OBESITY CLASSIFICATION USING MACHINE LEARNING

Prof. Rashmi Kulkarni[1], Miss Swapnali Doke[2], Miss Akanksha Gulekar[3], Mr. Anand Khonde[4] , Mr. Ayush Zade[5]

[1] Professor, Information Technology Engineering, Siddhant College Of Engineering, Maharashtra, India
[2] Student, , Information Technology Engineering, Siddhant College Of Engineering, Maharashtra, India
[3] Student, , Information Technology Engineering, Siddhant College Of Engineering, Maharashtra, India
[4] Student, , Information Technology Engineering, Siddhant College Of Engineering, Maharashtra, India
[5] Student, , Information Technology Engineering, Siddhant College Of Engineering, Maharashtra, India

ABSTRACT

The escalating global concern surrounding obesity underscores the urgent need for precise classification methodologies to address this pressing health issue effectively. This innovative project endeavors to pioneer a sophisticated system for obesity detection and classification, leveraging a fusion of cutting-edge image processing techniques and machine learning algorithms. Embracing a holistic approach, the system meticulously incorporates essential parameters such as body mass index, waist circumference, body fat percentage, and demographic variables like age, gender, and ethnicity to ensure comprehensive and accurate assessments. At its core, the system integrates a camera module to capture high-resolution full-body images, harnessing the power of advanced machine learning algorithms, including Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), to discern obesity levels with unparalleled precision. Moreover, the system seamlessly incorporates Logistic Regression (LR) to enable real-time obesity detection directly through a webcam interface, facilitating prompt intervention and support. Throughout the system's design and implementation, stringent privacy protocols and informed consent mechanisms take precedence, ensuring utmost respect for individual privacy and autonomy. This concerted effort amalgamates cutting-edge technology with ethical considerations to forge a transformative solution in the fight against obesity on a global scale.

Keyword : - *Support Vector Machine,Logistic Regression ,Decision Tree, Random Forest*

1. INTRODUCTION

Detecting and managing obesity using machine learning is crucial in healthcare. By analyzing patient data like characteristics and medical history, machine learning models can predict and diagnose obesity, aiding in treatment and prevention. With obesity becoming a global health crisis, accurate classification methods are essential for effective intervention. Obesity classification categorizes individuals based on weight and fat distribution, guiding treatment strategies. Common metrics like body mass index (BMI) and waist circumference are used for classification. Combining machine learning with these metrics offers a data-driven approach to address obesity effectively.

2. OBJECTIVES

2.1 Precise Detection:

Develop machine learning models that accurately spot and diagnose obesity by analyzing various data inputs like patient details and medical records.

2.2 Sorting Into Categories:

Use machine learning algorithms to group individuals into different obesity levels or risk groups, aiding healthcare providers in understanding the seriousness of obesity and planning treatments accordingly.

2.3 Clear Understanding:
Make sure machine learning models used for detecting obesity are easy to understand, allowing healthcare professionals to grasp why certain predictions or recommendations are made.

2.4 Integration with Healthcare Systems:
Incorporate machine learning models into existing healthcare setups to simplify obesity detection processes, enhance patient outcomes, and lower healthcare costs linked to obesity-related issues.

2.5 Strengthening Public Health Measures:
Utilize machine learning to analyze population data, uncovering trends, risk factors, and disparities in obesity rates. This information can guide public health policies and interventions aimed at tackling obesity on a larger scale.

2.6 Tailored Healthcare Approaches:
Employ machine learning to craft personalized healthcare plans for individuals with obesity, considering their unique traits, medical histories, and lifestyles.

2.7 Early Identification:
Implement algorithms to detect obesity early on or predict its likelihood based on risk factors, enabling timely interventions and preventive measures.

2.8 Verification and Assessment:
Thoroughly assess the effectiveness of machine learning models by validating their performance against independent datasets and employing meticulous evaluation metrics, guaranteeing their dependability and practicality when deployed in real-world healthcare environments**.**

3. LITERATURE SURVEY

In [1] In this study, we tackle the significant challenge of evaluating the postoperative health condition of patients with obesity, a pivotal aspect in determining the success of surgical interventions. To streamline this assessment process, we propose a novel approach for extracting key indicators and analyzing weight fluctuations from outpatient medical records, enabling prompt identification of potential complications and readmission risks. By juxtaposing conventional machine learning techniques with a modified version of attentive recurrent neural networks, our findings showcase the efficacy of an ensemble classifier comprising binary attentive bi-LSTM with data rebalancing through conditional generative adversarial networks (CGAN), achieving an impressive F1 score of 86.5% across eight distinct categories of obesity-related parameters. We advocate for the prioritization of data balance, particularly crucial for neural network applications operating on smaller datasets, coupled with meticulous fine-tuning strategies to further enhance model performance and robustness.

