH)

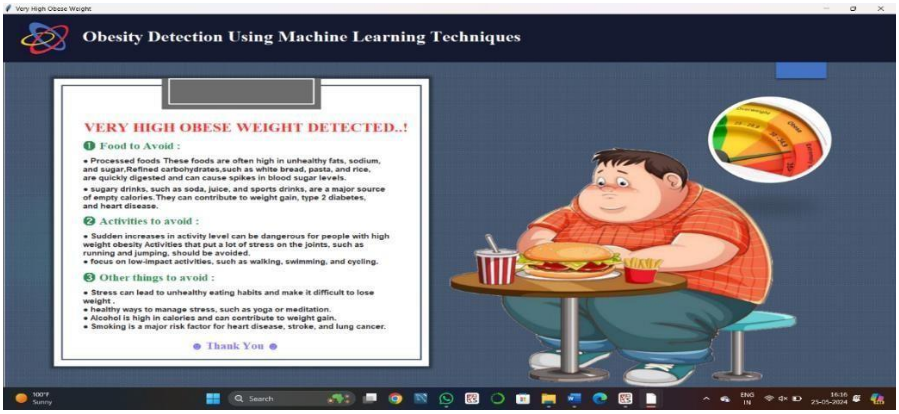

Figure 6.5.9. Prediction Result (VERY HIGH) ODS

6.4.9. Detection Result

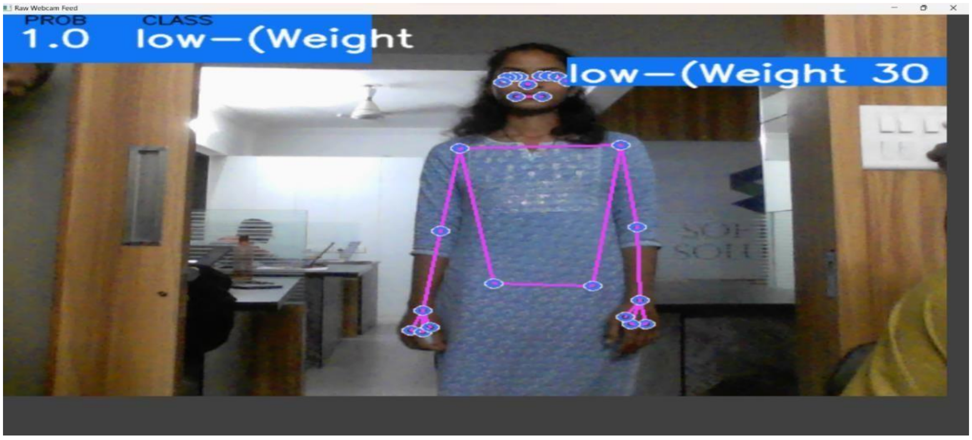

Figure 6.5.10. Detection Result ODS

6.5. Database Schema

6.5.1. Registration

registration		CREATE TABLE registration(Fullname TEXT, address TEXT, username TEXT, Email TEXT, Phoneno TEXT,Gender TEXT,age TEXT , password TEXT)
Fullname	TEXT	"Fullname" TEXT
address	TEXT	"address" TEXT
username	TEXT	"username" TEXT
Email	TEXT	"Email" TEXT
Phoneno	TEXT	"Phoneno" TEXT
Gender	TEXT	"Gender" TEXT
age	TEXT	"age" TEXT
password	TEXT	"password" TEXT

