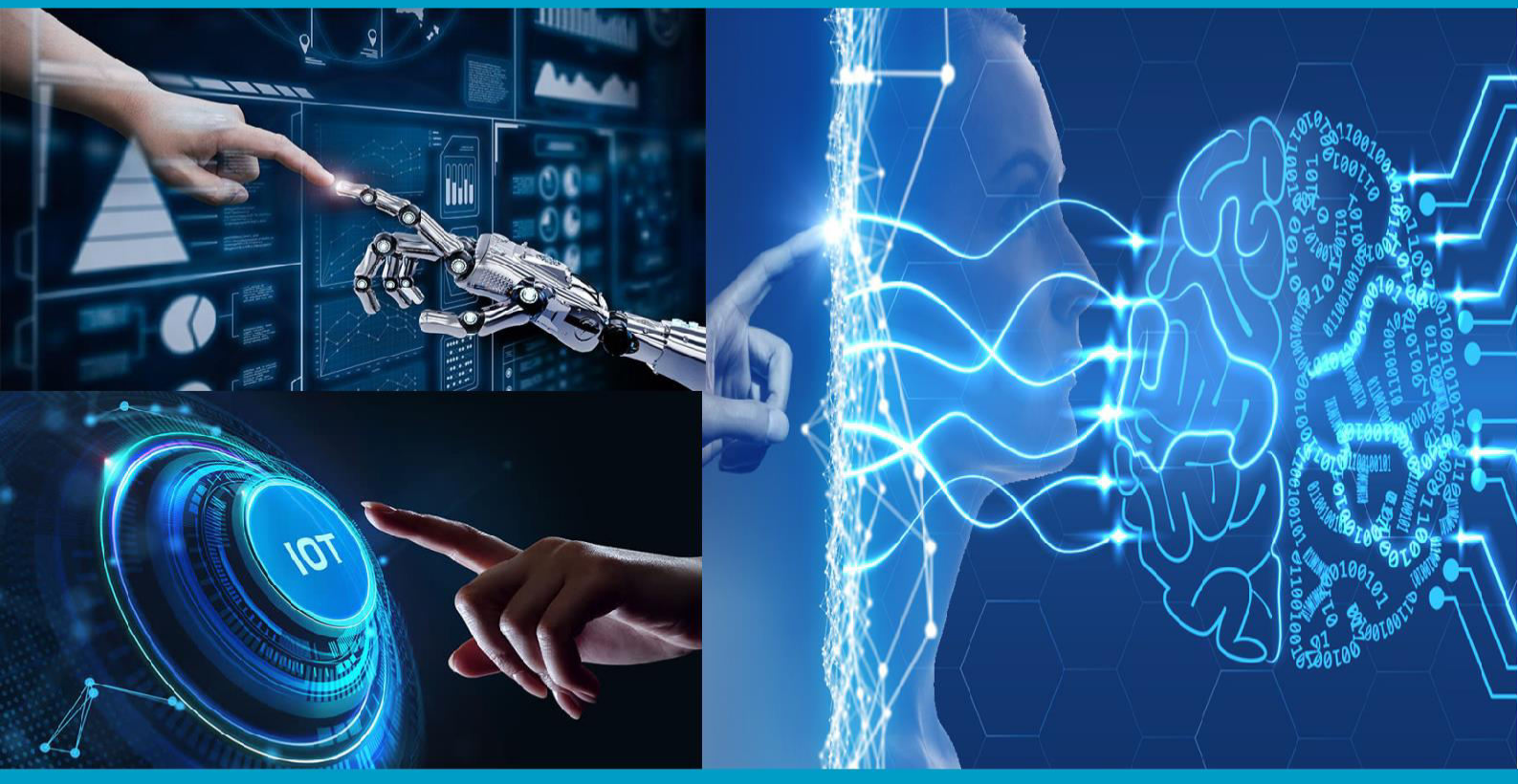
IJMRSETM

e-ISSN: 2395 - 7639



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

Volume 11, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.802

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

A Comprehensive Survey of Obesity Detection Methods

Prajwal Dhadge[1], Yusuf Khan[2], Dalvi Sourabh[3], Prof. Pachhade.R.C[4]

