"LiveScan: Automated Body Composition Analysis for Weight Monitoring"

[1]Snehal Nimase, [2]Manjusha Narhare, [3]Neetu Chauhan, [4]Abhijit Nikat,
[5]Prof N.H.Deshpande
Department of Information Technology,
Savitribai Phule Pune University,
Sinhgad College Of Engineering,
District: Pune, State: Maharashtra,
Country: India.

Abstract:- **Globally, around 2 billion people are dealing with the adverse health implications of "Corpulence," a term encompassing obesity and other weight-related issues. This widespread concern affects individuals of all ages and backgrounds, significantly impacting global health statistics. According to the World Health Organization (WHO), as of 2023, there are approximately 650 million obese adults, 340 million adolescents, and 39 million children worldwide, alongside a 3-5% prevalence of underweight adults. These conditions contribute to an array of health complications, from type 2 diabetes and cardiovascular diseases in obese individuals to anaemia and osteoporosis in those underweight. In response to this pressing issue, our project utilizes a comprehensive dataset from Kaggle, featuring key indicators such as age, gender, height, weight, and Body Mass Index (BMI). Our goal is to create a user-centric tool that seamlessly integrates dataset analysis and live camera footage for enhanced user experience and accuracy. We employ various machine learning algorithms, including Support Vector Machines (SVM), Random Forest, Decision Trees, and Logistic Regression for live feed analysis. Each algorithm undergoes extensive training and evaluation based on performance metrics like accuracy, precision, recall, and F1-score. The final step involves selecting the most effective algorithm to tackle the complex health challenges associated with corpulence, thereby paving the way for a healthier future through innovative technological solutions.**

Keywords *:- Machine Learning, Corpulence, SVM, Random Forest, Decision Tree, Logistic Regression.*

I. INTRODUCTION

Obesity, a complex health issue marked by excessive weight gain, arises from a variety of factors, including diet and genetics. The World Health Organization (WHO) highlights obesity and overweight as significant health risks for adults, attributing these conditions to the accumulation of too much body fat, influenced by lifestyle, heredity, and metabolic differences. Traditional studies on obesity often depend on self-reported data, which can suffer from inaccuracies. However, the advent of machine learning (ML) and live camera technologies presents a novel avenue for precise, real-time observation and analysis of obesity-related factors, enhancing the potential for timely prevention and intervention measures.

This research investigates the integration of ML algorithms with live camera feeds to predict obesity, aiming to revolutionize our understanding and management of its risks through instantaneous monitoring and predictive analytics. We examine the entire process, from data gathering and feature engineering to model creation, while also considering the ethical implications of using live camera footage. Our system stands out for its dual-input capability, accepting both live visual data and structured CSV files, thus offering users flexibility in data submission according to their convenience.

Live camera functionality enables the direct capture of visual data, providing instant and ongoing analysis, while the CSV option allows for the import of pre-collected structured data, supporting comprehensive batch analysis. This dual approach ensures versatility and ease of use, accommodating different user preferences and application scenarios. Through meticulous data processing and the application of ML algorithms to each input type, our model not only classifies individuals' weight status but also proposes customized dietary recommendations, leveraging technology to address the global challenge of obesity.

II. LITERATURE REVIEW

A systematic literature review spanning 2010-2020 explores obesity research and machine learning techniques for prevention and treatment. Identified factors influencing adult obesity, health consequences, and suitable machine learning methods underscore the critical role of early intervention in mitigating obesity's impact on public health [1]. The work addresses the rapid growth of Machine Learning in biomedical and healthcare communities, focusing on the accurate prediction and prevention of obesity and related diseases in the Indian population. The system integrates Random Forest and AdaBoost algorithms to consider various factors contributing to obesity. This comprehensive approach aims to benefit public health by offering insights into disease risk factors and suggesting preventative measures [2]. This study introduces a mobile application for early obesity screening, utilizing a Convolutional Neural Network (CNN) classifier model. The CNN model, based on MobileNetV2, achieves an accuracy of 87.50% in identifying obesity risk through thermal imaging. The application, deployed on the CAT S62 Pro smartphone, enables efficient and early diagnosis, revolutionizing healthcare accessibility [3]. The article employs dimension reduction techniques, specifically Principal Component Analysis (PCA), to predict obesity levels. Various machine learning methods, including Support Vector Machine (SVM) and Decision Tree, achieve prediction accuracies exceeding 90%. Family history of obesity emerges as a significant feature, emphasizing genetic or dietary influences [4] .This research addresses obesity prediction through machine learning algorithms, including Logistic Regression, Random Forest, and Ensemble techniques like Ada Boost and Voting Classifier. The Logistic Regression model exhibits the highest accuracy, emphasizing its effectiveness in predicting obesity levels [5] .Focusing on the rising global epidemic of obesity, this study employs machine learning methods to predict obesity risk. The Gradient Boosting technique outperforms other classifiers, achieving an accuracy of 97.08%. The study emphasizes the impact of sedentary lifestyles on obesity, especially among young individuals and children in Bangladesh [6].

