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.

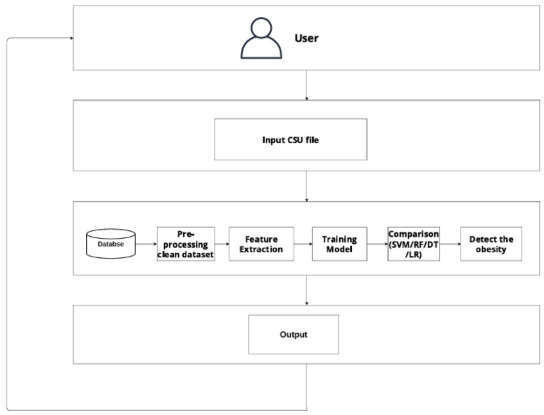

Fig 1. Methodology

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.