Intelligent Methods in Engineering Sciences 4(3): 100-113, 2025

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International Open Access

December, 2025

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2025.157

Classification of Obesity Levels Using Machine Learning Algorithms

Ahmet Can CITAK [a], ***Sumeyra SAHIN BAYRAM*** [b], ***Murat KOKLU*** [a, *]

[a] *Technology Faculty, Department of Computer Engineering, Selcuk University, Konya, Türkiye*
[b] *Faculty of Health Sciences, Department of Nutrition and Dietetics, Selcuk University, Konya, Türkiye*

ARTICLE INFO

Article history:
Received 16 November 2025
Accepted 29 December 2025

Keywords:
Artificial Neural Networks,
Classification,
Ensemble Learning,
Machine Learning,
Obesity,
Random Forest

ABSTRACT

Obesity has become a critical public health issue on a global scale due to the serious comorbidities and economic burden it brings. The aim of this study is to develop an effective machine learning model that can accurately determine obesity levels based on data including individuals' demographic characteristics and dietary habits, and to compare the performance of tree-based ensemble learning algorithms and Artificial Neural Network (ANN) approaches. In this context, classification was performed using Random Forest, XGBoost, CatBoost, and ANN (Artificial Neural Network) algorithms based on the open-source "Obesity Dataset" obtained from 1,610 participants and containing 14 different attributes. The models' performance was tested using a 5-fold cross-validation method and evaluated based on accuracy, f-score, precision, and recall using a confusion matrix. Experimental results show that tree-based ensemble models outperform the ANN approach in this dataset. The Random Forest algorithm was the most successful model with an accuracy rate of 94.34% and an F-score of 94.36, followed by XGBoost with an accuracy rate of 92.80%. In contrast, YSA remained at an accuracy rate of 82.98% and spent approximately 93 times more time in terms of training duration compared to Random Forest. When considering the obtained outputs, this study demonstrates that ensemble learning methods such as Random Forest are more efficient than ANN models in terms of both prediction accuracy and computational cost in the analysis of tabular health data, and that the developed model can be used as a reliable tool in clinical decision support systems.

1. INTRODUCTION

Obesity, which arises from the interaction of genetic and environmental factors, is a serious and chronic disease characterized by the accumulation of fat in the body to a level that threatens health. This condition, which has fundamental risk factors such as social, psychological, and dietary habits, is a critical health issue worldwide, regardless of age. Although it is known that more than 2 billion people worldwide are overweight or obese today, research shows that this situation is preventable. Despite being preventable, the prevalence of obesity continues to rise at an alarming rate. According to World Health Organization (WHO) data and large-scale epidemiological studies, obesity rates worldwide have nearly tripled since 1975, and this increase is not limited to developed countries but has also become a significant public health crisis in low- and middle-income countries [1]. When examining the underlying reasons for this increase, it is evident that the dominant role is played by "obesogenic environment" factors, which are characterized by sedentary behaviors brought about by modern life and the increased accessibility of energy-dense, nutrient-poor foods [2]. Indeed, recent studies conducted on young adult groups such as university students confirm that individuals are moving away from healthy eating patterns (e.g., the Mediterranean diet) and that the quality of their diets is alarmingly low [3]. The clinical significance of obesity is not limited to an increase in adipose tissue but also stems from the serious comorbidities it brings with it. Comprehensive meta-analyses in the literature demonstrate that obesity has a strong causal relationship with Type 2 diabetes, cardiovascular diseases, hypertension, and various types of cancer [4]. Furthermore, it has been reported that there is a linear relationship between increasing body mass index and mortality rates from all causes, and that obesity significantly shortens life expectancy [5]. Beyond individual health, obesity also places an unsustainable economic burden on healthcare systems. When indirect costs such as lost productivity and early retirement are

* **Corresponding Author:** mkoklu@selcuk.edu.tr

factored in alongside direct medical costs, the impact of obesity on the global economy is highlighted as being comparable to that of armed conflicts or tobacco use [6]. Therefore, combating obesity is not only a medical necessity but also a socio-economic imperative in the literature.

2. LITERATURE REVIEW

Representing the latest developments in the literature, Shaban and colleagues, in their study published in the journal Scientific Reports, proposed a comprehensive machine learning framework that not only classifies the current state of obesity but also predicts individuals' susceptibility to obesity. Researchers have gone beyond standard modeling thanks to this comprehensive system that integrates data preprocessing, advanced feature selection, and classification algorithms. The findings of the study indicate that the developed framework has high prediction accuracy and enables early intervention in preventive healthcare services by identifying individuals at risk before the onset of disease [7].

In obesity prediction models, the importance of data processing strategies, not just algorithm performance, is increasingly growing. In this context, Al Khushi Joshi has added a new dimension to the literature by examining gender-disaggregated data sets, where men and women are modeled separately, rather than the classic "generalized data" approach where all individuals are evaluated in a single pool. Researchers, working from the premise that metabolism and lifestyle habits differ between genders, reported that when data was separated, algorithms were able to identify gender-specific risk factors more accurately. For example, snacking between meals and family history were identified as risk factors for women, while physical activity and alcohol consumption were identified as risk factors for men. This analysis, led by methods such as Random Forest and Decision Tree, emphasizes the necessity of precision medicine approaches that account for biological sex differences, rather than a one-size-fits-all model for obesity prediction [8].

The non-linear and multi-layered structure of health data is a decisive factor in selecting the algorithm to be used. In this context, Ölçer compared fundamental machine learning algorithms such as Random Forest, SVM, k-NN, and Naive Bayes from a contemporary perspective in order to model the complexity of obesity data with minimal error. As a result of analyses conducted on the UCI dataset, the Random Forest algorithm maintained its top position by achieving the highest success rate thanks to its ability to analyze complex relationships within the dataset. The most striking finding of the study is the poor performance of the Naive Bayes algorithm. This situation proves that obesity parameters are not independent events but rather have a cyclical structure that triggers each other. Ölçer's findings confirm that probability-based simple models that disregard this multifaceted interaction between variables are inadequate for predicting multifactorial diseases such as obesity [9].

Suwarno and colleagues compared different machine learning algorithms (Decision Tree, Random Forest, SVM, Logistic Regression) that go beyond biological markers in obesity prediction and focus on dietary and lifestyle habits. The striking results of the study prove that the phenomenon of obesity is better modeled by decision mechanisms rather than linear equations. Indeed, while classic Logistic Regression remained at an accuracy rate of 60-75%, the Decision Tree algorithm was determined to be the most effective method with a very high success rate of 97-98%. The study also revealed that behavioral characteristics such as family history, frequency of snacking between meals, and even mode of transportation are the strongest mathematical predictors in determining individuals' obesity levels, rather than just calorie intake [10].

Another study conducted by Musa and Basaky revealed the dramatic effect of algorithm selection on prediction accuracy in obesity classification. When researchers analyzed the obesity dataset using different methods such as Logistic Regression, Naive Bayes, and Decision Trees, they found significant differences in performance. According to the numerical data of the study, the probability-based Naive Bayes algorithm remained at a very low and unacceptable accuracy rate of 54.7%, while the rule-based Decision Tree algorithm achieved a success rate of 97.4%, providing a significant advantage over its competitors. This enormous performance difference of approximately 43% mathematically proves that the phenomenon of obesity has a complex structure that can be modeled using hierarchical decision mechanisms rather than simple probability calculations [11].

When examining the technological evolution of studies in the literature, it is seen that methods that go beyond classical machine learning algorithms and mimic the working principle of the human brain are coming to the fore. In this context, Kıvrak achieved a record accuracy rate of 98.51% in his study using a multi-layer ANN architecture to detect obesity levels, surpassing the performance of classical methods. This ANN model, capable of analyzing complex and non-linear relationships between data, has confirmed the decisive role of variables such as daily meal frequency, water consumption, and technology usage time on obesity. The study demonstrates that appropriately trained YSA models have the potential to make diagnoses with the accuracy of a specialist physician using only lifestyle data, without the need for clinical tests [12].

