STM JOURNALS

Research & Reviews: A Journal of
Medical Science and Technology

https://journals.stmjournals.com/rrjomst/

ISSN: 2319-3417 (Online)
ISSN: 2349-1272 (Print)
Volume 14, Issue 2, 2025
May–August
DOI (Journal): 10.37591/RRJoMST

Research RRJoMST

Machine Learning-Driven Early Prediction and Prevention of Obesity and Overweight

Vikash Kumar[1], Anushri Kulkarni[2,*], Vaishnavi Pagote[3], Vanshika Mungelwar[4]

Abstract

Obesity has become a global health concern, with its prevalence reaching alarming levels in recent years. By classifying obesity-level, healthcare professionals can assess an individual's risk and develop appropriate treatment and prevention strategies. Healthcare professionals can customize interventions and create personalized treatment plans based on individual needs. This paper delivers a system provides an overview of obesity, highlighting the importance of accurate and standardized categorization for effective management and treatment strategies. Various classification systems have been proposed, primarily relying on body mass index (BMI), which relates weight to height. However, since BMI doesn't consider differences in body composition, its limitations have prompted the use of other measurements like waist circumference (WC), waist-to-hip ratio (WHR), and body fat percentage to provide a more accurate assessment of health and body weight. Obesity is associated with a wide range of health problems and chronic diseases, such as heart disease, diabetes, certain types of cancer, and musculoskeletal disorders. This paper outlines a cutting-edge system aimed at developing an automated obesity detection system utilizing machine learning techniques.

Keywords: Obesity, Machine Learning, BMI, WC, WHR

INTRODUCTION

Obesity presents itself as a one major public health concern, multifaceted and enduring health issue marked by the excessive build-up of body fat, resulting in substantial adverse impacts on an individual's overall health and quality of life. Obesity is a widespread global health concern that has reached epidemic levels in many parts of the world. Classifying obesity plays a crucial role in helping healthcare professionals evaluate its severity and choose suitable treatment approaches [1]. Various classification methods are commonly used, including body mass index (BMI), waist circumference, and body fat percentage, each offering a different perspective on an individual's health status.

***Author for Correspondence**
Anushri Kulkarni
E-mail: anushri.kulkarni_skncoe@sinhgad.edu

[1]Professor, Department of Electronics & Communication Engineering, Smt. Kashibai Navale College of Engineering, Vadgaon SKNCOE, SPPU, Pune, Maharashtra, India
[2]Principal, Department of Electronics & Communication Engineering, Smt. Kashibai Navale College of Engineering, Vadgaon SKNCOE, SPPU, Pune, Maharashtra, India
[3,4]Researcher, Department of Electronics & Communication Engineering, Smt. Kashibai Navale College of Engineering, Vadgaon SKNCOE, SPPU, Pune, Maharashtra, India

Received Date: July 17, 2024
Accepted Date: March 29, 2025
Published Date: April 25, 2025

Citation: Vikash Kumar, Anushri Kulkarni, Vaishnavi Pagote, Vanshika Mungelwar. Machine Learning-Driven Early Prediction and Prevention of Obesity and Overweight. Research & Reviews: A Journal of Medical Science and Technology. 2025; 14(2): 31–40p.

The classification of obesity is motivated by several factors, including the need to understand and address the growing global health concern of obesity. Here are some key motivations behind the classification of obesity: Obesity can lead to a multitude of health complications and contribute to the development of chronic health issues and various long-term health concerns, including heart complications, blood sugar imbalances, and some forms of cancer, and musculoskeletal disorders [2]. By classifying obesity, healthcare professionals can assess an individual's risk and develop appropriate treatment and prevention strategies.

Body Mass Index (BMI) is a tool used to assess weight status and potential health risks. It considers an individual's weight relative to their height. The calculation involves dividing weight in kilograms by height in meters squared. While BMI isn't the most precise indicator of body fat percentage, it remains a common tool to assess the risk of obesity and related health concerns [3].

To combat obesity proactively, prioritizing the identification of at-risk individuals for obesity is crucial. Focus on healthy eating patterns, incorporation of regular exercise, and development of positive health routines can be implemented. Addressing obesity in its early stages helps mitigate or prevent the onset of health complications, ultimately enhancing overall health and well-being.