Figure 6.6.1. Registration Schema

6.6. Code Snippet

6.7.1 Login Page

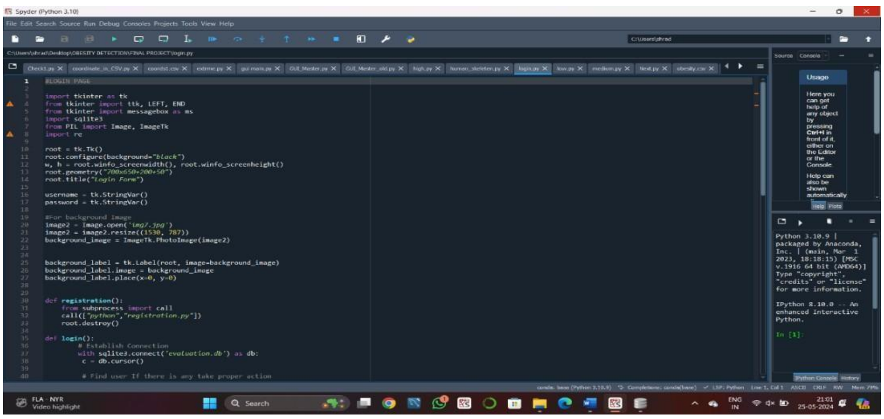

Figure 6.7.1 Login Page

6.7.2 Registration Page

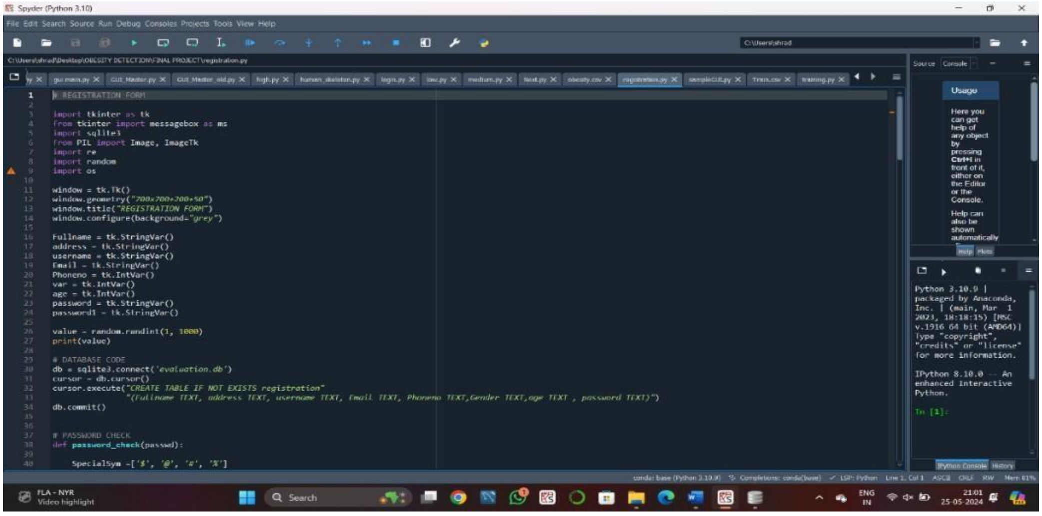

Figure 6.7.2 Registration Page

6.7.3 Dataset Training

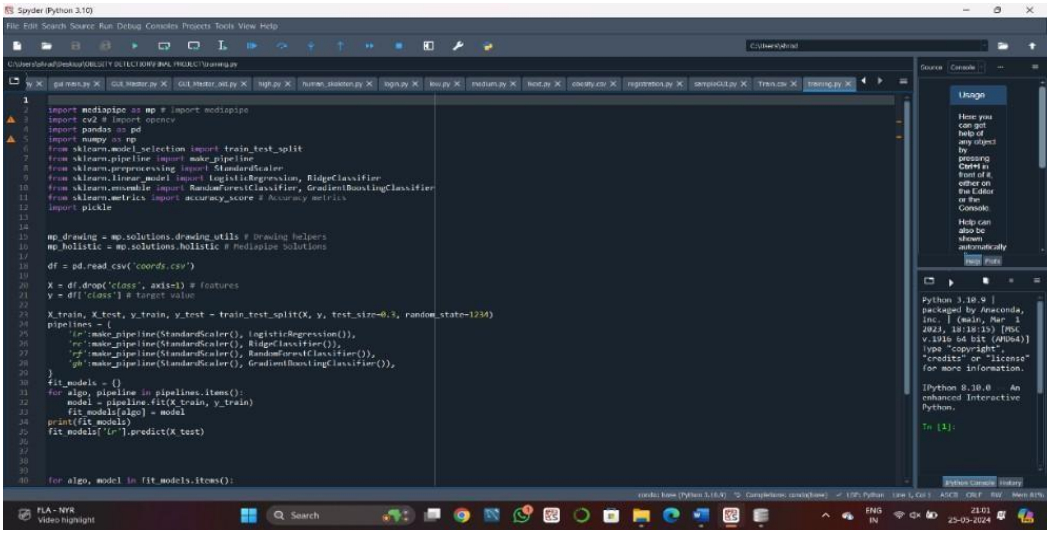

Figure 6.7.3 Dataset Training

6.7.4 Algorithm Page

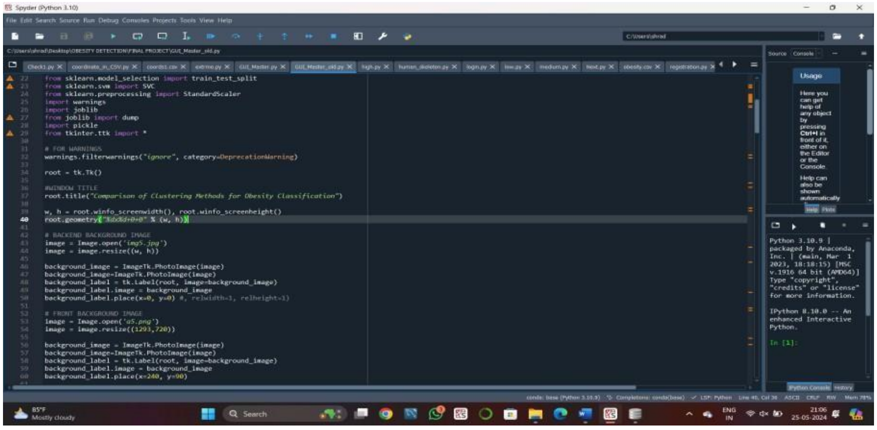

Figure 6.7.4 Algorithm Page

6.7.5 Prediction Page

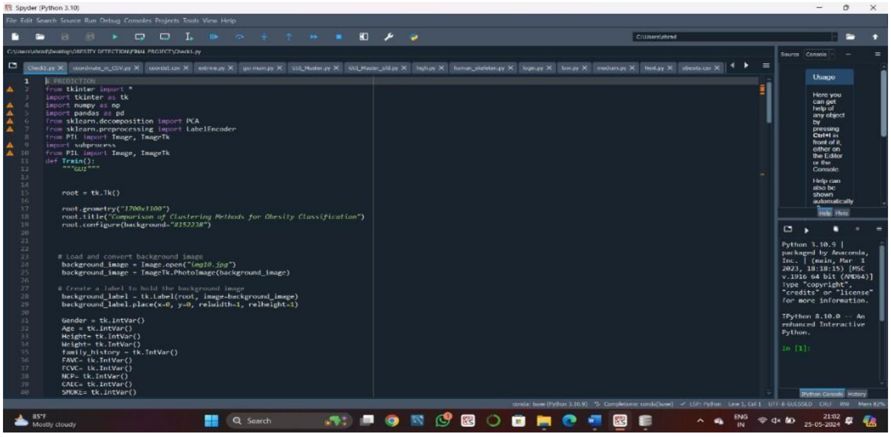

Figure 6.7.5 Prediction Page

6.7.6 Prediction Result (LOW)

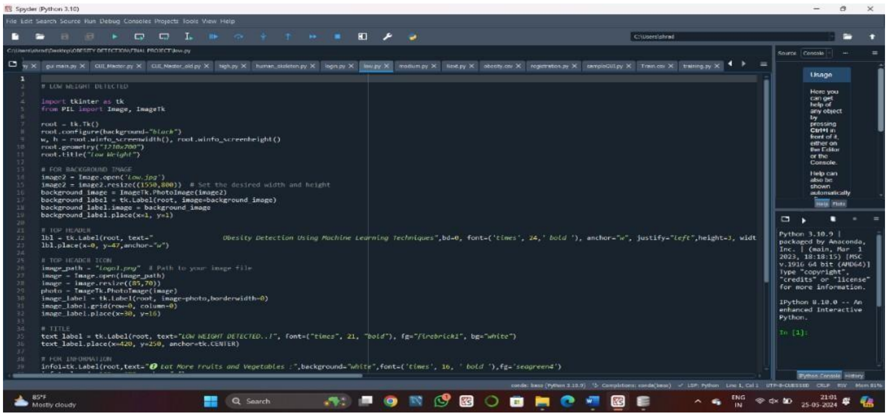

Figure 6.7.6 Prediction Result (LOW)

6.7.7 Prediction Result (HIGH)

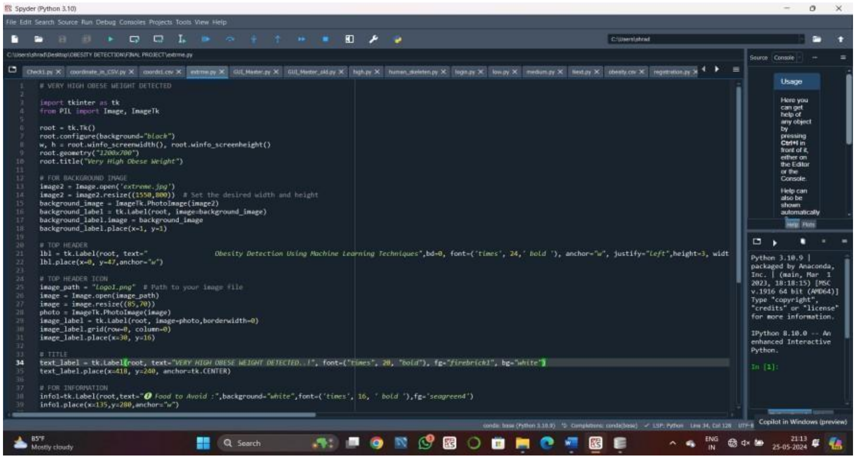

Figure 6.7.7 Prediction Result (HIGH)

6.7.8 Detection Page

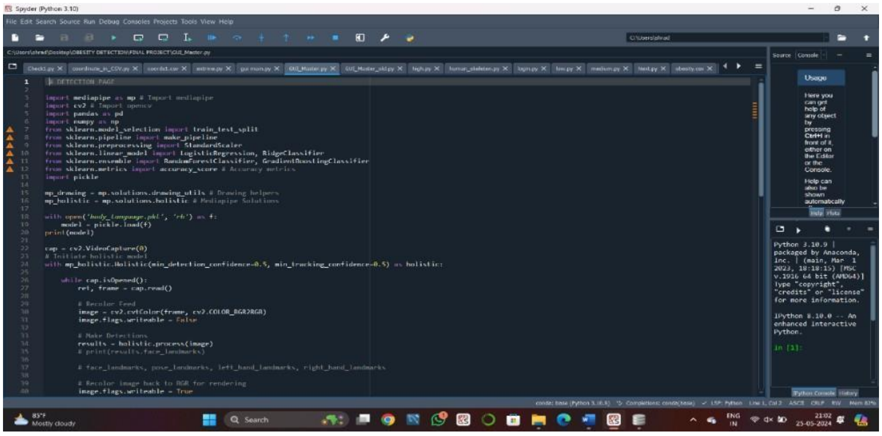

Figure 6.7.8 Detection Page

6.7.9 Human Skeleton

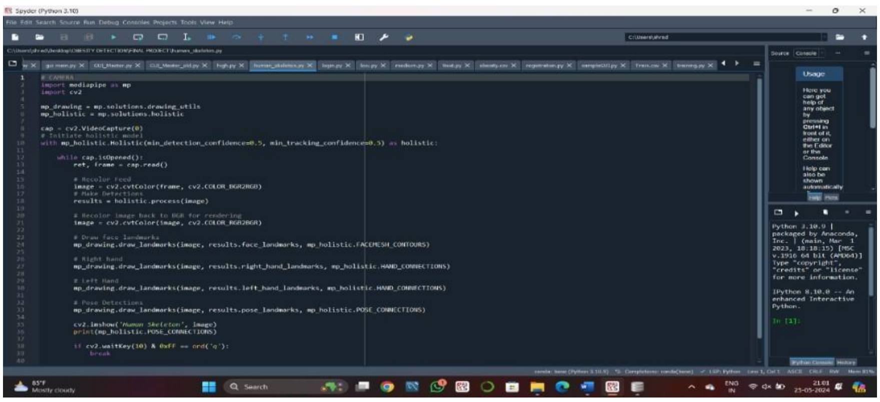

Figure 6.7.9 Human Skeleton

7. TESTING

Testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software Testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks in implementation of the software. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs. Software Testing, depending on the testing method employed can be implemented at any time in the development process. However, most of the test effort occurs after the requirements have been defined and the coding process has been completed. As such, the methodology of the test is governed by the Software Development methodology adopted.

7.1. Testing Approach: Manual Testing

Manual testing is a software testing process in which test cases are executed manually without using any automated tool. All test cases executed by the tester manually according to the end user's perspective. It ensures whether the application is working, as mentioned in the requirement document or not. Test cases are planned and implemented to complete almost 100 percent of the software application. Test case reports are also generated manually.

Manual Testing is one of the most fundamental testing processes as it can find both visible and hidden defects of the software. The difference between expected output and output, given by the software, is defined as a defect. The developer fixed the defects and handed it to the tester for retesting.

Manual testing is mandatory for every newly developed software before automated testing. This testing requires great efforts and time, but it gives the surety of bug-free software. Manual Testing requires knowledge of manual testing techniques but not of any automated testing tool.

Manual testing is essential because one of the software testing fundamentals is "100% automation is not possible.

7.1.1. Features

1. **Human Intuition and Creativity:** Leverage tester intuition for exploratory and ad-hoc testing to find unexpected defects.

2. **Flexibility:** Adapt and modify test cases dynamically during execution.

3. **User Interface (UI) and Usability Testing:** Effectively assess UI design and user-friendliness.

4. **Exploratory and Ad-Hoc Testing:** Conduct informal testing to discover unexpected issues.

5. **Accessibility Testing:** Manually evaluate accessibility features like screen readers and keyboard navigation.

6. **Interpersonal Communication:** Facilitate direct collaboration and feedback with development teams.

7. **Visual and Graphical Verification:** Identify visual discrepancies and layout issues.

8. **Non-Scripted Testing:** Test with varied inputs and scenarios, beyond predefined scripts.

7.1.2. Reason for Selection

- **Custom Scenarios:** Easily tests custom and one-off scenarios that may not be covered by automation.
- **Non-Technical Team:** Suitable for teams without advanced technical skills needed for automation.
- **Rapid Prototyping:** Useful in early prototyping stages where functionality changes rapidly.
- **Understanding Application Flow:** Helps testers gain a deeper understanding of the application's workflow.
- **Intermittent Issues:** Better at identifying and diagnosing intermittent issues that may be missed by automated tests.

7.2. Test Plan

A test plan is a document that outlines the strategy, objectives, resources, schedule, and scope of testing activities for a software project. It serves as a guide to ensure that testing is conducted systematically and efficiently.

Table 7.1. Test Plan

Sr. No.	Name of Tester	Test Item (Module/Webpage/Function etc.)	Planned Date
1.	Shraddha Borase	Login Page	03/04/2024
		Registration Page	
2.	Vaishnavi Bhoi	Algorithm Page	20/04/2024
		Next Button Navigation	
3.	Umesh Patel	Prediction	01/05/2024
		Submit Button Navigation	
4.	Vaishnavi Patil	Detection	20/05/2024

7.3. Test Case Design

Test case design is a critical step in the software testing process, ensuring that all aspects of the application are covered and tested thoroughly. Well-designed test cases help in identifying defects, validating functionality, and ensuring the software meets its requirements.

