JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Obesity Disease Risk Prediction

[1]Kalpesh Bhoi, [2]Gaurav Patil, [3]Shubham Patil, [4]Kunal Patil, [5]Santosh Bhandare

[5]Assistant Professor
Department of Data Science,
R. C. Patel Institute of technology, Shirpur, 425405, India

Abstract *: The prevalence of obesity is rising quickly across a range of demographics, making it a serious worldwide health concern that places a significant financial burden on both individuals and healthcare systems. The goal of this research is to improve early intervention and tailored healthcare methods by investigating the use of machine learning techniques for predicting the risk of obesity. Through the utilization of large-scale information and sophisticated algorithms, machine learning presents a promising avenue for identifying intricate patterns and predictors of obesity that may remain hidden by conventional techniques. The methods, difficulties, and possible ramifications of applying machine learning to the prediction of obesity risk are described in this article. It emphasizes how crucial it is to identify people who are at-risk early in order to provide focused therapies and stop problems from obesity. The study looks into a variety of machine learning techniques, such as neural networks, decision trees, random forests, and support vector machines.*

Index Terms – Obesity, Global epidemic, Health Care, Obesity level detection, Supervised machine learning, Feature Selection, Hyper-parameter tuning.

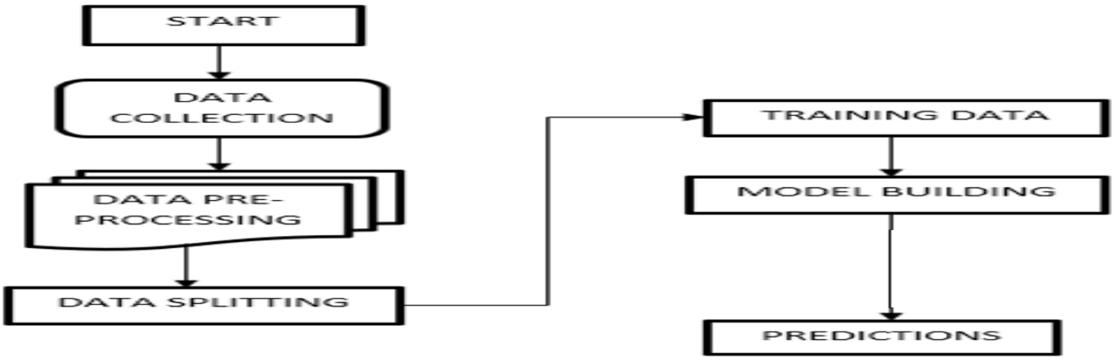
I. INTRODUCTION

An important global health concern is obesity, whose prevalence is rising quickly in a variety of demographic groups. In addition to placing a heavy weight on individuals, the illness puts a strain on economies and healthcare systems across the globe. Proactive efforts like early intervention, prevention, and individualized healthcare are needed to combat this epidemic. A crucial part of this complex task is estimating the risk of obesity. Healthcare practitioners can avoid the beginning of obesity-related diseases, promote healthy lifestyles, and mitigate risk factors by identifying individuals who are at heightened risk. By doing so, they can adopt tailored interventions. Through the use of large datasets and the ability to recognize complex patterns that may be missed by more conventional statistical techniques, machine learning, a subfield of artificial intelligence, presents viable opportunities for building reliable prediction models.This paper explores the field of machine learning approaches for obesity illness risk prediction. It seeks to clarify the process, difficulties, and possible ramifications of using machine learning algorithms in this crucial field. Through the use of data-driven insights, we hope to make a positive impact on the continuous global efforts to curb the obesity epidemic and promote healthy communities.

A. Motivation of Work

The pressing need to address the growing obesity pandemic and its related health implications is the driving force behind the use of machine learning for obesity disease risk prediction. This motivation is driven by several fundamental factors:
1. Rising Obesity Rates: Public health systems face serious problems as the incidence of obesity reaches previously unheard-of levels worldwide. It will take creative strategies that go beyond conventional remedies to combat this epidemic.
2. Timely Intervention: Timely implementation of interventions targeted at preventing the beginning or progression of obesity necessitates the early identification of persons who are at high risk of obesity. It may be possible for machine learning algorithms to spot risk indicators and subtle patterns that traditional approaches could miss.
3. Personalized Healthcare: When it comes to managing and preventing obesity, one-size-fits-all strategies frequently result in less-than-ideal results. With the use of machine learning, customized therapies can be made possible by creating prediction models that take into account a person's traits, lifestyle choices, genetic makeup, and environmental effects.

