v7.n9.2026

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

AUTOML AND DYNAMIC FEATURE SELECTION FOR OBESITY LEVEL CLASSIFICATION: A COMPARATIVE ANALYSIS OF TREE-BASED MODELS

AUTOML E SELEÇÃO DINÂMICA DE ATRIBUTOS NA CLASSIFICAÇÃO DE NÍVEIS DE OBESIDADE: UMA ANÁLISE COMPARATIVA DE MODELOS BASEADOS EM ÁRVORES

AUTOML Y SELECCIÓN DINÁMICA DE ATRIBUTOS EN LA CLASIFICACIÓN DE NIVELES DE OBESIDAD: UN ANÁLISIS COMPARATIVO DE MODELOS BASADOS EN ÁRBOLES

Rhyan Ximenes de Brito[1], Janaide Nogueira de Sousa Ximenes[2]

e799078

https://doi.org/10.47820/recima21.v7i9.9078

PUBLISHED: 09/2026

ABSTRACT

The classification of obesity levels based on anthropometric, dietary, and behavioral characteristics can contribute to more consistent analyses in health-related contexts. This study aimed to comparatively evaluate tree-based Machine Learning models through an AutoML approach combined with dynamic feature selection. The Estimation of Obesity Levels Based on Eating Habits and Physical Condition dataset was used, comprising 2,111 records, 16 predictive attributes, and seven nutritional status classes. Decision Tree, Random Forest, Extra Trees, Bagging Trees, AdaBoost, XGBoost, LightGBM, and CatBoost were evaluated using stratified 10-fold cross-validation. In each fold, attributes with importance values above the mean were selected, and the same classifier was retrained using the resulting subset. Performance was assessed through multiple metrics, including Accuracy, Balanced Accuracy, F1-score, Kappa, and MCC. LightGBM achieved the best overall performance, with an Accuracy of 0.977 ± 0.008 and an MCC of 0.974 ± 0.009, followed by CatBoost (MCC = 0.969 ± 0.010) and Extra Trees (MCC = 0.966 ± 0.010). In the dynamic selection process, Weight was selected in 100% of the folds by all models, while Height, Age, Gender, and FCVC also showed high recurrence. The results indicate that integrating systematic classifier comparison with dynamic feature selection makes it possible to identify consistent models and relevant variables for multiclass obesity level classification.

KEYWORDS: *Obesity. AutoML. Machine Learning. Feature Selection.*

RESUMO

A classificação dos níveis de obesidade a partir de características antropométricas, alimentares e comportamentais pode contribuir para análises mais consistentes em saúde. Este estudo teve como objetivo avaliar comparativamente modelos de Aprendizado de Máquina baseados em árvores por meio de uma abordagem AutoML associada à seleção dinâmica de atributos. Utilizou-se o conjunto *Estimation of Obesity Levels Based on Eating Habits and Physical Condition*, composto por 2.111 registros, 16 atributos preditores e sete classes de estado nutricional.

[1] Master's degree in Electrical and Computer Engineering (UFC), professor at IFCE.
[2] Master's degree in Computer Science (UECE), professor at IFPA.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Foram avaliados *Decision Tree*, *Random Forest*, *Extra Trees*, *Bagging Trees*, *AdaBoost*, *XGBoost*, *LightGBM* e *CatBoost*, empregando validação cruzada estratificada em 10 *folds*. Em cada partição, os atributos com importância superior à média foram selecionados e o mesmo classificador foi re-treinado com o subconjunto resultante. O desempenho foi analisado por múltiplas métricas, incluindo *Accuracy*, *Balanced Accuracy*, F1-*score*, Kappa e MCC. O *LightGBM* apresentou o melhor desempenho global, com *Accuracy* de 0,977 ± 0,008 e MCC de 0,974 ± 0,009, seguido por *CatBoost* (MCC = 0,969 ± 0,010) e *Extra Trees* (MCC = 0,966 ± 0,010). Na seleção dinâmica, *Weight* foi selecionado em 100% dos *folds* por todos os modelos, enquanto *Height*, *Age*, *Gender* e FCVC também apresentaram elevada recorrência. Os resultados indicam que a integração entre comparação sistemática de classificadores e seleção dinâmica de atributos permite identificar modelos consistentes e variáveis relevantes para a classificação multiclasse dos níveis de obesidade.

PALAVRAS-CHAVE: Obesidade. AutoML. Aprendizado de Máquina. Seleção de atributos.