In addition to the experimental studies in the current literature, the comprehensive systematic literature review conducted by Safaei and colleagues, covering 87 qualified articles from 2010 to 2020, is critical in revealing the general trend in the field. This study, which addresses obesity not only in terms of nutrition but also as a complex interaction of lifestyle, sociodemographic characteristics, genetic, and psychological factors, has evaluated the performance of artificial intelligence methods. The analysis results confirm that basic statistical methods such as logistic regression are insufficient to solve the complexity of modern data sets, whereas ANN and hybrid models provide the highest accuracy rates. The authors emphasize that advanced artificial intelligence architectures capable of analyzing complex relationships within the dataset are essential for solving multifactorial problems such as obesity, rather than relying on singular and simplistic models [13].

In addition to the predictive success of artificial intelligence algorithms, how these technologies can be integrated into the clinical stages of obesity management is also of vital importance. DeGregory and colleagues, in their study published in the journal Obesity Reviews, addressed the use of artificial intelligence under four main pillars: risk prediction, behavioral monitoring, treatment response prediction, and causal inference. Researchers suggest image processing technologies that analyze food consumption through plate photographs rather than questionnaires in order to eliminate self-reporting errors, which are the biggest limitation in obesity studies. Furthermore, it emphasizes that in order for physicians to trust algorithmic decisions, the black box problem must be overcome by transitioning to explainable artificial intelligence models, and that data should not be limited to lifestyle factors but should be enriched with genetic, proteomic, and microbiome data [14].

To overcome the limitations of individual algorithms and maximize predictive power, ensemble learning approaches are frequently used in the literature. Jindal and colleagues developed a ensemble model that combines the strengths of different classifiers rather than relying on a single model for obesity prediction. Research findings indicate that this hybrid structure produces more stable results by reducing variance compared to standalone algorithms and achieves a high accuracy rate of 89.68% in obesity risk prediction, demonstrating the superiority of ensemble methods in complex datasets [15].

3. MATERIAL AND METHODS

This study aims to classify and predict individuals' obesity levels with high accuracy. The 'Obesity Dataset', which includes genetic characteristics and dietary habits, was used as the data source for the study. Random Forest, XGBoost, CatBoost, and YSA algorithms were used in the modeling phase. This section covers the structural characteristics of the dataset used, the data preprocessing steps applied, and the theoretical basis of the models established.

3.1. Dataset

The dataset used in this study is the open-source 'Obesity Dataset' created and contributed to the literature by Niğmet Köklü and Süleyman Alpaslan Sulak to analyze individuals' obesity status based on their social and physical activities [16]. The data was collected through an online survey administered to volunteer participants living throughout Turkey. The dataset contains data from a total of 1,610 participants aged between 18 and 54. When examining the demographic distribution of participants, 898 (55.8%) were female and 712 (44.2%) were male. There are 14 variables in the dataset representing factors affecting obesity risk. These variables include factors such as demographic characteristics, dietary habits, and social life [17]. These variables are shown in Figure 1.

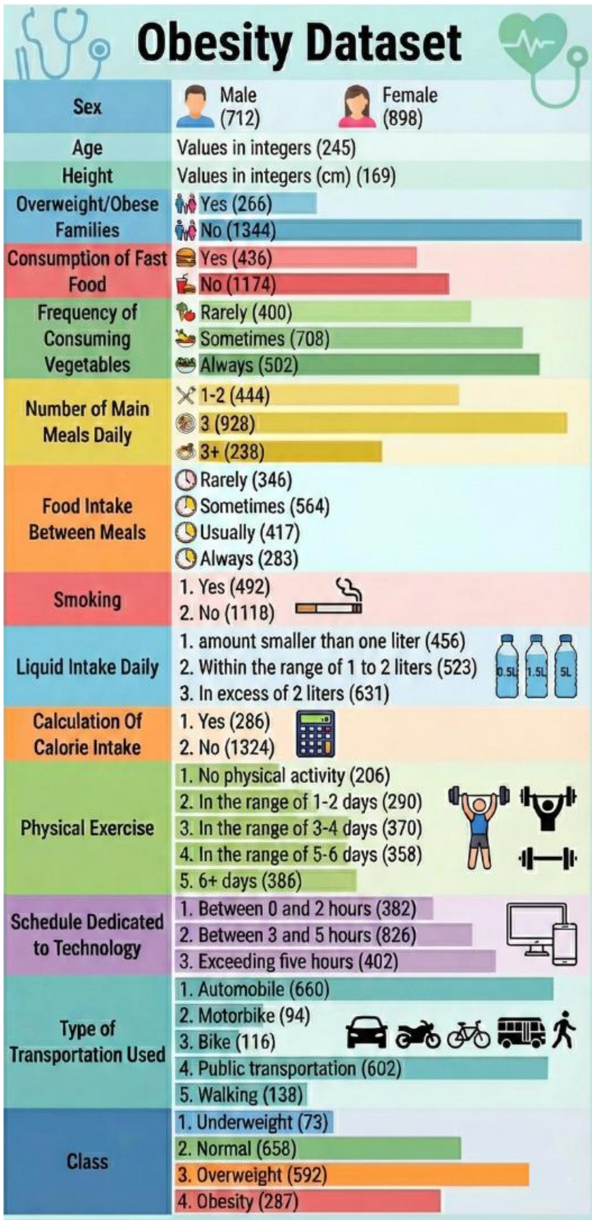

Figure 1. Distribution of Attributes Belonging to the Data Set

3.2. Exploratory Data Analysis and Visualization

In this study, Exploratory Data Analysis (EDA) procedures were performed to understand the structural characteristics of the data set, analyze the relationships between variables, and identify data distribution issues that could affect model performance. The analysis process was visualized using the Python programming language through the Matplotlib and Seaborn libraries. First, the distributions of the variables were examined to understand the structural characteristics of the data set used in the study. Histograms summarize the overall frequencies of both continuous and categorical variables in a dataset. In the visualization, it can be seen in Figure 2 that most of the variables are categorical or ordinal in nature, while the 'Height' variable exhibits a near-normal distribution.

Subsequently, an outlier analysis, as shown in Figure 3, was performed to minimize the model's sensitivity to noise. Upon examining the graphs, outliers exceeding the upper quartile limit in the 'Physical_Exercise' variable were identified, which could potentially be considered noise. Box plots also summarize the distribution characteristics of scaled features in the range [0, 1] [18]. Certain variables in the distribution visualizations (Consumption_of_Fast_Food, Smoking) indicate that the data exhibits a positively skewed structure, as evidenced by the median lines being positioned close to the lower boundary of the box.

