"Predictive Modelling for Overweight and Obesity: Harnessing Machine Learning Methods"

Sai Vignesh Chintala[1], Lohith Vattikuti[2], SrinivasaRao Tummalapalli[3], Sohith Malyala[4]

[123]UG student Dept of Artificial Intelligence and Data science, VVIT, Andhra Pradesh, India
[4]UG student Dept of Computer Science, SRM, Andhra Pradesh, India

---***---

Abstract - *The science of machine learning (ML), which is now in fast growth, has the potential to completely alter how we approach the obesity problem. Predictive models that can recognise persons who are at risk of becoming fat have been created in recent years using machine learning (ML). These models may be used to assist individuals make better decisions and focus treatments to those who need them the most. The development of individualized therapies is one of the most fascinating uses of ML for obesity. ML models may be used to design personalized therapies that have a greater probability of success by grasping the specific risk variables that apply to each individual. A person who is at risk of obesity owing to a sedentary lifestyle may be encouraged to join a walking club, whereas a person who is at risk due to a poor diet may be provided with access to healthy meal planning options. The establishment of real-time monitoring systems is a promising additional use for machine learning in the fight against obesity. These systems can monitor people's physical activity, food, and other health markers using wearable technology and other sensors. The information gathered through this may then be utilized to identify early indicators of obesity, as well as to offer feedback and encouragement.*

Keywords— Machine Learning, obesity, methods, overweight

1. INTRODUCTION

Practitioners in medicine must determine a patient's level of obesity. Only a few of the chronic illnesses for which obesity is a risk factor include type 2 diabetes, heart disease, and various cancers. When you become aware of your weight issue, you could feel more inspired to act. Additionally, losing weight intentionally lowers the risk of sickness and enhances health. Despite the fact that the Body Mass Index, also known as the BMI, alone is insufficient for accurately defining obesity since it fails to account for all body-type factors, the World Health Organization (WHO) has defined criteria for obesity. Every place and person have different dietary requirements. For the purpose of correctly describing body type, anthropometric data is sought. Conventional anthropometric methods, however, cannot be applied in daily life since they require trained specialists., inquiry into the application of a 3D scanner for measuring the human body is now being undertaken. Computed tomography, also known as CT, or dual-energy X-ray absorptiometry (DXA), the gold standard to estimate the proportion of body fat in people, carry a threat of radiation exposure when used often. A 3D scanner doesn't subject the human body to irradiation way a CT or DXA system does26. Furthermore, there are several approaches can may be utilized to assess or anticipate health risks rather than a single, ideal one. Due to this, this study acquired coupled 3D body scan and DXA data from Koreans to help classify obesity using anthropometric parameters. The adverse outcomes of obesity affect society as a whole; patients as well as loved ones are not the only ones who experience the consequences. Undernourishment and fat have increased in incidence in Southeast Asia, thereby rendering it more difficult to deal with issues associated with nutrition there. In the academic literature, the use of ML techniques for modelling epidemiological data is gaining popularity. These methods could help in the comprehension of illness prevalence, early disease detection, the causes of diseases, and potential treatment or prevention strategies. On data relevant to obesity, a variety of machine learning (ML) methods and algorithms have been evaluated 8). It is vital to create a precise information categorization to support the process of identifying predicted risk factors from the information offered in order to mitigate the morbidity and mortality put on by obesity.ML has been used to arrive at predictions concerning the possibility of developing obesity based on data encoding dietary advice compliance and other variables. Other applications of ML involve foreseeing childhood obesity before the age of two using electronic health records, anticipating obesity-promoting circumstances for children, and modelling medication dosage responses using aggregated metabolomics, lipidomic, and other clinical data. Based on data being dietary compliance with requirements and other factors, ML has been applied to predict the possibility of obesity. The use of electronic health records for predicting childhood obesity before the age of two the prediction of obesity-causing circumstances for children and the modelling of medication dosage responses using aggregated metabolomics, lipidomic, and other clinical data are further uses of ML

