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 is a major global epidemic, impacting over 2.1 billion people, with 41% of the world's population projected to be overweight or obese by 2030. This issue poses serious health risks, including diabetes and heart problems. To tackle this crisis, a comprehensive dataset has been created using data from educational institutions, focusing on Body Mass Index (BMI) as an indicator. Early detection is crucial for intervention, including lifestyle changes. A framework has been designed to forecast BMI and body fat percentage, as well as to provide personalized preventative strategies. Real time detection is also inculcated for determining whether the person is obese or not. The program is based on data collected from schools and colleges to create effective models for obesity detection and prevention. The results are compiled and shown on a web application that also contains numerous preventative measures and calculators.**

Keywords: **Obesity; Overweight; Prediction; Prevention; Web application; Machine Learning.**

I. INTRODUCTION

Childhood obesity poses a significant public health concern due to its association with various adverse health effects. Obese children face an increased risk of developing chronic conditions, including type 2 diabetes, cardiovascular disorders, and certain forms of cancer. Another contributing factor to the heightened risk of these illnesses in overweight children is the potential for them to transition into obese adults. Over the past few decades, the prevalence of childhood obesity has risen, not only in affluent nations but also emerging as a growing issue in underdeveloped nations.

Body mass index (BMI) serves as a metric for evaluating an individual's weight relative to their height. Computed by multiplying a person's height in square meters by their weight in kilograms, BMI is a commonly used method to assess the likelihood of obesity and associated health risks, despite its imperfections as an indicator of body fatness.

In addition to utilizing BMI as a metric, our system offers a real-time obesity detection feature through the individual's camera. By employing essential algorithms and conducting model training, we have integrated this supplementary capability to categorize individuals into either the obese or non-obese classification. This enhancement adds a dynamic and immediate dimension to our approach, contributing to a more comprehensive and proactive method for assessing and addressing obesity.

Early identification of individuals at risk of obesity is crucial for implementing preventive measures. These interventions may involve changes in diet, increased physical activity, and behavior modification. Addressing obesity in its early stages can potentially mitigate or reduce the risk of negative health outcomes, fostering overall health and well-being.

To establish effective preventive measures, a web application predicting BMI or obesity levels, body fat percentage, and recommending personalized interventions is necessary. A program has been developed and integrated into the framework to delineate the required preventative actions tailored to each individual.

Data collection was conducted through online forms and emails. The dataset was amalgamated with newly acquired data to create a highly efficient and realistic model.[1]

II. LITERATURE REVIEW

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

In their research, Yihan Deng, Peter Dolog, Markus Gass, and Kerstin Denecke delve into the multifaceted factors contributing to obesity, with a particular focus on the role of genetics and Single Nucleotide Polymorphisms (SNPs). This study aims to lay the groundwork for personalized medication by thoroughly examining the effects of SNPs on obesity. To achieve this, they employ real outpatient records and apply advanced machine learning techniques to analyse and extract meaningful 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 a comprehensive paper that delves into the prediction of obesity-related diseases within healthcare communities. This research revolves around the application of machine learning methods to offer valuable insights and predictive tools to healthcare professionals in primary health care settings. By focusing on the conditions associated with obesity, this paper aims to provide 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 delve into the intriguing intersection of obesity and insulin resistance. Their work likely employs FT-IR (Fourier-transform infrared spectroscopy) profiling to scrutinize patients with both obesity and insulin resistance—a condition closely related to obesity. This research seeks to offer a comprehensive understanding of the connections between obesity and insulin resistance through 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 contribute to the field of obesity research by focusing on the classification of Single Nucleotide Polymorphisms (SNPs) using FARNEM modelling. This paper's unique approach strives to provide insights into the genetic factors contributing to obesity, ultimately offering a systematic framework for analysing these crucial genetic elements.[10]

Paper 5: Fuzzy Chest Pain Assessment for Symptomatic and Obesity-Related Health Conditions
Thiago Orsi, Ernesto Araujo and Ricardo Simoes present a noteworthy paper centred on the assessment of chest pain symptoms in individuals facing health conditions related to obesity. This research offers an alternative and insightful approach for healthcare professionals in the evaluation and management of chest pain among symptomatic patients with obesity-related health concerns. It emphasizes a proactive approach to healthcare in addressing the intersection of obesity and chest pain symptoms.[11]