III. OBJECTIVES

1. To define problem of obesity and its increasing threats in society.
2. To conduct a detailed literature review on obesity classification method and related Technologies.
3. To implement image processing technique for obesity classification using live camera Data.
4. To integrate trained model into the live camera for practical implementation.
5. To evaluate the system’s performance in the live time scenarios and compare results.
6. To establish project as a valuable contribution for the classification of obesity using innovative approach.

IV. ARCHITECTURE

Creating a machine learning model is a structured process that begins with identifying the problem and collecting appropriate data. This is followed by data preprocessing, where issues like missing values and feature scaling are addressed. The dataset is then divided into training and testing segments to test the model's accuracy. Choosing an algorithm that fits the data's nature comes next, leading to the training phase where the model learns from the training data. After training, the model is tested and its performance evaluated using metrics like accuracy. If successful, the model is implemented for real-world predictions, with ongoing adjustments necessary to keep it current and effective.

A) Preprocessing is crucial in machine learning, involving steps to clean, transform, and organize data before model training. Tasks include handling missing values, scaling features, and encoding variables, enhancing the model's robustness.

B) Feature extraction reduces data dimensionality by capturing essential information, aiding model training by focusing on relevant characteristics while minimizing computational complexity and overfitting risk.

C) Classification assigns labels to input data based on learned patterns, constructing models to distinguish between predefined categories. Algorithms like Decision Trees and Support Vector Machines are commonly used for this task.

