“Detection Of Obesity Using machine Learning”

A Project

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

Submitted To

Dr. BABASAHEB AMBEDKAR TECHNOLOGICAL UNIVERSITY, LONERE

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 2023-24

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-I entitled

"Detection Of Obesity Using machine Learning"

As prescribed by Dr. Babasaheb Ambedkar Technological University, Lonere as a part of syllabus for the partial fulfillment in Bachelor of Computer Engineering for Academic year 2023-24.

GUIDE
Prof. A. P. Khan

H.O.D.
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
We have tested the performance of prototype
"Detection of Obesity Using Machine Learning"

Developed By

It has been successfully operated.

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

Date: 30/11/2023
Place: Shahada

GUIDE	**H.O.D.**
Prof. A. P. Khan	Prof. V. S. Mahajan

PROJECT IN-CHARGE

Prof. V.S. Mahajan

Prof. S.H.Shaikh

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 all the teachers and staff of the Department of Computer Engineering and Information Technology for their cooperation and support.

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

(Final Year Computer)

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: SVM, DT, RF, Diagnosis, Machine Learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 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 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

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 (BMI) and other relevant features.

1.3 Available Similar Systems

Healthcare researchers and institutions often use open-source machine learning libraries like scikit-learn, TensorFlow, and PyTorch to develop custom obesity detection models. Many research institutions and hospitals develop their in-house machine learning models and systems for disease detection, including obesity, as part of their research and clinical practice. Google has been involved in various healthcare-related projects, and its machine learning and AI capabilities can be applied to health data analysis and prediction tasks, including obesity detection.

1.4 Objectives of The Proposed System

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.

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 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.

1.5.2 System Architecture

- A system architecture is the conceptual model that defines the structure, behavior, and more views of a system.[1] 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.

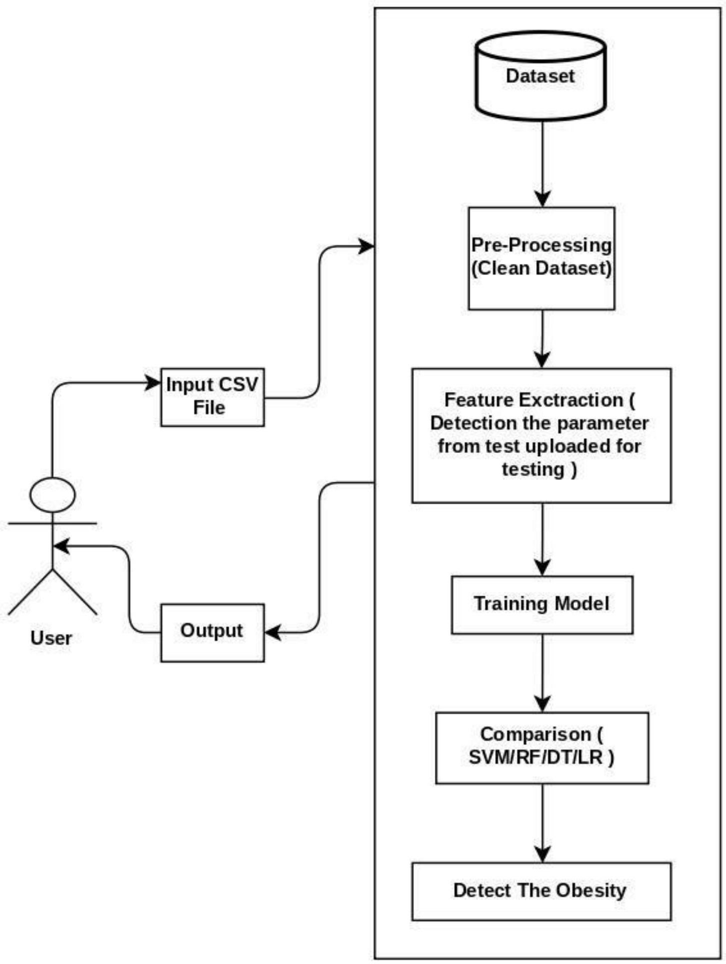

Fig.1.5.2: System Architecture of Detection of Obesity

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.

Chapter 2: LITERATURE SURVEY

In [1] paper, The postoperative health status of an obesity patient indicates the outcome of the surgical treatment. By each postoperative revisit, physicians need to go through the previous patient records to recall the patient status and to evaluate the postoperative risk of readmission. To support this process, we develop a method to extract indicators and to analyze weight changes, so that potential complications and risks of clinical readmission can be recognized timely. In this paper, we will compare two approaches that are based on traditional machine learning and neural networks. The performance of traditional machine learning on the task of obesity related entity extraction is compared with one variation of attentive recurrent neural networks. We conclude that for processing a small data set using neural networks, a data balancing method should firstly be applied to achieve an extended corpus and a general representation, which can apparently increase the differentiability of the input data.