RESUMEN

La clasificación de los niveles de obesidad a partir de características antropométricas, alimentarias y conductuales puede contribuir a análisis más consistentes en contextos relacionados con la salud. Este estudio tuvo como objetivo evaluar comparativamente modelos de Aprendizaje Automático basados en árboles mediante un enfoque AutoML asociado a la selección dinámica de atributos. Se utilizó el conjunto de datos Estimation of Obesity Levels Based on Eating Habits and Physical Condition, compuesto por 2.111 registros, 16 atributos predictivos y siete clases de estado nutricional. Se evaluaron Decision Tree, Random Forest, Extra Trees, Bagging Trees, AdaBoost, XGBoost, LightGBM y CatBoost mediante validación cruzada estratificada de 10 folds. En cada partición, se seleccionaron los atributos con valores de importancia superiores a la media y el mismo clasificador fue reentrenado utilizando el subconjunto resultante. El desempeño se evaluó mediante múltiples métricas, entre ellas Accuracy, Balanced Accuracy, F1-score, Kappa y MCC. LightGBM presentó el mejor desempeño global, con una Accuracy de 0,977 ± 0,008 y un MCC de 0,974 ± 0,009, seguido de CatBoost (MCC = 0,969 ± 0,010) y Extra Trees (MCC = 0,966 ± 0,010). En el proceso de selección dinámica, Weight fue seleccionado en el 100% de los folds por todos los modelos, mientras que Height, Age, Gender y FCVC también mostraron una alta recurrencia. Los resultados indican que la integración entre la comparación sistemática de clasificadores y la selección dinámica de atributos permite identificar modelos consistentes y variables relevantes para la clasificación multiclase de los niveles de obesidad.

***PALABRAS CLAVE**: Obesidad. AutoML. Aprendizaje Automático. Selección de atributos.*

INTRODUCTION

Obesity represents a major public health concern due to its high prevalence, multifactorial nature, and association with chronic diseases. In 2022, approximately one in eight people worldwide were living with obesity, including more than 890 million adults, while about 16% of the global adult population had this condition (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2025). This increasing prevalence has been observed across different countries and age groups, reinforcing

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

the need for strategies aimed at understanding, identifying, and preventing the problem (NCD RISK FACTOR COLLABORATION, 2024).

The analysis of obesity involves anthropometric, dietary, physical, and behavioral characteristics, favoring the application of computational methods capable of examining multiple attributes simultaneously. In this context, Palechor and Manotas (2019) provided a dataset designed to estimate obesity levels among individuals from Colombia, Peru, and Mexico, comprising 2,111 records, 17 attributes, and seven nutritional status categories. This dataset has subsequently been used in classification and data mining studies (PALECHOR; MANOTAS, 2019).

Machine Learning has shown relevant results in this task. De-La-Hoz-Correa *et al.* (2019), when comparing different classifiers, observed that the Decision Tree achieved the best performance on the analyzed dataset. Subsequent studies have also demonstrated the competitiveness of tree-based algorithms and ensemble methods, including *Random Forest*, *XGBoost*, *Extra Trees*, and *CatBoost*, although their performance depends on the characteristics of the data and the experimental protocol adopted (BAG *et al.*, 2023; GÖRMEZ *et al.*, 2025).

However, the diversity of models, configurations, and procedures involved in developing Machine Learning solutions increases the number of decisions required during experimentation. In this context, *Automated Machine Learning* (AutoML) seeks to automate stages such as data preparation, feature engineering and selection, model selection, and hyperparameter optimization, reducing repetitive manual procedures and enabling the systematic evaluation of different configurations (HE; ZHAO; CHU, 2021).

Feature selection also plays an important role, as redundant or poorly informative variables may increase computational complexity without necessarily improving performance. Identifying relevant subsets of features can help reduce dimensionality, improve computational efficiency, and enhance model interpretability and, under certain conditions, predictive performance (MOSLEMI, 2023; THENG; BHOYAR, 2024). In obesity classification, studies have combined feature selection, model optimization, and feature importance analysis, demonstrating that the contribution of individual features may vary according to the classifier employed (BAG *et al.*, 2023; GÖRMEZ *et al.*, 2025).