III. ALGORITHMS

1)Support Vector Machines (SVM): It stands out as a potent machine learning algorithm employed in both classification and regression tasks. The primary objective of SVM is to identify a hyperplane that effectively distinguishes data points, maximizing the margin between various classes. This algorithm finds practical applications in diverse fields such as text detection, character recognition, object identification, and the recognition of human activities. Renowned for its sturdy theoretical underpinnings and effective problem-solving capabilities, SVM is recognized for its versatility in addressing complex tasks.[1]

2) Decision Tree: Decision Trees represent a versatile machine learning algorithm applicable to both classification and regression tasks. The fundamental mechanism involves dividing the data into subsets based on feature values, constructing a tree-like structure for predictive purposes. An ensemble method known as Random Forest enhances accuracy and mitigates overfitting by amalgamating multiple decision trees. While Decision Trees are acknowledged for their interpretability, it's important to note that pruning may be necessary to prevent bias.

3) Random Forest: Random Forest stands as an ensemble learning method employing multiple decision trees (CART) to improve predictive accuracy, especially adept at managing intricate datasets. Through the amalgamation of outputs from numerous decision trees, Random Forest diminishes the likelihood of overfitting, ensuring robust and reliable results. Widely applied across diverse domains, including finance and healthcare, this technique proves valuable in addressing complex prediction challenges. [1],[8]

4) Logistic Regression: Logistic regression stands out as a commonly employed statistical and machine learning algorithm designed for binary classification. This model predicts the probability of an event happening based on input features. Despite the term "regression" in its name, logistic regression is predominantly utilized for tasks related to classification. Its appeal lies in its simplicity and interpretability, making it a favoured option for those new to machine learning endeavours. This algorithm proves effective in scenarios requiring the prediction of binary outcomes, such as spam detection or medical diagnosis. Our approach incorporates real-time obesity detection using the person's system camera. Through the implementation of LR algorithm and model training, we provide an extra layer of categorization, determining whether an individual falls into the obese or non-obese category.

These algorithms provide diverse solutions for classification and regression challenges, their appropriateness contingent on the unique data characteristics and task requirements. The choice depends on specific data traits and task needs.

IV. SYSTEM ARCHIETECTURE

System Architecture for Machine Learning-Based Obesity detection

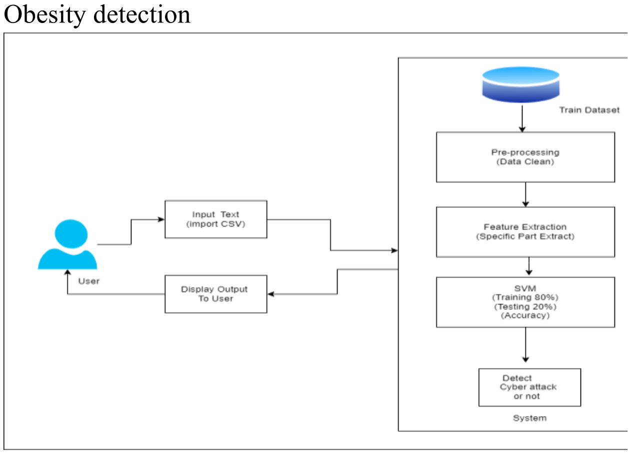

Figure 1: SYSTEM ARCHITECHTURE DIAGRAM

Data Collection Layer:

Sensors: various sensors and monitoring devices can be integrated into the healthcare infrastructure to continuously gather relevant health data. These sensors might include wearable devices for tracking vital signs, dietary intake monitors, and physical activity sensors. This real-time data collection provides essential inputs for assessing a person's health status.

Database: All the health data collected from sensors is securely stored in a centralized database system designed to meet healthcare data security standards. Regular backups and data retention policies are maintained to ensure data availability and integrity.

Data Preprocessing Layer:

Data Cleaning: Data cleaning in the obesity detection project involves the rigorous removal of outliers, noise, and redundant or erroneous health data. This step is essential to ensure the integrity and reliability of the health information used for analysis. Outliers in health data can be caused by measurement errors or anomalies in patient records.