In [2] paper, we present FREGEX, a method for automatically extracting features from biomedical texts based on regular expressions. Using Smith-Waterman and Needleman-Wunsch sequence alignment algorithms, tokens were extracted from biomedical texts and represented by common patterns. Three manually annotated datasets with information on obesity, obesity types, and smoking habits were used to evaluate the effectiveness of the proposed method. Features extracted using consecutive sequences of tokens (engrams) were used for comparison, and both types of features were mathematically represented using the TF-IDF vector model. Support Vector Machine and Naïve Bayes classifiers were trained, and their performances were ultimately used to assess the ability of the feature extraction methods. Results indicate that features based on regular expressions not only improved the performance of both classifiers in all datasets but also use fewer features than n-grams, especially in those datasets containing information related to anthropometric measures (obesity and obesity types). College Of Engineering 2023

In [3] paper, In addition to the rapid growth of Machine Learning in biomedical and healthcare communities, the accurate analysis of medical data benefits early disease detection, patient care and community services. However, the analyses accuracy of a disease is reduced when the intake and the quality of medical data is unexplored and incomplete. This work aims to developing a state-of-the-art system that streamlines machine learning algorithms for the effective prediction of Obesity and its related diseases considering the population of India. 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.

In [4] paper, Metabolic disorders such as type 2 diabetes mellitus, obesity and metabolic syndrome have a high incidence in the population of developed countries and require continuous clinical and pharmacological treatments throughout their progression. Obesity is associated with insulin resistance in over 90obese subjects are “protected” from this condition. Infrared spectroscopy has been investigated as a non–invasive tool on biofluids, together with signal processing techniques, in the research of novel and predictive biomarkers. In the following, a study based on saliva profiling using infrared spectroscopy on a population of metabolic abnormal and normal obese compared with control subjects is presented. Analysis has been carried out to design a consistent and standardized protocol for saliva profiling in different molecular regions of interest. Results obtained through an unsupervised classification technique allowed the grouping of patients belonging to a specific population based on the characteristic molecular signatures in the regions of Amide I, glucose and thiocyanate.

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.

In [10] paper, Overweight and obesity are growing health complications mostly associated with metabolic and musculoskeletal comorbidities.The purpose of this study was to assess the differences regarding plantar pressure distribution in participants with diabetic neuropathic feet who had a different body mass index (BMI). Peak plantar pressure was measured in 12 participants during level walking. The subjects were classified into three categories, each containing 4 participants, as nonobese, overweight and obese according to their BMI values. Peak plantar pressure was determined for the hindfoot, midfoot and forefoot regions using Padar-X in-shoe pressure measurement system. Lower peak plantar pressure was observed in midfoot compared to the hindfoot and forefoot. However, the obese group showed a significantly higher peak plantar pressure in midfoot compared to the non-obese and overweight diabetic foot with neuropathy. Therefore, the high peak plantar pressure at the midfoot can cause foot pain in obese diabetic neuropathic foot.

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) ($\geq$ 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 (Max |SSE|); 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 non-communicable 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 conside 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 behaviour 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 diseases.The present study conducted a systematic literature review to examine obesity research 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.Finally, this study seeks to support decision-makers looking to understand the impact of obesity on health in the general population and identify outcomes that can be used to guide health authorities and public health to further mitigate threats and effectively guide obese people globally.

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.

Chapter 3: ANALYSIS

3.1 System Development Requirements

3.1.1 Hardware Requirements

• **System:** Intel I3 Processor and above.

• **Hard Disk:** 20 GB

• **Ram:** 4GB

3.1.2 Software Requirements (Platform Choice)

• **Operating system:** Windows 7 or more.

• **Coding Language:** python

• **IDE:** Spyder

3.1.3 Database Requirements

SQLITE

3.2 Functional Requirements

3.2.1 System Feature1(Functional Requirement Admin):

Admin module will be on web module. Admin will verify user information and allow or reject to user. Load the Data set. User: User registers into system with personal information. Automatically user verification request sends to admin. After verification user can login into system.

3.2.2 System Feature2(Functional Requirement System):

By using SVM/RF/DT algorithm, Enhance Machine learning techniques like SVM for the Comparison of Clustering Methods for Obesity Classification.

3.3 External Interface Requirement

3.3.1 User Interface

• Application of Comparison of Clustering Methods for Obesity Classification.

3.4 Non-functional Requirement

3.4.1 Performance Requirements

The performance of the functions and every module must be well. The overall performance of the software will enable the users to work decently. Performance of encryption of data should be fast. Performance of the virtual environment should be fast Safety Requirement The application is designed in modules where errors can be detected and steadily. This makes it easier to install and update new functionality if required.

3.4.2 Safety Requirement

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

3.4.3 Software Quality Attributes

Our software has many quality attribute that are given below: -

Adaptability: This software is adaptable by all users.

Availability: This software is freely available to all users. The availability of the software is easy for everyone.

Maintainability: After the deployment of the project if any error occurs then itcan be easily maintained by the software developer.

Reliability: The performance of the software is better which will increase the reliability of the Software.

User Friendliness: Since, the software is a GUI application; the output generated is much user friendly in its behavior.

Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled

Security: Users are authenticated using many security phases so reliable security is provided.

Testability: The software will be tested considering all the aspects.

3.5 Data Flow Diagram

A data-flow diagram is a way of representing a flow of data through a process or a system. The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart. There are several notations for displaying data-flow diagrams. The notation presented above was described in 1979 by Tom DeMarco as part of structured analysis. For each data flow, at least one of the endpoints must exist in a process. The refined representation of a process can be done in another data-flow diagram, which subdivides this process into sub-processes.