Machine learning plays a pivotal role in revolutionizing obesity and overweight detection websites by offering accurate, personalized, and data-driven solutions for users. Obesity, a multifactorial and complex health condition, has reached epidemic proportions globally [4], presenting significant challenges to public health systems. This system offers a comprehensive examination of obesity, emphasizing the significance of precise and standardized classification for efficient implementation and intervention strategies. Various classification systems have been proposed, primarily relying on body mass index (BMI), which establishes a relationship between weight and height. However, Body mass index's limitations, such as its inability to account for variations in body composition, have led to the development of alternative classifications incorporating additional metrics, considering factors like midsection circumference, WHR (waist-to-hip ratio), and body fat percentage can provide a more comprehensive picture of health risks. Obesity can lead to a multitude of health complications and contribute to the development of chronic health issues and various long-term health concerns, including heart complications, blood sugar imbalances, and some forms of cancer, and musculoskeletal disorders. By classifying obesity, healthcare professionals can assess an individual's risk and develop appropriate treatment and prevention strategies [5, 6].

BACKGROUND STUDIES

This research examined the prevalence of overweight and obesity in English children (2015), analyzing variations by age, gender, and socioeconomic background. Additionally, it explored the alignment between parental and child weight perceptions compared to objective measurements. Finally, the study investigated historical factors contributing to childhood obesity [7, 8].

"This study aims to develop a machine learning model to predict obesity risk." A notable aspect of this work lies in its capacity to educate readers on the hazards and origins of obesity. Data collection involved a sample size exceeding 1,100 people across all age groups, including obese and non-obese. In this paper, the authors used ML techniques like KNN, XGB, logistic regression, and DT and compared them with obesity information or data. Obesity is defined as a BMI of 30 or higher, making it very difficult. Obesity can have negative effects on life, such as depression, decreased productivity and disability. Examine the most advanced ML techniques to study obesity and determine various pros and cons of these ML algorithms [9]. conducted a review of the literature on obesity. For machine learning in obesity, PubMed and Scopus databases were searched using different keywords. Additionally, only variables consistent with our controls (such as diet, dietary pattern, and physical condition) were effectively evaluated. The researchers investigated whether perceptions of SCT could predict whether African American caregivers' children would become obese. 128 caregivers completed a survey regarding caregivers' perceptions of childhood obesity [10]. Expectations were analyzed using horizontal multiple regression ($p < 0.05$). Utilizing extensive data sourced from the UK Millennium Cohort Study, the researchers constructed a ML based model to detect young individuals susceptible to developing overweight or obesity. To identify adolescents at risk of being overweight or obese at age 14, we calculated the child's BMI values at ages 3, 5, 7, and 11. The area of ML models for predicting obesity and its risk factors in children and adolescents; Including cutting-edge models using deep learning for EHRs. This model contrasts with many statistical models that often use logistic regression. Scope in the future is being analyzed with main features and applications emerging from this model. The authors collected and analyzed data on 272,826 children. A) All subjects, including child ages from

4 to 15, were examined. From 1999 to 2003, the prevalence of Ov/Ob increased significantly and from 2004 to 2008, it showed a decline. c) Group review. By 2004, Ov/Ob prevalence had increased in most of the groups studied [11, 12]. The authors of the study aimed to review the quality of the data and update knowledge on the consequences of being overweight in children. A computerized literature search was conducted, although it was only available for studies using prospective or retrospective samples. Data were collected independently by two authors, and the quality of the included studies was then assessed with four factors [13]. The authors examined the efficacy of different body mass index (BMI) thresholds in identifying excessive adiposity (determined by skin thickness), lipid irregularities, insulin levels, and elevated blood pressure in children, along with the heightened risk of obesity in adulthood. Authors note that false expectations will damage relationships. Abnormalities in childhood lead to many cardiovascular diseases in adults. High blood cholesterol, elevated blood pressure, and impaired blood sugar regulation are common in both young people and older adults [14].