Therefore, this study is justified by the need to investigate a strategy that combines experimental automation, systematic comparison of tree-based models, and dynamic feature selection. Unlike a fixed selection procedure, defining features according to the model and validation partitions makes it possible to examine both predictive performance and the recurrence of features considered relevant across different runs.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Accordingly, the research problem is expressed by the following question: to what extent can an AutoML approach combined with dynamic feature selection contribute to the comparison of tree-based models and to the consistent identification of relevant features for obesity level classification? The general objective is to comparatively evaluate tree-based Machine Learning models for obesity level classification through an AutoML approach combined with dynamic feature selection. Specifically, the study aims to compare classifiers under the same experimental protocol; perform feature selection according to the model and validation partitions; assess performance using metrics appropriate for a multiclass problem; analyze performance stability; identify the frequency with which features are selected; and determine which models and feature subsets provide the most consistent results.

This paper is organized as follows: Section 1 presents the Related Work; Section 2 discusses the Theoretical Background, covering obesity, tree-based models, AutoML, and feature selection; Section 3 describes the Materials and Methods, including the dataset, preprocessing, experimental strategy, models, and evaluation metrics; Section 4 presents the Results and Discussion; and Section 5 provides the Final Considerations and directions for future work.

1. RELATED WORK

This section presents studies on Machine Learning, AutoML, and feature selection applied to obesity classification. The analysis enables the comparison of existing approaches and the identification of limitations that support the proposal of this study.

Prasher and Nelson (2024) investigated obesity risk prediction using *XGBoost*, data preparation, feature engineering, and model optimization, achieving 96% accuracy. Although the results demonstrate the potential of the approach, the study focuses on a single classifier and a specific population, limiting comparative analysis across models and feature selection strategies.

Chauhan *et al.* (2024) compared *XGBoost*, *LightGBM*, and *CatBoost* for obesity level classification with hyperparameter optimization. *XGBoost* achieved the best performance, reaching 91.71% accuracy after tuning with *RandomizedSearchCV*. The authors also identified weight, height, and age among the most relevant features; however, feature selection was not performed specifically for each model and validation partition.

Shen *et al.* (2024) evaluated twelve classifiers on obesity data using SVM-SMOTE, data standardization, and SHAP for prediction interpretation. *AdaBoost* achieved the best performance, with 89.22% accuracy, precision, and F1-score. Despite the feature importance

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

analysis, SHAP was mainly employed after model construction rather than as a dynamic feature selection mechanism during validation.

Lopes *et al.* (2021) compared *Random Forest* and *Support Vector Machine* using the same dataset adopted in the present study. *Random Forest* achieved superior performance, with accuracy close to 95% and an F1-score of approximately 94.9%. However, the investigation was limited to two classifiers and did not incorporate automated feature selection mechanisms.

Yeswanth, Reddy, and Joshi (2025) evaluated *Random Forest*, *Gaussian Naive Bayes*, and *Decision Tree*, incorporating feature selection based on domain knowledge, correlation analysis, and importance estimated by *Random Forest*. *Random Forest* achieved the highest accuracy, at 85.7%, followed by *Decision Tree* at 80.1% and *Gaussian Naive Bayes* at 75.4%. Although feature selection was included, the procedure was not dynamically adapted to each classifier.

Modi *et al.* (2025) broadened the comparison by including Logistic Regression, *Random Forest*, *Gradient Boosting*, *Histogram Gradient Boosting*, *LightGBM*, *XGBoost*, and *stacking*. *XGBoost* achieved the highest accuracy, at 98.42%, followed by *LightGBM* at 97.79%, while *Random Forest* reached 93.69%. Hairani *et al.* (2025), in turn, focused on tree-based models and obtained 96.31% with *Gradient Boosting*, 94.59% with *Random Forest*, and 93.19% with C4.5.

Narayanan, Das, and Mirnalinee (2023) showed that AutoML frameworks can achieve performance comparable to manually developed biomedical classifiers, although limitations remain in feature selection and dimensionality reduction. Similarly, Kulkarni *et al.* (2021) reported accuracies of 81.2% for TPOT, 82% for H2O AutoML, and 100% for a manually configured pipeline, indicating that automated approaches may still overlook domain knowledge and relevant data preparation decisions.

Previous studies have shown the effectiveness of tree-based models, feature selection methods, and AutoML approaches in classification tasks. However, the joint use of multiple classifiers with dynamic feature selection tailored to each validation split remains insufficiently explored, representing the main methodological gap addressed in this study.