Table 7.2. Test Case (Manual Testing)

Sr. No.	Test Item	Purpose	Input(s)	Output(s)	Validation(s)
1.	Login Page	To make user log into system	- User Id - Password	Navigate to user account	YES

2.	Registration Page	Allow new users to create an account	-Full name -Address -E-mail - Phone num - Gender - User name - Password	success message confirming the creation of a new account or an error message indicating any issues encountered during registration.	NO
3.	Algorithm Page	Allows users to select a machine learning algorithm and see its accuracy	-	Confirmation messages, error notifications, and successful account creation	NO
4.	Next Button Navigation	correctly advances the user to the subsequent part of the application.	-	Properly display next page	NO
5.	Prediction	Collect user parameters to generate personalized predictions or calculations.	-Gender -Age - Height -Weight -Family History -FAVC -FCVC -NCP	Predicted health status or weight classification (e.g., high weight, low weight, overweight**)**.	YES

			-Smoke -CH20 -SCC -FAF -TUE -CALC -MTRANS		
6.	Submit Button Navigation	Display result on new page		shows the new page which display prediction results accurately based on the input data.	NO
7.	Detection	using a live camera feed to classify weight status into predefined ranges.	- Real time video feed capturing body images for analysis.	On-screen display of body weight classification (e.g., high weight, low weight, overweight) based on analysis results.	YES

7.4. Test Results

Table 7.3. Test Results (Manual Testing)

Sr. No.	Test Item	Errors	Bugs	Remark
1.	Login Page	None	None	Pass
2.	Registration Page	None	None	Pass
3.	Algorithm Page	None	None	Pass
4.	Next Button Navigation	None	None	Pass
5.	Prediction	None	None	Pass
6.	Submit Button Navigation	None	None	Pass
7.	Detection	None	None	Pass

8. PROJECT COST AND EFFORT

8.1. Estimation Technique: Detailed COCOMO

For the initial estimation of our project, we have used the first stage of COCOMO i.e., Basic COCOMO, now since our work is completed, we have all the necessary and actual information required for the cost calculation, hence here we will use Detailed COCOCMO. Detailed COCOMO incorporates all characteristics of the intermediate version with an assessment of the cost driver's impact on each step (analysis, design, etc.) of the software engineering process. The detailed model uses different effort multipliers for each cost driver attribute. These Phase Sensitive effort multipliers are used to determine the amount of effort required to complete each phase. In detailed COCOMO, the whole software is divided in different modules and then we apply COCOMO in different modules to estimate effort and then sum the effort.

Detailed COCOMO incorporates the set of "cost drivers" that include subjective assessment of product, hardware, personnel, and project attributes. The 17 cost drivers which are multiplicative factors that determine the effort required to complete our software project. Each of the 17 attributes receives a rating on a six-point scale that ranges from "very low" to "extra high" (in importance or value).

Table 8.1. Cost Drivers for Detailed COCOMO

Personnel Factors	
Analyst Capability (ACAP)	Platform Experience (PLEX)
Applications Experience (APEX)	Language and Tool Experience (LTEX)
Programmer Capability (PCAP)	Personnel Continuity (PCON)
Project Factors	
Use of Software Tools (TOOL)	Development Schedule (SCED)
Multisite Development (SITE)	
Platform Factors	
Execution Time Constraint (TIME)	Database Size (DATA)
Main Storage Constraint (STOR)	Product Complexity (CPLX)
Platform Volatility (PVOL)	Required Reusability (RUSE)

Required Software Reliability (RELY)	Documentation Match to Lifecycle Needs (DOCU)

After assigning rating to each of the cost drivers the ratings are multiplied together to yield **Effort Adjustment Factor (EAF)**.

The Detailed COCOMO formula takes the form:

$$\text{Effort, } E = a(KLoC)^{b} * EAF \text{ person months}$$

$$\text{Duration, } D = c(E)^{d} \text{ months}$$

$$\text{Person, } P = E/D \text{ persons}$$

where **E** is the effort applied in person-months, **KLoC** is the estimated number of thousands of delivered lines of code for the project, **D** is total time duration to develop the system in months, and **P** is number of persons required to develop that system.

The coefficients **a, c** and the exponent **b, d** are given in the following table.

Table 8.1.2 Coefficient/Exponent Values of Detailed COCOMO

Project Type	A	b	C	D
Organic	3.2	1.05	2.5	0.38
Semi-Detached	3.0	1.12	2.5	0.35
Embedded	2.8	1.20	2.5	0.32

8.2. Effort and Cost Calculation

8.2.1. Project Size

Table 8.2. Final Project Size

Software Module	LOC
Login	300
Registration	400
GUI Main	350
Check1	300
Human Skeleton	250

High	300
Low	350
Medium	350
Next	200
Coordinate_in_CSV	400
Sample_GUI	500
Extreme	350
GUI_Master	400
Training	450
GUI_Master_old	450
Total Lines of Code (LOC)	5350

8.2.2. Cost Drivers Selection

Table 8.3. Cost Drivers Selection

Cost Drivers	**Ratings**					
	Very Low	**Low**	**Usual**	**High**	**Very High**	**Extra High**
Personnel Factors						
Analyst Capability**(ACAP)**	1.46	**1.19**	1.00	0.86	0.71	---
Applications Experience**(APEX)**	1.29	**1.13**	1.00	0.91	0.82	---
Programmer Capability**(PCAP)**	1.42	1.17	**1.00**	0.86	0.70	---
Platform Experience **(PLEX)**	1.21	1.10	1.00	**0.90**	---	---
Language and Tool Experience**(LTEX)**	1.14	1.07	1.00	**0.95**	---	--
Personnel Continuity**(PCON)**	1.29	**1.12**	1.00	0.90	0.81	---
Project Factors						
Use of Software Tools**(TOOL)**	1.24	1.10	1.00	**0.91**	0.83	---
Multisite Development**(SITE)**	**1.24**	1.10	1.00	0.91	0.82	---
Development Schedule**(SCED)**	1.23	1.08	1.00	**1.04**	1.10	---

Platform Factors						
Execution Time Constraint**(TIME)**	---	---	**1.00**	1.11	1.30	1.66
Main Storage Constraint**(STOR)**	---	---	**1.00**	1.06	1.21	1.56
Product Factors						
Platform Volatility**(PVOL)**	---	**0.87**	1.00	1.15	1.30	---
Required Software Reliability**(RELY)**	0.75	0.88	1.00	1.15	**1.40**	---
Database Size **(DATA)**	---	0.94	**1.00**	1.08	1.16	---
Product Complexity**(CPLX)**	0.70	0.85	1.00	**1.15**	1.30	1.65
Required Reusability**(RUSE)**	---	**0.95**	1.00	1.07	1.15	1.24
Documentation Match to Life cycle Needs**(DOCU)**	0.81	0.91	**1.00**	1.11	1.23	---
Effort Adjustment Factor (EAF)	1.19*1.13*1.00*0.90*0.95*1.12*0.91*1.24*1.04*1.00 *1.00*0.87*1.40*1.00*1.15*0.95*1.00 = **2.01**					

8.2.3. Effort Calculation

The system falls into the **Semi_detached** category. The value a and b according to embedded system is a =**3.0** and b =**1.12**.

Total LOC (approx.) of project is**5350LOC = 5.350KLOC**

Effort (E) =$a(\text{KLoC})^{b}$ * EAF

$E=3.0*(5.350)^{1.12}*2.01$

E = 39.45≈ 39Person Months

8.2.4. Duration Calculation

The value c and d according to embedded system is c =**2.5** and d =**0.35**.

Duration (D) = $c\,(E)^{d}$

$D = 2.5*(39.45)^{0.35}$

D = 9.04 ≈ 9 Months

8.2.5. Person Required

Person Required = Effort Applied (E) / Development Time (D)

= 39/9

= 4.33≈4 Persons

8.2.6. Total Cost

Each team member has charged₹500/- per month with ₹1000/- spent for other resources & miscellaneous purposes in each month, in additional Microphone is purchased for ₹500/-. Thus,

Total Cost of System = ((Person Charges * Person Required) + Resource Charges) * Duration

=((500 * 5) + 1000) * 8 = **₹28,000/-**

8.3. Calculation Summary

Table 8.4. Summary of different calculations.

Calculation	Value
Size of the Project	5350 LoC
Effort Required	39 Person Months
Duration Required	9 Months
Person Required	4
Total Cost	₹28,000/-

9. CONCLUSION

Diagnosing obesity is difficult, as it is a complex disease that varies in nature. Improvement of the diagnosis of obesity is needed in the health sector to help reduce the risk/implications to the barest minimal. To detersssmine the obesity status of a patient, a physician must carry out some physical assessment on the patient and examine the results of a patient's test to assess prior judgments because it is subject to the doctor's interpretation. This project with the development of the Obesity prediction model would solve these issues of validity of diagnosis, time-consuming factor and also provide a reliable diagnosis system that can be used for all gender. A reliable and time-saving obesity prediction model has been developed in this research. The machine learning model was designed using the python programming language. Algorithms that were used to achieve the aim of this research are random forest classifier ,Support vector machine , Naive Bayes classifier ,Decision Tree Classifier. This prediction model is highly recommended for hospitals, clinics, diagnostic centers and the health sector in general, as it will help them to accurately predict obesity status in patients before it gets to a complex stage.

10. FUTURE SCOPE

Healthcare professions found it hard to find healthcare data and perform analysis on them due to lack of tools, resources. But using ML, we can overcome this and can perform analysis on real-time data leading to better modeling, predictions. This enhances and improves overall healthcare services. Now, IoTs being integrated with ML in order to make smart healthcare devices that sense if there is any change in the person's body, health data when he uses the device, and this will notify the person regarding this through an app. This helps in easy monitoring, advanced prediction, and analysis there by reducing errors, saving time and life of people.

REFERENCES

[1] *Yihan Deng, Peter dolog, Jorn-Markus gass ,"Obesity entity extraction from real outpatient records: When learning based methods meet small imbalanced medical data sets",2019.*

[2] *Christopher A. Flores, Rosa L. Figueroa "FREGEX: A Feature Extraction Method for Biomedical Text Classification using Regular Expressions",2019.*

[3] *Naomi Christianne Pereira,Supriya solaskar,"OBESITY RELATED DISEASE PREDICTION FROM HEALTHCARE COMMUNITIES USING MACHINE LEARNING",2020.*

[4] *S. A. Pullano, M. Greco, D.P. Foti,"FT–IR Saliva Profiling in Patients with Obesity and Obesity–Related Insulin Resistance",2018.*

[5] *Phaik-Ling Ong, Yun-Huoy Choo,Nurul A. Emran,"Classification of SNPs for Obesity Analysis using FARNeM Modeling",2013.*

[6] *Thiago Orsi, Ernesto Araujo, Ricardo Simoes, "Fuzzy Chest Pain Assessment for Unstable Angina based on Baranwal Symptomatic and Obesity Clinical Conditions",2015.*

[7] *CASIMIRO A. CURBELO MONTANEZ, PAUL FERGUS, "SAERMA: Stacked Autoencoder Rule Mining Algorithm for the Interpretation of Epistatic Interactions in GWAS for Extreme Obesity",*

[8] *Susana Abe Miyahira, Ernesto Araujo, Member, IEEE "Fuzzy Obesity Index for Obesity Treatment and Surgical Indication",2020.*

[9] *Kunal Rajput Faculty of Science & Technology University of Canberra, "Obesity and Co-Morbidity detection in Clinical text using Deep Learning and Machine Learning techniques",2018.*

[10] *Muhammad Nouman, Surapong Chat pun Institute of biomedical engineering, Faculty of Medicine, Prince of Songkhla University, "Plantar pressure distribution in non-obese, overweight and obese subjects with diabetic neuropathy while walking",2018.*

[11] *Miguel Castorena∗, Nguyen Doan†, Benjamin Gillmore‡ and Jimmy Lahn§, Joshua Lorenzen, "Overweight Object Transportation with a Set of*

Collaborative Robots",2020.