APPROACH AND DESIGN

Childhood obesity remains a notable health concern both in the United States and globally. The habits and behaviors children adopt from their parents and other adults play crucial roles in contributing to childhood obesity. You can avoid childhood obesity by teaching your children about proper nutrition and encouraging them to be active. Our system is designed to find obesity and give needed protection against it [15–17]. For children in this age group, weight categories are defined as follows:

- Obesity presents as BMI, which is a variant of the Association Fit Global Health's growth rate and is average for people's age.
- Obesity rates are higher than the World Health Organization's average growth rate based on more than two different standards.
- Children and young people aged 14–21 from 2017 to 2020:
- 14.7 million children and young people are affected by obesity; The rate of obesity is 19.7%.
- Among youth between 6 and 11, the obesity rate is 20.7%, while it rises to 22.2% for adolescents from 12 to 19. Some populations are also more likely to have childhood obesity [18, 19].
- Elevated blood pressure, heightened cholesterol levels, type 2 diabetes, and respiratory ailments are all linked with obesity. In the present scenario, it's notable that researchers primarily rely on BMI for detecting obesity. However, in our project, we aim to utilize 11 variables to ascertain whether young individuals are overweight.

The dataset we employ is in .csv format, a standardized file format utilized for implementing this model. It comprises 17 attributes or columns and contains 2112 rows or data entries. These attributes can be either numerical or textual in nature. The dataset is approximately divided into four segments.

NUTRITIONAL HABITS

These habits briefly describe how well a person knows his nutritional habits and the macros he eats. Consumption of high calorie (string consisting of 2 categories, "yes" or "no").

Vegetable consumption frequency. This parameter provides insight into the frequency of vegetable consumption per day. (Range is 1 to 4).

Number of main foods. Calculate how many healthy meals a person eats per day. (Range is 1 to 4).

Food consumption at meals is categorized into four groups: "Never," "Occasional," "Frequent," and "Regular." Daily water intake (H_2O) is quantified in liters.

Alcohol consumption is classified into four categories: "Abstinent," "Occasional," "Regular," and "Excessive".

Physical Conditions Symptoms

These following characteristics gives a short description of a person's physical or behavioral characteristics or the basics of a person's lifestyle.

Calorie Consumption Tracking (SCC) (String type has 2 groups, “YES” and “NO”).

Functional activity frequency of the body (FAF) (a number in the range of 0 to 3).

Equipment lifetime (TUE) (ranging from 0–2).

Common Attributes

- Gender (String type, has 2 values: “Male” and “Female”).
- Age (various species, various 15–26).
- Height (various types and measured in meters).
- Weight (quantity in kilograms).
- Family background of obesity (divided into 2 groups: “YES” or “NO”).
- Smoking (category YES or NO 2).

The data demonstrates completeness with no missing entries. The variable of interest, classified into 8 distinct categories, reflects the outcome.

- Low_Underweight,
- Medium_Normal_Weight,
- Medium_Over_Weight,
- High_Over_Weight,
- Very_High_Over_Weigt,
- Obesity_Level_I,
- Obesity_Level_II,
- Obesity_Level_III.

To delve deeper into understanding the delineation of these distinct groups and ascertain their independence, we adopt the K-7 algorithmic approach. This methodology enables us to discern clear delineations among four distinct groups: Underweight, Overweight, Obesity, and Normal-Weight. Notably, the categories of Obesity and Overweight are further stratified into three and two subtypes, respectively. Our dataset serves as an exemplar of classification, initially comprising an exhaustive array of features, totalling approximately 16. However, in a bid to mitigate data dimensionality and associated costs, we opt to prune insignificant values and eliminate outliers from the dataset. As a result of this pre-processing, our dataset, whether pre-training or post-training, is streamlined to encompass 11 pertinent features.

Following the pre-processing stage, our dataset undergoes a meticulous refinement process aimed at enhancing its efficacy and reliability for subsequent analyses. Leveraging advanced statistical techniques and machine learning methodologies, we employ feature selection algorithms to identify and retain the most informative attributes while discarding redundant or noise-inducing variables. Additionally, we apply normalization techniques to standardize the range and distribution of the data, ensuring consistency across features and facilitating more robust model training. Furthermore, we conduct exploratory data analysis (EDA) to gain deeper insights into the underlying patterns and relationships within the dataset, aiding in the identification of potential confounding factors or biases (Figure 1).