Table 1. Comparison between related studies and the proposed approach

Author(s)	Techniques used	Weaknesses/limitations
Kulkarni *et al.* (2021)	TPOT, H2O AutoML, and a manually configured AdaBoost pipeline.	Diabetes-focused study; limited incorporation of domain knowledge in AutoML.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Narayanan, Das, and Mirnalinee (2023)	Five AutoML frameworks with multiple feature selection/reduction methods.	Focused on ADMET screening; limited feature selection support in some frameworks.
Prasher and Nelson (2024)	XGBoost with data preparation, feature engineering, and optimization.	Single classifier and specific older-adult population.
Chauhan *et al.* (2024)	XGBoost, LightGBM, CatBoost, feature importance, and hyperparameter tuning.	Limited to three boosting models; no dynamic feature selection.
Shen *et al.* (2024)	Twelve classifiers, SVM-SMOTE, standardization, and SHAP.	SHAP used mainly for post-training interpretation, not dynamic selection.
Yeswanth, Reddy, and Joshi (2025)	Random Forest, Naive Bayes, Decision Tree, feature selection, and grid search.	Only three classifiers; feature selection not adapted to each model.
Lopes *et al.* (2021)	Random Forest, SVM, One-Hot Encoding, and multiclass evaluation.	Only two classifiers; no AutoML or dynamic feature selection.
Modi *et al.* (2025)	Multiple ensemble and boosting models, including LightGBM and XGBoost.	No dynamic feature selection; limitations in dataset size, imbalance, and interpretability.
Hairani *et al.* (2025)	Gradient Boosting, Random Forest, and C4.5.	Only three models; no dynamic feature selection.
Present study	AutoML, multiple tree-based models, dynamic feature selection, and multiclass evaluation.	Integrates model-specific feature selection and comparison under a common protocol.

Source: Prepared by the authors.

Table 1 shows that the analyzed studies provide important advances in obesity classification but address separately several aspects that the present proposal seeks to integrate. Lopes *et al.* (2021) and Hairani *et al.* (2025) highlight the strong performance of tree-based models, while Modi *et al.* (2025) broaden this comparison through ensemble and *gradient boosting* methods. Other studies incorporate feature selection or feature analysis, such as Yeswanth, Reddy, and Joshi (2025), Shen *et al.* (2024), and Chauhan *et al.* (2024), but do not perform feature selection dynamically and specifically for each model and validation partition.

In this context, the proposed approach differs by integrating AutoML, systematic comparison of different tree-based models, and dynamic feature selection dependent on the model and experimental partitions. Although studies such as Kulkarni *et al.* (2021) and Narayanan, Das, and Mirnalinee (2023) demonstrate the potential of automation, they also point to limitations related to the incorporation of domain knowledge and feature selection. Therefore,

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

there remains room for an approach that simultaneously evaluates classifier performance, the recurrence of selected features, and the stability of results throughout the validation process.

2. THEORETICAL BACKGROUND

This section presents the theoretical foundations that support the approach adopted in this study for obesity level classification, covering concepts related to obesity, tree-based algorithms, AutoML, and feature selection.

2.1. Obesity

Obesity is a chronic and multifactorial condition associated with excessive body fat accumulation and the interaction of biological, behavioral, social, and environmental factors. Although BMI is widely used for classification, complementary measures, such as waist circumference, can improve the assessment of metabolic and cardiovascular risks. In Brazil, the increasing prevalence of obesity has accompanied changes in dietary patterns, urbanization, and levels of physical activity, with consequences for both population health and the healthcare system, reinforcing its relevance as a public health problem (FERREIRA; MAGALHÃES, 2006; ARRAIS *et al.*, 2026; SILVA *et al.*, 2026).

The contemporary understanding of obesity extends beyond explanations based exclusively on the balance between energy intake and expenditure. Factors such as income, education, access to healthy foods, environmental conditions, the structure of food systems, and access to healthcare services also influence its occurrence and outcomes. In addition, obesity is associated with several chronic conditions, including hypertension, diabetes, cardiovascular diseases, and some types of cancer, further increasing its clinical, social, and economic impacts (NILSON; OLIVEIRA E SILVA, 2026; ARRAIS *et al.*, 2026).

Given this complexity, addressing obesity requires integrated strategies for prevention, monitoring, and treatment, combined with public policies aimed at promoting healthy eating, increasing physical activity, reducing inequalities, and improving access to healthcare services. In parallel, computational techniques have been investigated as auxiliary tools for analyzing and classifying obesity levels. In this context, Machine Learning models can simultaneously consider different anthropometric, dietary, and behavioral characteristics, helping to identify relevant patterns and support assessment processes, provided that their results are properly validated in representative populations (LOPES *et al.*, 2021).