[12] *Chellaswamy C Department of Electronics and Communication Engineering Kings Engineering College Chennai, "Machine Learning Based Condition Recognition System for Bikers",2020.*

[13] *Juan Montenegro1, Lilia Muñoz1,2, Mel Nielsen1, Vladimir Villarreal1,2 1 Grupo de Investigación en Tecnologías Computacionales Emergentes, Universidad Tecnológica de Panamá, Panamá 2 Centro de Estudios Multidisciplinarios en Ciencias, Ingeniería y Tecnología – CEMCIT AIP, Panamá "Plataforma para el monitoreo de la obesidad y el sobrepeso Platform for monitoring obesity and overweight",2020.*

[14] *Eliana Marina Alvarez Di Fino Esc. de Nutricion, FCM, UNC, "Spatial analysis applied to nutritional epidemiology",2019.*

[15] *Ramy El Jalbout, "Carotid artery noninvasive elastography (NIVE) to detect early changes of cardiovascular diseases in overweight and obese children".2019.*

[16] *Vasile Manoliu1, Member IEEE, Cristina Maria Stanescu 2 1 Faculty of Electrical Engineering, Politehnica University of Bucharest, Romania, "A Study of the Cardiovascular Function and Aerobic Capacity in Overweight and Obese Subjects",2019.*

[17] *Carlos Martinez-Millana, Antonio Martinez-Millana*, Carlos Fernandez-Llatas ITACA Universitat Politecnica de Val ` encia ` Valencia, Spain ,"CompArin-g data base engines for building big data analytics in obesity detection",2019.*

[18] *Saptashwa Mitra∗ , Yu Qiu∗ , Haley Moss†, Kaigang Li† and Sangmi Lee Pallickara, "Effective Integration of Geotagged, Ancilliary Longitudinal Survey Datasets to Improve Adulthood Obesity Predictive Models" ,2018.*

[19] *Naomi Christianne Pereira1, Jessica D'souza2, Parth Rana3, Supriya Solaskar,"OBESITY RELATED DISEASE PREDICTION FROM HEALTHCARE COMMUNITIES USING MACHINE LEARNING",2019.*

[20] *UM Chaskar Department of Instrumentation and Control Engineering College of Engineering Pune, "Machine Learning Approach to Study the Impact of Obesity on Autonomic Nervous System using Heart Rate Variability*

Features",2021.

[21] *Shivam Garg Dept. of CSE IIIT Kota Jaipur, India,"Obesity Risk Prediction using Wi- Fi based Human Activity Recognition Techniques",2021.*

[22] *Özgü Taçyıldız, 2 Duygu Çelik Ertuğrul, 3 Yıltan Bitirim, 4 Neşe Akcan, 5 Atilla Elçi, "Ontology-based Obesity Tracking System for Children and Adolescents" ,2018.*

[23] *Haofeng Yan, Yu Wang, Tingdong Fang, Xiaofei Chen, "Evaluation Method of Operation State in Overweight Rejection based on Unary Linear Regression Model",2023.*

[24] *Hera siddiqui 1, ajita rattani 1, nikki k. Woods2, laila cure3, rhonda k.lewis4,"A Survey on Machine and Deep Learning Models for Childhood and Adolescent Obesity",2021.*

[25] *GRAZIA V. AIOSA1, MAURIZIO PALESI1, (Senior Member, IEEE),AND FRANCESCA SAPUPPO 2, (Member, IEEE), "Explainable AI for Decision Support to Obesity Comorbidities Diagnosis",2023.*

[26] *LEIRE BASTIDA, GLORIA CEA, ANA MOYA, ALBA GALLEGO, "Promoting Obesity Prevention and Healthy Habits in Childhood: The OCARIoT Experience",2022.*

[27] *Rodolfo Canas ˜ Cervantes, Ubaldo Martinez Palacio, "Estimation of obesity levels based on computational intelligence",2020.*

[28] *Mahmood Safaei a, Elankovan A. Sundararajan a, Maha Driss b,c, Wadii Boulila b,c, Azrulhizam Shapi'i d, "A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learningapproaches used to predict obesity",2021.*

[29] *Marie-Eve Piché, André Tchernof, Jean-Pierre Després, "Obesity Phenotypes, Diabetes, and Cardiovascular Diseases",2020.*

[30] *Ayan Chatterjee, Martin W. Gerdes and Santiago G. Martinez, "Identification of Risk Factors Associated with Obesity and Overweight—A Machine Learning Overview",2019.*

[31] *NileshP.Sable, Anay Indani, Raj Mehta, Ratmmala Bhimanpallewar, Sara Shaikh, Soham Jadhav "A Machine Learning approach for Early Detection and Preventionof Obesity and Overweight",2023.*

APPENDIX - A

bias: could mean the intercept (e.g. in neural nets), typically refers to the bias in bias-variance decomposition

classifier: specific model or technique (i.e. function) that maps observations to classes

confusion matrix: a table of predicted class membership vs. true class membership3

hypothesis: a specific model h(x)of all possible in the hypothesis space H input, feature, attribute: independent variable, explanatory variable, covariate, predictor variable, column instance, example: observation, row learning: model fitting

machine learning: a form of statistics utilizing various algorithms with a goal to generalize to new data situations

regularization, penalization, shrinkage: The process of adding a penalty to the size of coefficients, thus shrinking them towards zero but resulting in less overfitting (at an increase to bias)

supervised: has a dependent variable

target, label: dependent variable, response, the outcome of interest

unsupervised: no dependent variable (or rather, only dependent variables); think clustering, PCA etc.

weights: coefficients, parameters

Accuracy decrease (classification) : mean decrease of prediction accuracy after XjX is permuted

Gini decrease (classification) : mean decrease in the Gini index of node impurity (i.e. increase of node purity) by splits on XjX

Mse increase (regression) : mean increase of mean squared error after XjX is permuted

Node purity increase (regression) :mean node purity increase by splits on XjX as measured by the decrease in sum of squares

Mean minimal depth :mean minimal depth calculated in one of three ways specified by the parameter mean_sample

No of trees : total number of trees in which a split on XjX occurs

No of nodes : total number of nodes that use XjX for splitting (it is usually equal to no_of_trees if trees are shallow)

Times a root : total number of trees in which XjX is used for splitting the root node (i.e., the whole sample is divided into two based on the value of XjX)

P value : p-value for the one-sided binomial test using the following distribution:

Hyperplane: Hyperplane is the decision boundary that is used to separate the data points of different classes in a feature space. In the case of linear classifications, it will be a linear equation i.e. wx+b = 0.

Support Vectors: Support vectors are the closest data points to the hyperplane, which makes a critical role in deciding the hyperplane and margin.

Margin: Margin is the distance between the support vector and hyperplane. The main objective of the support vector machine algorithm is to maximize the margin. The wider margin indicates better classification performance.

Kernel: Kernel is the mathematical function, which is used in SVM to map the original input data points into high-dimensional feature spaces, so, that the hyperplane can be easily found out even if the data points are not linearly separable in the original input space. Some of the common kernel functions are linear, polynomial, radial basis function(RBF), and sigmoid.

Hard Margin: The maximum-margin hyperplane or the hard margin hyperplane is a hyperplane that properly separates the data points of different categories without any misclassifications.

Soft Margin: When the data is not perfectly separable or contains outliers, SVM permits a soft margin technique. Each data point has a slack variable introduced by the soft-margin SVM formulation, which softens the strict margin requirement and permits certain misclassifications or violations. It discovers a compromise between increasing the margin and reducing violations.

C: Margin maximisation and misclassification fines are balanced by the regularisation parameter C in SVM. The penalty for going over the margin or misclassifying data items is decided by it. A stricter penalty is imposed with a greater value of C, which results in a smaller margin and perhaps fewer misclassifications.

Hinge Loss: A typical loss function in SVMs is hinge loss. It punishes incorrect classifications or margin violations. The objective function in SVM is frequently formed by combining it with the regularisation term.