Machine Learning Classification Methods Used in Model Development:
SVM (Support Vector Machine)

- This learning method uses training data with predefined outcomes to perform tasks like sorting data into groups (classification) or predicting values (regression).
- In the context of obesity detection, SVM can be trained on labelled data (e.g., health records, demographic information, lifestyle habits) to classify individuals as either obese or non-obese.

- SVM operates by identifying the hyperplane that most effectively divides the data into distinct classes, maximizing the margin between classes.
- Support Vector Machines (SVMs) exhibit remarkable efficacy in processing high-dimensional datasets, offering a robust solution particularly adept at navigating nonlinear relationships by leveraging kernel functions.

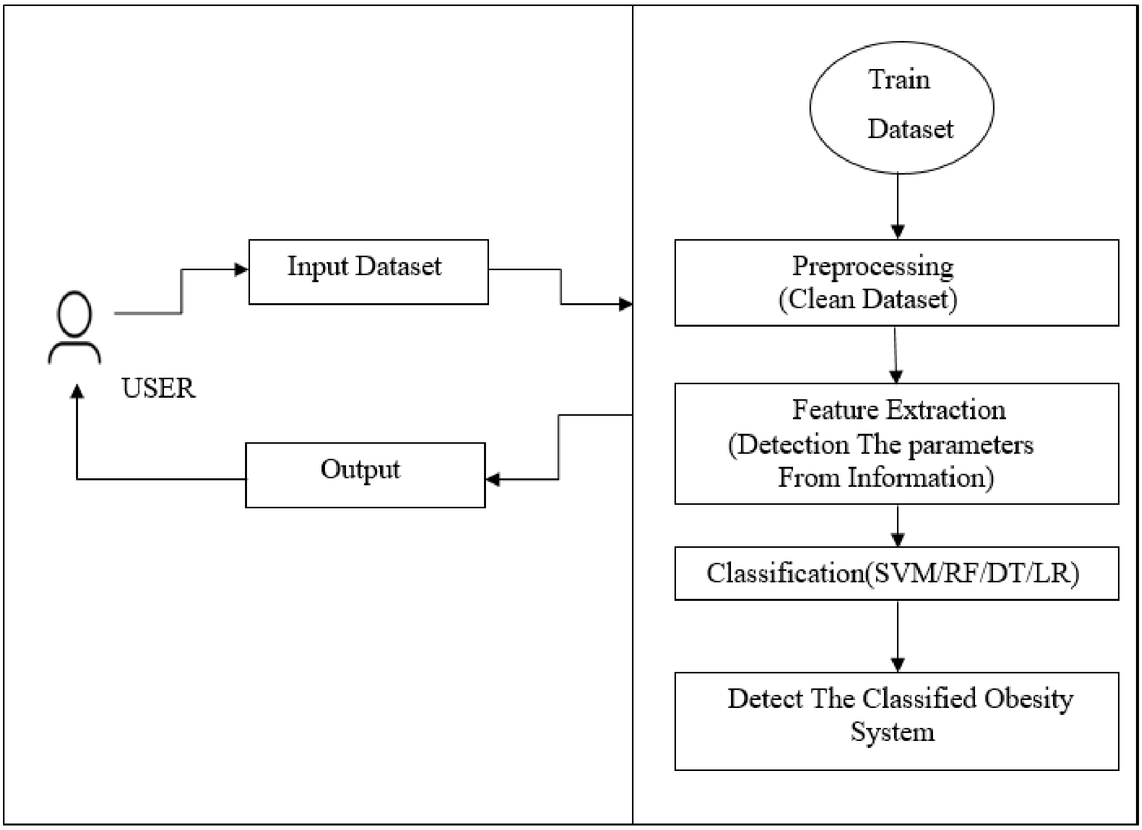

Figure 1. Approach and Design.

RF (Random Forest)

- Random Forests leverage decision trees to make predictions. During training, they build multiple trees, and for classification tasks, choose the most common output among the trees. For regression problems, they average the predictions from all the trees. In this context here obesity detection, RF can be utilized to analyze various features (e.g., BMI, dietary habits, physical activity levels) and their importance in predicting obesity.
- RF is robust against overfitting and can handle large datasets with high dimensionality.
- It can provide insights into feature importance, which is valuable for understanding the factors contributing to obesity.