The data-flow diagram is a tool that is part of structured analysis and data modelling. When using UML, the activity diagram typically takes over the role of the data-flow diagram. A special form of data-flow plan is a site-oriented data-flow plan. Data-flow diagrams can be regarded as inverted Petri nets, because places in such networks correspond to the semantics of data memories.

DFD Level 0

A Data Flow Diagram (DFD) at level 0, also known as DFD0, represents the highest abstraction level of a system or a process. It provides an overview or context-level

representation of the entire system's functionality, showing the interactions between the system and external entities. Input represents the initial stage where raw data or information is received by the system. This could involve various sources such as user input, sensors, databases, or external systems. The data is collected and entered into the system for further processing. The Processing stage embodies the core operations performed on the received input. Here, the system manipulates, analyzes, organizes, or transforms the data as required by the system's functionalities or business rules. This phase involves various algorithms, calculations, validations, and logical operations to process the input data effectively. Following the Processing stage, the processed information moves to the Output phase. This phase involves the dissemination or presentation of the processed data to users, other systems, or databases.

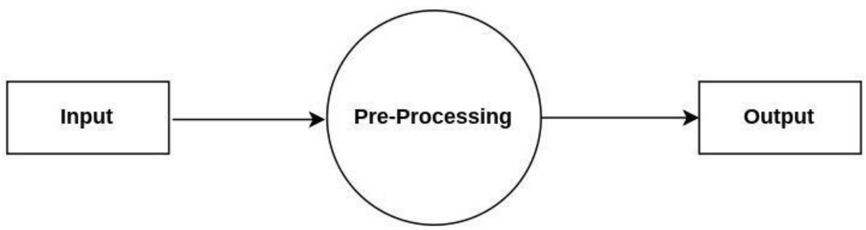

Fig.3.5.1: DFD Level 0 for Detection of Obesity Using Machine Learning

DFD Level 1

A Data Flow Diagram (DFD) at Level 1 is a more detailed breakdown of the processes depicted in the higher-level DFD0 (context-level diagram). DFD Level 1 provides a deeper understanding of the system by expanding on the processes identified in the context-level diagram and showing their sub-processes, inputs, outputs, and data stores. Input: This initial stage illustrates the diverse sources through which data enters the system. Each source, represented as a sub-process, showcases distinct methods of data acquisition such as user inputs. Pre-processing:Following data ingestion, the Pre-processing phase details the sequential steps involved in refining and preparing the raw data for analysis Feature Extraction: Once the data is cleansed and prepared, the Feature Extraction stage involves extracting essential patterns or attributes from the pre-processed data Output: Finally, the Output stage demonstrates how processed information or insights are disseminated or utilized.

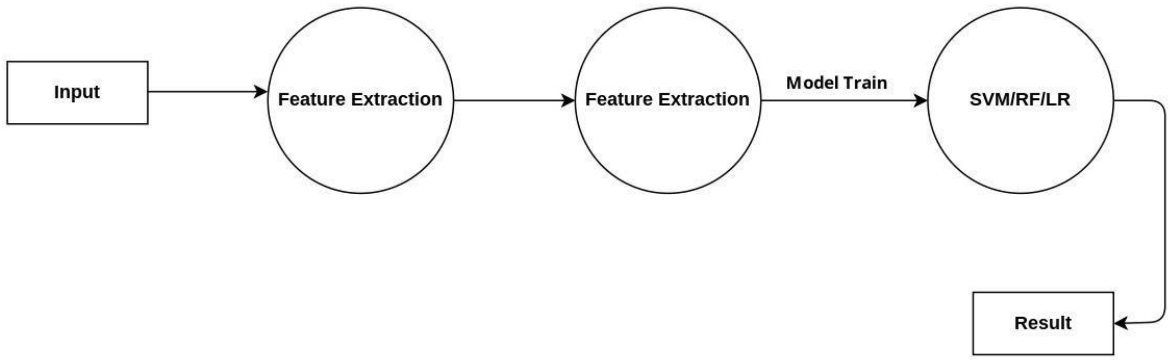

Fig.3.5.2: DFD Level 1 for Detection of Obesity Using Machine Learning

Chapter 4: PLANNING

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 "**Co**nstructive **C**ost **Mo**del". 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:

$$E = a_b (KLoC)^{(b_b)} \qquad \text{[person-months]}$$

$$D = c_b (E)^{(d_b)} \qquad \text{[months]}$$

$$P = E / D \qquad \text{[persons]}$$

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 coefficient $\mathbf{a_b}$, $\mathbf{c_b}$ and the exponent $\mathbf{b_b}$, $\mathbf{d_b}$ are given in the next table.

Table 4.2.4 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 Estimations

Size Estimation

Table 4.2.5 Size Estimation of Current System.

Software Module	**LOC**
Login	400
Registration	1000
Home page	2000
Input text	1000
Result	1000
Total Estimated Lines Of Code (LOC)	5400

Line of code approximately will be 5400.

