A Survey on Machine Learning Approach for Early Detection and Prevention of Obesity and Overweight

PROF. M. K. NIVANGUNE[1], SIMRAN SAYYAD[2], MRUNAL SHEWALE[3], PRACHI RANE[4], PRERANA MADHAVI[5]
[1, 2, 3, 4, 5] *Computer Engineering, SAE, Kondhwa*

Abstract— Obesity stands as a pressing global crisis, affecting more than 2.1 billion individuals, with projections indicating that 41% of the global population could be overweight or obese by 2030. This widespread issue poses severe health hazards, including diabetes and cardiovascular complications. To combat this epidemic, a comprehensive dataset has been meticulously crafted utilizing information from educational institutions, centering on Body Mass Index (BMI) as a key metric. Early identification holds paramount importance for implementing timely interventions, such as lifestyle modifications. A systematic framework has been devised to forecast BMI and body fat percentage, alongside offering personalized preventive strategies. Furthermore, real-time detection mechanisms have been integrated to swiftly ascertain an individual's obesity status. Leveraging data gathered from schools and colleges, effective models for obesity detection and prevention have been developed. The outcomes are consolidated and showcased through a desktop application, enriched with an array of preventive measures and calculators.

Indexed Terms- Obesity; Overweight; Prediction; Prevention; Desktop application; Machine Learning.

I. INTRODUCTION

Childhood obesity is a pressing public health concern due to its link with various detrimental health outcomes. Obese children are at a higher risk of developing chronic conditions like type 2 diabetes, cardiovascular diseases, and certain cancers. Additionally, overweight children are more likely to become obese adults, further exacerbating their susceptibility to these illnesses. The prevalence of childhood obesity has surged in recent decades, impacting not only affluent nations but also emerging economies.

Body mass index (BMI) is a standard measure used to assess an individual's weight relative to their height. Calculated by multiplying a person's height in square meters by their weight in kilograms, BMI is widely employed to gauge the risk of obesity and associated health complications, despite its limitations in accurately measuring body fatness.

Besides relying on BMI as a metric, our system incorporates a real-time obesity detection feature utilizing the individual's camera. Through the implementation of crucial algorithms and rigorous model training, we have seamlessly integrated this additional capability to classify individuals into either obese or non-obese categories. This augmentation introduces a dynamic and instant aspect to our approach, enhancing its comprehensiveness and proactiveness in assessing and tackling obesity.

The timely detection of individuals prone to obesity is essential for initiating preventive actions. These interventions might encompass dietary adjustments, heightened physical activity, and behavioral changes. Tackling obesity during its initial phases has the potential to alleviate or diminish the likelihood of adverse health consequences, promoting general health and wellness.

To implement efficient preventive strategies, a web application that forecasts BMI or obesity levels, body fat percentage, and suggests personalized interventions is essential. A specialized program has been created and incorporated into the system to outline the necessary preventive measures customized for each person.

Data gathering was carried out via online forms and email submissions. The dataset was combined with newly obtained information to construct a highly effective and realistic model.

II. LITERATURE REVIEW

Paper 1: Obesity Risk Factors Extraction from Real Outpatient Records Using Medical Data and Machine Learning

Yihan Deng, Peter Dolog, Markus Gass, and Kerstin Denecke explore the diverse factors influencing obesity in their research, emphasizing the significance of genetics and Single Nucleotide Polymorphisms (SNPs). Their study seeks to pave the way for personalized medicine by comprehensively investigating the impact of SNPs on obesity. To accomplish this objective, they utilize actual outpatient records and employ sophisticated machine learning methods to analyze and derive valuable insights from the data.[7]

Paper 2: Obesity-Related Disease Prediction from Healthcare Communities Using Machine Learning

Naomi Christianne Pereira, Jessica D'Souza, Parthi Hanai, and Supriya Solanki present an in-depth paper that explores the prediction of diseases related to obesity within healthcare communities. Their research centers on employing machine learning techniques to furnish healthcare professionals in primary healthcare settings with valuable insights and predictive tools. By concentrating on conditions linked with obesity, this paper seeks to introduce an innovative and proactive approach to healthcare management.[8]

Paper 3: FT-IR Sales Profiling in Patients with Obesity and Obesity-Related Insulin Resistance

S.A. Pullano, M. Greco, M.G. Bianco, D.P. Foti, A. Brunetti, and A.S. Fiorillo explore the fascinating intersection of obesity and insulin resistance. Their study likely utilizes FT-IR (Fourier-transform infrared spectroscopy) profiling to examine patients with both obesity and insulin resistance, a condition intimately linked to obesity. This research aims to provide a thorough comprehension of the correlations between obesity and insulin resistance through the application of advanced analytical techniques.[9]

Paper 4: Classification of SNPs for Obesity Analysis using FARNeM Modelling

Phaik-Ling Ong, Yun-Huoy Choo, and Nurul A. Emran advance obesity research by concentrating on the classification of Single Nucleotide Polymorphisms (SNPs) using FARNEM modeling. Their paper introduces a distinctive approach aimed at uncovering insights into the genetic factors underlying obesity, with the ultimate goal of furnishing a systematic framework for analyzing these pivotal genetic components.[10]