DT (Decision Tree)

- DT represent a straightforward and easily understandable supervised type of machine learning technique. This approach makes predictions by uncovering easy-to-understand decision rules hidden within the data.
- In the context of obesity detection, DT can be employed to create a hierarchical structure of decisions based on features, such as age, gender, diet, and physical activity, to classify individuals into obese or non-obese categories.
- DT can easily handle both numerical and categorical data and is particularly useful for visualizing decision-making processes.

- However, Decision Trees (DTs) often encounter challenges with overfitting, particularly as the tree depth increases, which can be mitigated using techniques like pruning.

LR (Logistic Regression)

- Logistic Regression (LR) is a statistical technique well-suited for tasks where data can be classified into two categories. It analyzes the data and generates a probability score, indicating the likelihood of a new observation belonging to a specific class.
- In the context of obesity detection, LR can be applied to model the probability of an individual being obese based on various features.
- LR is interpretable, providing insights into the influence of each feature on the likelihood of obesity.
- It's relatively efficient to train and can handle large datasets.
- However, Linear Regression (LR) operates under the assumption of a linear association between the features and the logarithm of the odds of the outcome, a premise that may not consistently align with the intricate nature of complex datasets.

RESULTS AND DISCUSSION

We built a user-friendly web-application where visitors can input their details and learn about their obesity class. We transitioned to a Python-based website for the frontend. We opted to use SQLite as our database. The dataset, which continues to grow as new users enter their information, is stored in SQLite.

To interact with the SQLite database, we utilized Python's built-in SQLite module for database operations. We retrieve the data from SQLite in the form of a dictionary using SQLite's query capabilities. The retrieved data is then interpreted into an executable format and provided to the machine learning model for prediction.

After processing the data through the machine learning model, the predicted outcome triggers a web page request, which is then returned to the user through the Python-based web framework used for the frontend. This framework handles the routing and rendering of the web pages based on the model's predictions, effectively facilitating the prediction and intervention of obesity and overweight at early stage.

IMPLEMENTED UIS

In addition to the features mentioned, the automated obesity detection system incorporates camera detection technology for enhanced user experience and accuracy. By integrating camera detection capabilities into the web application, users have the option to capture images for further analysis and assessment of their body composition.

Upon accessing the web application, users are prompted to utilize their device's camera to capture a full-body image or specific areas of interest, such as the waistline or hip region. The captured images are then processed using image recognition algorithms to extract relevant metrics, such as waist circumference, hip circumference, and overall body shape (Figures 2–5).

These extracted metrics are integrated into the machine learning model alongside traditional input parameters like height, weight, and age. By incorporating visual data from camera detection, the system enhances the accuracy of obesity classification by considering additional factors, such as body shape and distribution of adipose tissue.

Here, Table 1 represents various machine learning (ML) models including Support Vector Machines (SVM), Random Forest, Decision Trees (DT), and Logistic Regression (LR) were evaluated for their effectiveness in predicting obesity levels in individuals. Results revealed that SVM exhibited the highest accuracy at 97.4%. Consequently, we opted to utilize SVM as the primary algorithm for training and testing our model. Subsequent sections present the outcomes derived from this selected approach.