Dual Problem: A dual Problem of the optimisation problem that requires locating the Lagrange multipliers related to the support vectors can be used to solve SVM. The dual formulation enables the use of kernel tricks and more effective computing

A Machine Learning approach for Early Detection and Prevention of Obesity and Overweight

[1]Nilesh P. Sable
Bansilal Ramnath Agarwal Charitable Trust's,
Vishwakarma Institute of Information Technology, Pune, India
Savitribai Phule Pune University, Pune, India
drsablenilesh@gmail.com

[2]Ratnmala Bhimanpallewar
Bansilal Ramnath Agarwal Charitable Trust's,
Vishwakarma Institute of Information Technology, Pune, India
Savitribai Phule Pune University, Pune, India
ratnmalab@gmail.com

[3]Raj Mehta
Bansilal Ramnath Agarwal Charitable Trust's,
Vishwakarma Institute of Information Technology,
Pune.
raj.21910186@viit.acin

[4]Sara Shaikh
Bansilal Ramnath Agarwal Charitable Trust's,
Vishwakarma Institute of Information Technology,
Pune
sara.21910006@viit.ac.in

[5]Anay Indani
Bansilal Ramnath Agarwal Charitable Trust's,
Vishwakarma Institute of Information Technology,
Pune.
anay.21911075@viit.ac.in

[6]Soham Jadhav
Bansilal Ramnath Agarwal Charitable Trust's,
Vishwakarma Institute of Information Technology,
Pune
Soham.21910380@viit.acin

***Abstract*- More than 2.1 billion people worldwide are shuddering from overweightness or obesity, which represents approximately 30% of the world's population. Obesity is a serious global health problem. By 2030, 41% of people will likely be overweight or obese, if the current trend continues. People who show indications of weight increase or obesity run the danger of contracting life-threatening conditions including type 2 diabetes, respiratory issues, heart disease, and stroke. Some intervention strategies, like regular exercise and a balanced diet, might be essential to preserving a healthy lifestyle. Thus, it is crucial to identify obesity as soon as feasible. We have collected data from sources like schools and colleges within our organization to create our dataset. A vast range of ages is considered and the BMI value is examined in order to determine the level of obesity. The dataset of people with normal BMI and those at risk has an inherent imbalance. The outcomes are collected and showcased via a website which also includes various preventive measures and calculators. The outcomes are promising, and clock an accuracy of about 90%.**

Keywords- Obesity, Data Driven Model, Classification, Prediction, Prevention, Webpage, Machine Learning

I. INTRODUCTION

Due to the numerous detrimental health effects it is linked to, childhood obesity is a severe public health problem. [1] Obese children are more likely to acquire chronic diseases including type 2 diabetes, cardiovascular disorders, and some forms of cancer. [2] A second factor that may enhance the risk of these illnesses developing in fat children is the likelihood that they will grow up to be obese adults. In recent decades, childhood obesity has become more common, especially in affluent nations, but it has also become a bigger issue in underdeveloped nations. [3,4]

Body mass index, or BMI, is a metric used to assess a person's weight in relation to their height. It is determined by multiplying a person's height in square meters by their weight in kilos. [4] BMI is a regularly used technique to determine the likelihood of acquiring obesity and related health issues, despite the fact that it is not a perfect indicator of body fatness. [5]

In order to put preventive measures in place, it is crucial to identify those who are at risk of becoming obese as early as feasible. This may include interventions such as dietary changes, increased physical activity, and behavior modification. By addressing obesity early on, it is possible to prevent or reduce the risk of negative health outcomes and improve overall health and well-being. [6]

It is necessary to have a solid framework that forecasts BMI or obesity levels, body fat percentage, and suggests personalized preventative measures. A programme is created and included into the framework that explains the preventative actions that are required for every individual.

Data was collected through 2 schools and our college via online forms and emails. The data collected was nearly —---. We combined the data collected via the dataset and our newly collected data to create the most efficient and realistic model possible.

The website created to display the findings and outcomes is created using HTML, CSS and Javascript frameworks. It is then integrated with the machine learning model using Node.js . The entire data collected is also stored on cloud computing using MongoDB in order to ease the entire process and adapt to increasing data volume.

II. RELATED WORK

Here are some reviews made by the authors of various papers. In [7], The study looks at how common overweight and obesity were among kids in England in 2015. By age, sex, and wealth, it describes variations between various groups of kids. Comparisons are made between parents' and kids' perceptions of their weight and actual measurements of body mass. There is also discussion of historical trends in childhood obesity. The goal of [8] is to develop a machine-learning-based approach for estimating the risk of obesity. The amazing thing about this article is that it will inform readers about the dangers of obesity and its causes. They gather more than 1100 data from a wide range of individuals of all ages, including both those who are obese and those who are not. In [9], the authors implemented and compared machine learning classification methods such KNN, XGB, Logistic Regression, and DT to obesity data. Obesity is defined as a BMI of 30 or more, making it an extremely complicated condition. Obesity will negatively impact quality of life in ways including sadness, decreased productivity, and disability. [10] examines the state of using machine learning algorithms to studies of obesity and identifies the advantages and disadvantages of these techniques. An analysis of papers that specifically addressed obesity was done. Different keywords were used to search the PubMed and Scopus databases for machine learning applications in obesity. Additionally, only studies that concentrated on modifiable factors (such as nutrition intake, dietary patterns, and/or physical activity) under our control were thoroughly examined. The authors of [11] investigated if the SCT ideas might predict whether or not children of African American caregivers will become obese. The paper-based survey on caregiver opinions of childhood obesity was completed by 128 caregivers. Expectancies were examined using a multiple linear regression (p .05). Using the vast amount of data made available by the UK's Millennium Cohort Study, the authors of [12] develop a machine learning-driven model to identify young people at risk of becoming overweight or obese. The childhood BMI values from years 3, 5, 7, and 11 are taken into consideration for the goal of identifying adolescents who will be at risk of becoming overweight or obese by the age of 14. Authors of [13] discover a thorough and critical evaluation of the field of machine learning models to forecast childhood and adolescent obesity and associated consequences, including the most recent ones utilising deep learning with EHR. These models are contrasted with more established statistical ones, which mostly made use of logistic regression. Future potential are explored along with the key characteristics and applications resulting from these models. The writers of [14] gathered and evaluated data from 272 826 kids. A) The whole study population, which included children ages 4 to 16, was observed. Between 1999 and 2003, ov/ob prevalence showed a large increase trend, and between 2004 and 2008, a significant decreasing trend. c) Analyses of subgroups. Ov/Ob prevalence rose in the majority of the examined subgroups until 2004. The authors' [15] study set out to conduct a thorough literature review and update the data on the prevalence of childhood overweight. A computerised bibliographical search was done, although it was only for research that used either a prospective or retrospective longitudinal design. Data was independently gathered by two writers, who then evaluated the included research' methodological quality across four aspects. The authors of [16] investigated the effectiveness of several body mass index (BMI) cut points in detecting children with excess adiposity (based on skinfold thicknesses), unfavourable lipid, insulin, and blood pressure levels, and a high risk for severe adult obesity. The author of [17] [18] [19]states that the false expectations that result might be detrimental to their socialisation. Childhood abnormalities cause many of the cardiovascular consequences that are typical of adult-onset obesity. Hyperlipidemia, hypertension, and impaired glucose tolerance are more prevalent in obese children and adolescents.

III. METHODOLOGY

In the United States and across the world, childhood obesity is still a severe health problem. The habits that kids take up from their parents and other adults are a major factor in childhood obesity. By educating your kid about appropriate eating practices and motivating them to keep active, you can avoid childhood obesity. This model is determined to detect and then provide some preventive measures over obesity. For children of that age, overweight and obesity are characterized as follows:

- Overweight is defined as having a BMI for one's age that is more than one standard deviation over the WHO Growth Reference median.
- Obesity exceeds the WHO Growth Reference median by more than two standard deviations.

For children and adolescents aged 14-21 years in 2017-2020[1]:

- 14.7 million kids and teenagers were afflicted by the 19.7% obesity prevalence rate.
- 20.7% of children aged 6 to 11 and 22.2% of children aged 12 to 19 were obese. Additionally, certain communities are more likely to have children who are obese.

High blood pressure, high cholesterol, type 2 diabetes, and respiratory issues are among illnesses linked to obesity.

Considering the current scenario, we can see that researchers have only considered BMI for obesity detection but whereas in our project we would have 11 different parameters for finding out whether a given young adult is obese or not.

The Dataset we are using is in the format .csv and it is a predefined dataset that we are using to implement this model.

The dataset we are using has 17 attributes or 17 columns and 2112 rows or data tuples in it. The attributes are both numeric as well as string type. The Dataset is broadly divided into 4 parts.

A. Attributes related to eating habits:

These attributes give a brief description about the eating habits of the individual and how aware he is of the macros he is consuming.

Frequent consumption of high caloric food (FAVC) (String types with 2 classes yes or no).

Frequency of consumption of vegetables (FCVC). This parameter basically gives information about how many meals in a day are vegetables. (Numeric Value in the range 1 to 4).

Number of main meals (NCP). Count of how many proper clean meals an individual eats in a day. (Numeric Value in the range 1 to 4).

Consumption of food between meals (CAEC). (Divided into 4 classes No, Sometimes, Frequent, Always).

Consumption of water daily (CH20) (Numeric Value measured in liters).

Consumption of alcohol (CALC). (Divided into 4 classes No, Sometimes, Frequent, Always).

B. Attributes related to physical condition.

The below given attributes provide a brief idea about physical condition or the movements of a person in general or basically the lifestyle of a person.

Calories consumption monitoring (SCC) (String types with 2 classes yes or no).

Physical activity frequency (FAF) (Numeric Value in the range 0 to 3).

Time using technology devices (TUE) (In the range 0 to 2).

Transportation used (MTRANS) (String type divided into 4 classes.).

C. General Attributes.

Gender (String type with 2 values Male and Female).

Age (Numeric Type in the range from 16 to 25).

Height (Numeric Type and measured in metres).

Weight (Numeric Type and measured in Kilogram).

Family History with overweight (String types with 2 classes yes or no).

SMOKING (Divided into 2 classes YES or NO).

The dataset is free of missing values or any NULL values. The target attribute is divided into 7 different classes.

Insufficient_Weight, Overweight_Level_I, Normal Weight, Obesity_Type_III, Obesity_Type_I, Overweight_Level_II, Obesity_Type_II. For Deep Understanding of these classes and are they independent or not we used K-Means clustering algorithm for 7 clusters. With this we got a proper clarity of 4 independent classes namely Insufficient_Weight, Overweight, Obesity, Normal_Weight. Further Obesity and Overweight is divided into 3 and 2 types respectively. The Dataset is an example of multiclass classification. Initially we took each and every attribute into consideration for determining obesity which was about 16 attributes. In order to reduce the data dimensionality and cost we choose to remove the values that were less impactful on the dataset and also remove the outliers from the dataset. The pre-processed dataset or the final training dataset consisted of 11 attributes.