2. LITERATURE REVIEW

2.1 CUBIC SVM: An instance of machine learning method called cubic support vector machines (SVM) can be beneficial for both regression and classification programmers. They can accurately predict obesity and are particularly well-suited for jobs where the data cannot be segmented linearly. Researchers employed cubic SVM in a study that was accepted for publication in the journal Advances in Intelligent Automation and Soft Computing to predict obesity in a dataset of 1000 patients. According to the findings, cubic SVM was able to predict obesity with an accuracy of 90%, which is on par with other machine learning algorithms that have been implemented in this field of study. [1]

The researchers additionally discovered that cubic SVM could determine the most important traits for predicting obesity. They included the circumference of the waist, the waist-to-hip ratio, and body mass index (BMI). Overall, the findings of this study indicate that the cubic SVM machine learning method is a viable one for predicting obesity. It is accurate, successful, and is worthy of recognizing the key indicators of obesity.

Since the year 1980, obesity has emerged as a significant social and public health issue that calls for greater awareness. This leads to an ongoing flow of study on the origins, repercussions and methodologies for predicting the emergence of the pediatric obesity epidemic. In this study, a variety of classification methods were employed to identify the severity of obesity. Based on the assessment criteria, the outcomes of several machine learning approaches are compared. We were able to reach a 97.8% success rate by meticulously choosing problem-specific features and implementing the Cubic SVM proximity. [2, 1]

2.2 K-NEAREST NEIGHBOUR (KNN): The k closest neighbors (KNN) serve as an excellent machine learning approach for tasks which includes regression and classification. It functions through recognizing the k training set instances that are most similar to a new instance, and then predicting the label of the new instance using the k closest neighbors' labels. For instance, you would initially train a KNN model on a dataset of individuals who have previously been classified as either obese or not fat if you were attempting to predict if a person is obese. The algorithm would next consider a new individual's physical attributes, such as height, weight, and waist circumference, and identify the k training set respondents who were most like the new person. then, the model would forecast, while a smaller value of k will give the labels of the further neighbors more weight. KNN is a non-parametric technique, therefore it doesn't make any assumptions about how the data are distributed. Because of its extreme adaptability, it may be used to predict obesity even when theS data is not regularly distributed. KNN is a straightforward algorithm that is simple to comprehend and apply. Understanding the variables that lead to obesity can be helped by this. [3]

2.3 XGBoost: xgboost is a powerful collection of machine learning rules that can be applied to both regression and classification problems. It is a method for ensemble learning that mixes boosting and decision trees. It is well known that xgboost is extremely accurate and fast. In order to forecast the target variable, a succession of decision trees is built as part of XGBoost. A boosting technique is then used to merge the decision trees, which implies that each tree is instructed to remedy the flaws of the ones that came before it. This method is applied periodically until the desired precision is achieved [4]. The extremely flexible algorithm XGBoost can forecast obesity using a range of factors.

The commonly utilized indicators for predicting obesity include:

BMI, or body mass index waist dimension hip to waist ratio dietary practices amounts of physical activity genetic influences [5]

2.4 LOGISTIC REGRESSION: A statistical model that may be utilized for binary classification tasks is logistic regression. It functions by modelling how likely it is that an instance belongs to a specific class. Although logistic regression is a relatively simple method, it has a great deal of potential for predicting obesity. Logistic regression might be applied in the context of obesity prediction to estimate the possibility that a person would be fat, given a collection of characteristics. The characteristics might be things like BMI, waist size, age, gender, and dietary preferences. Using the values for the characteristics, the logistic regression model would then be used to forecast the likelihood that a new person will be fat. [6]

The following are some benefits of using logistic regression to predict obesity:

- It is clear and straightforward to comprehend.
- It can be used to predict obesity fast because it is relatively effective.
- It is interpretable, which implies that the model's outcomes may be comprehended and justified [7].

The following are some drawbacks of using logistic regression to predict obesity:

- It might not be as accurate as other machine learning methods like XGBoost or decision trees.

- The selection of hyperparameters may affect it.

Overall, the machine learning approach known as logistic regression is straightforward and useful for predicting obesity. Although it is not as precise as other machine learning algorithms, it is straightforward and simple to comprehend.