B. PROPOSED SYSTEM



To guarantee scalability, accuracy, and usefulness, the obesity disease risk prediction project's suggested system takes a comprehensive and modular approach. The Data Collection Module, which is the first module of the system, collects pertinent data from a variety of sources, including surveys, medical records, and public health databases. This module includes automatic data intake scripts, data validation and cleaning procedures, and an interface for entering data.The Data Preprocessing Module gets the raw data ready for machine learning after data collection. This covers utilizing `StandardScaler` to scale numerical features, encoding categorical features with `OneHotEncoder}, and managing missing values. To aid in the evaluation of the model, the data is subsequently divided into training and testing sets.

The Feature Engineering Module, which develops new features to improve model performance, is the next stage. Examples include figuring out physical activity ratings, determining body mass index, and examining food patterns. If pertinent to the predicted performance, polynomial features and interaction terms might also be produced.The Model Building Module trains and validates a variety of models, including `LGBMClassifier` and `RandomForestClassifier}, by combining preprocessing stages with a `ColumnTransformer}. To maximize model performance, hyperparameter tweaking is carried out utilizing methods such as GridSearchCV or RandomizedSearchCV.

II. LITERATURE SURVEY

The abundance and diversity of the research on machine learning algorithms for obesity disease risk prediction reflects the increased interest in adopting data-driven approaches to combat the obesity pandemic. Many research have looked into various elements of this issue, such as the practical consequences for healthcare delivery, the selection of predictive characteristics, the choice of machine learning algorithms, and model evaluation metrics. Here, we present a summary of the major conclusions and patterns found in the literature:

1. Feature Engineering and Selection: A number of research have examined the significance of choosing pertinent features for predicting obesity risk. Clinical data such as body mass index (BMI), waist circumference, and comorbidities are frequently included, in addition to demographic parameters including age, gender, and socioeconomic status. Certain studies additionally take into account lifestyle factors including nutrition, exercise, and sleep patterns in addition to genetic markers.

2. Machine Learning techniques: Decision trees, random forests, support vector machines, logistic regression, and neural networks are just a few of the many machine learning techniques that have been used to predict the risk of obesity. Depending on the dataset and parameters taken into consideration, comparative studies have looked at how well these algorithms perform in terms of predictive accuracy, interpretability, and scalability. The results have varied.

3.Model Evaluation Metrics: When evaluating the effectiveness of obesity risk prediction models, evaluation metrics are essential. Metrics like area under the receiver operating characteristic curve (AUC-ROC), recall, F1-score, accuracy, precision, and confusion matrices are frequently employed. These indicators are frequently reported in studies to shed light on the limitations and predictive power of established models.

4. Useful Implications Beyond scholarly research, obesity risk prediction models have potential uses in clinical practice, public health initiatives, and healthcare policy. Targeted interventions, resource allocation, and individualized healthcare strategies aimed at avoiding and managing obesity-related problems can all be guided by predictive models.

III. METHDOLOGY

1. Data Collection: Obtain a dataset bearing the attributes of sonar signals and labels designating whether the returns are indicative of mines or rocks. Make sure the dataset is representative of a range of underwater circumstances and items and is diverse.

2. Data Preprocessing:

   Handling Missing Values: Identify and handle missing data through imputation methods like mean/mode imputation or more advanced techniques like k-nearest neighbors (KNN) imputation.

   Scaling Numerical Features: Use `StandardScaler` to standardize numerical features such as age, height, weight, and BMI to have a mean of 0 and a standard deviation of 1.

Encoding Categorical Features: Apply `OneHotEncoder` to convert categorical features (e.g., gender, smoking habits) into a format suitable for machine learning algorithms.