In [2] In this work, we introduce FREGEX, a method designed to automatically extract features from biomedical texts using regular expressions. By leveraging Smith-Waterman and Needleman-Wunsch sequence alignment algorithms, tokens are extracted and represented by common patterns found in the texts. We evaluated the effectiveness of FREGEX using three manually annotated datasets focusing on obesity, obesity types, and smoking habits. Comparing features extracted from consecutive token sequences (n-grams), we mathematically represented both types of features using the TF-IDF vector model. Training Support Vector Machine and Naïve Bayes classifiers, we found that features derived from regular expressions not only enhanced classifier performance across all datasets but also required fewer features compared to n-grams, particularly in datasets containing anthropometric measures such as obesity and obesity types.

In [3] This paper addresses the critical issue of obesity and its associated diseases in India. Emphasizing the importance of accurate medical data analysis for early disease detection, the authors highlight the multifaceted nature of obesity, influenced by lifestyle choices, dietary habits, and socio-psychological factors. By developing a state-of-the-art system using machine learning algorithms, the paper aims to predict obesity and its related diseases effectively. Factors such as government regulations on food accessibility and changing lifestyles are identified as significant contributors to the rising obesity rates in India. Through proactive healthcare approaches and advanced predictive modeling techniques, the authors advocate for preventing and managing obesity-related diseases.

In [4] In this paper Authors investigates the use of infrared spectroscopy for saliva profiling in individuals with metabolic disorders. This non-invasive approach aims to identify biomarkers associated with obesity and insulin resistance. Through signal processing techniques and unsupervised classification, the study explores molecular signatures in saliva samples, focusing on regions like Amide I, glucose, and thiocyanate. The research contributes to understanding the molecular mechanisms of these conditions and may improve diagnostic and therapeutic strategies in clinical practice.

In [5] This paper delves into the role of genetics in obesity risk analysis, specifically focusing on Single Nucleotide Polymorphisms (SNPs). While lifestyle factors are known contributors to obesity, recent research highlights the importance of genetic factors. Analyzing SNPs can aid in personalized medication, but the large and noisy nature of SNPs data poses challenges in accuracy and computational complexity. Efficient data reduction methods are crucial to enhance analysis results and reduce computational burdens. The study investigates the application of Forward attribute reduction based on neighborhood rough set model (FARNeM) in obesity-related SNPs analysis. While FARNeM did not achieve optimal reduction rates in the experimental dataset, the analysis underscores the value of incorporating feature selection processes before employing learning algorithms. This research contributes to advancing methodologies for SNPs analysis and underscores the significance of feature selection in improving the accuracy and efficiency of obesity-related genetic studies.

4. EXISTING SYSTEM

The system proposed involves the utilization of Support Vector Machines (SVMs) to classify obesity based on specific features from a dataset. However, there are limitations to this approach that need to be taken into account.

Firstly, SVMs are sensitive to noisy data, which can adversely impact their performance. Noisy or mislabeled data points within the training set can lead to sub-optimal model outcomes, compromising its ability to generalize effectively to unseen data. Addressing this limitation necessitates robust data preprocessing techniques to identify and mitigate the effects of noise in the dataset.

Secondly, the performance of SVMs hinges on the selection of appropriate features. The relevance and quality of features directly influence the model's ability to discern patterns and make accurate predictions. Consequently, a comprehensive feature selection process is imperative to ensure that only the most informative features are included, thereby enhancing the model's performance and interpretability.

Lastly, SVMs may encounter scalability issues when confronted with larger datasets. As the dataset size increases, the computational complexity associated with training and testing SVM models escalates accordingly. This scalability challenge necessitates the implementation of efficient algorithms and optimization strategies to mitigate computational burdens and ensure timely model training and evaluation.