Effort Estimation

The value a_b and b_b 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:5400LOC=5.40KLOC

Effort (E) =a_b(KLoC)$^{(b_b)}$ [Person -Month]

E=3.0*(5.40)$^{1.12}$

E= PM (Calculate)

Person-Month=19PM (approx)

Duration Estimation

Duration (D) = c_b(E)$^{(d_b)}$[months]

= 2.5*(19.83)$^{0.35}$

= 2.5 *2.844

= 7.11

Duration ≈7 [months]

Person Required

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

= 19.83/7

= 3.83[count]

Person Required =4 [Persons]

Cost Estimation

We take the assumption each person charges 7000 rupees per month.

Total Estimation= 7,000*7= Rs.49,000/- +Two Mobile phone Cost

Total Estimation=Rs.49,000+15000=64,000/-

Total Estimation=Rs.49,500 /-

4.3 TEAM STRUCTURE

Table 4.3 Team Structure

Sr. No.	Name of task	Sub task	Period
1	Feasibility Study	1.1 Problem Definition: • Collecting detailed problem definition of the system to be implemented	09/09/2023 To 14/09/2023
		1.2 Initiation: • Visiting different websites. • Studying existing system with its limitation • Going through Journals, magazines • Studying the reference books	15/09/2023 To 20/09/2023
2	Requirement Analysis and Specification	2.1 Project Plan: • Preparing for complete project plan • Verify and Validate Requirement	21/10/2023 To 26/10/2023
		2.2 Requirement Analysis: • Software requirements • Hardware requirements • Database	27/10/2023 To 30/10/2023
3	Design	3.1 Design: • Describing relationships between modules and sub modules • Decompose project work packages into schedule activities.	01/11/2023 To 07/11/2023
		3.2 UML documentation: • UML Diagram • Data Flow Diagrams	8/11/2023 To 18/11/2023

		3.3 Form Designs: • Showing relationship among different menus and sub menus	19/11/2023 To 27/11/2023

Chapter 5. DESIGN

5.1 Introduction

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 important, 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.

5.2 UML Modelling

The unified modelling language (UML) is a general-purpose visual modelling that is intended to provide a standard way to visualize the design of a system.

UML provides a standard notation for many types of diagrams which can be roughly divided into three main groups: behaviour diagrams, interaction diagrams, and structure diagrams.

5.2.1 Use Case Diagram

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.

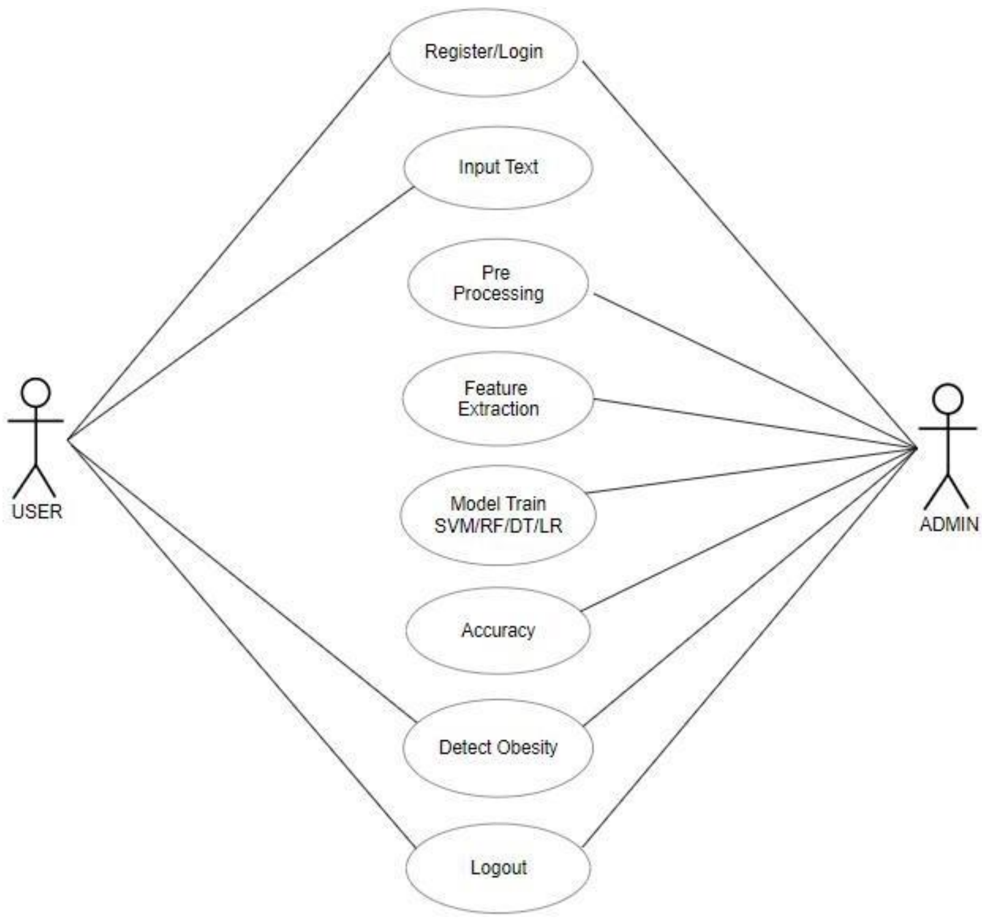

Fig.5.2.1: Use Case Diagram for Detection of Obesity Using Machine Learning

Table 5.2.1: Use Case Description

Use case	Description
Register/Login	User/Admin Login If registered.
Input Text	User Input The text.
Preprocessing	Admin Preprocessing the input information.
Feature Extraction	Extract The parameters.
Model-Train SVM/RF/DT/LR	Comparisons of Algorithm Accuracy.
Accuracy	Show The Accuracy.
Detect Obesity	Detect The Obesity of the user.
Logout	User/Admin Can Logout.