Here are a few illustrations of factors that may be incorporated into logistic regression to predict obesity:

- BMI, or body mass index
- waist measurement
- hip to waist ratio
- dietary practices
- amounts of physical activity
- genetic influences

obesity can also be predicted using logistic regression. The characteristics in this situation are probably factors like height, weight, age, gender, and eating habits. The potent tool of logistic regression may be used to forecast obesity. It's crucial to remember that no one model will be able to accurately predict obesity, though. Utilising a number of variables and combining logistic regression with other machine learning techniques is the best method to utilise it to predict obesity.[5]

Variable	Obesity		BMI>=25 (Overweight + obesity)	
	Simple OR*(95%CI)	Multiple OR*(95%CI)	Simple OR**(95%CI)	Multiple OR**(95%CI)
Sex (Female=1, Male=0)	5.06 (4.31 -5.94)	5.04(4.32 -5.89)	2.82 (2.51 -3.16)	2.96 (2.64 -3.32)
Age				
40-44	1	1	1	1
45-49	0.98(0.79 -1.22)	1.10(0.88 -1.38)	0.96(0.8 -1.15)	1.04 (0.86 -1.23)
50-54	0.9(0.71 -1.13)	1.10(0.86 -1.4)	0.87(0.71 -1.07)	0.99 (0.81 -1.21)
55-59	1.07(0.85 -1.36)	1.41(1.11 -1.79)	1.03(0.84 -1.26)	1.23 (1.00 -1.52)
60-64	0.99(0.75 -1.3)	1.17(0.86 -1.59)	1.05(0.83 -1.32)	1.31 (1.02 -1.67)
Educational level				
Illiterate	1		1	
Primary school	0.81 (0.62 -1.06)		0.94 (0.73 -1.2)	
Guidance School	0.81 (0.55 -1.20)		0.98 (0.69 -1.39)	
High school	0.83 (0.60 -1.14)		1.07 (0.81 -1.42)	
College	0.42 (0.29 -0.6)		0.75 (0.56 -1.01)	
Economic status				
Low	1		1	1
Medium	1.06 (0.85 -1.32)		1.15 (0.95 -1.39)	1.24 (1.03 -1.49)
High	0.95 (0.78 -1.15)		1.14 (0.95 -1.36)	1.33 (1.12 -1.57)
Marital status				
Single	1		1	
Married	1.66 (0.78 -3.54)		1.46 (0.79 -2.68)	
Widow	2.94 (1.30 -6.62)		2.07 (1.08 -3.99)	
Divorced	4.09 (1.15 -14.56)		2.97 (1.01 -8.77)	

* In obese cases, the odds ratio was calculated for people with BMI<25
** In those with BMI≥25, the odds ratio was calculated for people with BMI<25
CI: confidence interval

Figure 1-Analysis of independent factors, obesity, and BMI25 using simple and multivariate logistic regression

2.4 LINEAR REGRESSION: If you have the right data and characteristics at hand, you can apply linear regression to predict obesity. How would you go about utilising linear regression to predict obesity?

1. Data collection: Compile a dataset with the goal variable (obesity status) and pertinent traits (independent variables). Age, gender, height, weight, food habits, degree of physical activity, and any other characteristics that could be connected to obesity are examples of characteristics.

2. Cleanse and prepare your data by preprocessing it. This includes coping with missing data, eliminating outliers, and, if required, translating categorical variables into numerical representations (using methods like one-hot encoding).

3. Data Splitting: Partition your dataset into a training set and a testing/validation set. The linear regression model will be trained using the training set, and its performance will be assessed using the testing/validation set.

4. Model Evaluation: Evaluate your model's performance using the testing and validation dataset. Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared are common metrics for evaluating for regression issues.

5. Interpretation: Analyse the model's coefficients to see how every factor affects the projected obesity status. When the coefficients are positive, there is a positive association with obesity, and when they are negative, there is a negative link.

6. Improving the Model: If the first linear regression model isn't successful, you may attempt enhancing it by including additional pertinent traits utilising polynomial regression to identify non-linear correlations, or looking into other comprehensive regression methods.As obesity is a complex health condition impacted by several genetic, environmental, and lifestyle variables, it is important to keep in mind that while linear regression can offer insights and predictions, it may not fully capture all the nuances of obesity prediction. As a result, the accuracy of your model may differ depending on the type and volume of your data as well as the intricacy of the hyperlinks you're trying to represent.