In summary, while SVMs offer promising capabilities for obesity classification, it is essential to acknowledge and address the inherent limitations of the system. By employing rigorous data preprocessing, feature selection, and optimization techniques, these challenges can be mitigated, paving the way for more robust and effective classification models.

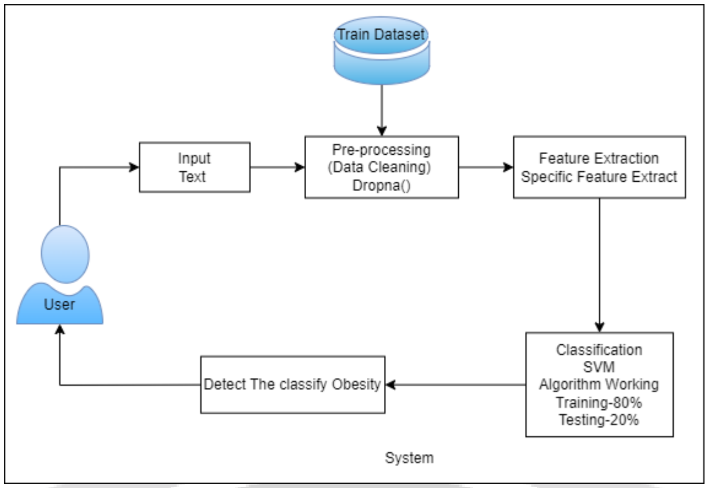

Fig: Existing System Architecture

5. PROPOSED SYSTEM

5.1 Task:

The proposed system aims to revolutionize obesity classification by integrating multiple machine learning algorithms into an efficient and user-friendly platform. Comprising three main modules - registration/login, training and testing dataset creation, and a camera module for real-time obesity detection - the system ensures seamless user identification, personalized interactions, and accurate obesity assessment.

5.2 Corpus:

The registration/login module allows users to input essential information, while the training and testing dataset module generates comprehensive datasets, including physical descriptors such as age, height, weight, and BMI. The camera module captures full-body images in real-time, employing advanced machine learning algorithms like SVM, DT, RF, and LR for pre-processing and feature extraction.

5.3 Preprocessing:

These algorithms utilize logistic regression classifiers to predict obesity status accurately. The system establishes a connection between modules, utilizing the dataset created for training machine learning models. Additionally, user input, including age and gender, enhances the system's personalization capabilities.

5.4 Model:

Dataset creation involves both user-provided information and randomly generated data, ensuring model generalization to diverse body types. With an algorithmic approach incorporating SVM, DT, RF, and LR, the system conducts a thorough analysis of image features for obesity classification. Privacy considerations are paramount, with user consent obtained for image processing, ensuring ethical usage. A Tkinter GUI-based interface enhances user-friendliness, displaying real-time classification results and facilitating seamless interaction. Overall, the proposed system offers a comprehensive, accurate, and privacy-conscious solution for obesity classification using machine learning.

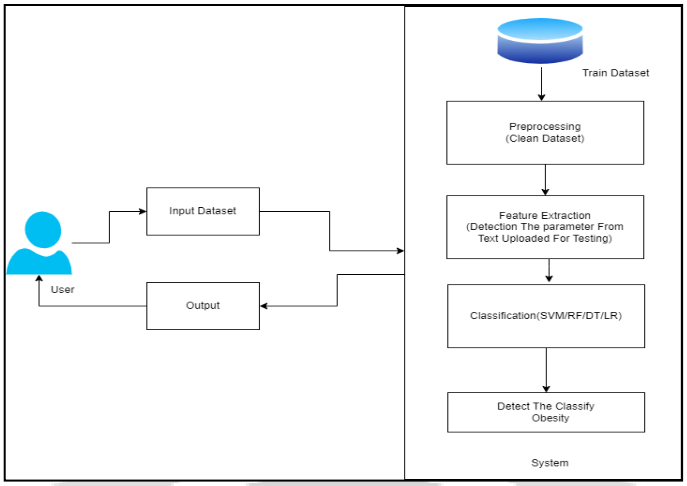

Fig: Proposed System Architecture

5.5 Used Algorithm:

1. Support Vector Machine (SVM):

Support Vector Machine (SVM) is a powerful supervised learning algorithm utilized for classification and regression tasks. It identifies an optimal hyperplane in a high-dimensional feature space, effectively separating different classes. SVM excels when clear boundaries exist between classes and is particularly suitable for high-dimensional data. By maximizing the margin between classes, it enhances generalization and robustness. SVM's versatility and effectiveness make it a popular choice in various domains for tasks requiring precise classification or regression.