Paper 5: Fuzzy Chest Pain Assessment for Symptomatic and Obesity-Related Health Conditions

Thiago Orsi, Ernesto Araujo, and Ricardo Simoes present a significant paper focused on evaluating chest pain symptoms in individuals with health conditions associated with obesity. Their research introduces an alternative and insightful approach for healthcare professionals to assess and manage chest pain among symptomatic patients with obesity-related health issues. It underscores a proactive healthcare approach to addressing the convergence of obesity and chest pain symptoms.[11]

III. ALGORITHMS

1)Support Vector Machines (SVM):

The support vector machine (SVM) is a powerful machine learning algorithm utilized in both classification and regression tasks. Its main goal is to discover a hyperplane that effectively separates data points, thereby maximizing the margin between different classes. SVM finds practical applications in various domains including text detection, character recognition, object identification, and human activity recognition. Known for its robust theoretical foundations and efficient problem-solving abilities, SVM is renowned for its versatility in tackling complex tasks.[1]

2) Decision Tree:

Decision Trees stand as a versatile machine learning algorithm applicable to both classification and regression tasks. Their core mechanism involves partitioning the data into subsets based on feature values, thereby constructing a tree-like structure for predictive purposes. An ensemble method called Random Forest enhances accuracy and addresses overfitting by combining multiple decision trees. While Decision Trees are valued for their interpretability, it's crucial to recognize that pruning may be required to mitigate bias.

3) Random Forest:
Random Forest is an ensemble learning technique that utilizes multiple decision trees (CART) to enhance predictive accuracy, particularly effective for handling complex datasets. By aggregating outputs from numerous decision trees, Random Forest reduces the risk of overfitting, thereby ensuring robust and dependable results. Its widespread application across various domains, such as finance and healthcare, highlights its significance in addressing intricate prediction tasks.[1],[8]

4) Logistic Regression:
Logistic Regression: emerges as a widely used statistical and machine learning algorithm tailored for binary classification tasks. Despite its name including "regression," logistic regression primarily serves classification purposes, predicting the probability of an event based on input features. Its simplicity and interpretability make it a preferred choice, particularly for newcomers to machine learning. This algorithm excels in scenarios necessitating binary outcome predictions, such as spam detection or medical diagnosis. Our approach integrates real-time obesity detection utilizing the person's system camera. By employing the Logistic Regression algorithm and conducting model training, we introduce an additional layer of categorization, determining whether individuals belong to the obese or non-obese category.

4)Naïve Bayes:
In machine learning, the Naive Bayes algorithm is a probabilistic classification technique based on Bayes' theorem with an assumption of independence between predictors. It is commonly used for classification tasks, particularly in text classification and spam filtering. The algorithm calculates the probability of each class given a set of input features and selects the class with the highest probability as the predicted outcome. Despite its simplicity and the "naive" assumption of feature independence, Naive Bayes often performs well in practice, especially with large datasets. It is computationally efficient and requires minimal training data, making it a popular choice for various classification problems in real-world applications.

These algorithms offer diverse solutions for classification and regression challenges, with their suitability depending on unique data characteristics and task requirements. The selection of the appropriate algorithm hinges on specific data attributes and task objectives.

IV. SYSTEM DESIGN

Architecture for a Machine Learning-Based Obesity Detection System:

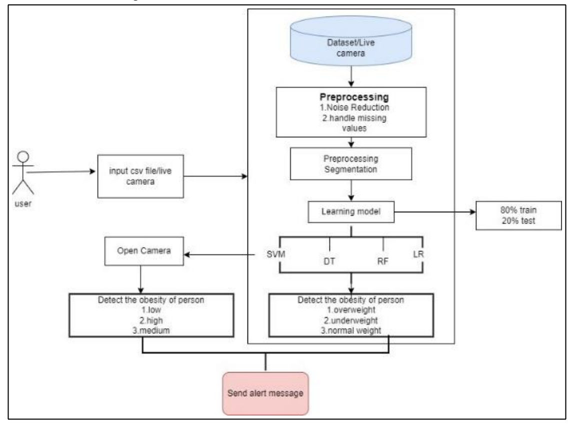

Figure 1: SYSTEM ARCHITECHTURE DIAGRAM

Admin Module:
Within this module, the administrator is required to log in using valid credentials to access administrative privileges. Once logged in successfully, the administrator can perform various operations, including viewing all registered users and authorizing their access to the system.

User Management:
This module enables the administrator to oversee user registration and authorization processes. The administrator has the capability to view a comprehensive list of registered users, encompassing details such as usernames, emails, and addresses. Additionally, the administrator can authorize user access, ensuring that only authorized individuals gain entry into the system.