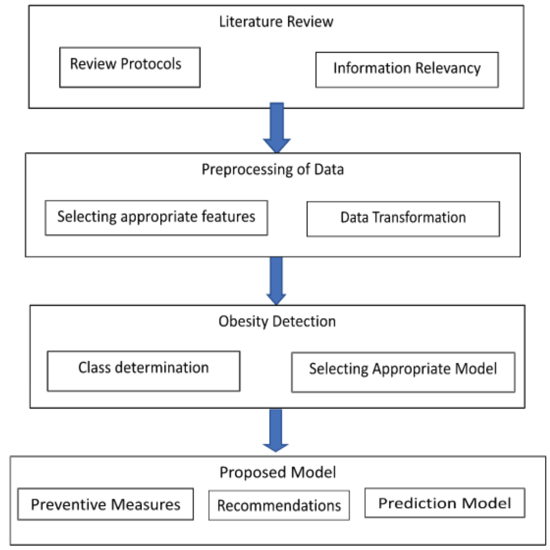

Fig. 1. Methodology

D. Algorithms

ML Classification Techniques used for the model:

1) K Nearest Neighbours

One data mining technique, the K-nearest neighbour (KNN) method, separates a collection of data based on learning previously labelled or classified data. KNN is a type of supervised learning that identifies newly discovered query instances based mostly on how closely they resemble preexisting categories.

2) SVM

SVM delivers evidence and a distinct theoretical foundation. SVM has been employed by several agents to address issues with text detection, character recognition object identification, and human activity recognition . For situations requiring regression, the statistical learning systems offered the segmentation depending on the "A" theoretical equations. A great element of SVM is the availability of credible methodologies and tools for addressing issues promptly and efficiently.

3) Random Forest

Random forest employs many decision trees (CART), which are then put together to produce the much more accurate result available.

Two applications that rely on decision tree design during training are regression and classification; a different ensemble learning strategy is random decision forest.

The decision tree approach generates classification probabilities using the Gini index technique, which assesses the likelihood that a certain piece will in fact be incorrectly categorised. The system's capacity to forecast is improved by this technique.

The decision tree model's bias has to be removed for the system to be effective.

4) Gradient Boosting

The specific needs of the algorithm can be met by heavily tailoring machines that enhance gradients. Intricate non-linear function connections can be captured well using support vector machines. This set of models has proven to be quite effective in a variety of practical situations. The GBMs are also appropriate and relevant and easy to customize to satisfy a broad range of needs in the real world.

IV. RESULTS AND DISCUSSION

We created a simplistic website where users can enter their data and can gain information about the health class. The front end was created using html CSS and JavaScript. We used mongo Db as our database where we stored our dataset which keeps on increasing as and when new users enter their information. The data is sent to mongo db using mongoose and it is fetched in python as a dictionary by using Py-mongo. Then it is converted into executable format and given to a model for prediction. Depending upon model prediction a web page request is sent back to the user.

In Table No.1, we tested different ML models such as KNN, Random Forest, Gradient Boosting, and SVM for determining the obesity level in a person. And we found out that Gradient Boosting gives us the best accuracy of 97% and hence considering that we trained and tested our model using that algorithm. And then the output is presented.

TABLE I. COMPARATIVE ANALYSIS BETWEEN DIFFERENT ML MODELS

Model	Accuracy
KNN	0.85
Random Forest	0.84
Gradient Boosting	0.97
SVM	0.75

In this Table 2, we can see that we compared previous year papers for the algorithm they have used and their accuracy, where we found out that our paper had an accuracy of 97% as compared to the paper “A Machine Learning Approach for Obesity Risk Prediction” having second highest accuracy of 93%.

TABLE II. COMPARISON ANALYSIS WITH PREVIOUS WORK

Name of Paper	Accuracy
MOFit: A Framework to reduce Obesity using Machine learning and IoT	0.73
Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research	0.72
Obesity Level Estimation Software based on Decision Trees	0.90
A Machine Learning Approach for Obesity Risk Prediction	0.93
Proposed Model	**0.97**

V. CONCLUSION

Obesity is a condition that affects individuals everywhere, regardless of their socioeconomic or cultural status. Since 1980, the number of adults who suffer from obesity has quadrupled, reaching more than 1900 million in 2014. Many methods and solutions have been created to be able to identify or forecast the onset of the disease in order to aid in the battle against it.

In this study, we utilized many strategies to reach the best precision rates to diagnose obesity, which had 97% precision using gradient boosting algorithms. Data mining is a crucial tool that allows us to uncover information. The website created is an excellent tool in aiding users to get information about the preventive measures and implement new, fresh objectives in their health plan.

REFERENCES

[1] Singh, A.S., Mulder, C., Twisk, J.W.R., Van Mechelen, W., Chinapaw, M.J.M.: Tracking of childhood overweight into adulthood: a systematic review of the literature. Obes. Rev. 9, 474–488 (2008)

[2] Freedman, D.S., Mei, Z., Srinivasan, S.R., Berenson, G.S., Dietz, W.H.: Cardiovascular risk factors and excess adiposity among overweight children and adolescents: the Bogalusa heart study. J. Pediatr. 150, 12–17 (2007)

[3] Raj M, Sundaram KR, Paul M, Deepa AS, Kumar RK. Obesity in Indian children: Time trends and relationship with hypertension. Natl Med J India. 2007;20:288–93. [PubMed] [Google Scholar]

[4] Laxmaiah A, Nagalla B, Vijayaraghavan K, Nair M. Factors affecting prevalence of overweight among 12 to 17 year old urban adolescents in Hyderabad, India. Obesity (Silver Spring) 2007;15:1384–90. [PubMed] [Google Scholar]

[5] Subramanyam V, R,J, Rafi M. Prevalence of overweight and obesity in affluent adolescent girls in Chennai in 1981 and 1998. Indian Pediatr. 2003;40:332–6. [PubMed] [Google Scholar]

[6] Panjikkaran ST, Kumari K. Augmenting BMI and Waist-Height Ratio for establishing more efficient obesity percentiles among school children. Indian J Community Med. 2009;34:135–9. [PMC free article] [PubMed] [Google Scholar]

[7] Health Survey for England 2015 Children’s body mass index, overweight and obesity Published 14th December 2016

[8] A Machine Learning Approach for Obesity Risk Prediction August 2021 DOI:10.1016/j.crbeha.2021.100053

[9] Comparative evaluation of machine learning classifiers with Obesity dataset A. Ramya; K. Rohini

[10] A review of the application of machine learning in adult obesity studies Mohammad Alkhalaf1,2, Ping Yu1,3, *, Jun Shen1 and Chao Deng3,4

[11] Examining Whether the Social Cognitive Theory Concepts Predict Childhood Obesity Prevention Outcome Expectation Dayna S. Alexander Chunhua Cao, Moya L. Alfonso

[12] Machine Learning Approach for the Early Prediction of the Risk of Overweight and Obesity in Young People Balbir Singh & Hissam Tawfik

[13] Machine Learning Models to Predict Childhood and Adolescent Obesity: A Review by Gonzalo Colmenarejo.

[14] Age-specific stabilization in obesity prevalence in German children: A cross-sectional study from 1999 to 2008: Susann Blüher,Christof Meigen,Ruth Gausche,berhard Keller,Roland Pfäffle,Matthew Sabin,George Werther,Rasha Odeh &Wieland Kiess

[15] Tracking of childhood overweight into adulthood: a systematic review of the literature A. S. Singh,C. Mulder,J. W. R. Twisk,W. Van Mechelen,M. J. M. Chinapaw.

[16] Cardiovascular Risk Factors and Excess Adiposity Among Overweight Children and Adolescents: The Bogalusa Heart Study from Freedman DS, Mei Z, Srinivasan SR, Berenson GS, Dietz WH.

[17] Health Consequences of Obesity in Youth: Childhood Predictors of Adult Disease William H. Dietz, MD

[18] Ghodke, Sayali Vijay, Arti Sikhwal, And Dr Nilesh Sable. "Novel Approach of Automatic Disease Prediction And Regular Check-Up System Using Ml/Dl." Design Engineering (2021): 2885-2896.

[19] Mahalle, P.N.; Sable, N.P.; Mahalle, N.P.; Shinde, G.R. Predictive Analytics of COVID-19 Using Information, Communication and Technologies. Preprints 2020, 2020040257 (doi: 10.20944/preprints202004.0257.v1).

Obesity Detection Using Machine Learning Techniques

Ashphak Khan[1], Shraddha Borse[2], Vaishnavi Bhoi[3], Vaishnavi Patil[4], Umesh Patel[5]
Department of Computer Engineering, P.S.G.V.P. Mandal's D. N.Patel College of Engineering, Shahada, Dist.-Nandurbar (Maharashtra)

***Abstract*—Obesity has become a global health concern, with its prevalence reaching alarming levels in recent years. Effective management of obesity requires a clear understanding of its classification, which enables healthcare professionals to tailor interventions and develop personalized treatment strategies. This review aims to provide a comprehensive overview of obesity classification systems, highlighting their strengths, limitations, and implications for clinical practice. The review begins by discussing the commonly used anthropometric measures for assessing obesity, such as body mass index (BMI), waist circumference (WC), and waist-to-hip ratio (WHR). It explores the advantages and drawbacks of these measures, including their inability to accurately account for variations in body composition and distribution of adipose tissue. For this research, we apply prominent machine learning algorithms. We used the algorithm of random forest, logistic regression, Decision Tree, support vector machine (SVM), and we have measured the performance of each of these classifications in terms of some prominent performance metrics. From the experimental results, we determine the obesity of high, medium, and low.**

Keywords —Diagnosis, DT, Machine Learning, RF, SVM.

I. INTRODUCTION

Machine Learning Obesity detection using machine learning is an important application in the health care domain. Machine learning models can be used to predict and diagnose obesity based on various input data, including patient characteristics, medical history, and other relevant features. Obesity classification refers to the categorization of individuals based on their body weight and fat distribution. It is an important concept in healthcare and research because obesity is associated with various health risks and can inform treatment and prevention strategies. There are several methods and criteria used to classify obesity, with the most common being based on body mass index (BMI) and waist circumference.