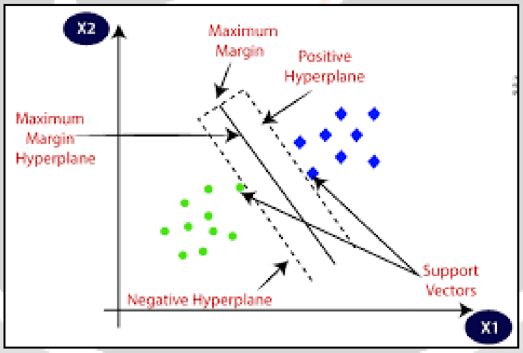

Fig: Support Vector Machine

2. Decision Tree (DT):

Decision Trees are tree-like structures that recursively split the dataset based on features to make decisions. They ask a series of questions to divide the data into subsets, ultimately reaching a leaf node with a final decision. Decision Trees are intuitive, easy to understand, and suitable for both classification and regression problems.

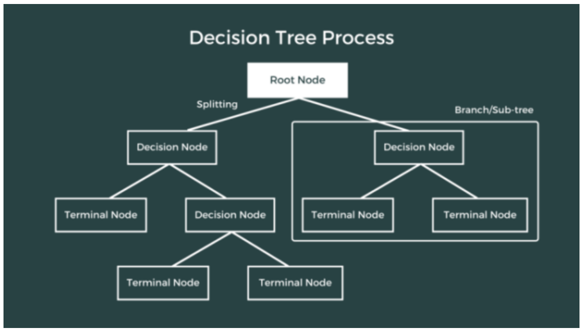

Fig:Decision Tree

3. Random Forest (RF):

Random Forest is an ensemble learning method that constructs multiple decision trees during training and merges their predictions for more accurate and robust results. By combining predictions from diverse trees, Random Forest reduces over-fitting and improves generalization, making it a powerful tool for classification tasks.

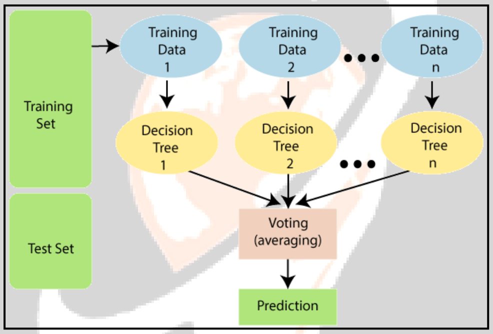

Fig: Random Forest

4. Logistic Regression (LR):

Logistic Regression is a fundamental statistical technique primarily employed in binary classification tasks. It models the relationship between input features and the probability of a binary outcome using the logistic function, which transforms linear regression output into probabilities ranging from 0 to 1. By estimating parameters through methods like maximum likelihood estimation, Logistic Regression offers a straightforward yet effective approach for predicting binary outcomes. Its simplicity, interpretability, and broad applicability across diverse domains make it a widely used tool in data analysis and machine learning.

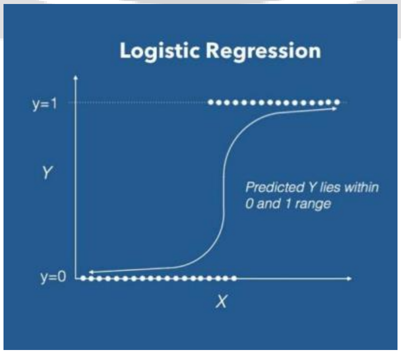

Fig: Logistic Regression

6. TECHNOLOGY USED

6.1 Python

Python, renowned for its user-friendly syntax and extensive library support, serves as a versatile tool across diverse domains, ranging from web development and data science to artificial intelligence and automation. Its innate simplicity and readability empower developers to express complex ideas concisely, fostering rapid prototyping and seamless collaboration. Python's ecosystem boasts a plethora of libraries and frameworks tailored to specific needs, augmenting its utility in various applications. Whether crafting dynamic web applications, unraveling intricate data patterns, training sophisticated machine learning models, or orchestrating automated workflows, Python's adaptability and robustness make it a stalwart ally in the ever-evolving landscape of modern technology.