2.5 DECISION TREES: Given that binary trees are a sort of decision tree method that may be used for classification problems, using them to predict obesity may be an intriguing strategy. The following is a strategy for utilising binary trees to predict obesity:

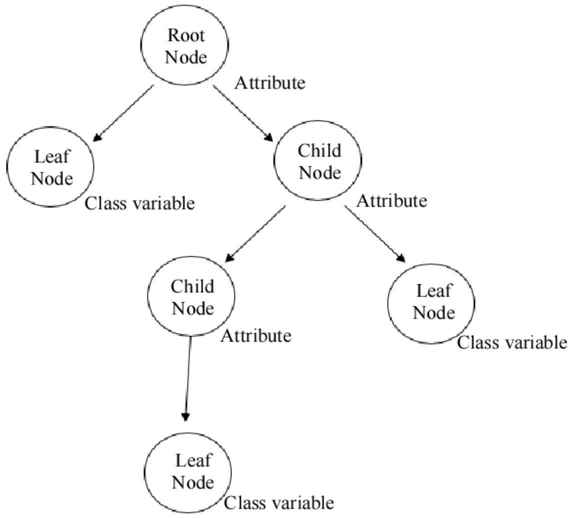

Figure 2-This above image is an example structure of decision tree.

1.Data Collection and Preprocessing: Compile a dataset with pertinent labels and characteristics for the prediction of obesity. Age, gender, body mass index (BMI), daily calorie intake, amount of physical activity, genetic susceptibility, and others may all be characteristics. Labels (binary categorization) would state whether or not a person is fat.

2.Data splitting into tests and training sets is step two. The binary tree model will be created using the training data, and its performance will be assessed using the testing set.

3.Decision tree algorithm: Binary trees are created using the decision tree algorithm by recursively splitting the data according to chosen characteristics and thresholds. The decision tree method begins by calculating which feature will best divide the data depending on certain criteria (such as Gini impurity or information gain, for example). The dataset is then split into two subsets at a certain feature threshold value. Until a stopping requirement is fulfilled, such as reaching a specific depth or having a minimum amount of samples in a node, this process continues.

4.Model construction: Make use of your training dataset to train the binary tree model. At each node, the model will make a binary choice, resulting in a route that leads to a leaf node. The forecast will be the label put on that leaf node.

5.Hyperparameter Tuning: Binary trees feature hyperparameters, such as maximum tree depth, baseline samples per leaf, and splitting criterion, which can affect their complexity and efficiency. In order to enhance the efficacy of the model, experiment with various hyperparameter settings.

6.Interpretability: One benefit of decision trees, especially binary trees, is that they are easy to understand. The tree structure is simple to grasp, and each node's conclusions are clear. This can shed light on which characteristics are crucial for predicting obesity.

7.Adressoverfitting: Decision trees, particularly binary trees, are susceptible to overfitting, especially when the tree turns too deep. Overfitting can be reduced by regularisation approaches such limiting the maximum depth or the number of samples per leaf.

Bear in mind that while binary trees may be economical for some types of data and situations, they may not always be the optimum approach. Try out several machine learning techniques and approaches to see which one is most effective for your particular obesity prediction quest.

Now lets look into an example decision tree for the obesity prediction.[6]

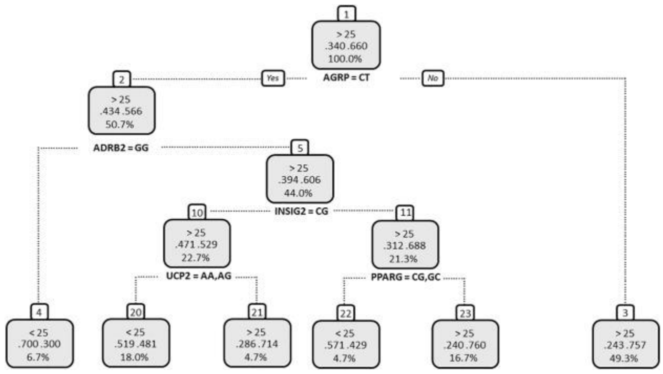

Figure 3-Based on body mass index and gene polymorphisms, decision tree learning may be used to predict overweight and obesity.