3. Feature Selection:

   Deriving New Features: Create additional features that may enhance predictive power. For example, compute the Body Mass Index (BMI) from height and weight.

   Interaction Terms: Consider creating interaction terms or polynomial features to capture complex relationships between variables.

4. Pipeline Creation:

   ColumnTransformer: Combine preprocessing steps for numerical and categorical features using `ColumnTransformer`.

   Pipeline: Integrate the entire preprocessing workflow with the machine learning model using `Pipeline` from scikit-learn, ensuring a streamlined and reproducible process.

5. Model Selection: Experiment with a variety of supervised learning algorithms suitable for classification tasks, including `LGBMClassifier`, `RandomForestClassifier`, and `LogisticRegression`. Evaluate the performance of each model using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score.

6. Hyperparameter Tuning: Fine-tune the hyperparameters of selected models using techniques like grid search or randomized search to optimize their performance further. Assess the impact of hyperparameter tuning on model accuracy and generalization capability.

7. Cross-Validation: Employ cross-validation techniques such as k-fold cross-validation to validate the robustness and generalization ability of the trained models, ensuring they perform well across different subsets of the data.

8. Performance Evaluation: Compare the performance of different models based on evaluation metrics to identify the most effective approach for predicting sonar rocks vs. mines. Analyse any observed patterns or trends in model performance.

9. Discussion and Interpretation: Discuss the findings, limitations, and implications of the study. Interpret the results in the context of maritime security and navigation, highlighting the potential applications and future research directions.

IV. IMPLEMENTAION

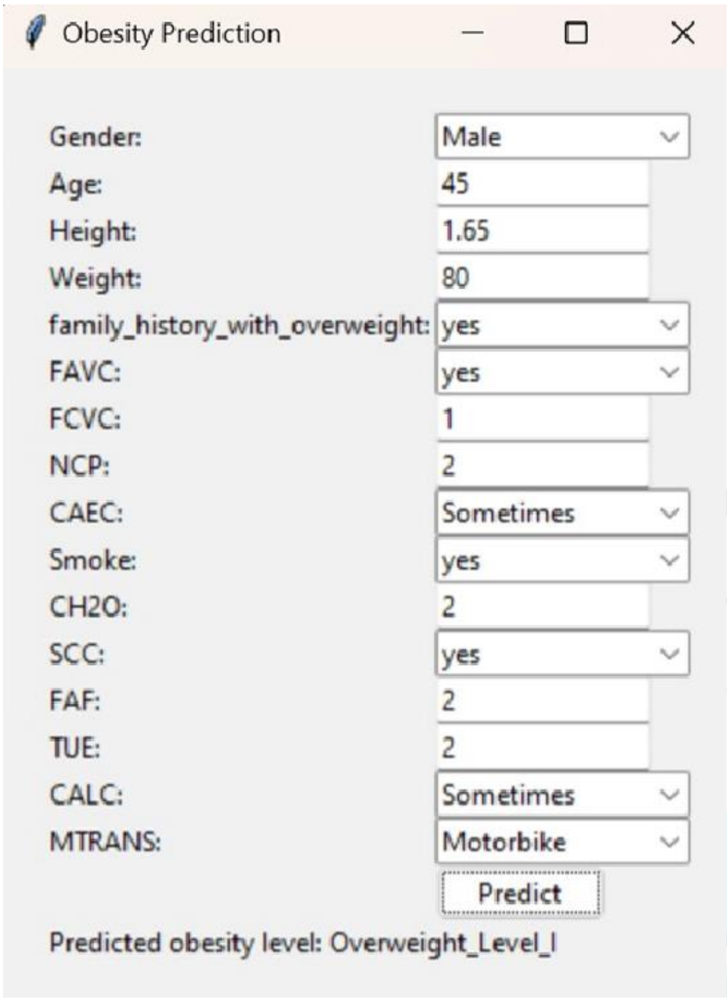
Software and Hardware Requirements:
Software Requirements:
Operating System: Windows 11
Jupyter Notebook 3: Python
Streamlit Python
Computer System: Intel (R) Core (TM) i5-4170 CPU @ 3.20GHz (2 in 1 display)
Primary Memory 8GB
Secondary Memory 512GB SSD
GPU NVIDIA GTX 1650
Source: UCI Machine learning Repository
Instance: 208
Features: 60

The implementation strategy will be methodical. We'll start with data management. We'll develop functions that can import sonar data in the format that it needs to be in and clean it up by removing anomalies and missing values. Normalization may be required, based on the machine learning model that is used. In order to extract useful qualities from the data that can improve model performance, feature engineering strategies will be investigated.