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

2.2. Tree-Based Algorithms

Tree-based algorithms constitute a family of supervised learning methods widely used in classification problems. Their structure represents the decision-making process through successive splits of the dataset, in which features are used to form increasingly homogeneous subsets with respect to the analyzed classes. Decision Trees offer advantages such as the ability to handle different types of features and the relative ease of interpreting the generated rules, although they may be sensitive to variations in the training data and prone to overfitting when their complexity is not properly controlled (XIMENES; BRITO, 2024).

One strategy for reducing the limitations of a single tree is the use of ensemble methods, which combine multiple trees to produce a collective decision. *Random Forest* is one of the main representatives of this approach, constructing multiple trees while incorporating randomization mechanisms that increase diversity among estimators. This strategy seeks to reduce model variance and improve generalization compared with the use of a single tree (BREIMAN, 2001). Such characteristics have shown relevant results in different problems involving tabular data. In the study by Ximenes and Brito (2024), for example, *Random Forest* achieved 98.62% accuracy in the classification of divorce-related data, while the Decision Tree reached 97.09%. In another application, Brito and Ximenes (2026) observed competitive performance of *Random Forest* in the classification of clinical data, indicating that tree ensemble methods can maintain predictive stability across different domains.

In addition to tree aggregation methods, boosting strategies construct classifiers sequentially, with each new estimator contributing to the reduction of errors produced by previous models. *AdaBoost* is one of the classical approaches in this category, combining low-complexity classifiers to form a more robust final model (FREUND; SCHAPIRE, 1997). Subsequent techniques, such as *XGBoost* and *CatBoost*, incorporated additional mechanisms for training optimization, regularization, and improved generalization (CHEN; GUESTRIN, 2016; PROKHORENKOVA *et al.*, 2018).

In the present study, this methodological diversity is explored through the evaluation of *Decision Tree*, *Random Forest*, *Extra Trees*, *Bagging Trees*, *AdaBoost*, *XGBoost*, *LightGBM*, and *CatBoost*, covering approaches ranging from individual trees to different aggregation and boosting mechanisms. The joint evaluation of these algorithms is particularly relevant for obesity level classification because it enables the assessment of different ways of representing relationships among anthropometric, dietary, and behavioral features. Lopes *et al.* (2021), when analyzing this type of problem, found that *Random Forest* outperformed SVM, achieving accuracy close to 95% in obesity level estimation. Therefore, comparing different tree-based algorithms

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

makes it possible to empirically identify which strategies provide the best balance among predictive performance, stability, and generalization ability for the investigated data.

2.3. AutoML and Feature Selection

Automated Machine Learning (AutoML) is an approach aimed at automating different stages of the machine learning model development process, including data preprocessing, feature and algorithm selection, and hyperparameter optimization. This approach systematizes modeling decisions and enables the automated exploration of different configurations, thereby reducing reliance on manual choices (HUTTER; KOTTHOFF; VANSCHOREN, 2019; FEURER *et al.*, 2019; HE; ZHAO; CHU, 2021).

Thus, AutoML is designed to automate key decisions involved in the development of machine learning models, including data preprocessing, feature transformation and selection, algorithm selection, and hyperparameter optimization. In this context, Hutter, Kotthoff, and Vanschoren (2019) note that automation may encompass multiple components of the machine learning workflow, particularly feature selection and transformation. This broader perspective positions AutoML beyond the mere comparison of classifiers, as it supports the automated construction of pipelines capable of adapting both data representation and predictive models to the characteristics of the problem under investigation. Feurer *et al.* (2019) further support this view by emphasizing that AutoML systems should integrate decisions concerning algorithm selection, feature preprocessing, and hyperparameter configuration.

Feature selection aims to identify the features most relevant to the predictive task by removing redundant or poorly informative variables that may increase computational complexity and training time without improving predictive performance. In AutoML workflows, excessively high-dimensional feature sets may also increase computational cost and impair model performance. In this context, statistical and feature-importance-based methods can help focus the learning process on the most representative features, reduce the risk of overfitting, and improve computational efficiency (NARAYANAN; DAS; MIRNALINEE, 2023).