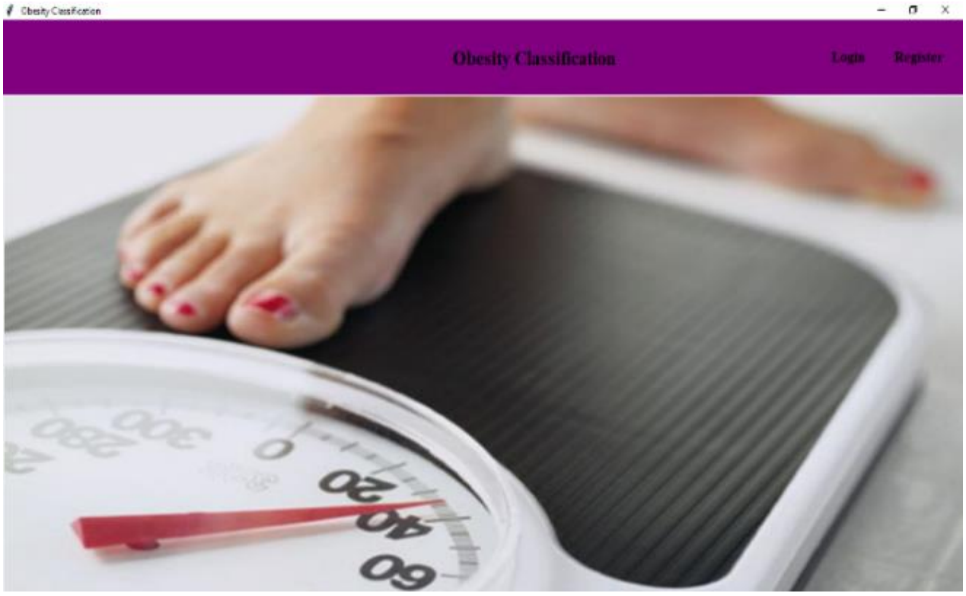

Figure 2. Home Page.

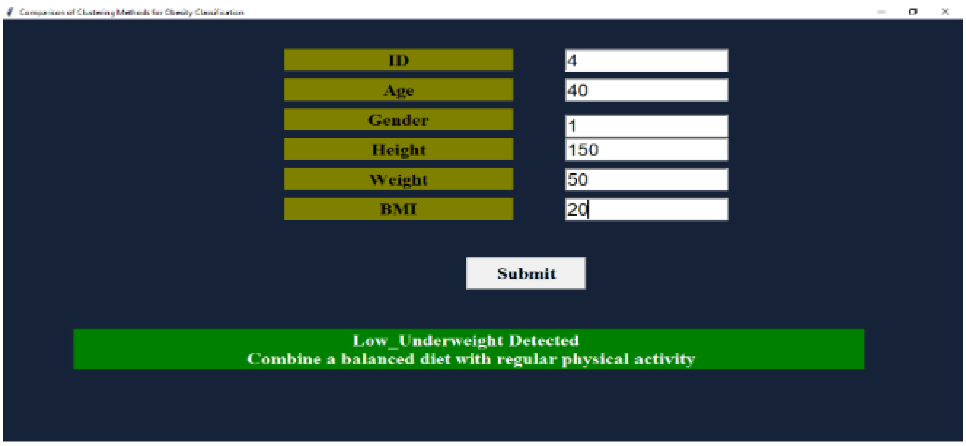

Figure 3. User Data Entry and Prediction 1.

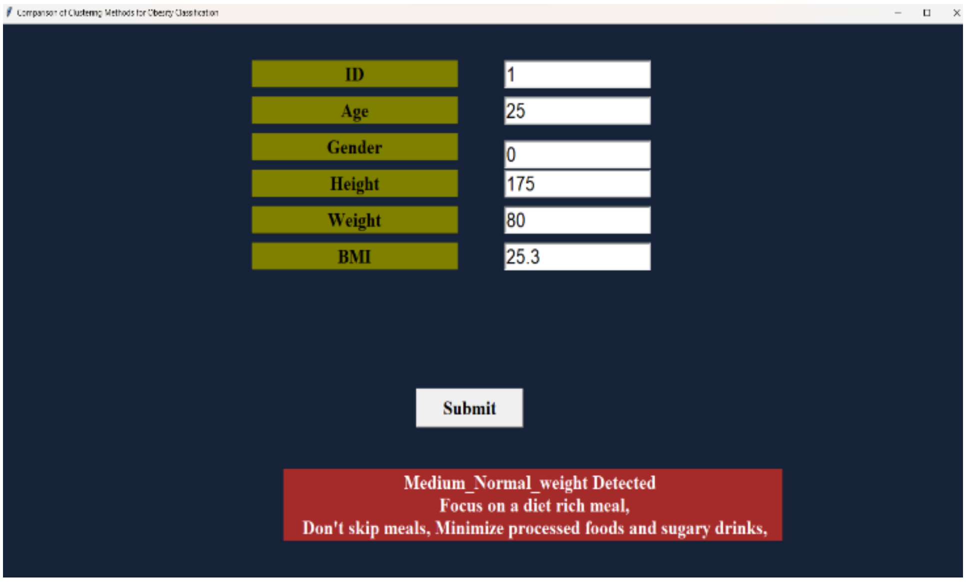

Figure 4. User Data Entry and Prediction 2.