5.2.2 Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions[1] with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e., workflows), as well as the data flows intersecting with the related activities. Although activity diagrams primarily show the overall flow of control, they can also include elements showing the flow of data between activities through one or more data stores.

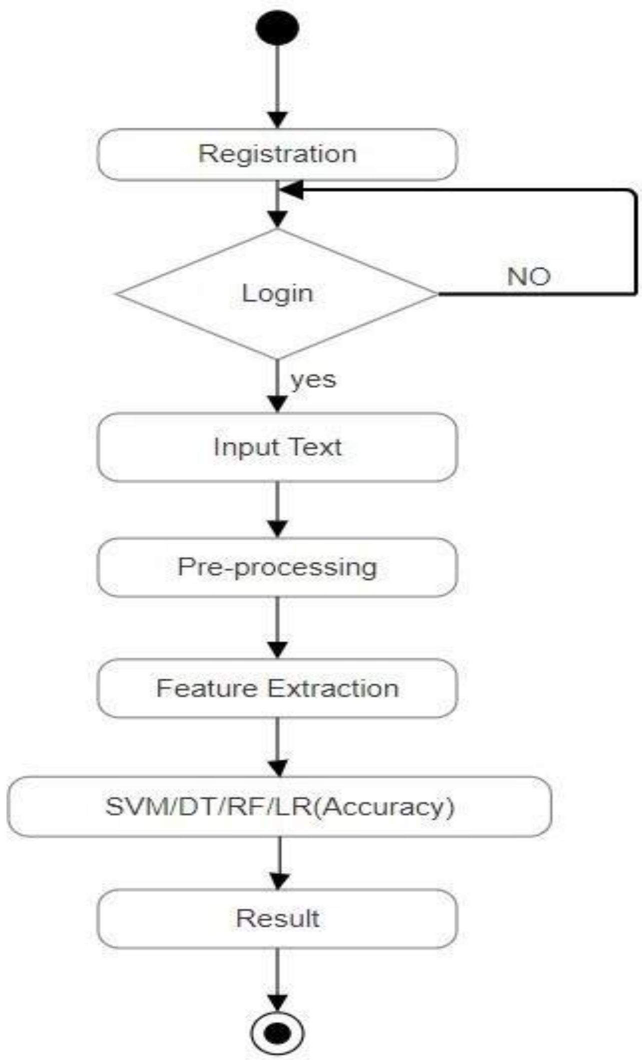

Fig.5.2.2: activity Diagram for Detection of Obesity Using Machine Learning

5.2.3 Sequence Diagram

Sequence diagram is a kind of interaction diagram. 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.