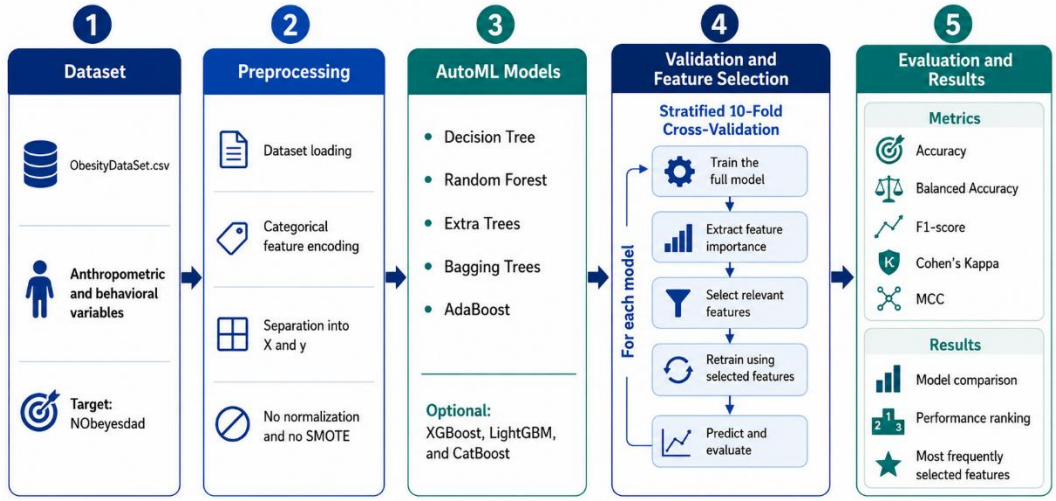
In this study, feature selection is integrated into the automated evaluation of tree-based models. In each cross-validation fold, the classifier is initially trained using all available features, after which only features with importance values above the mean feature importance calculated for the respective model are retained. The classifier is then retrained using the selected feature subset, relying on native feature importance measures or, when necessary, permutation importance. In addition, the selection frequency of each feature is recorded, enabling the analysis of feature recurrence and stability across different cross-validation folds and models.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

3. MATERIALS AND METHODS

Figure 1 summarizes the methodological workflow of the study, from dataset preparation to model comparison. Tree-based classifiers were employed within an AutoML approach and evaluated using stratified 10-fold cross-validation, with selection of the most relevant features followed by model retraining. The use of tree-based algorithms eliminates the need for feature normalization because their operation does not directly depend on feature scale. SMOTE was also not applied as the primary resampling strategy, with the aim of preserving the dataset distribution and avoiding the introduction of additional synthetic samples. Model performance was assessed using metrics such as Accuracy, Balanced Accuracy, F1-score, Cohen's Kappa, and MCC, which were selected to provide a more robust evaluation and reduce interpretative bias in scenarios involving potential class imbalance.

Figure 1. Methodological workflow



Source: Prepared by the authors.

3.1. Dataset Description

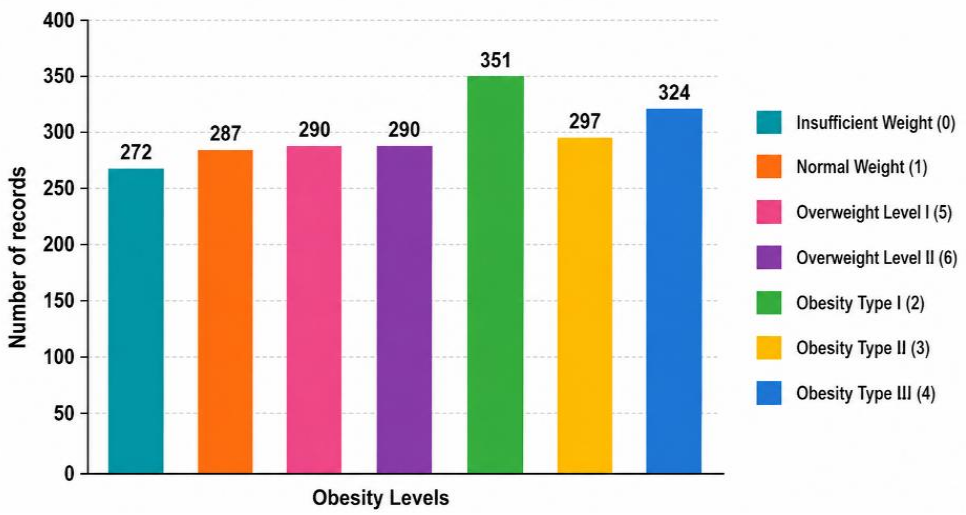
This study used the *Estimation of Obesity Levels Based on Eating Habits and Physical Condition* dataset, available in the UCI *Machine Learning Repository* and originally proposed by Palechor and Manotas (2019). The dataset contains 2,111 records, 16 predictive features, and one target variable, representing individuals from Mexico, Peru, and Colombia. The features include demographic and anthropometric information, such as gender, age, height, and weight,