6.2 Spyder

Spyder stands out as a robust integrated development environment (IDE), uniquely tailored to cater to the demanding requirements of scientific computing and data analysis within the Python ecosystem. Boasting an intuitive user interface coupled with a comprehensive suite of tools, Spyder offers developers and researchers an unparalleled environment for tackling complex computational tasks. Its seamless integration with essential scientific libraries such as NumPy, SciPy, and Matplotlib streamlines workflows, enabling swift prototyping and seamless exploration of data sets. With features including a multi-language editor, interactive console, and variable explorer, Spyder empowers users to delve deep into data analysis, statistical modeling, and visualization, making it a go-to choice for professionals in fields spanning data science, machine learning, and scientific research. Whether unraveling intricate data patterns or conducting cutting-edge research, Spyder provides the essential tools and capabilities needed to excel in the ever-evolving landscape of scientific computing.

6.3 Anaconda

Anaconda serves as an open-source platform specialized in meeting the demands of scientific computation, data analysis, and machine learning endeavors, offering comprehensive support for both Python and R programming languages. It simplifies package management and deployment by offering a comprehensive collection of pre-installed libraries and tools commonly used in these fields. Anaconda includes Conda, a package manager for easy installation and management of packages, Anaconda Navigator for graphical management of environments and projects, and Anaconda Cloud for sharing and distributing packages and environments. Additionally, it provides Spyder IDE, an integrated development environment designed for scientific computing and data analysis. Overall, Anaconda is widely favored by data scientists and researchers for its simplicity, extensive library support, and ease of use in setting up Python and R environments for various data-related tasks.

6.4 SQLite

SQLite, a lightweight and self-contained relational database management system, stands as a versatile solution for various applications requiring data storage and management. Unlike traditional client-server databases, SQLite operates directly on local files, eliminating the need for a separate server process. Its simplicity and ease of use make it an ideal choice for embedded systems, mobile applications, and small to medium-scale web projects where low overhead and minimal configuration are paramount. Despite its lightweight nature, SQLite supports a rich set of SQL features, including transactions, indexes, and triggers, facilitating efficient data manipulation and retrieval. Whether powering mobile apps, managing website content, or serving as a reliable backend for desktop applications, SQLite offers a robust and efficient solution for storing and accessing structured data across a myriad of platforms and use cases.

6.5 Python Libraries

1) Tkinter

2) Numpy

3) Matplotlib

4) Keras

5)Pandas

6) PyTorch

7) Pillow

8) OpenCV

9) TensorFlow

7. CONCLUSION

In conclusion, Our corpulence classification venture utilizes a multifaceted approach to weight classification, coordination progressed machine learning calculations such as SVM, DT, RF, and LR. The framework utilizes a camera module to capture full-body pictures, prepared through pre-processing and highlight extraction. Calculated Relapse is connected for real-time weight location. The cooperative energy of these calculations encourages exact classification into underweight, ordinary weight, and overweight categories. This all encompassing approach, combining cutting-edge innovation and algorithmic differences, guarantees a comprehensive arrangement for corpulence appraisal, with the potential for far reaching affect on well being observing and intervention.

8. REFERENCE

[1]. Yihan Deng, Peter Dolog, Jorn-Markus Gass,Kerstin Denecke, Title: Obesity entity extraction from real outpatient records: When learning-based methods meet small imbalanced medical data sets

[2]. Christopher A. Flores, Rosa L. Figueroa, Title: A Feature Extraction Method for Biomedical Text Classification using Regular Expressions

[3]. Naomi Christianne Pereira1, Jessica D'souza, Parth Rana, Supriya Solaskar, Title: Obesity Related Disease Prediction from Healthcare Communities using Machine Learning.

[4]. S.A. Pullano, M. Greco, M.G. Bianco, D.P. Foti, A. Brunetti, A.S. Fiorillo, Title: FT–IR Saliva Profiling in Patients with Obesity and Obesity–Related Insulin Resistance.

[5]. Kunal Rajput,Girija Chetty,Rachel Davey,Title: Obesity and Co-Morbidity detection in Clinical text using Deep Learning and Machine Learning techniques.