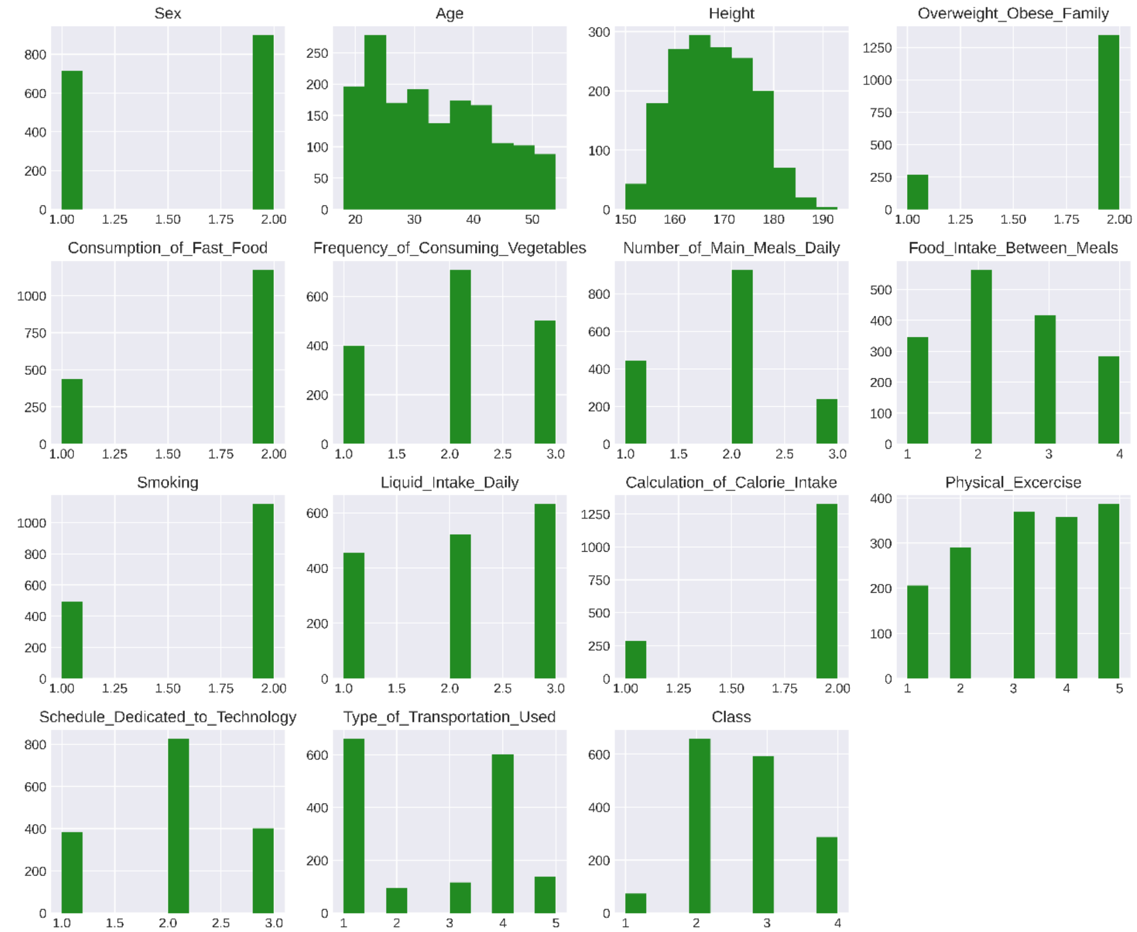

Figure 2. Frequency distribution graphs of the attributes in the dataset

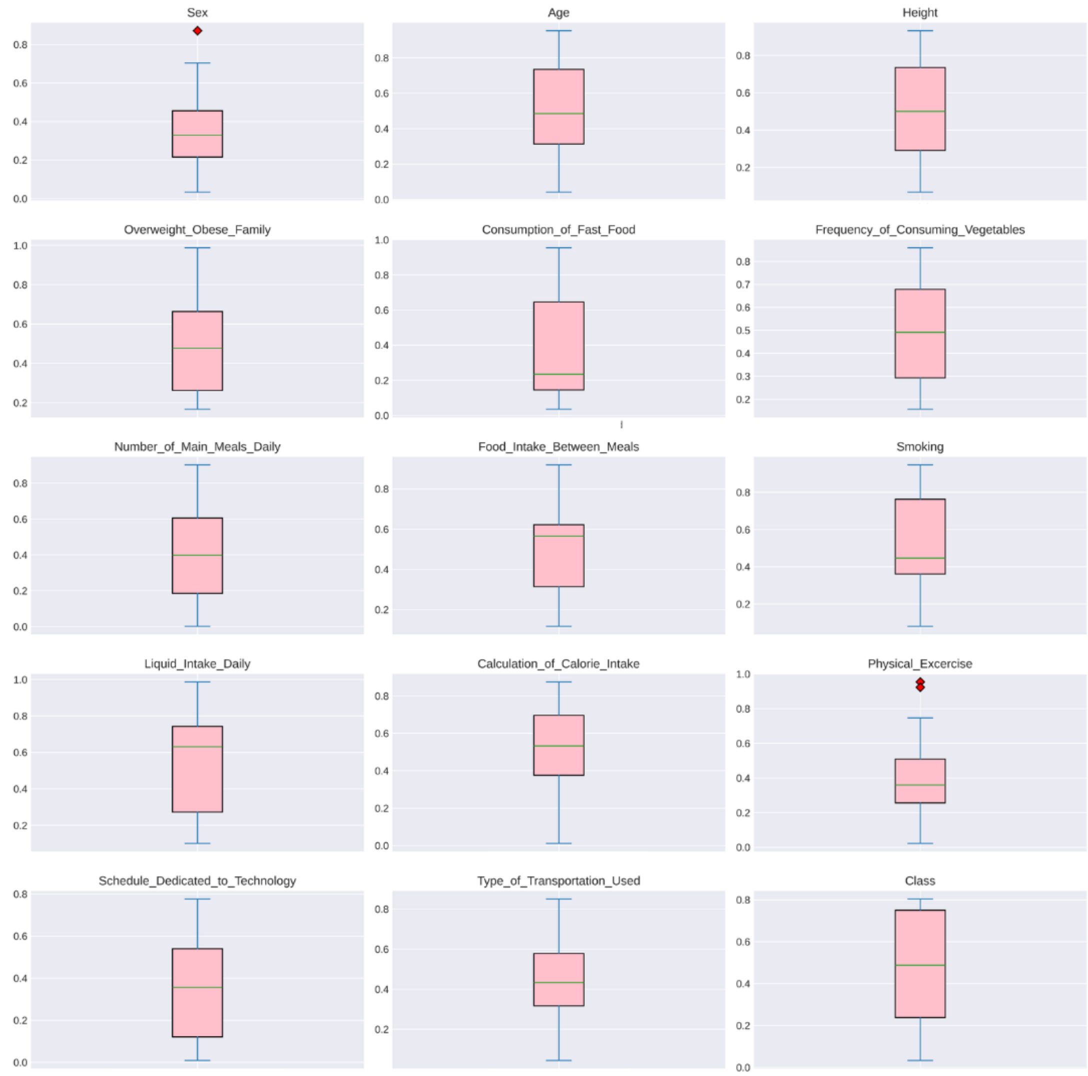

Figure 3. Outlier analysis and distribution characteristics of variables (Box Plots)

Finally, to reveal the direction and strength of linear relationships between numerical and coded categorical variables in the dataset, an analysis based on Pearson correlation coefficients was performed using the heat map shown in Figure 4 [19]. When examining the heat map, a moderate positive relationship is observed between the target variable ‘Class’ and ‘Gender’ (r=0.57) and ‘Age’ (r=0.49). On the other hand, when examining the risk of multicollinearity among attributes, it was determined that the vast majority of variables have low correlation with each other, which is considered a positive indicator in terms of the model's generalization ability. The high generalization capability of such behavioral and lifestyle datasets has also been demonstrated in recent studies where artificial intelligence algorithms were successfully used to classify individuals' environmental attitudes [20].