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

as well as dietary habits, family history of overweight, water and alcohol consumption, physical activity, time spent using technological devices, and mode of transportation. The dataset is publicly available in the *UCI Machine Learning Repository – Obesity Dataset*.

The target variable classifies individuals into seven categories: Insufficient Weight (0), Normal Weight (1), Obesity Type I (2), Obesity Type II (3), Obesity Type III (4), Overweight Level I (5), and Overweight Level II (6), characterizing the problem as a multiclass classification task. The original dataset comprises approximately 23% of records collected directly through a web-based platform and 77% of records synthetically generated using SMOTE, a procedure performed by the dataset authors before its public release. Therefore, this balancing procedure was not carried out in the present study but represents an intrinsic characteristic of the dataset used in the experiments, as shown in Figure 2.

Figure 2. Imbalanced dataset



Source: Adapted from Lopes *et al.* (2021).

3.2. Preprocessing

Data preprocessing was defined according to the characteristics of the tree-based algorithms employed in this study. Initially, the *ObesityDataSet* dataset was imported, and categorical variables were converted into numerical representations using *LabelEncoder*. This procedure was applied to the categorical columns, including the target variable *NObeyesdad*, allowing the data to be processed in a format compatible with the classifiers. The data were then

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

organized into a feature matrix X and a target vector y, preserving the structure required for obesity level classification.

Unlike approaches that directly depend on distance measures or feature magnitude, no normalization or standardization techniques, such as *StandardScaler* or *MinMaxScaler*, were applied. This decision was based on the nature of the evaluated models, since Decision Trees and tree-based ensemble methods perform splits according to rules derived from feature values and therefore do not require feature scaling to define decision boundaries. Accordingly, the original representation of the numerical features was preserved, avoiding additional transformations that would provide no direct benefit to the considered family of classifiers.

The methodological choice was to preserve the observed data distribution and avoid introducing additional synthetic samples that could alter the composition of the analyzed dataset. To mitigate the effects of possible differences in class distribution during evaluation, stratified cross-validation and complementary performance metrics, including Balanced Accuracy, Cohen's Kappa, and MCC, were adopted. Validation was performed using stratified 10-fold cross-validation with data shuffling and a random seed of 42, ensuring that class proportions were preserved across folds and that the experimental procedure remained reproducible. Thus, preprocessing was deliberately kept concise, prioritizing only the transformations required for computational data representation while avoiding unnecessary interventions that could alter the dataset distribution before model training.

3.3. AutoML Strategy and Dynamic Feature Selection

In the present study, AutoML is applied to obesity level classification through the evaluation of different tree-based algorithms combined with feature selection strategies, enabling the comparison of predictive configurations and the identification of those that achieve the best performance according to the criteria established in the experimental protocol. The adopted strategy was structured as a lightweight and reproducible AutoML approach in which different tree-based classifiers are evaluated under the same experimental protocol. The workflow includes *Decision Tree*, *Random Forest*, *Extra Trees*, *Bagging Trees*, *AdaBoost*, and, when available, *XGBoost*, *LightGBM*, and *CatBoost*. Evaluation is performed using stratified 10-fold cross-validation with data shuffling and a random seed of 42, allowing the models to be compared under equivalent conditions.

Feature selection is performed dynamically for each model and in each cross-validation fold. First, the classifier is trained using all available features, after which their respective importance values are computed. Features with importance values above the mean are retained;

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

if no feature satisfies this criterion, the three most relevant features are preserved. Feature importance is obtained through *feature_importances_*, the specific *CatBoost* feature-importance method, or, when necessary, permutation importance. The same model type is then retrained using only the selected feature subset.

In addition to predictive performance, the procedure records how frequently each feature is selected across folds, enabling the identification of recurrent features among the evaluated models. Finally, the mean and standard deviation of the performance metrics are calculated, and the classifiers are ranked according to mean MCC, combining performance comparison with an analysis of the stability of the selected features.