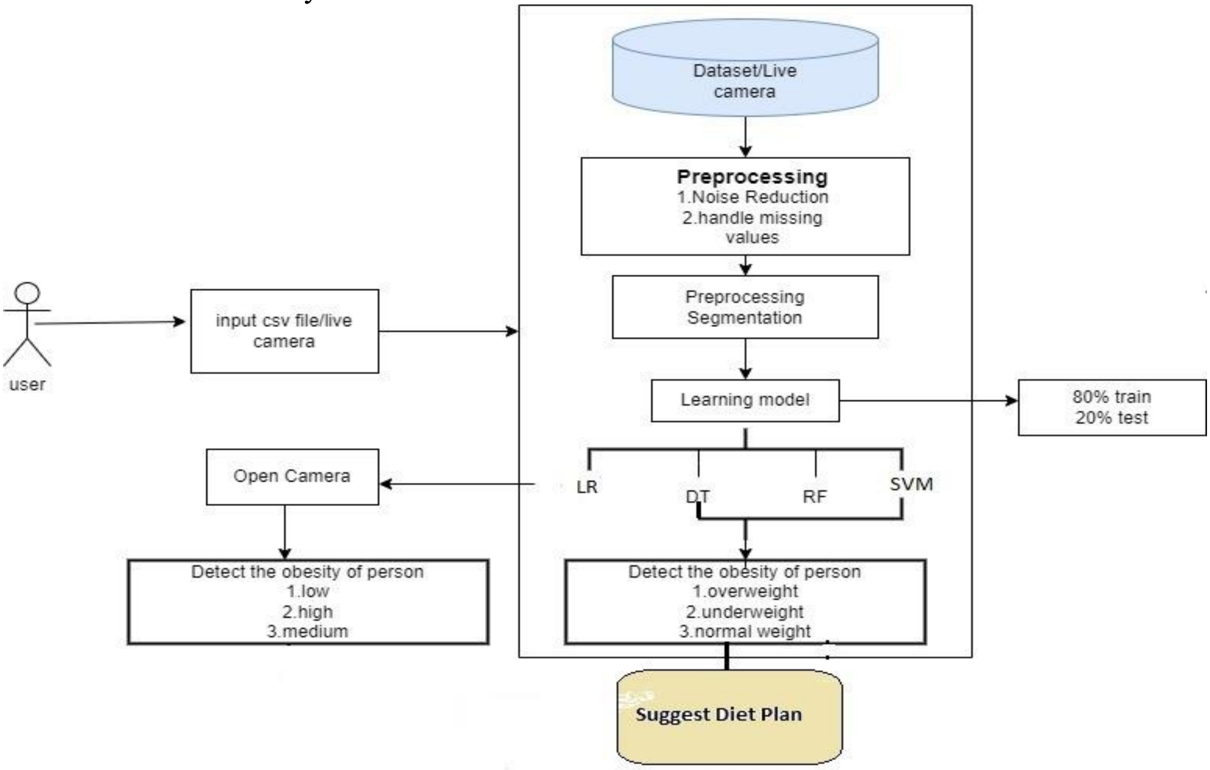


fig .1. System Architecture for Obesity Classification and Detection.

Precision:- It is calculated as the ratio of true positive predictions to the sum of true positives and false positives.

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

Recall:-It is calculated as the ratio of true positive predictions to the sum of true positives and false negatives.

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

F1-Score:-It is a harmonic mean of precision and recall.

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

Accuracy :- It is a performance metric used in machine learning to evaluate the overall correctness of predictions made by a model.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

V. METHODOLOGY

This project aimed to address the global health concern of obesity and malnutrition by leveraging machine learning (ML) algorithms and live camera analysis for prediction and intervention strategies. Here's a breakdown of what was done:

1. Problem Identification: Recognizing the significant health impact of obesity and malnutrition affecting nearly 2 billion people worldwide, the project focused on developing a system to predict and address these conditions.

2. Data Collection and Input Methods: Utilizing datasets from sources like Kaggle, the project incorporated features such as age, gender, height, weight, and BMI. It also integrated live camera technology to capture real-time visual data for analysis.

3. Algorithm Selection and Training: The project employed various ML algorithms including Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression. These algorithms were trained using both dataset inputs and live camera inputs to classify individuals into categories such as overweight, underweight, or normal weight.

4. Evaluation Metrics: The performance of each algorithm was assessed using metrics like accuracy, precision, recall, and F1-score to determine the most effective model for testing.

5. Model Testing and Output: The selected model was used to classify individuals based on their obesity status, providing insights into their health risks. For dataset inputs, the model determined obesity levels (overweight, underweight, normal weight) and recommended personalized diet plans. For live camera inputs, it categorized body types into low, medium, or high body fat levels.

6. Dual-Input System: The project implemented a versatile system allowing users to choose between live camera or CSV file inputs based on their preferences and data availability. This dual-input functionality enhanced usability and flexibility in data processing.

7. Ethical Considerations: The project considered ethical implications, particularly regarding the collection and use of live camera data. Measures were taken to ensure privacy, consent, and responsible data handling throughout the process.

8. Discussion and Future Directions: The paper discussed the integration of ML algorithms with live camera analysis for obesity prediction, emphasizing its potential for real-time monitoring and intervention strategies. Future directions may include refining models, expanding datasets, and exploring additional features for more accurate predictions.

Overall, the project aimed to leverage advanced technologies to address the global challenge of obesity and malnutrition, offering insights and interventions to improve public health outcomes.

5.1 How was research performed?

The systematic literature review conducted from 2010 to 2023 aimed to explore the intersection of obesity research and machine learning techniques for prevention and treatment. The research identified various factors influencing adult obesity, including health consequences, and highlighted the critical importance of early intervention in mitigating obesity's impact on public health. Several studies were analyzed, each employing different methodologies to address the challenge of obesity prediction and prevention:

1. Integration of Machine Learning Algorithms: One study focused on the rapid growth of machine learning in biomedical and healthcare communities. It integrated Random Forest and AdaBoost algorithms to consider multiple factors contributing to obesity, aiming to provide insights into disease risk factors and suggest preventative measures.

2. Development of Mobile Application for Screening: Another study introduced a mobile application for early obesity screening. It utilized a Convolutional Neural Network (CNN) classifier model based on MobileNetV2. The application, deployed on a smartphone with thermal imaging capabilities, achieved an accuracy of 87.50% in identifying obesity risk. This approach revolutionized healthcare accessibility by enabling efficient and early diagnosis.