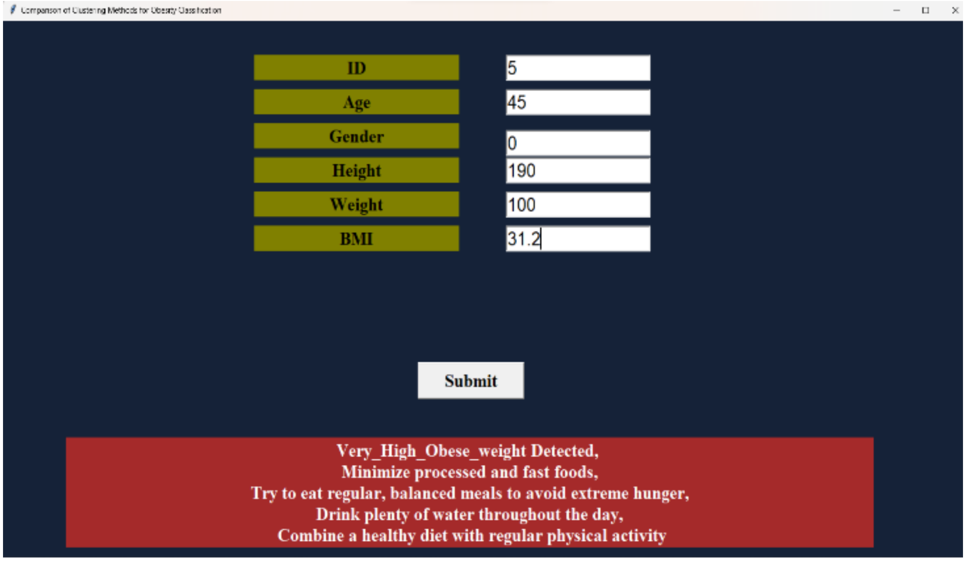

Figure 5. User Data Entry and Prediction 3.

Table 1. Comparison between used different Ml Models.

Model	Accuracy
SVM (Support Vector Machine)	0.97
RF (Random Forest)	0.87
DT (Decision Tree)	0.91
LR (Logistic Regression)	0.78

Here, in Table 2, a comparative analysis of previous research papers was conducted, focusing on the algorithms employed and their corresponding accuracy rates. Our study achieved an accuracy of 97%, surpassing other works, such as “A Machine Learning Approach for Obesity Risk Prediction,” which attained the second-highest accuracy of 93%. This synthesis underscores the efficacy of our approach in accurately predicting obesity risk.

Table 2. Synthesis of previous research.

Research Paper’s Name	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
Our Introduced Model	*0.97*

CONCLUSIONS

Exploration of ML-driven this introduced model underscores the immense potential of technology in addressing this global health challenge. Through this these are achieved:

- *Accurate Categorization:* Demonstrated the capability of machine learning to accurately categorize individuals based on their body weight and composition. This categorization enables us to identify and comprehend the prevalence and severity of obesity within diverse populations, paving the way for more targeted interventions.
- *Clear Definition:* By employing measurable criteria, such as Body Mass Index, waist size, and fat of body in percentage, developed a clear and consistent definition of obesity. This standardized approach ensures that healthcare professionals and researchers worldwide can work from a common foundation, facilitating better collaboration and understanding.

- *Effectiveness:* This classification system to assess the effectiveness of interventions and treatment strategies for obesity. By measuring outcomes and comparing results across different patient populations, we gain valuable insights into which approaches yield the best results, ultimately enhancing the quality of care.

Throughout this investigation, a multitude of methodologies were employed to achieve optimal precision rates in diagnosing obesity. Leveraging the SVM algorithm, our study achieved an impressive precision rate of 97%. Machine Learning emerges as a pivotal instrument, enabling us to unveil valuable insights. The website developed serves as a valuable resource, empowering users to access information on preventive measures and establish novel health objectives within their wellness regimen.