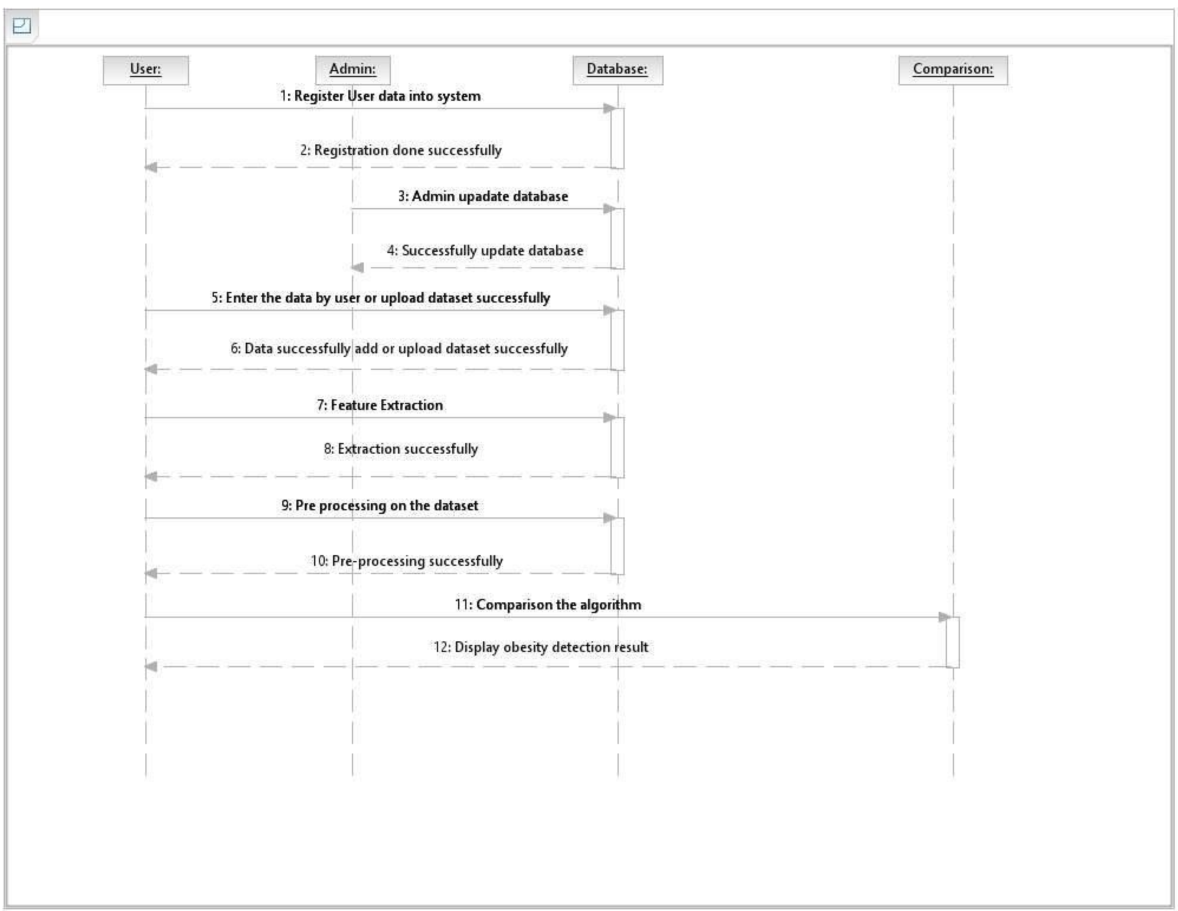

Fig.5.2.3: Sequence diagram for Detection of Obesity Using Machine Learning

5.2.4 Class Diagram

A class diagram in the Unified Modelling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

The class diagram is the main building block of object-oriented modelling. It is used for general conceptual modelling of the structure of the application, and for detailed modelling, translating the models into programming code. Class diagrams can also be used for data modelling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

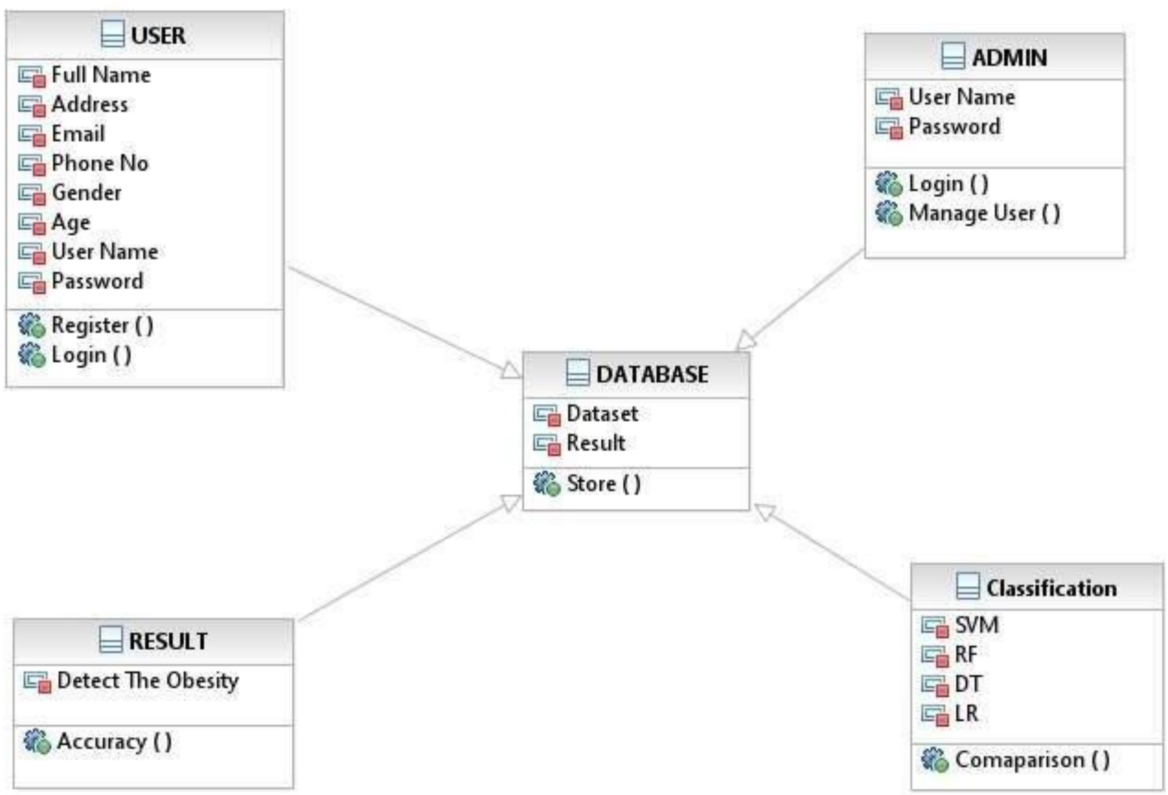

Fig.5.2.4: class diagram for Detection of Obesity Using Machine Learning

5.2.5 Component Diagram

component diagrams show the structure of the software system, which describes the software components, their interfaces, and their dependencies. You can use component diagrams to model software systems at a high level or to show components at a lower, package level. This type of diagram supports component-based development in which a software system is divided into components and interfaces that are reusable and replaceable.