U.G. Student, Department of Computer Engineering, Vishwabharati Academy's College of Engineering, Ahmednagar, Maharashtra, India[1 2 3]

Associate Professor, Department of Computer Engineering, Vishwabharati Academy's College of Engineering, Ahmednagar, Maharashtra , India[4]

ABSTRACT: Obesity is a medical condition characterized by a fast rise in body fat. Obesity is becoming increasingly common. The worldwide obesity rate has proved that this is a serious health issue over the last decade. We collected data from specific urban and rural areas containing various risk factors of our daily food habits such as fast food per week, cold drinks per week, tea with sugar, vegetable per week, and physical records such as TV shows per day, sleep per day, daytime sleep, daily physical exercise, and so on. We were able to get 819 data points based on 18 variables. Then, according to the BMI score, we labelled the class of each record in the data set as normal, overweight, and obese. Our data set contains a total of 249 normal, 203 overweight, and 367 obese data. It includes 235 females and 584 males ages between 10 to 82. Weka is used to pre-process the data, which includes deleting duplicate instances and partial data, fixing missing values, and double-checking the data. And also, to determine the accuracy and error measurement of a model to examine the obesity risk factor class. We employ seven widely used classification methods, including BN, NB, HT, LR, MP, NB Tree, and RS. We use metrics such as accuracy, kappa, precision, recall, and f-measure to assess their performance. The findings show that the Bayes Net method classification model works better than other research with the same antecedents, having the highest weighted value with 98.78% accuracy, 99.7% ROC, 98.82% precision, 98.82% recall, and 98.82% f-measure. The analysis of the results focuses on controlling this cardiovascular disease in Bangladeshi rural and urban people.

KEYWORDS: Obesity, Overweight, Classification Models, Prevention, Machine Learning.

I. INTRODUCTION

Obesity refers to a state of excessive fat accumulation throughout the body. There are several factors, including diet and environmental ones that contribute to obesity. Today, obesity is a major health concern all over the world. People are progressively adopting an unhealthy lifestyle, as seen by their consumption of excessive junk food, late-night sleeping, and prolonged sitting. In particular, adolescents are affected by their unconscious attitudes. Malignancies, diabetes, metabolic syndrome, and cardiovascular disease are only a few of the many chronic diseases that are made worse by the prevalence of overweight and obesity as lifestyle conditions, metabolic syndrome, and cardiovascular disease are only a few of the many chronic diseases that are made worse by the prevalence of overweight and obesity as lifestyle conditions.

Researchers have worked hard to find out factors that influence the generation of obesity, even developing web tools like the calculation of body mass index (BMI) (World Health Organization, «Body mass index calculator,» WHO), where one can determine a person's level of obesity. However, these tools only calculate BMI and don't take into account other important factors, like whether the person has a family history of obesity, how much time they spend exercising, or their gene expression profiles. So, we need a smart tool that can find the risk of obesity quickly and accurately.

This study will analyse obesity datasets using a variety of machine learning algorithms to create better results for early obesity detection that are more accurate. In particular, the ML models are investigated in this work to indicate numerous risk factors for the rising incidence of obesity among persons living in urban and rural areas of Bangladesh. This article contributes

-To collect an obese dataset with random responders.

- To evaluate the performance of different machine learning algorithms on this obese dataset.

-To find out the best classification algorithm based on evaluation matrix accuracy, roc, precision, recall, and f-measure

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

II. LITERATURE SURVEY

As Pervious Study [1], In this paper, we will compare two approaches that are based on traditional machine learning and neural networks. Relevant aspects referring to a health status change or treatment-relevant aspects are extracted from the outpatient medical records as they are generated for each postoperative revisit. The performance of traditional machine learning on the task of obesity-related entity extraction is compared with one variation of attentive recurrent neural networks. The ensemble classifier of binary attentive bi-LSTM with the data balancing using conditional generative adversarial networks (CGAN) has achieved F1 measure of 86.5% on the task of classification of eight classes of obesity-related entities.

Another Study by Christopher A. Flores, Rosa L. Figueroa, and Jorge E. Pezoa [2], In this work, we present FREGEX a method for automatically extracting features from biomedical texts based on regular expressions. Using Smith-Waterman and Needleman-Wunsch sequence alignment algorithms, tokens were extracted from biomedical texts and represented by common patterns. Three manually annotated datasets with information on obesity, obesity types, and smoking habits were used to evaluate the effectiveness of the proposed method.