3. Utilization of Dimension Reduction Techniques: A different approach employed dimension reduction techniques, specifically Principal Component Analysis (PCA), to predict obesity levels. Various machine learning methods, including Support Vector Machine (SVM) and Decision Tree, were utilized, achieving prediction accuracies exceeding 90%. Family history of obesity emerged as a significant feature, highlighting genetic or dietary influences.

4. Evaluation of Machine Learning Algorithms: Another study focused on obesity prediction using machine learning algorithms such as Logistic Regression, Random Forest, and Ensemble techniques like AdaBoost and Voting Classifier. The Logistic Regression model demonstrated the highest accuracy among the models, emphasizing its effectiveness in predicting obesity levels.

5. Emphasis on Sedentary Lifestyles and Obesity: Lastly, a study addressed the rising global epidemic of obesity by employing machine learning methods to predict obesity risk, particularly focusing on the impact of sedentary lifestyles. The Gradient Boosting technique outperformed other classifiers, achieving an accuracy of 97.08% and highlighting the importance of addressing sedentary behaviors, especially among young individuals and children.

Overall, the research encompassed a diverse range of methodologies, including algorithm integration, mobile application development, dimension reduction techniques, and evaluation of machine learning algorithms, to tackle the complex issue of obesity prediction and prevention. Each study contributed valuable insights into understanding obesity risk factors and developing effective interventions.

5.2 Algorithms

We are using mainly four types of algorithms named: SVM, Random Forest, Decision Tree, Logistic Regression to get better result according to their corresponding input. All they fall under “Supervised Machine Learning Algorithm”.

A) SVM:

SVM (Support Vector Machine) is one of the most popular Supervised Machine learning Algorithm. It is not only used for classification but also for the regression. But, the main work of this algorithm is to solve the classification problem in ML. SVM works by finding the optimal hyperplane that best separates data points belonging to different classes in a high-dimensional space. The goal is to maximize the margin, which is the distance between the hyperplane and the nearest data points (called support vectors) of each class

B) Random Forest:

It is most popular Supervised Machine Learning Algorithm. It helps to quickly build any model .The training can be done early to evaluate the model’s performance. Random Forest belongs to the ensemble learning family, which combines multiple individual models to improve predictive performance and reduce overfitting.

c)Decision Tree:

It is the one of the most important Supervised Machine Learning Algorithms.It doesn’t require much computation to perform the classification.It mainly used for the data analysis as it break down complex data into more manageable parts. A Decision Tree organizes data into a hierarchical structure of nodes, where each node represents a feature and each branch represents a decision based on that feature.

D)Logistic Regression:

It falls under the Supervised Machine Learning Algorithm. It will do training on Live Camera as an input .It is used to draw the relationship between one dependent variable and one or more independent variables. Logistic Regression models the probability that a given input belongs to a particular class. It uses the logistic function (also called the sigmoid function) to map the output of a linear combination of input features to a probability value between 0 and 1

5.3 Technologies Used

The technologies used in the "Automated Obesity Detection and Classification via Live Camera Analysis" project include:

1. Python: Chosen as the primary programming language for its versatility and extensive support in the machine learning and deep learning domains.

2. TensorFlow and Keras: Utilized for deep learning implementation, offering user-friendly interfaces, pre-trained models, and efficient GPU acceleration.

3. Scikit-Learn: Employed for machine learning tasks, specifically for harnessing Decision Trees (DT) and Support Vector Machines (SVM) for classification.

4. OpenCV (Open Source Computer Vision Library): Critical for image processing, including image pre-processing, feature extraction, and manipulation.

5. Natural Language Toolkit (NLTK): Utilized for handling text data, facilitating text analysis, tokenization, and feature extraction.

6. Pandas and NumPy: Essential for data manipulation and analysis, including data preprocessing, handling structured data, and performing statistical analysis.

7. Spyder IDE: Chosen as the Integrated Development Environment for code development due to its fast data loading and real-time code suggestions.
8. Tkinter: Utilized as a Python library for creating graphical user interfaces (GUIs), allowing interaction between the user and the application.
9. Windows 10: Selected as the operating system for its compatibility and familiarity, providing a stable environment for developing and deploying the solution.

These technologies collectively enable the development of an automated obesity detection and classification system via live camera analysis, integrating machine learning, deep learning, computer vision, and natural language processing techniques to address the project objectives.