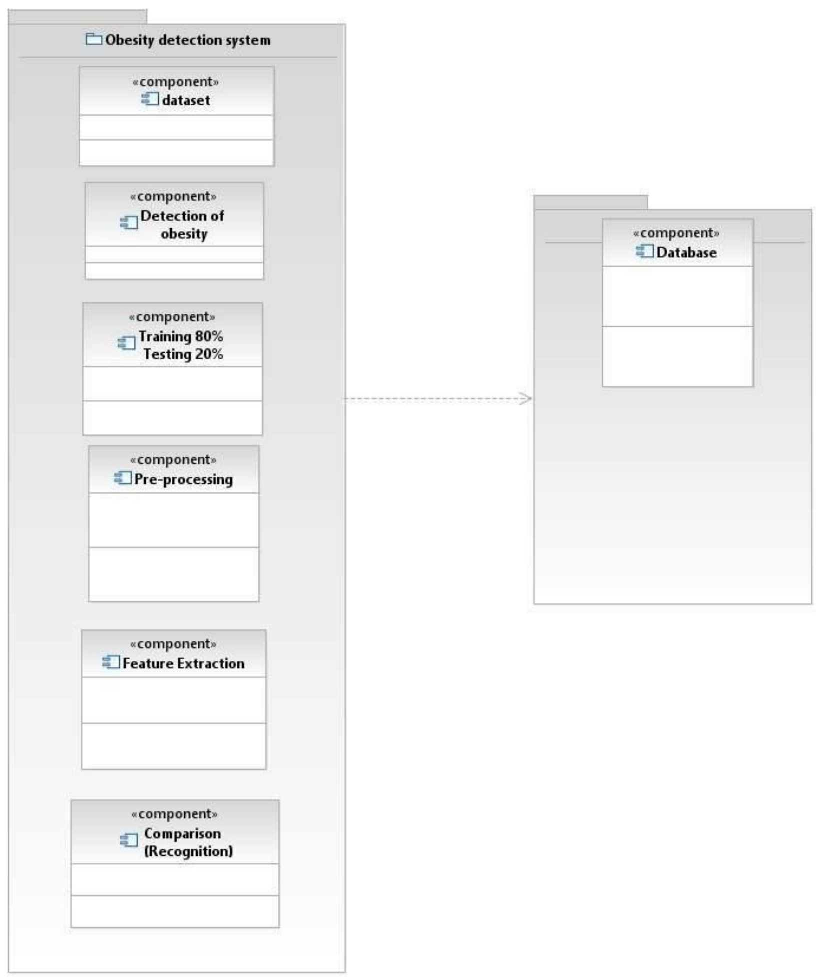

Fig.5.2.5: component diagram for Detection of Obesity Using Machine Learning

5.2.6 Deployment Diagram

Deployment diagram in the Unified Modeling Language models the *physical* deployment of artifacts on nodes. To describe a web site, for example, a deployment diagram would show what hardware components exist what software components run on each node and how the different pieces are connected.

The nodes appear as boxes, and the artifacts allocated to each node appear as rectangles within the boxes. Nodes may have sub nodes, which appear as nested boxes. A single node in a deployment diagram may conceptually represent multiple physical nodes, such as a cluster of database servers.

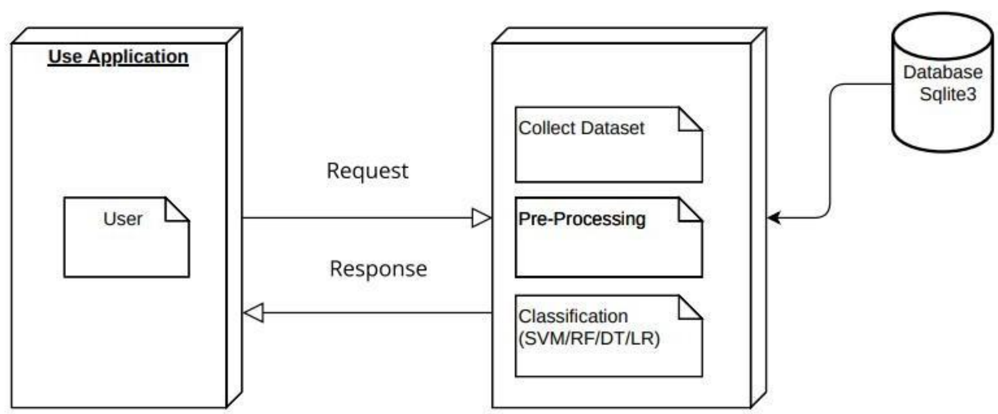

Fig.5.2.6: deployment diagram for Detection of Obesity Using Machine Learning

5.3 Data Modelling

5.3.1 E-R Diagram

An entity–relationship model (or ER model) describes interrelated things of interest in a specific domain of knowledge. A basic ER model is composed of entity types (which classify the things of interest) and specifies relationships that can exist between entities (instances of those entity types).