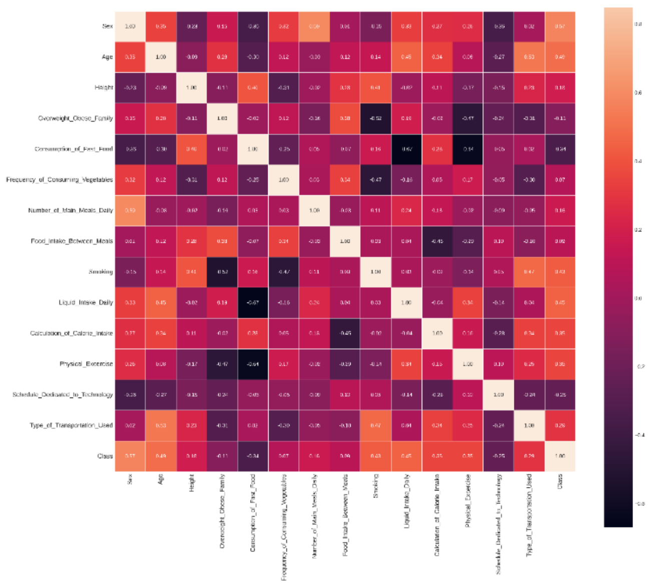

Figure 4. Pearson correlation coefficient matrix between variables

3.3. Data Preprocessing

Data preprocessing was performed systematically to prepare the "Obesity Dataset" for the learning phase.

Given the architectural differences between tree-based ensemble methods and Artificial Neural Networks, specialized preprocessing steps were implemented for each model group.

3.3.1. General Preprocessing for All Models

Numerical Normalization: Numerical attributes, including Age and Height, were normalized into the [0, 1] range using MinMaxScaler. This step ensures that features with different magnitudes contribute equally to the model's decision-making process.

Label Encoding: Categorical features such as Gender, Smoking, and Transportation Type were converted into numerical values using LabelEncoder to make them machine-readable.

3.3.2. Preprocessing for Tree-Based Models (Random Forest, XGBoost, CatBoost)

Feature Handling: These models were trained on the label-encoded dataset. Since decision trees are naturally robust to the scale of input features, the hierarchical nature of encoded categorical data was directly utilized for splitting nodes.

Handling Class Imbalance: For the Random Forest algorithm, the class_weight='balanced' parameter was applied during the training phase to compensate for the minority classes in the dataset.

3.3.3. Preprocessing for Artificial Neural Networks (ANN)

Feature Scaling Necessity: Unlike tree-based models, ANN requires consistent numerical scales for the backpropagation algorithm to converge efficiently.

Input Layer Configuration: A unified feature matrix was constructed to feed a 14-dimensional input layer. All categorical inputs were encoded and then combined with normalized numerical inputs to ensure the mathematical stability of the gradient descent optimizer.

3.4. Machine Learning Algorithms

The ability to automatically extract meaningful patterns from data and generalize them to new situations forms the basis of machine learning, an important part of artificial intelligence. In this process, numerous techniques have been developed to meet needs such as classification, prediction (regression), grouping (clustering), and data simplification (dimension reduction) [21, 22, 23]. In this study, four different machine learning algorithms known for their high success rates in the literature—Random Forest, XGBoost, CatBoost, and ANN—were used comparatively to estimate obesity levels [24, 25, 26]. During the model implementation, the class distribution of the dataset was analyzed. Although the four obesity levels showed a relatively balanced distribution, the Random Forest algorithm was configured with the class_weight='balanced' parameter to further stabilize its decision-making process against minor frequency variations. Conversely, for XGBoost and CatBoost, explicit weighting was not applied; these gradient-boosting frameworks naturally address class distribution by iteratively focusing on misclassified samples from underrepresented groups through their sequential error-correction mechanisms. This differentiated approach ensures that each algorithm operates according to its inherent architectural strengths while maintaining high generalization across all classes.

3.4.1. Random Forest

It is a ensemble learning algorithm that improves prediction accuracy by combining the results produced by multiple decision trees developed by Breiman [27]. This method, which uses the bagging technique, creates random subsets from the data set and trains a separate decision tree for each subset. The majority decision based on the voting results is taken as the basis for all trees created in the classification process. The main reason the algorithm is called ‘Random’ is that a randomly selected subset of attributes is used when constructing each tree, and the splitting operation is performed using the most suitable attribute within this subset [28]. The Random Forest algorithm was chosen for this study because it is resistant to overfitting and successfully manages imbalances in the dataset. Figure 5 shows the algorithm architecture.

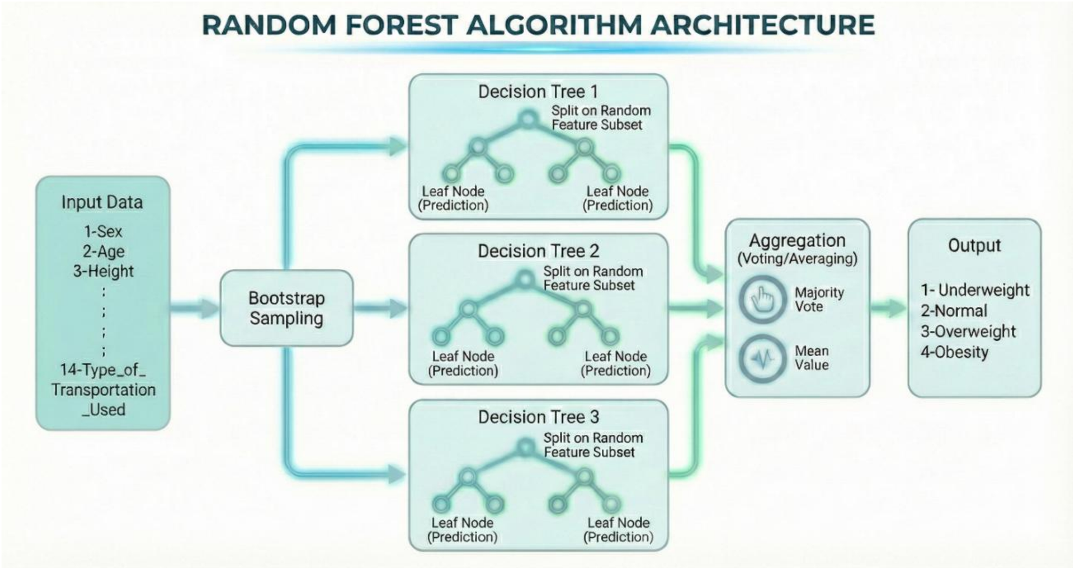

Figure 5. The Architecture of the Random Forest Algorithm

3.4.2. XGBoost (eXtreme Gradient Boosting)

XGBoost is a scalable tree-based algorithm developed under the gradient boosting framework that focuses on speed and performance. It progresses by sequentially adding weak learners, each new tree aims to correct the errors made by the previous tree. XGBoost prevents overfitting by controlling model complexity through its L1 and L2 regularization terms. It was selected as the comparison algorithm in this study due to its high performance, particularly in structural data [29]. Figure 6 shows the algorithm architecture.