VI. RESULT

6.1 Data Collection

The data for this study comprised both live camera feeds and a dataset obtained from Kaggle. The dataset included features such as age, gender, height, weight, and BMI. Live camera feeds were utilized to capture real-time visual information for analysis.

6.2 Analysis Procedure

The analysis procedure involved several steps:

Data Preprocessing: Both the live camera data and the dataset were subjected to preprocessing techniques such as image cropping, resizing, normalization, and handling missing values.

Feature Extraction: Relevant features, including body coordinates and facial features, were extracted from the live camera data. From the dataset, features such as BMI, age, and gender were extracted.

Model Training: Different machine learning algorithms, including SVM, Random Forest, Decision Tree, and Logistic Regression, were employed for training. SVM, Random Forest, and Decision Tree were trained using the dataset, while Logistic Regression was trained using the live camera data.

Evaluation Metrics: Various evaluation metrics such as accuracy, precision, recall, and F1-score were calculated to assess the performance of the trained models.

6.3 Statistical Tests

Statistical tests were performed to evaluate the effectiveness of the trained models in classifying obesity. These tests included:

Accuracy: The overall correctness of predictions made by the models.

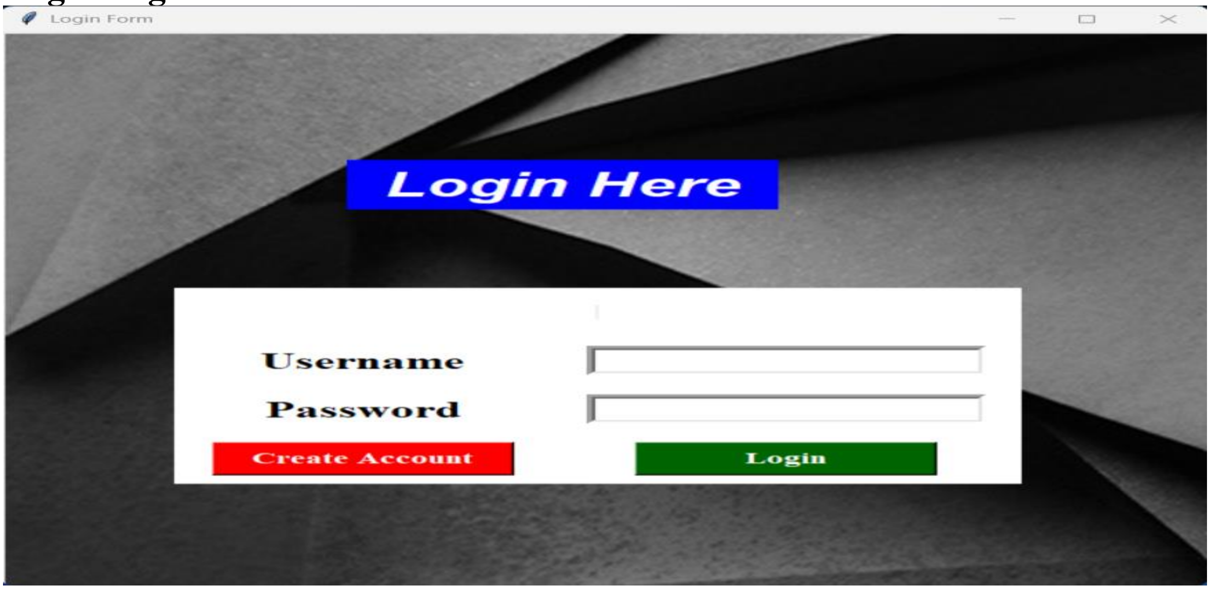
Precision: The ratio of true positive predictions to the sum of true positives and false positives.

Recall: The ratio of true positive predictions to the sum of true positives and false negatives.