The main concern of this paper is to analyze people for obesity and make them aware of the obesity risk factor. This paper aims to predict the obesity risk. The analysis is conducted into two parts where firstly it read the data and then checks the data if it matches the factor with obesity, and then it will show the result. For our analysis, first, we collect raw data sets for our analysis depend on some factors. In addition, we preprocess those data, then we applied machine learning supervised algorithms to check the accuracy, sensitivity, specificity, precision, recall, and F1 Score. Then we found which algorithm works more optimal and detect the actual outcome

II. RELATEDWORK

Here are some reviews made by the authors of various paper. In the paper [1], An obese patient's postoperative health status provides insight into how their surgical treatment is performed. Physicians must review prior patient records at every postoperative visit to review the patient's status and assess the postoperative risk of readmission. To facilitate this procedure, we devise a technique for identifying indicators and evaluating weight fluctuations, enabling prompt identification of possible complications and clinical readmission hazards. Two methods that are based on neural networks and conventional machine learning will be compared in this paper. One variant of attentive recurrent neural networks is compared with traditional machine learning on the task of obesity-related entity extraction. We conclude that to achieve an extended corpus and a general representation—which may enhance the differentiability of the input data—a data balancing method should be used first when processing a small data set using neural networks.In paper [2], the author presents FREGEX, an automatic feature extraction technique based on regular expressions for biomedical texts. Tokens were extracted from biomedical texts using the Smith-Waterman and Needleman Wunsch sequence alignment algorithms, and they were represented by common patterns. To assess the efficacy of the suggested approach, three manually annotated datasets containing data on obesity, obesity types, and smoking habits were used. For comparison, features extracted from successive token sequences (engrams) were utilized, and the TF-IDF vector model was used to mathematically represent both kinds of features. According to the [4] paper, metabolic disorders—like type 2 diabetes mellitus, obesity, and metabolic syndrome—are highly prevalent in developed nations' populations and necessitate ongoing medical care as they advance. Over 90 percent of obese individuals are "protected" against insulin resistance, a condition that is linked to obesity. Along with signal processing techniques, infrared spectroscopy has been studied as a non-invasive tool on biofluids in the search for new and predictive biomarkers. According to a recent study [5], genetics is a significant factor in obesity risk analysis in addition to lifestyle choices. Numerous academic works examine how Single Nucleotide Polymorphisms (SNPs) relate to obesity in order to enable tailored medicine. SNP data are typically noisy and large, though, which has an impact on the computational complexity and accuracy of data processing and analysis. In the paper [13], Information and communication technologies (ICT) have emerged as a viable alternative for tracking and managing obesity and overweight in recent years. ICT can also save costs and shorten the time between doctor and nutritionist visits. An idea for a smartphone application to track and monitor issues related to obesity and overweight is presented in this article. The prototype offers timely information to patients, doctors, and nutritionists. It creates a channel of communication between the parties so that regular updates on patients' progress can be obtained, and nutritionists can provide precise and tailored recommendations for long-term behaviour change, which will aid in the fight against obesity.

The study in this paper [27] uses a data set related to the primary causes of obesity torefer to high calorie intake, a decrease in energy expenditure from inactivity, dietary disorders, genetics, socioeconomic factors, anxiety, and depression, as well as high calorie intake. Since obesity is a global illness that affects people of all ages and genders, scientists have worked very hard to pinpoint the early causes of the condition. This study develops an intelligent method based on supervised and unsupervised data mining techniques to detect obesity levels and support health professionals and the public in adopting healthier lifestyles in the fight against this worldwide epidemic. It is suggested [29] that obesity, which is generally characterized as an excess of body fat that negatively impacts health, can no longer be assessed primarily using the body mass index due to its heterogeneous nature. The cardiovascular disease risk associated with overweight and moderate obesity is primarily determined by the presence of excessive amounts of visceral adipose tissue and ectopic fat. More aggressive treatments, such as obesity surgery, also known as metabolic surgery, have been studied in this subgroup of patients due to the challenges in normalizing body fat content in patients with severe obesity. More aggressive treatments, like obesity surgery, also known as metabolic surgery, have been studied in this subgroup of people due to the challenges in normalizing body fat content in patients with severe obesity. The main goal of this research [30] is to review different machine learning (ML) techniques and how they are applied to publicly accessible sample health data linked to lifestyle diseases like type II diabetes and obesity. Rather than presenting a risk prediction model. We excluded pregnancy and genetic factors and focused on people in the age range of >20 to <60, both male and female. This work meets the criteria for a tutorial article on applying various machine learning techniques to the identification of possible obesity risk factors.

III. METHODOLOGY

The primary objective is to detect and diagnose obesity in individuals at an early stage. Early detection allows for timely intervention and management of the condition, which can improve health outcomes. Develop a model that can accurately assess the presence and severity of obesity. This helps healthcare professionals make informed decisions and recommendations for patients. Educate individuals about the risks associated with obesity and promote healthy lifestyle choices. The model can serve as a tool for health education and awareness.

High blood pressure, high cholesterol, type 2 diabetes, and respiratory issues are among illnesses linked to obesity.

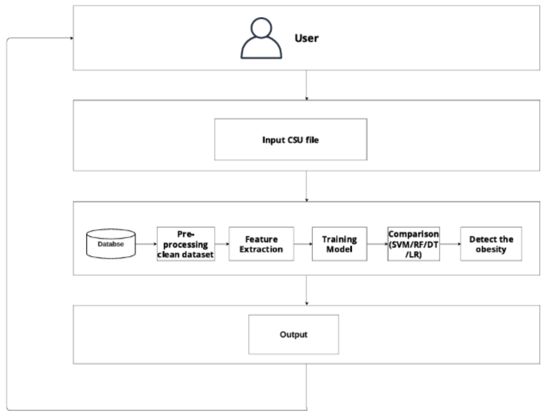

Fig 1. Methodology

IV. MODULE

- Admin

In this module, the admin must log in by using a valid username and password. After login successful he can do some operations, such as View All Users and Authorize,

• View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, username, email, address, and admin authorize the users.

• End User

In this module, there are n numbers of users are present. Users should register before doing any operations. Once the user registers, their details will best or to the database. After registration successfully, he has to login by using authorized username and password. Once Login is successful user will do some operations like Manage Account.

V. DATASET

We are utilizing a predefined dataset in the.csv format to carry out this model's implementation. The dataset we are using consists of 2112 rows, or data tuples, and 17 attributes, or 17 columns. Both string and numeric types are present in the attributes. There are four main sections to the dataset.

A. Attributes related to eating habits:

Characteristics associated with eating patterns: These characteristics provide a succinct overview of the person's eating patterns and level of macronutrient awareness.

Frequent intake of foods high in calories (FAVC) (yes/no string types with two classes). Frequency of vegetable consumption (FCVC).

This parameter essentially reports on the number of vegetable-based meals a person has each day. (Number within the interval of 1 to 4) Count of the main courses (NCP).the number of healthy, appropriate meals a person consumes in a day. (Number within the interval of 1 to 4). food intake in between meals (CAEC). Split up into four classes No, Occasionally, Often, Consistently). Daily water consumption (CH20) (numerical value expressedinlitres). Consumption of alcohol (CALC). (Divided into 4 classes No, Sometimes, Frequent, Always).

B. Attributes related to physical condition.

The characteristics listed below give a general idea of a person's physical state, manner of movement, and overall way of life. Monitoring of calorie consumption (SCC) (two-class string types with yes/no options). Frequency of physical activity (FAF): a numeric value between 0 and 3.
Time spent utilizing technological devices (TUE) (between 0 and 2). Use of transportation (MTRANS) (four classes of strings).

C. General Attributes.

Gender (String type with 2 values Male and Female). Age (Numeric Type in the range from 16 to 25). Height (Numeric Type and measured in meters). Weight (Numeric Type and measured in Kilogram). SMOKING (Divided into 2 classes YES or NO).

There are no NULL values or missing values in the dataset. There are seven classes for the target attribute. Types of Obesity: Normal Obesity Type I, Overweight Level II, Type II Obesity, Insufficient Weight, & Weight. K-Means clustering algorithm was applied for 7 clusters in order to gain a deep understanding of these classes and whether or not they are independent. This gave us a clear understanding of the four distinct classes—insufficient weight, overweight, obesity, and normal weight. Additionally, there are two types of overweight people and three types of obese people. One instance of multiclass classification is the Dataset.

To determine obesity, we first considered each of the roughly 16 attributes. We decide to eliminate the values to lower the data's dimensionality and cost.We decided to eliminate the values that had less of an impact on the dataset as well as the outliers to decrease the cost and dimensionality of the data. There were eleven attributes in the final training dataset, also known as the pre-processed dataset.

D. Algorithms

ML Classification Techniques used for the model:

1. SVM

SVM is a supervised learning algorithm used for both classification and regression tasks. SVM aims to find an optimal hyper-plane that separates the data points of different classes or predicts continuous target values.

2. Decision Tree

Decision Trees are versatile supervised learning algorithms used for classification and regression tasks. Decision Trees create a tree-like model of decisions and their possible consequences based on the features in the input data.

3. Random Forest

Random Forest is a popular machine learning algorithm used for classification and regression tasks due to its high accuracy.

4. Neave Bias

To apply Naive Bayes for obesity detection, first need a dataset containing relevant features such as age, gender, weight, height, BMI.Each instance in the dataset should be labeled with whether the individual is obese or not.

Then, you would preprocess the data, handling missing values, outliers, and encoding categorical variables if necessary. Next, split the data into training and testing sets.

After preprocessing, train a Naive Bayes classifier using the training data. The classifier would learn the probability distribution of the features given the class labels (obese or not obese). Naive Bayes is a probabilistic machine learning algorithm based on Bayes' Theorem.

VI. RESULTS

In Table No.1, we tested different ML models such as SVM, Random Forest, Decision Tree, and Neave Bias for determining the obesity level in a person. And we found out that Random Forest gives us the best accuracy of 95% and hence considering that we trained and tested our model using that algorithm. And then the output is presented.

Table I. COMPARATIVE ANALYSIS BETWEEN DIFFERENT ML MODELS

No	Algorithm	Accuracy
1	SVM	75.4137%
2	Decision Tree	93.6170%
3	Random Forest	95.2830%
4	Neave Bias	84.5670%

VII. CONCLUSION

Obesity classification is a complex topic that involves categorizing individuals based on their body mass index and associated health risks. We have conducted in-depth research using various machine learning techniques to predict the risk of obesity. The risk forecast for obesity has been completed by four explicit classifications. The merits of those classifiers have been measured in terms of conspicuous performance metrics. The relative merits of the results achieved have been assessed by analyzing the results of similar works thereafter. The accuracy came out from logistic regression with a value of 95.09%. Our future plan is to make this work more rigorous with a bigger data set to cover as much a wider range of low-obese and medium-obese and high-obese people.