Feature Extraction: These features may include patient demographics, Body Mass Index (BMI), physical activity levels, dietary patterns, genetic information, and family medical history. These features are crucial for creating a comprehensive patient profile for accurate obesity risk assessment.

Normalization/Standardization: Normalization or standardization is a vital step in the obesity detection project. It ensures that all the health data features are on a consistent scale.

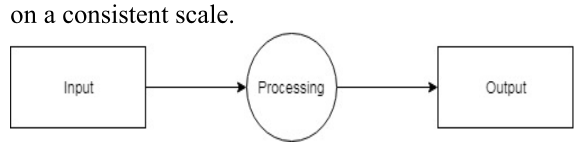

Figure 2: DFD LEVEL 0

Spatiotemporal Analysis Layer:

Graph Laplacian: Utilized for structured analysis of health data, capturing spatial and temporal dependencies relevant to obesity trends.

Temporal Analysis Engine: Analyses time-series health data for obesity-related trends and seasonality.

Spatial Analysis Engine: Examines spatial relations to identify geographical variations in obesity prevalence.

Machine Learning Layer:

Training Module: Uses historical data to train machine learning models like SVM, Decision Trees, and Random Forests.

Detection Module: Applies trained models to real-time health data, detecting obesity risk factors in individuals.

Prediction Module: Forecasts future obesity trends and risk patterns based on historical data and trends.

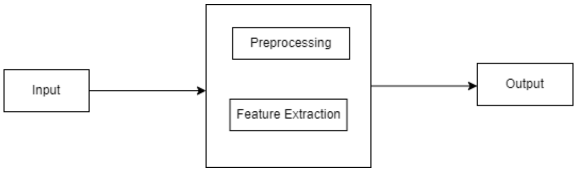

Figure 3: DFD LEVEL 1

Decision & Response Layer:
Threshold-based Detector: Establish dynamic thresholds, adapting to evolving health data, to flag individuals at risk of obesity when certain criteria are met.
Alert System: In the event of detecting obesity risk factors, an automated alert system can instantly notify healthcare professionals or individuals regarding potential health concerns.
Auto-response System: Implement automated responses, like suggesting dietary changes or exercise routines, for individuals at risk of obesity. This system can also assist in creating personalized prevention plans.

Feedback Loop:
Re-training System: Periodically retrain machine learning models with new health data and emerging obesity patterns to ensure the system maintains high accuracy in detecting obesity risks.
Update & Maintenance: Regular updates and maintenance are crucial to keep the system up to date with the latest health information and obesity detection techniques.

Reporting & Visualization:
Dashboard: Offer a user-friendly interface where healthcare professionals and individuals can access real-time health data, view obesity risk assessments, and personalized health plans. Utilize graphs, charts, and tables for a comprehensive view of health status.
Logs: Maintain detailed logs of all activities, including health data analysis, predictions, and preventive actions taken to manage obesity risk.

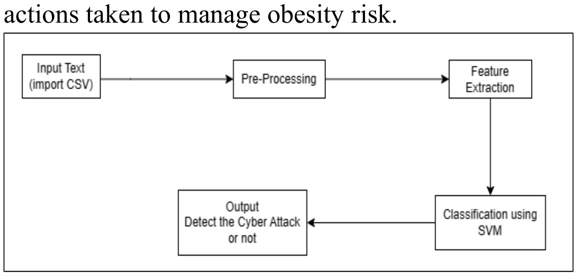

Figure 4: DFD LEVEL 2

Security Layer:
Encryption: Implement robust data encryption protocols to secure all health data, whether in transit or at rest, ensuring the privacy of patient information.
Authentication & Authorization: Implement security measures such as multi-factor authentication and role-based access controls to guarantee that only authorized individuals, including healthcare providers and patients, can access confidential health information within the system.

In the context of the obesity detection project, these layers and security measures contribute to the accurate identification of obesity risk factors and personalized responses while maintaining the privacy and security of health data.

V. KEY ASPECTS OF OBESITY DETECTION PROJECT

In our project aimed at obesity detection using machine learning, we emphasize key aspects that will drive the success of our endeavour. These foundational components are central to our project's goals and objectives:
Aspect 1: BMI as a Vital Feature
Overview:
Body Mass Index (BMI) serves as a fundamental feature in our obesity detection model. It plays a pivotal role in assessing an individual's weight relative to their height and is a critical predictor of obesity risk. We will explore the incorporation of BMI data into our machine learning algorithms to enhance the accuracy of our system.