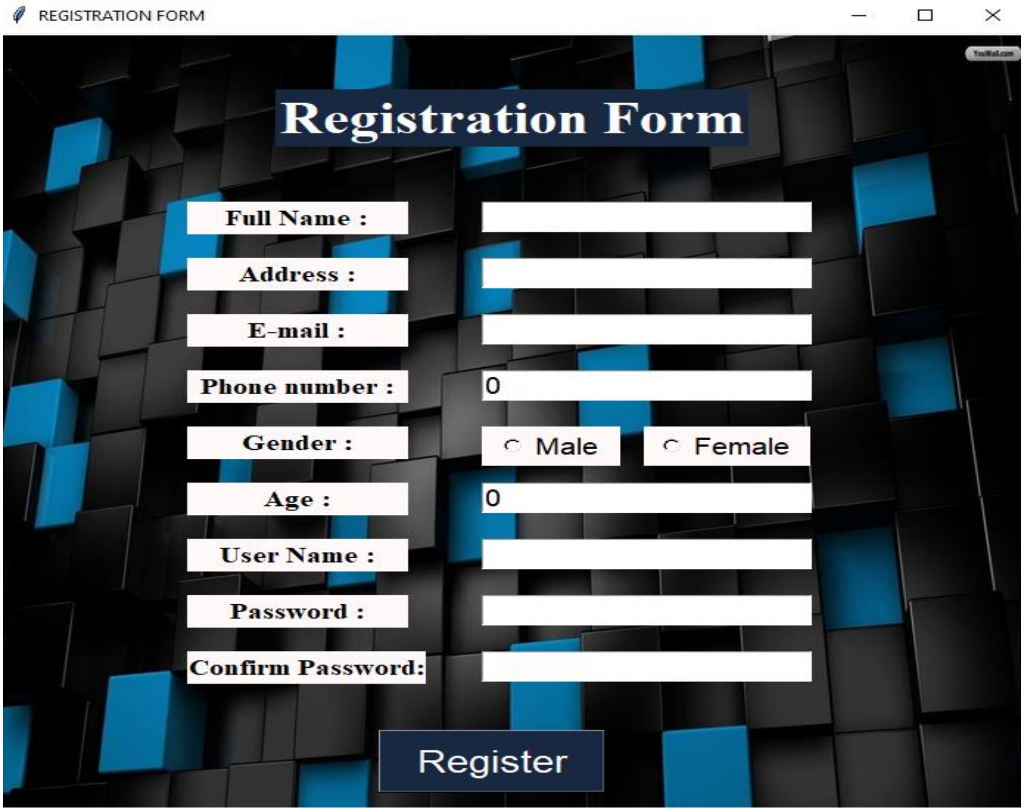
F1-Score: The harmonic mean of precision and recall.

The results of the statistical tests indicated the performance of each machine learning algorithm in classifying obesity. Additionally, the system's ability to provide real-time monitoring and predictive modeling of obesity risk factors was evaluated based on the analysis of live camera data.

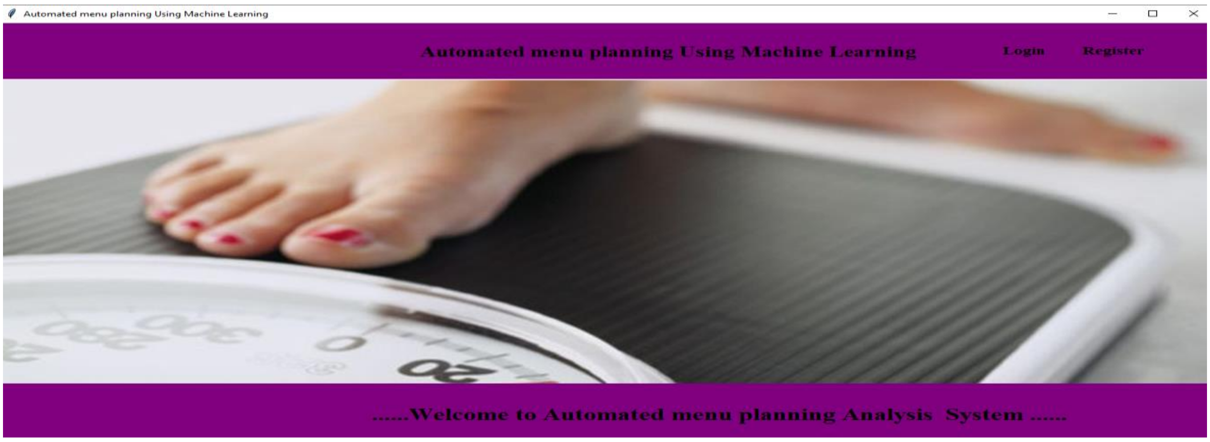
Login Page:



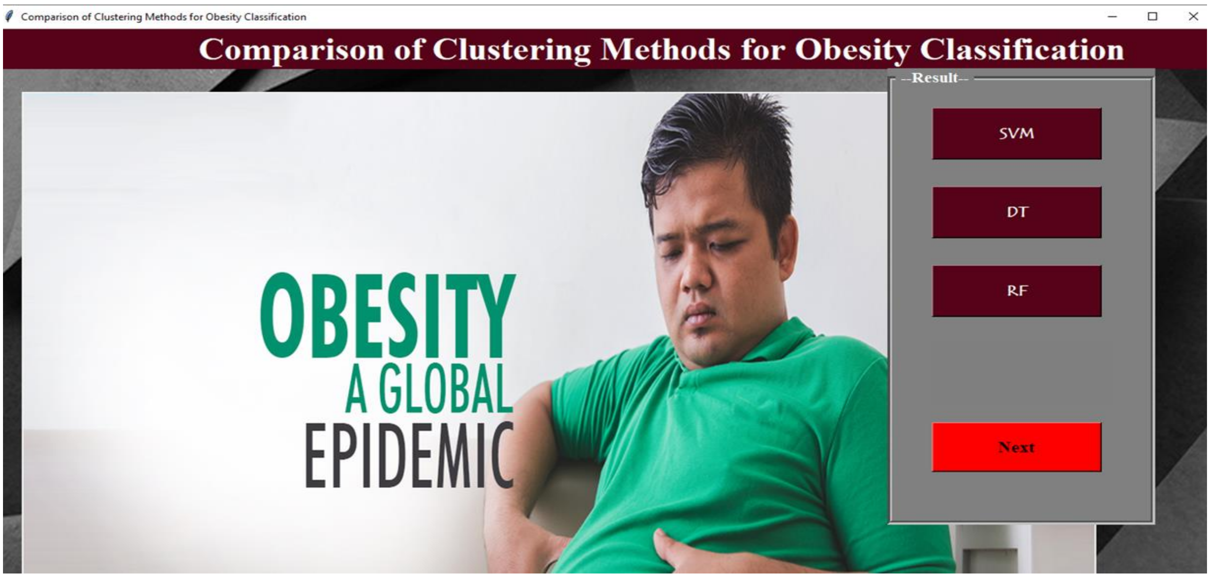
Registration Page:



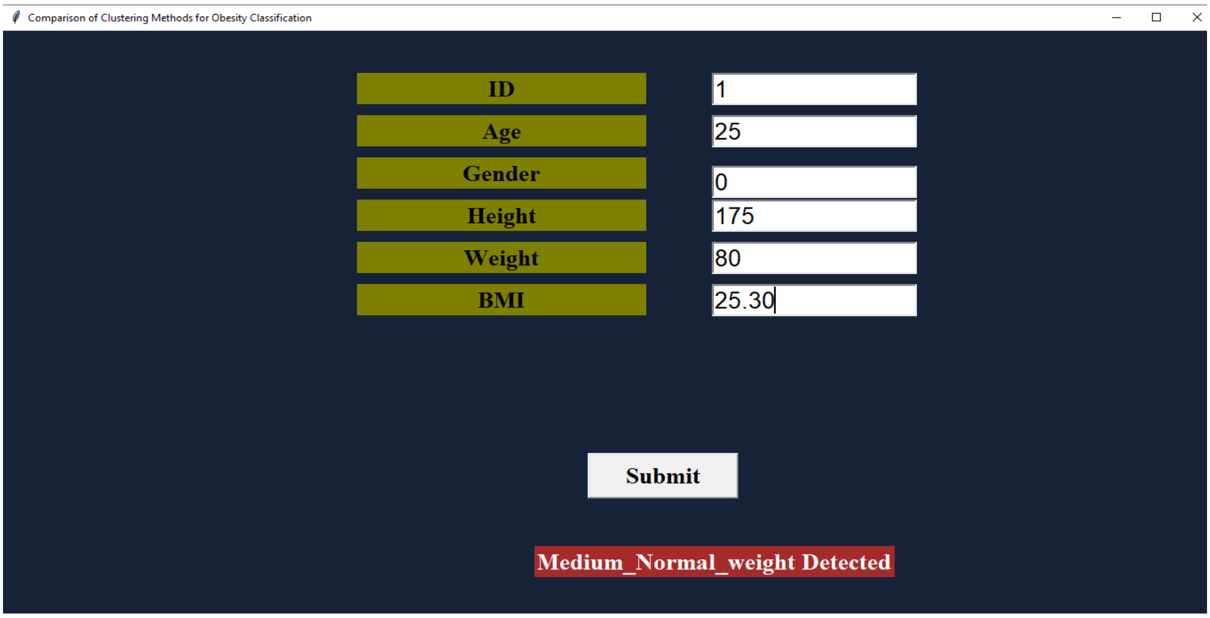
GUI Page:



Master Page:



Check Page:



RESULT:-

VII. DISCUSSION

7.1 Significance of the Results

Our study's results underscore the potential of integrating machine learning algorithms with live camera to enhance the precision and applicability of obesity classification systems. The application of different algorithms SVM, Random Forest, Decision Tree, and Logistic Regression allowed us to evaluate various models' efficiency in real-time.

The use of live camera data, in conjunction with a structured dataset, provided a unique approach to capturing and analyzing real-time visual information against established biomedical indicators (e.g., BMI, age, and gender).

7.2 Limitations

Despite the promising outcomes, this study is not without its limitations. One significant constraint is the reliance on available datasets, which may not comprehensively represent the global population due to demographic, geographical, and socioeconomics dissimilarities. Additionally, the variability in camera quality, lighting conditions, and angles could affect the accuracy of real-time data capture and analysis, potentially introducing bias or errors in feature extraction and model predictions.

7.3 Directions for Future Research

Future research aims at addressing these limitations by incorporating a more diverse dataset that reflects a broader demographic and testing the models in varied environmental conditions to enhance their robustness and generalizability.

Moreover, integrating additional types of data, such as physical activity levels, and genetic markers, could provide a more holistic view of obesity risk factors and improve the predictive power of the models.

In conclusion, our study highlights the potential of leveraging machine learning for obesity classification and providing a foundation for future research aimed at refining these approaches for broader, more effective application in public health and personalized medicine.

VIII. CONCLUSION

Employing machine learning alongside real-time video analytics for predicting obesity heralds a promising direction in enhancing our approach to addressing, preventing, and managing this significant public health challenge. This paper explores the synergistic potential of these advanced technologies to offer instantaneous tracking, prognostic assessments, and a more profound comprehension of obesity's complexities. The fusion of machine learning with real-time video analytics in predicting obesity marks a groundbreaking pathway in both research and practical deployment. As technological advancements progress and methods for gathering data evolve, the prospects for preemptive action in the obesity epidemic are set to increase. Harnessing the capabilities of these sophisticated technologies paves the way for monumental strides towards fostering healthier individuals and a more resilient society

IX. REFERENCES

[1] H. Rohana et al., "Worldwide epidemic of obesity", Obesity and obstetrics, pp. 3-8, 2020.

[2] Naomi Christianne Pereira; Jessica D'souza; Parth Rana; Supriya Solaskar “OBESITY RELATED DISEASE PREDICTION FROM HEALTHCARE COMMUNITIES USING MACHINE LEARNING”,2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

[3] “A Mobile Application for Obesity Early Diagnosis Using CNN-based Thermogram Classification”, 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).

[4] Zhenghao He , “Comparison Of Different Machine Learning Methods Applied To Obesity Classification”, 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE).

[5] K.Nirmala Devi; N. Krishnamoorthy; P. Jayanthi; S. Karthi; T. Karthik; K. Kiranbharath ,“Machine Learning Based Adult Obesity Prediction”, 2022 International Conference on Computer Communication and Informatics (ICCCI).

[6] A.S Maria; R Sunder; R.Satheesh Kumar, “OBESITY RISK PREDICTION USING MACHINE LEARNING APPROACH” , 2023 International Conference on Networking and Communications (ICNWC).

[7] "Obesity and Overweight", World Health Organization World Health Organization, Apr. 2020, [online] Available: https://www.who.int/en/news-room/fact-sheets/detail/obesityand-overweight.

[8] X. Pang, C.B. Forrest, F. Le-Scherban and A.J. Masino, "Prediction of early childhood obesity with machine learning and electronic health record data", International Journal of Medical Informatics, vol. 150, pp. 104454, 2021.

[9] Ferdowsy Faria et al., "A machine learning approach for obesity risk prediction", Current Research in Behavioral Sciences, vol. 2, 2021.