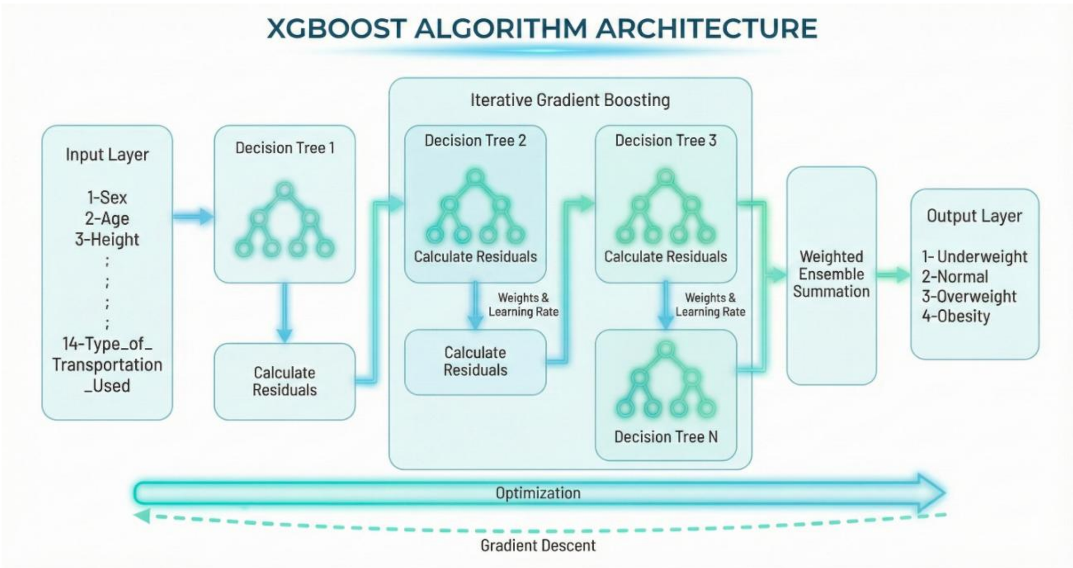

Figure 6. Architecture of the XGBoost Algorithm

3.4.3. CatBoost (Categorical Boosting)

Developed by Yandex, CatBoost is a gradient boosting algorithm optimized for datasets with a high number of categorical variables. Unlike traditional methods, it uses the sequential target statistic method to prevent information loss that may occur when converting categorical data into numerical data [30]. Due to the obesity dataset used in the study containing numerous categorical variables (gender, dietary habits, mode of transportation etc.), the CatBoost algorithm constitutes one of the most critical models in this study. Figure 7 shows the algorithm architecture.

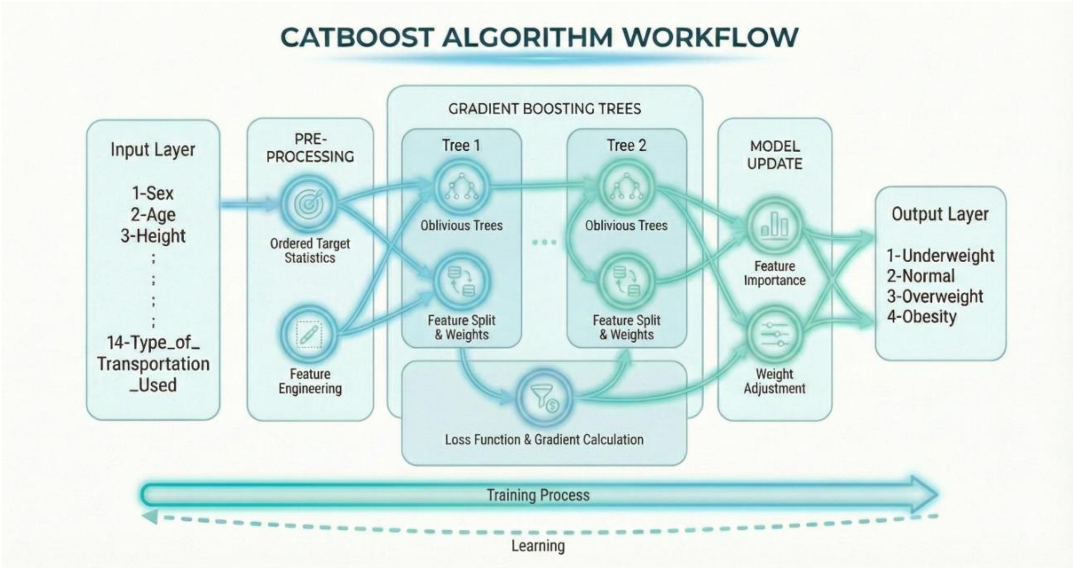

Figure 7. Architecture of the CatBoost Algorithm

3.4.4. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are computational models inspired by the biological neural structures of the human brain, designed to capture complex, non-linear relationships within data. In this study, a multi-layer perceptron (MLP) architecture was implemented using the Keras Sequential API. The proposed model consists of an input layer corresponding to the 14 features of the obesity dataset, followed by two hidden layers with 64 and 32 neurons, respectively [31, 32, 33]. To ensure effective learning and prevent the vanishing gradient problem, the Rectified Linear Unit (ReLU) activation function was employed in the hidden layers. The output layer comprises 4 neurons with a Softmax activation function to provide probability distributions for the four obesity classes. The model was compiled using the Adam optimizer and sparse categorical cross-entropy loss function. Training was conducted for 100 epochs with a batch size of 8, and a 20% validation split was utilized to monitor the model's generalization performance during the learning process. Figure 8 illustrates this architecture.

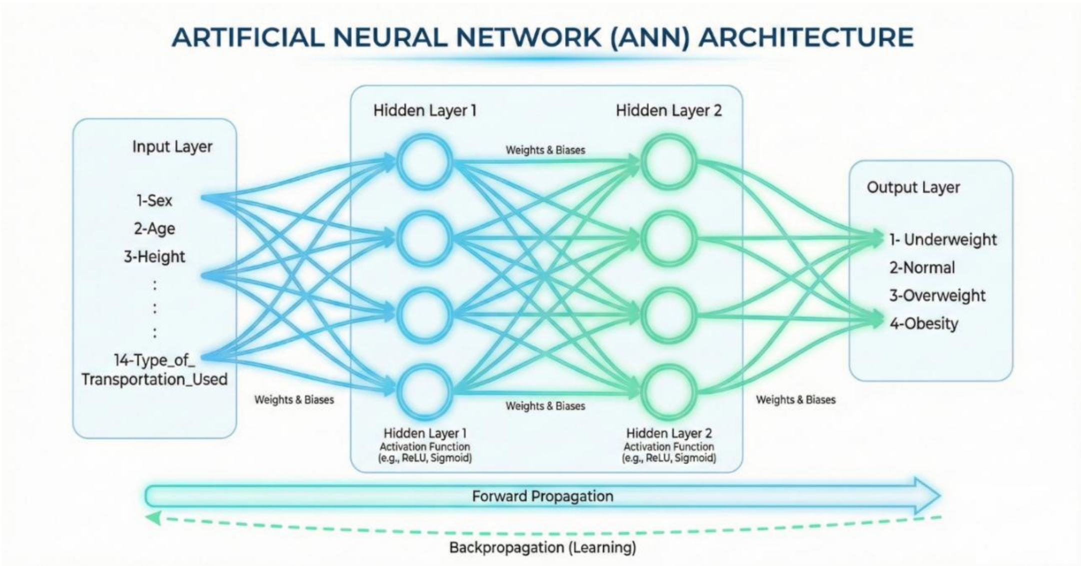

Figure 8. Architecture of the ANN Algorithm

3.5. Cross Validation

To objectively evaluate the performance of the machine learning models (Random Forest, XGBoost, CatBoost, and ANN) developed in this study and to prevent the problem of overfitting, the K-Fold Cross Validation method, whose diagram is given in Figure 9, was used. The most common and accepted approach in the literature, k=5 was preferred [34]. During this process, the dataset was randomly divided into 5 equal parts. In each iteration, one of the parts was separated as the test set, while the remaining four parts were used as the training set. This process was repeated 5 times, and a different subset of data was tested in each cycle. The overall performance of the model was determined by taking the arithmetic mean of the success scores obtained from each iteration. This method prevents biased results that are tied to a specific part of the data set and confirms the model's generalizability.