Aspect 2: Algorithm Selection and Implementation
Overview:
We have carefully chosen a set of machine learning algorithms to underpin our obesity detection system:
-**Support Vector Machine (SVM):** SVM offers robust classification capabilities. We will leverage SVM to build a model capable of accurately identifying obesity risk based on BMI and other relevant features.[1]

- Logistic Regression (LR): Logistic Regression is an essential algorithm for binary classification. We will explore its application in determining obesity risk, particularly with BMI as a central feature.

-Decision Tree: Decision Tree algorithms provide transparency in decision-making. We will assess their potential in explaining obesity risk assessments based on BMI and other factors.

-Random Forest: Random Forest, known for its ensemble learning capabilities, will be employed to create a comprehensive obesity detection model that incorporates BMI as a prominent factor.[1]

Aspect 3: Data Collection and Feature Engineering
Overview:
Data collection will be a crucial phase in our project. We will gather comprehensive datasets that encompass BMI measurements, lifestyle information, genetic markers, and more. Feature engineering will be employed to select and refine relevant attributes for obesity risk assessment.

Aspect 4: Model Evaluation and Validation
Overview:
We prioritize the rigorous evaluation and validation of our models. The performance of each algorithm, considering BMI and other features, will be thoroughly assessed to ensure that our system provides accurate and reliable obesity risk predictions.

Aspect 5: User Interface Development
Overview:
We aim to create a user-friendly interface that allows individuals to input their BMI and related data. The project will also provide real time obesity detection using individual's camera. The interface will provide personalized obesity risk assessments based on the selected algorithms.

Aspect 6: Ethical Considerations and Privacy
Overview:
The ethical aspects of our project are paramount. We will ensure the privacy and consent of individuals while collecting and handling health data. Ethical guidelines and privacy safeguards will be integrated into our project.
By focusing on these key aspects, including the role of BMI as a critical feature and the implementation of SVM, LR, Decision Tree, and Random Forest algorithms, we aim to develop a robust and effective obesity detection system that will contribute to the promotion of health and well-being.

VI. MODEL PERFORMANCE METRICS:

1.Obesity Detection Accuracy:
In our study, we evaluated the model's performance in accurately identifying individuals as obese or overweight based on their BMI and relevant indicators. The accuracy of obesity detection was calculated as follows:
Accuracy = (TP + TN) / (TP + TN + FP + FN)

2.Precision and Recall for Obesity:
Precision, which measures the percentage of true positive obesity predictions among all positive predictions, was computed as follows:
Precision = TP / (TP + FP)

Recall, indicating the percentage of true positives among all actual obese individuals, was calculated as:
Recall = TP / (TP + FN)

3.F1-Score for Obesity:
The F1-Score, which balances precision and recall for obesity detection, was determined using the formula:
mathematical
F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

4.AUC-ROC for Obesity:
To evaluate the model's ability to distinguish between obese and non-obese individuals, we used the Area Under the Receiver Operating Characteristic (ROC) curve.

5.Body Fat Percentage Prediction Accuracy:
If your model predicts body fat percentage, you can assess the accuracy of these predictions using metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE).

MAE = Σ |Actual - Predicted| / Number of Instances
MSE = Σ (Actual - Predicted) ^2 / Number of Instances

6.Recommendation System Effectiveness:
We evaluated the effectiveness of our recommendation system by measuring the percentage of personalized recommendations that were successfully followed by individuals.

VII. CONCLUSION

Our project, "A Machine Learning Approach for Early Detection and Prevention of Obesity and Overweight," has delved into the intricate subject of obesity classification. Recognizing the complexity of this issue, we have highlighted the significance of categorizing individuals according to their Body Mass Index (BMI), real time obesity detection through individual's camera, and the associated health risks. We have referenced the widely accepted classification system provided by the World Health Organization (WHO), which segregates obesity into distinct classes based on specific BMI ranges. This classification framework is crucial for not only identifying but also addressing the health concerns associated with obesity. Our project underscores the importance of this approach in the ongoing battle against the global epidemic of obesity and overweight conditions, emphasizing the need for early detection and preventive measures.

REFERENCE

[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.