REFERENCES

[1] Yihan Deng, Peter dolog, Jorn-Markus gass ,“Obesity entity extraction from real outpatient records: When learning based methods meet small imbalanced medical data sets”,2019.

[2] Christopher A. Flores, Rosa L. Figueroa “FREGEX: A Feature Extraction Method for Biomedical Text Classification using Regular Expressions”,2019.

[3] Naomi Christianne Pereira,Supriya solaskar,“OBESITY RELATED DISEASE PREDICTION FROM HEALTHCARE

COMMUNITIES USING MACHINE LEARNING",2020.

[4] S. A. Pullano, M. Greco, D.P. Foti,"FT–IR Saliva Profiling in Patients with Obesity and Obesity–Related Insulin Resistance",2018.

[5] Phaik-Ling Ong, Yun-Huoy Choo,Nurul A. Emran,"Classification of SNPs for Obesity Analysis using FARNeM Modeling",2013.

[6] Thiago Orsi, Ernesto Araujo, Ricardo Simoes, "Fuzzy Chest Pain Assessment for Unstable Angina based on Baranwal Symptomatic and Obesity Clinical Conditions",2015.

[7] CASIMIRO A. CURBELO MONTANEZ, PAUL FERGUS, "SAERMA: Stacked Autoencoder Rule Mining Algorithm for the Interpretation of Epistatic Interactions in GWAS for Extreme Obesity",

[8] Susana Abe Miyahira, Ernesto Araujo, Member, IEEE "Fuzzy Obesity Index for Obesity Treatment and Surgical Indication",2020.

[9] Kunal Rajput Faculty of Science & Technology University of Canberra, "Obesity and Co-Morbidity detection in Clinical text using Deep Learning and Machine Learning techniques",2018. Detection of Obesity Using Machine Learning Dept. of Computer Engineering 44 D.N.Patel College of Engg. Shahada

[10] Muhammad Nouman, Surapong Chat pun Institute of biomedical engineering, Faculty of Medicine, Prince of Songkhla University, "Plantar pressure distribution in non-obese, overweight and obese subjects with diabetic neuropathy while walking",2018.

[11] Miguel Castorena∗ , Nguyen Doan†, Benjamin Gillmore‡ and Jimmy Lahn§, Joshua Lorenzen, "Overweight Object Transportation with a Set of Collaborative Robots",2020.

[12] Chellaswamy C Department of Electronics and Communication Engineering Kings Engineering College Chennai, "Machine Learning Based Condition Recognition System for Bikers",2020.

[13] Juan Montenegro1, Lilia Muñoz1,2, Mel Nielsen1, Vladimir Villarreal1,2 1 Grupo de Investigación en Tecnologías Computacionales Emergentes, Universidad Tecnológica de Panamá, Panamá 2 Centro de Estudios Multidisciplinarios en Ciencias, Ingeniería y Tecnología – CEMCIT AIP, Panamá "Plataforma para el monitoreo de la obesidad y el sobrepeso Platform for monitoring obesity and overweight",2020.

[14] Eliana Marina Alvarez Di Fino Esc. de Nutricion, FCM, UNC, "Spatial analysis applied to nutritional epidemiology",2019.

[15] Ramy El Jalbout, "Carotid artery noninvasive elastography (NIVE) to detect early changes of cardiovascular diseases in overweight and obese children".2019.

[16] Vasile Manoliu1, Member IEEE, Cristina Maria Stanescu 2 1 Faculty of Electrical Engineering, Politehnica University of Bucharest, Romania, "A Study of the Cardiovascular Function and Aerobic Capacity in Overweight and Obese Subjects",2019.

[17] Carlos Martinez-Millana, Antonio Martinez-Millana*, Carlos Fernandez-Llatas ITACA Universitat Politecnica de Val ` encia ` Valencia, Spain ,"Comparing data base engines for building big data analytics in obesity detection",2019.

[18] Saptashwa Mitra∗ , Yu Qiu∗ , Haley Moss†, Kaigang Li† and Sangmi Lee Pallickara, "Effective Integration of Geotagged, Ancilliary Longitudinal Survey Datasets to Improve Adulthood Obesity Predictive Models" ,2018.

[19] Naomi Christianne Pereira1, Jessica D'souza2, Parth Rana3, Supriya Solaskar,"OBESITY RELATED DISEASE PREDICTION FROM HEALTHCARE COMMUNITIES USING MACHINE LEARNING",2019.

[20] UM Chaskar Department of Instrumentation and Control Engineering College of Engineering Pune, "Machine Learning Approach to Study the Impact of Obesity on Autonomic Nervous System using Heart Rate Variability Features",2021.

[21] Shivam Garg Dept. of CSE IIIT Kota Jaipur, India,"Obesity Risk Prediction using WiFi based Human Activity Recognition Techniques",2021.

[22] Özgü Taçyıldız, 2 Duygu Çelik Ertuğrul, 3 Yıltan Bitirim, 4 Neşe Akcan, 5 Atilla Elçi, "Ontology-based Obesity Tracking System for Children and Adolescents" ,2018.

[23] Haofeng Yan, Yu Wang, Tingdong Fang, Xiaofei Chen, "Evaluation Method of Operation State in Overweight Rejection based on Unary Linear Regression Model",2023.

[24] Hera siddiqui 1, ajita rattani 1, nikki k. Woods2, laila cure3, rhonda k.lewis4,"A Survey on Machine and Deep Learning Models for Childhood and Adolescent Obesity",2021.

[25] GRAZIA V. AIOSA1, MAURIZIO PALESI1, (Senior Member, IEEE),AND FRANCESCA SAPUPPO 2, (Member, IEEE), "Explainable AI for Decision Support to Obesity Comorbidities Diagnosis",2023.

[26] LEIRE BASTIDA, GLORIA CEA, ANA MOYA, ALBA GALLEGO, "Promoting Obesity Prevention and Healthy Habits in Childhood: The OCARIoT Experience",2022. Detection of Obesity Using Machine Learning Dept. of Computer Engineering 46 D.N.Patel College of Engg. Shahada

[27] Rodolfo Canas ˜ Cervantes, Ubaldo Martinez Palacio, "Estimation of obesity levels based on computational intelligence",2020.

[28] Mahmood Safaei a, Elankovan A. Sundararajan a, Maha Driss b,c, Wadii Boulila b,c, Azrulhizam Shapi'i d, "A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity",2021.

[29] Marie-Eve Piché, André Tchernof, Jean-Pierre Després, "Obesity Phenotypes, Diabetes, and Cardiovascular Diseases",2020.

[30] Ayan Chatterjee, Martin W. Gerdes and Santiago G. Martinez, "Identification of Risk Factors Associated with Obesity and Overweight—A Machine Learning Overview",2019.

[31] NileshP.Sable, Anay Indani, Raj Mehta, Ratmmala Bhimanpallewar, Sara Shaikh, Soham Jadhav "A Machine Learning approach for Early Detection and Prevention of Obesity and Overweight",2023[2] J. U. Duncombe, ―Infrared navigation—Part I: An assessment offeasibility,‖*IEEETrans.ElectronDevices*,vol.ED-11,pp. 34-39,Jan.1959.

International Journal of Innovative Research in Technology

An International Open Access Journal Peer-reviewed, Refereed Journal
www.ijirt.org | editor@ijirt.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of International Journal of Innovative Research in Technology (ISSN 2349-6002) is hereby awarding this certificate to

SHRADDHA BORASE

In recognition of the publication of the paper entitled

OBESITY DETECTION USING MACHINE LEARNING TECHNIQUES

Published in IJIRT (www.ijirt.org) ISSN UGC Approved (Journal No: 47859) & 7.37 Impact Factor

Published in Volume 10 Issue 11, April 2024

Registration ID 162796 Research paper weblink:https://ijirt.org/Article?manuscript=162796

EDITOR

EDITOR IN CHIEF

ISSN 2349-6002

IJIRT
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY

International Journal of Innovative Research in Technology

An International Open Access Journal Peer-reviewed, Refereed Journal
www.ijirt.org | editor@ijirt.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of International Journal of Innovative Research in Technology (ISSN 2349-6002) is hereby awarding this certificate to

VAISHNAVI BHOI

In recognition of the publication of the paper entitled

OBESITY DETECTION USING MACHINE LEARNING TECHNIQUES

Published in IJIRT (www.ijirt.org) ISSN UGC Approved (Journal No: 47859) & 7.37 Impact Factor

Published in Volume 10 Issue 11, April 2024

Registration ID 162796 Research paper weblink:https://ijirt.org/Article?manuscript=162796

EDITOR

EDITOR IN CHIEF

ISSN 2349-6002
9 772349 600203

IJIRT

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY

Innovative Research in Technology

An International Open Access Journal Peer-reviewed, Refereed Journal

www.ijirt.org | editor@ijirt.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of International Journal of Innovative Research in Technology (ISSN 2349-6002) is hereby awarding this certificate to

VAISHNAVI PATIL

In recognition of the publication of the paper entitled

OBESITY DETECTION USING MACHINE LEARNING TECHNIQUES

Published in IJIRT (www.ijirt.org) ISSN UGC Approved (Journal No: 47859) & 7.37 Impact Factor

Published in Volume 10 Issue 11, April 2024

Registration ID 162796 Research paper weblink:https://ijirt.org/Article?manuscript=162796

EDITOR

EDITOR IN CHIEF

ISSN 2349-6002

9 772349 600203

International Journal of Innovative Research in Technology

An International Open Access Journal Peer-reviewed, Refereed Journal
www.ijirt.org | editor@ijirt.org An International Scholarly Indexed Journal

Certificate of Publication

The Board of International Journal of Innovative Research in Technology (ISSN 2349-6002) is hereby awarding this certificate to

UMESH PATEL

In recognition of the publication of the paper entitled

OBESITY DETECTION USING MACHINE LEARNING TECHNIQUES

Published in IJIRT (www.ijirt.org) ISSN UGC Approved (Journal No: 47859) & 7.37 Impact Factor

Published in Volume 10 Issue 11, April 2024

Registration ID 162796 Research paper weblink:https://ijirt.org/Article?manuscript=162796

EDITOR

EDITOR IN CHIEF

ISSN 2349-6002
9 772349 600203