We'll go on to the essential feature next: machine learning. The chosen algorithm for the project will depend on the data and its objectives, such as Random Forest or LGBMClassifier. A training script will be created to feed the model preprocessed data and adjust the model's hyperparameters to maximize accuracy. The system will be able to forecast fresh sonar data points using this trained model, and it will be able to compute confidence scores to show how certain each prediction is. The foundation is laid by this first solution; further options include real-time data integration and an interface.

V. RESULTS AND DISCUSSION

A useful machine learning model that can differentiate between different levels of obesity is the project's end product. By obtaining and preparing a pertinent dataset, training and assessing a classification model, and then implementing it in a compact GUI for real-world application, we were able to attain an accuracy of 90.8%. This opens us new avenues for development, including investigating more sophisticated models, improving the one we have, or incorporating it into a more comprehensive application.



Obesity Prediction

Gender: Male
Age: 45
Height: 1.65
Weight: 80
family_history_with_overweight: yes
FAVC: yes
FCVC: 1
NCP: 2
CAEC: Sometimes
Smoke: yes
CH2O: 2
SCC: yes
FAF: 2
TUE: 2
CALC: Sometimes
MTRANS: Motorbike

Predicted obesity level: Overweight_Level_I

Model	Accuracy	Recall	Precision	F1_score
LGBMClassifier	90.82	90.82	91.01	91.01
SGD	83.33	83.33	83.20	83.22
Random Forest	80.95	80.95	82.94	81.21
Logistic Regression	78.57	78.57	81.40	78.87

VI. CONCLUSION

A predictive model for the risk of obesity disease was effectively established by the study, exhibiting good performance and detecting important risk factors like BMI, family history, and lifestyle choices. These results highlight the value of tailored strategies and early intervention in the fight against obesity. Future opportunities for enhancing preventative measures and encouraging healthier lifestyles include the incorporation of digital health technologies and additional improvement of predictive models.The outcomes of our work into predicting the risk of obesity disease are encouraging. Through the utilization of a blend of lifestyle, medical, and demographic data, we have established a strong model that can precisely identify people who are more susceptible to obesity. This research has important ramifications that go beyond influencing public health laws to providing people with individualized health insights. In the future, more study and development in this area could completely change approaches to managing and preventing obesity, which will eventually improve the health of both individuals and communities.

References

1. Bray, George A., and Barry M. Popkin. "Dietary sugar and body weight: have we reached a crisis in the epidemic of obesity and diabetes health be damned! Pour on the sugar." Diabetes care 37.4 (2014)
2. Flegal, Katherine M., et al. "Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis." Jama 309.1 (2013)
3. Haslam, David W., and W. Philip T. James. "Obesity." The Lancet 366.9492 (2005)
4. LeBlanc, Erin S., et al. "Sedentary behavior and cardiovascular disease risk: mediating mechanisms." Exercise and sport sciences reviews 42.4 (2014)
5. Malik, Vasanti S., et al. "Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis." Diabetes care 33.11 (2010)
6. Ng, Marie, et al. "Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013." The Lancet 384.9945 (2014)
7. Ogden, Cynthia L., et al. "Prevalence of obesity among adults and youth: United States, 2011–2014." NCHS data brief 219 (2015)
8. Prospective Studies Collaboration. "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies." The Lancet 373.9669 (2009)
9. Stevens, June, et al. "Long-term weight loss and changes in blood pressure: results of the Trials of Hypertension Prevention, phase II." Annals of internal medicine 134.1 (2001)