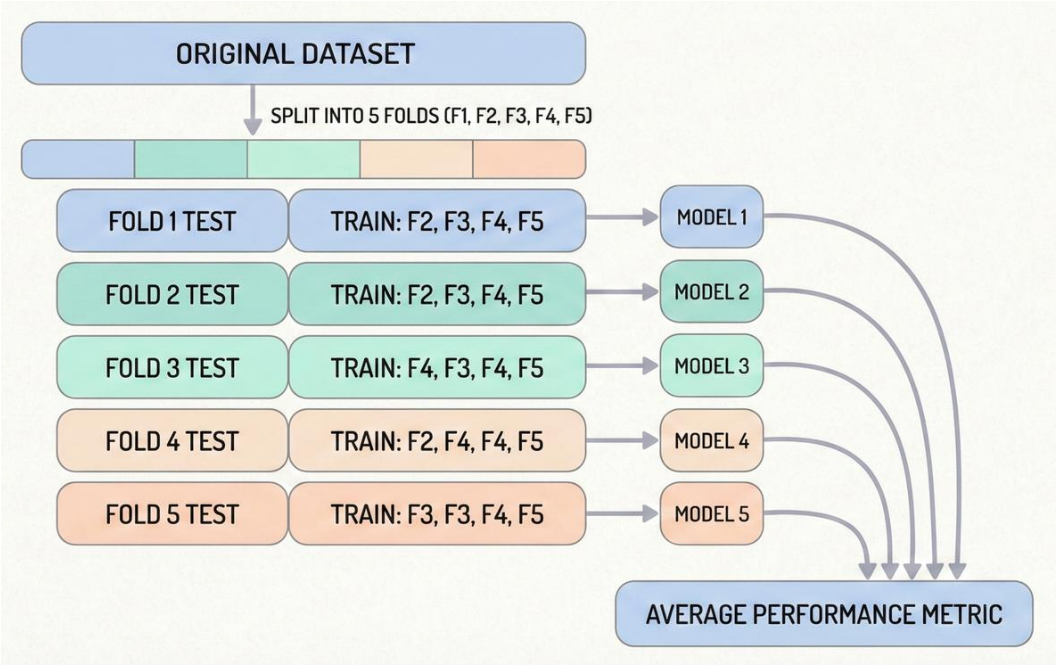

Figure 9. Diagram of the cross-validation method

3.6. Performance Metrics

Internationally recognized performance metrics have been used to quantitatively express the predictive capabilities of the developed machine learning models, to make an objective comparison between models, and to analyze classification errors in detail [35]. The Confusion Matrix shown in Figure 10 was used as the primary tool for evaluating the models. This matrix encompasses several key elements, including true positives (examples correctly classified as positive), false positives (examples incorrectly classified as positive instead of negative), true negatives (examples correctly classified as negative), and false negatives (examples incorrectly classified as negative instead of positive) [36].

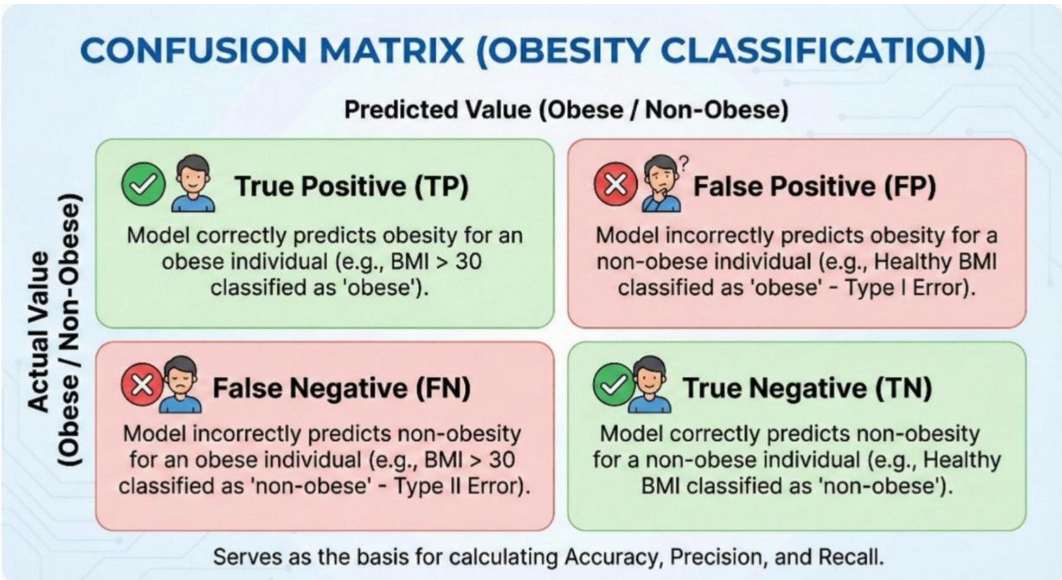

Figure 10. General structure and explanation of confusion matrix values

The metrics in Table 1 were calculated based on the confusion matrix values obtained.

Table 1. Defining Performance Metric Formulas

Measure	Description	Formula
Accuracy	It indicates the percentage of correctly classified examples within the total data. It is a reliable indicator when classes are evenly distributed or close to each other. However, if there is a significant imbalance between classes in the dataset (e.g., the positive class being rare), high scores may be misleading about the model's success [37, 38].	(TP+TN)/(TP+FP+FN+TN)×100
Precision	It indicates how many of the examples classified as positive by the model are actually correct. This criterion plays a decisive role, particularly in scenarios where the cost of false positive results is critical [38, 39].	TP/(TP+FP)×100
Recall	It represents the model's success in identifying positive data points. It shows how much of the system's current positive conditions it can cover. This metric is a decisive performance criterion in areas such as medical diagnosis [38, 39].	TP/(TP+FN)×100
F-Score	This metric, created by taking the harmonic mean of the precision and recall values, establishes a balance between the two variables. It is the most appropriate performance metric, especially in situations where both false positives and false negatives are critically important in the analytical process [38, 39].	(2×TP)/(2×TP+FP+FN)×100

The machine learning algorithms and models used in the study, along with the hyperparameter settings applied during the training process, are summarized in Table 2.

Table 2. Algorithm Parameters

ML Algorithm	Parameters
Random Forest	target_size=4, class_weight=balanced, random_state=42
XGBoost	n_estimators=100, random_state=55, learning_rate=0.1, verbosity=1, early_stopping_rounds=10
CatBoost	iterations=1000, depth=5, border_count=50, l2_leaf_reg=0.4, learning_rate=4e-2
ANN	Hidden_layers=2, activation_function=ReLU(hidden)-softmax(output), optimizer=adam, learning_rate=00.1, loss_function=sparse_categorical_crossentropy, max_epoch=100, batch_size=8, validation_split=0.2

4. Experimental Results

In this study, the performance of Random Forest, XGBoost, CatBoost, and Artificial Neural Network models developed to estimate obesity levels was tested using the 5-fold cross-validation method. Considering the class imbalance in the dataset, the success of the models was primarily evaluated using the Accuracy and F1-Score metrics obtained from the complexity matrices in Figure 11.

The experimental findings obtained from complexity matrices are detailed in Table 3.

Table 3. Comparison of Machine Learning Algorithm Training Results

ML Algorithm	Accuracy	Recall	F-Score	Precision
Random Forest	94.34	94.35	94.36	94.51
XGBoost	92.80	92.80	92.79	93.05
CatBoost	76.34	76.34	76.36	76.64
ANN	82.98	82.98	82.83	83.16

The analyses revealed that tree-based learning algorithms demonstrated superior performance on this dataset compared to the ANN approach. Figure 12 shows a graph that allows us to compare model performances.

Random Forest: The most successful model in the study was the Random Forest algorithm with 94.34% accuracy. Training the model with the class_weight='balanced' parameter played a decisive role in the accurate prediction of minority classes (e.g., 'Underweight' or 'Type I Obesity') and increased overall success.

XGBoost: Among gradient boosting-based methods, XGBoost has demonstrated the highest performance. The model closely followed Random Forest with an accuracy rate of 92.80%. XGBoost's high speed and regularization capability have ensured that the model produces stable results.

CatBoost: The CatBoost algorithm, known for its ability to process categorical variables, demonstrated acceptable but limited performance in the analyses conducted, achieving an accuracy rate of 76.34 and an F1-score of 76.36.