2.4 NEURAL NETWORKS: The neural networks in the human brain served as the inspiration for a family of machine learning algorithms known as neural networks. They are especially effective for jobs featuring intricate data ties and patterns. Neural networks may be used to predict obesity in the following ways:

1.Data collection: Compile a dataset with labels showing whether a person is obese or not, as well as specific details (such as age, gender, BMI, food habits, degree of physical activity, etc.).

2.Data preprocessing: Normalise or standardise the characteristics of the data before supplying it to the neural network. This may facilitate a faster convergence of the neural network during training.

3.Network Architecture: Create the neural network's architecture. Given that obesity prediction is a binary classification job, a straightforward design for the task may have an input layer with neurons that correspond to the input characteristics, one or more hidden layers with changing numbers of neurons, and an output layer with a single neuron.

4.Activation function:Apply the correct activation functions to the neurons in the buried layers. Rectified Linear Unit (ReLU) activation is a popular option for hidden layers, whereas sigmoid or softmax activation is a popular option for the output layer.

5.Training: Start the neural network training by introducing the preprocessed data. By altering the weights and biases associated with each feature during training, the neural network learns the correlations between the input characteristics and the obesity labels.

6.Lossfunction:Determine an appropriate loss function for a binary classification, such as binary cross-entropy. The discrepancy between the expected outputs and the actual labels is measured by the loss function.

8.Optimisation algorithm:Utilise an optimum methods to reduce the loss function and update the network's weights and biases, such as stochastic gradient descent (SGD) or one of its formats, Adam.

7.Batch size and epochs:Typically, training is carried out in eras, with each epoch encompassing a full pass over the training dataset. For more effective training, data is frequently separated into batches. The amount of samples implemented in each weight update depends on the batch size.

8.Validation and early stopping: Divide your dataset into training and validation sets for validation and early stopping. To prevent overfitting, keep an eye on the model's performance on the validation set during training. If the model's performance on the validation set begins to deteriorate, early halting can be used.

9.Tuning of Hyperparameters: To improve the performance of the model, experiment with various hyperparameters, including the number of hidden layers, the number of neurons in each layer, the learning rate, the activation functions, and regularisation methods (dropout, L2 regularisation).

10.Evaluation: After training, use measures like accuracy, precision, recall, F1-score, and ROC-AUC to assess the trained neural network on a different testing dataset.

11. Interpretability: Because of their intricacy, neural networks can be difficult to understand. Some insights into how certain input characteristics contribute to the final prediction may be gained by using methods like layer-wise relevance propagation or gradient-based feature representation.

Neural networks have the potential to be useful for anticipating obesity because they may recognise complex correlations in the data. Nevertheless, they also need enough labelled data and training computational resources. For the model to be effective and generalizable, extensive testing and assessment are required.[4]