In [3], This work aims to overcome the above-mentioned limitations by developing a state-of-the-art system that streamlines machine learning algorithms for the effective prediction of Obesity and its related diseases considering the population of India. The government indulgent regulation on food restriction provides easy accessibility to unhealthy, processed food. Also, India's growing career-oriented lifestyle gives rise to irregular biological patterns especially in younger generation who prefer indoor games rather than playing outside.

Another paper [4], Metabolic disorders such as type 2 diabetes mellitus, obesity and metabolic syndrome have a high incidence in the population of developed countries and require continuous clinical and pharmacological treatments throughout their progression. Obesity is associated with insulin resistance in over 90% of patients, but a small percentage of obese subjects are "protected" from this condition. Infrared spectroscopy has been investigated as a non–invasive tool on biofluids, together with signal processing techniques, in the research of novel and predictive biomarkers.

In [5], In this paper, we investigated feature selection process in obesity related SPNs analysis using Forward attribute reduction based on neighbourhood rough set model (FARNeM). The experimental results were compared against Correlation Feature Selection (CFS) method and ReliefF method. Classification accuracy, sensitivity, specificity, positive predictive value and negative predictive value were chosen to assess the performance of the comparison methods on error rate and validated by paired-sample T-test.

III. PROPOSED SYSTEM ARCHITECTURE

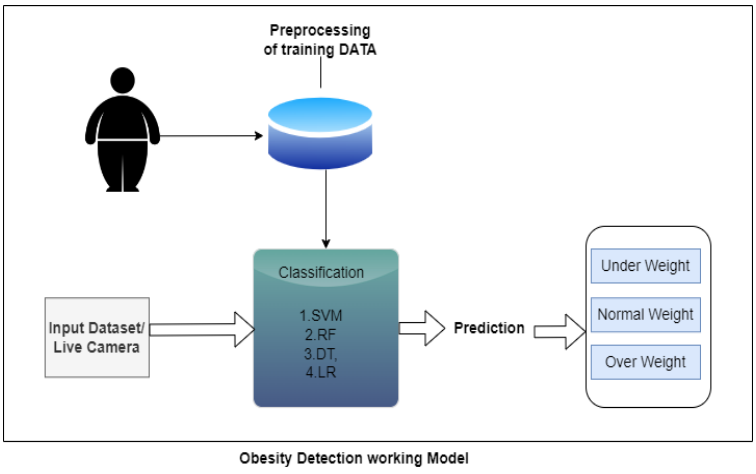


Fig 1. Obesity Detection System Architecture

1.Collect Data through Kaggle conducted specific to Indian Population and/or from other reliable sources.
2. Train and test the data with algorithms like Logistic Regression, Supervised Vector Machine, Decision Tree, Random Forest.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

3. Pick the most suitable algorithm with the best accuracy through SVM, RF, DT, LR and test it with various sampling schemes.
4. Implement the model based on the selected algorithm and create GUI for the system with a sign up and sign in page for user authentication.
5. Now, considering the attributes selected above, we take input for the various tests.
6. Perform predictions using the selected algorithms.
7. Display results and health status. Store the results and provide easy access to monitor the health of a person..