Artificial Neural Network (ANN): The ANN model has achieved an accuracy rate of 82.98%. In tabular data, the relatively limited size of the dataset for ANNs has constrained the model's ability to learn complex relationships compared to tree-based models.

a) Confusion Matrix of Random Forest **b)** Confusion Matrix of XGBoost

c) Confusion Matrix of CatBoost **d)** Confusion Matrix of ANN

Figure 11. Confusion matrix for all machine learning algorithms

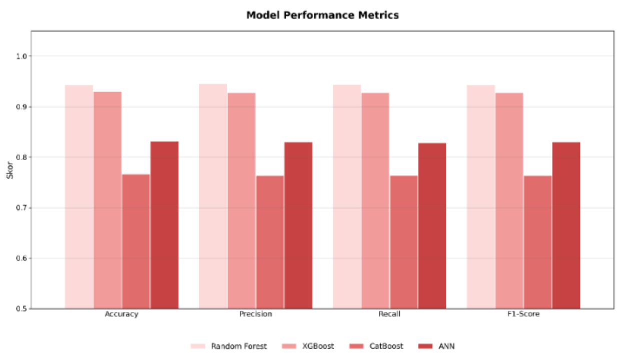

Figure 12. Bar Chart of the Machine Learning Algorithm Training Results

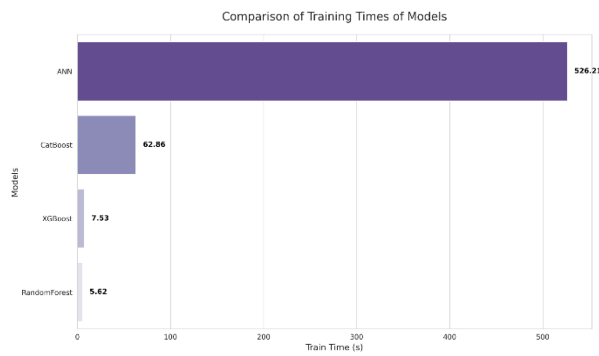

Figure 13. Comparison of Training Times for Machine Learning Algorithms

The `time` function in Python's standard library was used to measure the training times of the models. Random Forest stood out as the model with the shortest training time at 5.62 seconds, followed by XGBoost at 7.53 seconds. In contrast, the ANN model had the longest training time at 526.21 seconds, showing that the ANN required approximately 93 times more time than the Random Forest model. The graph comparing the training times of the models is shown in Figure 13.

5. Result and Discussion

In this study, the performance of machine learning algorithms (Random Forest, XGBoost, CatBoost, and ANN) was analyzed in the classification of a multifactorial public health problem such as obesity. The experimental findings obtained reveal that tree-based models produce superior and more stable results on tabular health data compared to ANN architecture.

Our research results show that the Random Forest algorithm outperforms all other methods with an accuracy rate of 94.34% and an F1-score of 94.36. This finding is fully consistent with the thesis reported in the literature by Suwarno and colleagues that "the phenomenon of obesity is better modeled by nonlinear decision mechanisms" [9]. The success of Random Forest can be explained by the algorithm's ability to reduce variance through the bagging method and its resilience to noisy observations in the dataset. Especially in cases where categorical and numerical data are mixed, as in our dataset, Random Forest's ability to hierarchically partition the feature space has maximized classification success.

In contrast, our ANN model showed the lowest performance among the models compared, with an accuracy rate of 82.98%. Although the literature reports over 98% success with Kıvrak YSA, this performance difference in our study is related to the volume and structure of the data set. While YSA models generally demonstrate their full potential on very large datasets and unstructured data, as emphasized in Ölçer's work, algorithms such as Random Forest typically show higher success on medium-sized and structured datasets. The results of our study support the view in the literature that "ensemble methods are superior for tabular data" [40].

Additionally, XGBoost closely followed Random Forest, achieving a competitive accuracy rate of 92.80%. The finding that the ensemble learning approach proposed by Jindal and colleagues is more reliable than individual models has been confirmed once again by the high performance of Random Forest and XGBoost in our study [14]. High accuracy alone is not sufficient for the clinical validity of artificial intelligence models used in the fight against obesity. At the same time, the Recall value must also be high. Because in medical diagnoses, missing a disease is far more costly than a false alarm.

In our study, the Random Forest algorithm's high Recall value of 94.51 indicates that the model is highly accurate in detecting actual obesity cases. This situation indicates that the developed model can be used as a reliable Decision Support System (DSS) for the early detection of individuals at risk within the framework of the proactive health approach proposed by Shaban and colleagues [7]. Furthermore, bibliometric analyses covering the last four years in the field of educational data mining demonstrate that these techniques have become increasingly critical as decision support systems for improving both individual performance and health outcomes [41]. The model's ability to accurately distinguish between the "Normal" and "Overweight" categories, where transitions are particularly frequent, is critically important for alerting individuals in the transition phase to obesity and taking preventive measures.

In addition to the strengths of the study, there are also some limitations. The data set used consists of survey data based on participants' self-reports. As DeGregory and colleagues point out, individuals' potential for bias when reporting their own height and weight may introduce potential noise into the dataset [13]. In future studies, the use of objectively measured data collected in a clinical setting and the expansion of the dataset to include different demographic groups will increase the model's generalizability.

Declaration of Ethical Standards

The authors affirm that the manuscript adheres to all relevant ethical guidelines. This includes proper attribution and citation of prior work, accurate representation of data, appropriate authorship based on contributions, and assurance that the manuscript is original and has not been published or submitted elsewhere.

Credit Authorship Contribution Statement

Conceptualization: ACC and SSB; Methodology: ACC and SSB; Software: ACC; Validation: SSB and MK Formal analysis: ACC and MK; Investigation: ACC and SSB; Resources: ACC; Data curation: ACC and SSB; Writing original draft preparation: ACC, SSB and MK; Writing review and editing: SSB and MK; Visualization: ACC; Supervision: MK; Project administration: MK. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

This study uses a dataset obtained from Kaggle (Koklu ve Sulak, 2024) and the data can be accessed via the following link: https://www.kaggle.com/datasets/suleymansulak/obesity-dataset/data

Acknowledgments

The authors would like to express their gratitude to the Department of Computer Engineering, Faculty of Technology, Selcuk University, for providing access to their laboratory facilities used in this study.

Disclosure Statement

Generative artificial intelligence tools were employed for grammar refinement, linguistic clarity, and improvements in academic writing quality. These tools served as language-editing assistance within the manuscript preparation process.

Conflicts of Interest

The authors declare no conflict of interest

References

[1] Blüher, M. (2019). Obesity: global epidemiology and pathogenesis. Nature reviews endocrinology, 15(5), 288-298. https://doi.org/10.1038/s41574-019-0176-8

[2] Swinburn, B. A., Sacks, G., Hall, K. D., McPherson, K., Finegood, D. T., Moodie, M. L., & Gortmaker, S. L. (2011). The global obesity pandemic: shaped by global drivers and

local environments. The lancet, 378 (9793), 804-814. https://doi.org/10.1016/S0140-6736(11)60813-1

[3] Bayram, S. Ş., & Aktaş, N. (2020). Selçuk Üniversitesi Öğrencilerinin Akdeniz Diyet Kalitelerinin Değerlendirilmesi. Beslenme ve Diyet Dergisi, 48(3), 65-75. https://doi.org/10.33076/2020.BDD.1386

[4] Guh, D.P., Zhang, W., Bansback, N. et al. (2009). The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. BMC Public Health 9, 88. https://doi.org/10.1186/1471-2458-9-88

[5] Di Angelantonio, E., Bhupathiraju, S. N., Wormser, D., Gao, P., Kaptoge, S., De Gonzalez, A. B., ... & Hu, F. B. (2016). Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. The lancet, 388 (10046), 776-786. https://doi.org/10.1016/S0140-6736(16)30175-1