End User Module:
In this module, users are required to register before gaining access to system functionalities. Upon successful registration, user details are securely stored in the database. Subsequently, authenticated users can log in using their authorized credentials to perform various operations such as managing their accounts and accessing personalized features.

V. IMPLEMENTATION DETAILS

1)System Architecture:
Overview:
The desktop application architecture has been meticulously crafted to accommodate the integration of sophisticated machine learning algorithms for the early detection and prevention of obesity. Through thoughtful design and implementation, the application seamlessly incorporates modules for real-time obesity detection via camera input, utilizing Logistic Regression (LR) algorithm. Additionally, a comparative analysis of obesity detection algorithms, including SVM, Decision Tree, Random Forest, and Naive Bayes, has been facilitated.

2)Data Collection and Integration:
Overview:
A robust data collection framework has been established, leveraging online forms and email surveys to gather comprehensive datasets. These datasets have been meticulously integrated, amalgamating existing data with newly acquired information to enrich the model's efficiency and efficacy.

3)Algorithm Implementation:
Overview:
The implementation of the LR algorithm for real-time obesity detection represents a significant milestone, showcasing the successful deployment and utilization of machine learning techniques in a practical setting. Furthermore, the individual implementation of SVM, Decision Tree, Random Forest, and Naive Bayes algorithms highlights the diversity of approaches employed for obesity prediction and classification.

4)Model Training and Evaluation:
Overview:
Rigorous model training procedures have been executed, utilizing the collected datasets to train machine learning models capable of predicting BMI, obesity levels, and body fat percentage. Subsequent performance evaluation of these models underscores the SVM's superior accuracy in predicting obesity status, providing valuable insights into algorithm performance.

5)User Interface and Recommendations:
Overview:
An intuitive and user-friendly interface has been meticulously designed, providing individuals with a seamless platform to input their data and receive personalized recommendations for preventive measures. The integration of the World Health Organization's (WHO) classification system for obesity further enhances the application's utility, categorizing individuals based on BMI ranges and offering tailored interventions.

6)Integration and Deployment:
Overview:
The successful integration of all application components into a cohesive framework ensures smooth interaction and interoperability between different modules. Following rigorous testing, the application has been deployed across desktop platforms, enabling widespread accessibility to early detection and preventive measures for obesity.

7)Testing and Validation:
Overview:
Extensive testing, including unit testing, integration testing, and user acceptance testing, has been conducted to validate the reliability and accuracy of the application. Real-world scenarios and user feedback have further validated the effectiveness of implemented algorithms and recommendations, ensuring alignment with user expectations and healthcare standards.

8)Continuous Improvement:
Overview:
A commitment to continuous improvement drives ongoing refinements to the application's features and algorithms. Incorporating feedback from users and healthcare professionals, iterative updates are implemented to integrate the latest research findings and advancements in machine learning techniques for obesity detection and prevention, ensuring the application remains at the forefront of innovation in healthcare technology.

CONCLUSION

Our project, titled "A Machine Learning Approach for Early Detection and Prevention of Obesity and

Overweight," has thoroughly explored the intricate domain of obesity classification. Recognizing the multifaceted nature of this issue, we have emphasized the importance of categorizing individuals based on their Body Mass Index (BMI), implementing real-time obesity detection through individual camera data, and assessing associated health risks. To establish a robust framework, we have referenced the classification system outlined by the World Health Organization (WHO), which categorizes obesity into distinct classes according to specific BMI ranges. This classification methodology is pivotal not only for identification but also for addressing the health implications linked with obesity. Our project underscores the significance of such an approach in combatting the global epidemic of obesity and overweight conditions, highlighting the necessity for early detection strategies and preventive interventions.

REFERENCES

[1] Nilesh P. Sable, Ratnmala Bhimanpallewar, 3 Raj Mehta, Sara Shaikh, Anay Indani, Soham Jadhav, "A Machine Learning approach for Early Detection and Prevention of Obesity and Overweight" in 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) Pune, India. Apr 7-9, 2023.

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

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

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

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

[6] Yihan Deng, "Obesity entity extraction from real outpatient records: When learning-based methods meet small imbalanced medical data sets" in 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).

[7] Naomi Christianne Pereira, Jessica D'Souza, Parthi Rana, Supriya Solaskar, "Obesity related Disease Prediction from Healthcare Communities using Machine Learning" in IEEE – 45670.

[8] S.A. Pullano, M. Greco, M.G. Bianco, D.P. Foti, A. Brunetti, and A.S. Fiorillo, "FT–IR Saliva Profiling in Patients with Obesity and Obesity–Related Insulin Resistance", This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

[9] Phaik-Ling Ong, Yun-Huoy Choo, and Nurul A. Emran, "Classification of SNPs for Obesity Analysis using FARNeM Modelling".

[10] Thiago Orsi, Ernesto Araujo and Ricardo Simoes, "Fuzzy Chest Pain Assessment for Unstable Angina based on Braun Wald Symptomatic and Obesity Clinical Conditions" in 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) July 6-11, 2014, Beijing, China.