IV. METHODOLOGY

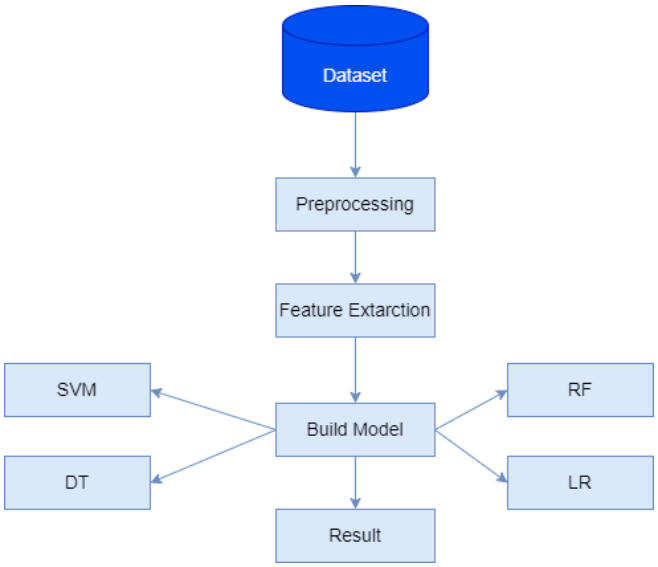


Fig 3. FlowChart of System

The amount of healthcare data generated through smart devices is enormous, in addition to the huge amount of patient data captured in hospital electronic medical records . Increasing availability of digital data due to increased computerized data capture has also accelerated the advancement of ML technology and generated needs for researchers to build more robust prediction models to support data-driven health care.

4.1. Database Collection:

The data set for this study was collected through a survey.

n°	Features	Questions	Answer
1	n_age	whats is you age?	numeric
2	n_gender	What is your gender?	numeric
3	n_height	What is your height?	numeric
4	n_weight	What is your weight?	numeric

Table.1. Physical Descriptions

Physical description Features. Includes features such as gender, age, height, weight.
The target feature was of the categorical type, and was calculated through the data labelling process, in which, for each instance of the dataset, the BMI was calculated using the weight (measured in kilograms) and height (measured in meters) information, according to Equation (1):

$$\text{BMI} = \frac{\text{Weight(kg.)}}{\text{Height(m.)} \times \text{Height(m.)}}$$

In the identification of classes for data labelling, the classification table of BMI.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

Classification	BMI
Under Weight	Below 18.5
Normal Weight	18.5 – 24.9
Over Weight	25.0 – 29.9

Table.2. Classification of Body Mass Index

4.2 Classification using Machine Learning:

4.2.1. **Support Vector Machines (SVM)**:SVM construct optimal separation limits between variables, applying the input data in a larger nonlinear space, called characteristic space .Furthermore, the algorithm uses different kernel functions to model different degrees of nonlinearity and efficiency.

4.2.2**. Decision Tree (DT):** DT is very popular for its simple structure, ease of interpretation, and for its efficiency.The construction of the tree presents an iterative process, starting from a training dataset ($\mathcal{T}$) with n observations, recursively partitioned, thus dividing into increasingly homogeneous data subsets .

4.2.3. **Random Forest (RF):** RF is a flexible algorithm and is an expansion of the Decision Tree. This algorithm creates randomness from a dataset and trains each of its trees with different random data, then the trees are grouped, and by combining their results the errors are calculated to have a more accurate prediction

4.2.4.**Logistic regression:** is used to test a predefined hypothesis and find a relationship between input and output variables when the output variables are categorical in nature (i.e., weight gain or loss) . Linear regression is similar to logistic regression in terms of examining the association between input and output variables. Its output is continuous, not binary variable (i.e., weight changes in kilograms). It also assumes a linear relationship between input and output variables.

4.3. **Performance analysis metrics**

Performance metrics such as precision, recall, F-1 score, and accuracy are derived from the confusion matrix.

A confusion matrix is a matrix that allows analysing the performance of each class by computing the true positive, true negative, false positive, and false negative of the prediction model, and it helps in the evaluation of the quality of the classification model.

$$\textbf{Precision} = \frac{TP}{TP+FP}$$

$$\textbf{Recall} = \frac{TP}{TP+FN}$$

$$\textbf{F1\_score} = 2 \times \frac{precision \times recall}{precision + recall}$$

$$\textbf{Accuracy} = \frac{no.of\ predictions}{total\ no.of\ prediction}$$

The comparisons of the ML models' performance are discussed in this section. The models are evaluated on the obesity dataset with different train test ratios.

V. CONCLUSION

The integration of machine learning and live camera analysis for obesity prediction represents the best approach with the potential to improve the way we understand, prevent, and combat this public health issue. As discussed in this research paper, this innovative combination of technologies offers real-time monitoring, predictive modelling, and a deeper understanding of the nature of obesity. The utilization of machine learning and live camera analysis for obesity prediction is an exciting and promising avenue for research and practical application. As technology advances and data collection methods become more sophisticated, the potential for early detection in the fight against obesity will only grow. By controlling the power of these cutting-edge technologies, we can take significant steps towards a healthier and more sustainable future for individuals and society.