[6] Tremmel, M., Gerdtham, U.-G., Nilsson, P. M., & Saha, S. (2017). Economic Burden of Obesity: A Systematic Literature Review. International Journal of Environmental Research and Public Health, 14 (4), 435. https://doi.org/10.3390/ijerph14040435

[7] Shaban, W.M., El-Din Moustafa, H. & El-Seddek, M.M. (2025). Machine learning framework for predicting susceptibility to obesity. Scientific Reports 15, 35040. https://doi.org/10.1038/s41598-025-20505-9

[8] Al Khushi Joshi, E. (2023). Comparison of Different Machine Learning and Self-Learning Methods for Predicting Obesity on Generalized and Gender-Segregated Data. International Journal of Recent Innovation in Clouds, Computing, Analytics and Networking, 11(10), 464-471. https://doi.org/10.17762/ijritcc.v11i10.8510

[9] Ölçer, E. Makine Öğrenmesi Temelli Obezite Durum Tahminlemesi. Bilgisayar Bilimleri ve Mühendisliği Dergisi, 17(2), 156-164. https://doi.org/10.54525/bbmd.1469701

[10] Suwarno, Murnaka, N. P., Prasetyo, P. W., & Arifin, S. (2023). Performance comparison of machine learning algorithms for predicting obesity level. AIP Conf. Proc. 2733 (1): 020002. https://doi.org/10.1063/5.0140856

[11] Musa, F., & Basaky, F. (2022). Obesity prediction using machine learning techniques. Journal of Applied Artificial Intelligence, 3 (1), 24-33. https://doi.org/10.48185/jaai.v3i1.470

[12] Kıvrak, M. (2021). Deep learning-based prediction of obesity levels according to eating habits and physical condition. The Journal of Cognitive Systems, 6 (1), 24-27. https://doi.org/10.52876/jcs.939875

[13] Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., & Shapi'i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. Computers in biology and medicine, 136, 104754. https://doi.org/10.1016/j.compbiomed.2021.104754

[14] DeGregory, K. W., Kuiper, P., DeSilvio, T., Pleuss, J. D., Miller, R., Roginski, J. W., ... & Thomas, D. M. (2018). A review of machine learning in obesity. Obesity reviews, 19 (5), 668-685. https://doi.org/10.1111/obr.12667

[15] Jindal, K., Baliyan, N., & Rana, P. S. (2018). Obesity Prediction Using Ensemble Machine Learning Approaches. In: Sa, P., Bakshi, S., Hatzilygeroudis, I., & Sahoo, M. (Eds.), Recent Findings in Intelligent Computing Techniques. Advances in Intelligent Systems and Computing, vol 708, pp. 397-405. Springer, Singapore. https://doi.org/10.1007/978-981-10-8636-6_37

[16] Sulak, S. A. (2024). Obesity Dataset. Kaggle. https://www.kaggle.com/datasets/suleymansulak/obesity-dataset (Access date: 12 November 2025).

[17] Koklu, N., & Sulak, S. A. (2024). Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities. Sinop Üniversitesi Fen Bilimleri Dergisi, 9(1), 217-239. https://doi.org/10.33484/sinopfbd.1445215

[18] Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician, 32 (1), 12–16. https://doi.org/10.1080/00031305.1978.10479236

[19] Gu, Z. (2022). Complex heatmap visualization. Imeta, 1 (3), e43. https://doi.org/10.1002/imt2.43

[20] Koklu, N., & Sulak, S. A. (2024). Classification of Environmental Attitudes with Artificial Intelligence Algorithms. Intelligent Methods In Engineering Sciences, 3(2), 54-62. https://doi.org/10.58190/imiens.2024.99

[21] Butuner, R., Cinar, I., Taspinar, Y. S., Kursun, R., Calp, M. H., & Koklu, M. (2023). Classification of deep image features of lentil varieties with machine learning techniques. European Food Research and Technology, 249(5), 1303-1316. https://doi.org/10.1007/s00217-023-04214-z

[22] Koklu, M., & Ozkan, I. A. (2020). Multiclass classification of dry beans using computer vision and machine learning techniques. Computers and Electronics in Agriculture, 174, 105507. https://doi.org/10.1016/j.compag.2020.105507

[23] Kursun, R., & Koklu, M. (2025). Classification of Eggplant Diseases Using Feature Extraction with AlexNet and Random Forest. Sinop University Journal of Natural Sciences, 10(1), 1-15.

[24] Cinar, I., & Koklu, M. (2019). Classification of rice varieties using artificial intelligence methods. International Journal of Intelligent Systems and Applications in Engineering, 7(3), 188-194. https://doi.org/10.18201/ijisae.2019355381

[25] Dong, X., Yu, Z., Cao, W. et al. A survey on ensemble learning. Frontiers of Computer Science. 14(2), 241–258 (2020). https://doi.org/10.1007/s11704-019-8208-z

[26] Ozkan, I. A., & Koklu, M. (2017). Skin lesion classification using machine learning algorithms. International Journal of Intelligent Systems and Applications in Engineering, 5(4), 285-289. https://doi.org/10.18201/ijisae.2017534420

[27] Breiman, L. (2001). Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

[28] Yasin, E., & Koklu, M. (2023, December). Utilizing Random forests for the classification of pudina leaves through feature extraction with inceptionV3 and VGG19. In Proceedings of the International Conference on New Trends in Applied Sciences (Vol. 1, pp. 1-8).

[29] Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Cornell University. https://doi.org/10.1145/2939672.2939785

[30] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31. https://doi.org/10.48550/arXiv.1706.09516

[31] Koklu, M., Sarigil, S., & Ozbek, O. (2021). The use of machine learning methods in classification of pumpkin seeds (Cucurbita pepo L.). Genetic Resources and Crop Evolution, 68(7), 2713-2726. https://doi.org/10.1007/s10722-021-01226-0

[32] Koklu, M., Unlersen, M. F., Ozkan, I. A., Aslan, M. F., & Sabanci, K. (2022). A CNN-SVM study based on selected deep features for grapevine leaves classification. Measurement, 188, 110425. https://doi.org/10.1016/j.measurement.2021.110425

[33] Krogh, A. (2008). What are artificial neural networks?. Nature Biotechnology 26, 195–197. https://doi.org/10.1038/nbt1386

[34] Berrar, D. (2019) Cross-Validation. In: Ranganathan, S., Gribskov, M., Nakai, K. and Christian Schönbach, C., Eds., Reference Module in Life Sciences Encyclopedia of Bioinformatics and Computational Biology, Vol. 1, Elsevier, Amsterdam, 542-545. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

[35] Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. Midwest Artificial Intelligence and Cognitive Science Conference, 710 (1), 120-127.

[36] Yasin, E. T., & Koklu, M. (2025). A comparative analysis of machine learning algorithms for waste classification: inceptionv3 and chi-square features. International Journal of Environmental Science and Technology, 22(10), 9415-9428. https://doi.org/10.1007/s13762-024-06233-z

[37] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining and Knowledge Management Process, 5 (2), 1. https://doi.org/10.5121/ijdkp.2015.5201

[38] Saritas, M. M., & Koklu, M. (2024). Classification of cauliflower leaf diseases using features extracted from Squeezenet with decision tree and random forest. In: Proceedings of the 4th International Conference on Frontiers in Academic Research (ICFARI 2024), Konya, Türkiye, pp. 563-572.

[39] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

[40] Koklu, M., Cinar, I., & Taspinar, Y. S. (2022). CNN-based bi-directional and directional long-short term memory network for determination of face mask. Biomedical signal processing and control, 71, 103216. https://doi.org/10.1016/j.bspc.2021.103216

[41] Koklu, N., & Sulak, S. A. (2024). Recent Developments in Educational Data Mining: A Four-Year Bibliometric Analysis. Advances in Education Sciences, M. Dalkılıç and O. Soslu, Eds. Platanus Publishing, 5-29.