Acknowledgment

I extend our deepest appreciation to Mrs. Anushri Kulkarni, Assistant Professor, for her invaluable guidance, support, and encouragement throughout this research. Her insights and expertise in the field of machine learning were instrumental in shaping the direction of this research.

We would also like to express our gratitude to the faculty and staff of Department of Electronics and Telecommunication, Smt. Kashibai Navale College of Engineering, for providing us with the necessary resources and facilities to carry out this project.

Finally, we extend our appreciation to all those who have directly or indirectly supported our work. This research would not have been possible without their encouragement and belief in our endeavor.

REFERENCES

1. Singh AS, Mulder C, Twisk JWR, van Mechelen W, Chinapaw MJM. Tracking of childhood overweight into adulthood: A systematic review of the literature. Obes Rev. 2008;9(5):474–88.
2. Freedman DS, Mei Z, Srinivasan SR, Berenson GS, Dietz WH. Cardiovascular risk factors and excess adiposity among overweight children and adolescents: The Bogalusa Heart Study. J Pediatr. 2007;150(1):12–7.
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(6):288–93.
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(6):1384–90.
5. Subramanyam V, Rafi M. Prevalence of overweight and obesity in affluent adolescent girls in Chennai in 1981 and 1998. Indian Pediatr. 2003;40(8):332–6.
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(2):135–9.
7. Health and Social Care Information Centre. Health Survey for England 2015: Children's body mass index, overweight and obesity. Leeds: NHS Digital; 2016.
8. Ferdowsy F, Rahi KSA, Jabiullah MI, Habib MT. A machine learning approach for obesity risk prediction. Curr Res Behav Sci. 2021;2:100053. doi:10.1016/j.crbeha.2021.100053.
9. Ramya A, Rohini K. Comparative evaluation of machine learning classifiers with obesity dataset. In: 2021 1st International Conference on Computational Sciences and Soft Computing (ICCS). IEEE; 2021. p. 72–6. doi:10.1109/ICCS54944.2021.00016.
10. Alkhalaf M, Yu P, Shen J, Deng C. A review of the application of machine learning in adult obesity studies. Adv Clin Inform. 2022;2(1):32–48. doi:10.3934/aci.2022002.
11. Alexander DS, Cao C, Alfonso ML. Examining whether the Social Cognitive Theory concepts predict childhood obesity prevention outcome expectations. Health Educ Behav. 2021;41(2):143–51. doi:10.1177/0272684X20915383.
12. Singh B, Tawfik H. Machine learning approach for the early prediction of the risk of overweight and obesity in young people. In: Hassanien AE, Slowik A, Snášel V, El-Deeb H, editors. Proceedings of the International Conference on Innovations in Bio-Inspired Computing and Applications. Springer; 2020. p. 471–81. doi:10.1007/978-3-030-50423-6_39.

13. Colmenarejo G. Machine learning models to predict childhood and adolescent obesity: A review. Nutrients. 2020;12(8):2466. doi:10.3390/nu12082466.
14. Blüher S, Meigen C, Gausche R, Keller B, Pfäffle R, Sabin M, et al. Age-specific stabilization in obesity prevalence in German children: a cross-sectional study from 1999 to 2008. Int J Pediatr Obes. 2010;6(Suppl 2):e199–206. doi:10.3109/17477166.2010.526305.
15. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. JAMA. 2005;293(15):1861–7.
16. Sidorov G, Velasquez F, Stamatatos E, Gelbukh A, Chanona-Hernandez L. Syntactic n-grams as machine learning features for natural language processing. Expert Syst Appl. 2014;41(3):853–60.
17. Hossain R, Mahmud SH, Hossin MA, Noori SRH, Jahan H. PRMT: Predicting risk factor of obesity among middle-aged people using data mining techniques. Procedia Comput Sci. 2018;132:1068–76.
18. Di Nardo F, Casagrande F, Boemi M, Fumelli P, Morosini P, Burattini R. Insulin resistance in hypertension quantified by oral glucose tolerance test: Comparison of methods. Metabolism. 2006;55(2):143–50.
19. Liang Y, Kelemen A. Sequential support vector regression with embedded entropy for SNP selection and disease classification. NIPHA. 2011;4(3):301–12.