ACKNOWLEDGMENTS

This project is developed with the help of our guide (Prof Pachhade.R.C) vishwabharati college of engineering.

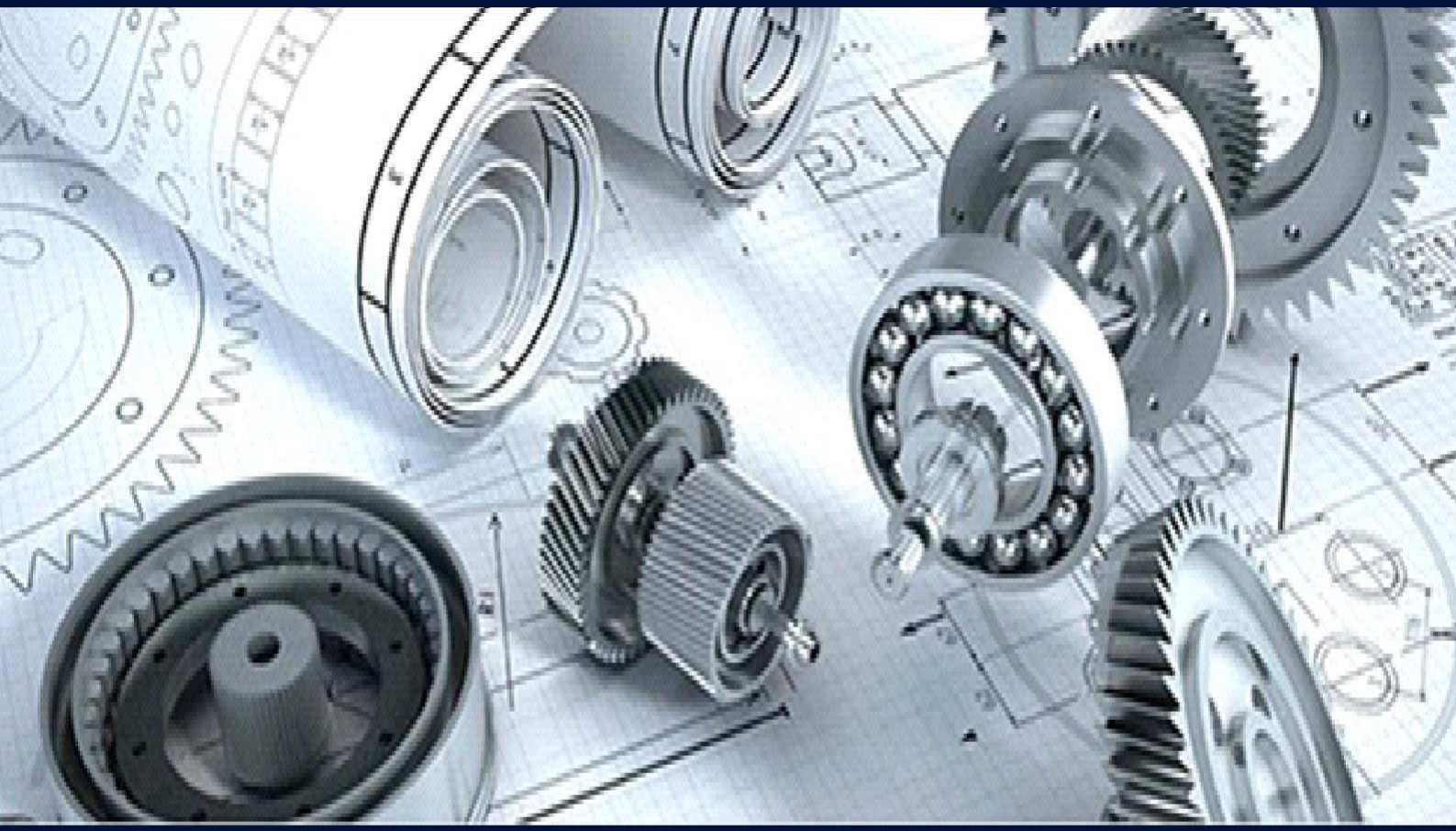
| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

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

,



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

+91 99405 72462 +91 63819 07438 ijmrsetm@gmail.com