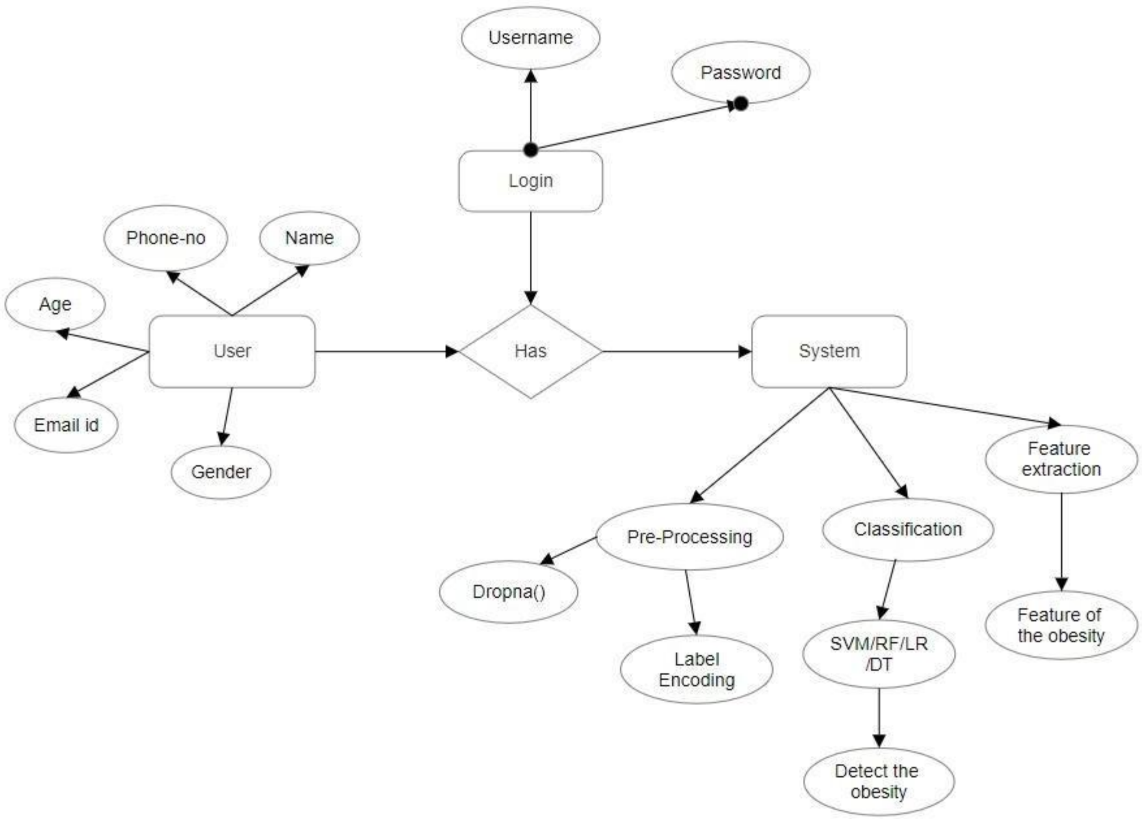

Fig.5.3.1: E-R Diagram for Detection of Obesity Using Machine Learning

Chapter 6: IMPLEMENTATION

6.1 Implementation Language

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 whitespace. 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.

6.2 Database

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.

6.3 Hardware Design

RAM: 4GB

As we are using Machine Learning Algorithm and Various High Level Libraries Laptop RAM minimum required is 8 GB.

Hard Disk: 40 GB

Processor: Intel i5 Processor

Spyder IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required.

IDE: Spyder Best

Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that makes typing feasible and fast.

Coding Language: Python Version 3.5

Highly specified Programming Language for Machine Learning because of availability of High-Performance Libraries.

Operating System: Windows 10

Latest Operating System that supports all type of installation and development Environment

6.4 Implementation Tool

Anaconda: 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., which was founded by Peter Wang and Travis Oliphant in 2012. As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Enterprise Edition, both of which are not free.

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.

In contrast, conda analyses the current environment including everything currently installed, and, together with any version limitations specified (e.g. the user may wish to have TensorFlow version 2,0 or higher), works out how to install a compatible set of dependencies, and shows a warning if this cannot be done. Open-source packages can be individually installed from the Anaconda repository, Anaconda Cloud (anaconda.org), or the user's own private repository or mirror, using the conda install command. Anaconda, Inc. compiles and builds the packages available in the Anaconda repository itself, and provides binaries for Windows 32/64 bit, Linux 64 bit and MacOS 64-bit. Anything available on PyPI may be installed into a conda environment using pip, and conda will keep track of what it has installed itself and what pip has installed.

Custom packages can be made using the conda build command and can be shared with others by uploading them to Anaconda Cloud, PyPI or other repositories. The default installation of Anaconda2 includes Python 2.7 and Anaconda3 includes Python 3.7. However, it is possible to create new environments that include any version of Python packaged with conda.

Spyder

Spyder is a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

Beyond its many built-in features, its abilities can be extended even further via its plugin system and API. Furthermore, Spyder can also be used as a PyQt5 extension library, allowing you to build upon its functionality and embed its components, such as the interactive console, in your own software.

Features

- **Editor**: Work efficiently in a multi-language editor with a function/class browser, realtime code analysis tools (pyflakes, pylint, and pycodestyle), automatic code completion (jedi and rope), horizontal/vertical splitting, and go-to-definition.
- **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.

- **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.

Chapter 7: CONCLUSION

7.1 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.*

[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 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 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.