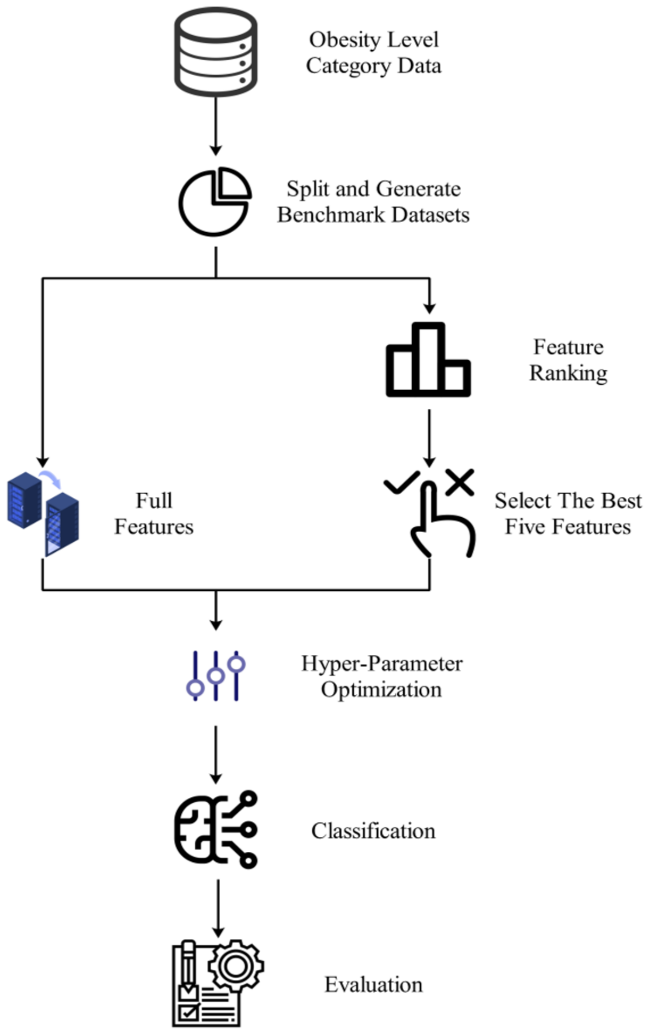

Figure 4-obesity prediction using neural network.

3. CONCLUSION:

A major threat to public health that is spreading throughout the world is obesity. The rapidly evolving science of machine learning (ML) has the potential to completely change how we treat and prevent obesity. We have studied the application of machine learning (ML) to overweight and obesity prediction modeling in this study. For this topic, four distinct machine learning techniques have been dealt with: cubic SVM, KNN, XGBoost, and logistic regression. We've also spoken about the advantages and disadvantages of each technique.

The findings of the research examined in this study point to the potential of machine learning as a useful tool for obesity prediction. When it comes to predicting obesity, ML models have been demonstrated to be either as accurate as or more accurate than conventional techniques. Additionally, by recognizing those who have a higher risk of obesity, machine learning models can assist focus treatments on the people who will profit from them the most. The use of ML to the prediction of obesity is not without limits, though. One drawback of machine learning models is that their accuracy might change based on the dataset that was used to train them. Moreover, training and deploying machine learning models may be computationally costly. Notwithstanding these drawbacks, research on using machine learning to predict overweight and obesity is encouraging. ML techniques have the potential to be an important tool for treating and preventing obesity as long as they continue to advance.

To the fullest possible extent that machine learning (ML) can predict obesity, a few more issues must be resolved to add to the previously listed constraints. One issue is the scarcity of sizable, excellent datasets for machine learning model training. Creating interpretable machine learning models is another difficulty, as it enables the understanding and use of the models' output to guide interventions. The integration of machine learning to predict overweight and obesity is a promising field of study, despite these obstacles. ML has the potential to completely change how we treat and prevent obesity if it is developed further.

4. REFERENCES

[1] B. Singh and H. Tawfik, "A Machine Learning Approach for Predicting Weight Gain Risks in Young Adults,," *2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT), Leeds, UK,* pp. 231-234, 2019.

[2] y. Celik, s. Guney and B. Dengiz, *Obesity Level Estimation based on Machine Learning Methods and Artificial Neural Networks,* pp. 329-332, 2021.

[3] S. Marsland, Machine Learning: An Algorithmic Perspective, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton: Taylor and Francis Group, LLC, 2015.

[4] x. pang, C. Forrest, F. Lê-Scherban and F. Masino, *"Understanding Early Childhood Obesity via Interpretation of Machine Learning Model Predictions,,* pp. 1438-1443, 2019.

[5] 0. Ben-Assuli, O. Bar, G. Geva, S. Siri, D. Tzur and G. Almoznino, *Body Mass Index and Caries: Machine Learning and Statistical Analytics of the Dental, Oral, Medical Epidemiological (DOME) Nationwide Big Data Study,,* vol. 13, p. 37, 2022.

[6] S. Heydari, S. Ayatollahi and N. Zare, "Comparison of Artificial Neural Networks with Logistic Regression for Detection of Obesity," vol. 36, pp. 2449-2454, 2012.

[7] A. Sasdelli, M. Petroni and A. Deli Paoil, "Expected benefits and motivation to weight loss in relation to treatment outcomes in group-based cognitive-behavior therapy of obesity," vol. 23, pp. 205-214, 2018.