3.4. Evaluated Tree-Based Models

Eight classifiers from the family of tree-based models were evaluated, covering individual approaches, bagging methods, and boosting techniques. The experimental set consisted of *Decision Tree*, *Random Forest*, *Extra Trees*, *Bagging Trees*, *AdaBoost*, *XGBoost*, *LightGBM*, and *CatBoost*, with the last three included when their respective libraries were available in the execution environment, as shown in Table 2. To ensure reproducibility, the models used *random_state*=42 whenever this parameter was available. The Decision Tree was used with its default configuration, while *Random Forest* and *Extra Trees* were configured with 200 estimators. *Bagging Trees* used a Decision Tree as the base estimator and 200 estimators, whereas *AdaBoost* retained its default parameters except for the random seed.

Among the gradient boosting methods, *XGBoost* was configured with 300 estimators, a maximum depth of 6, a learning rate of 0.1, *subsample*=0.9, and *colsample_bytree*=0.9, using a multiclass objective and *mlogloss* as the internal evaluation metric. *LightGBM* employed 400 estimators, a learning rate of 0.05, 31 leaves, and sampling ratios of 0.9 for both observations and features. *CatBoost* was configured with 400 iterations, a learning rate of 0.08, a depth of 8, and the *MultiClass* loss function. These configurations enabled the comparison of different tree construction and ensemble strategies under the same validation protocol, with the analysis focused on multiclass obesity level classification.

Classifier performance was measured using Accuracy, Balanced Accuracy, weighted Precision, weighted Recall, weighted Specificity, weighted F1-score, Cohen's Kappa, and MCC. All models were evaluated using stratified 10-fold cross-validation, allowing the metrics to be subsequently aggregated and compared consistently across the different algorithms.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Table 2. Hyperparameters of the models used in the AutoML approach

Model	Hyperparameters used
Decision Tree	*random_state* = 42; all other parameters kept at the library default values.
Random Forest	*n_estimators* = 200; *random_state* = 42; *n_jobs* = -1
Extra Trees	*n_estimators* = 200; *random_state* = 42; *n_jobs* = -1
Bagging Trees	*Base estimator*: *DecisionTreeClassifier(random_state=42)*; *n_estimators* = 200; *random_state* = 42; *n_jobs* = -1
AdaBoost	*random_state* = 42; all other parameters kept at the library default values.
XGBoost	*n_estimators* = 300; *max_depth* = 6; *learning_rate* = 0.1; *subsample* = 0.9; *colsample_bytree* = 0.9; *objective* = *multi:softmax*; *eval_metric* = *mlogloss*; *random_state* = 42; *n_jobs* = -1; *num_class* defined according to the number of classes.
LightGBM	*n_estimators* = 400; *learning_rate* = 0.05; *num_leaves* = 31; *subsample* = 0.9; *colsample_bytree* = 0.9; *random_state* = 42; *n_jobs* = -1
CatBoost	*iterations* = 400; *learning_rate* = 0.08; *depth* = 8; *loss_function* = *MultiClass*; *verbose* = *False*; *random_state* = 42

Source: Prepared by the authors based on the experimental results.

3.5. Experimental Protocol

The experimental protocol was designed to ensure equivalent evaluation conditions across all considered models. Stratified 10-fold cross-validation was employed with data shuffling and *random_state*=42. In each fold, the data were divided into training and test subsets while preserving class distribution, and all classifiers were evaluated using exactly the same partitions. This organization enabled the assessment of model behavior across different subsets of the dataset, reducing dependence on a single train-test split.

In each fold, the model was initially trained using all available features. Based on this training, feature importance values were computed, and features with importance values above the mean were selected. A new instance of the same classifier was then trained exclusively on the selected feature subset and used to generate predictions for the corresponding test set. The metrics obtained in each fold were subsequently stored for aggregation. This procedure integrates dynamic feature selection directly into the validation process, ensuring that feature selection is performed independently within each experimental fold.

Performance was measured using Accuracy, Balanced Accuracy, weighted Precision, weighted Recall, weighted Specificity, weighted F1-score, Cohen's Kappa, and MCC. After the ten folds, the mean and standard deviation of each metric were calculated for every classifier. In addition, the models were ranked according to mean MCC, while feature selection frequency was

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

recorded across folds, enabling the analysis of both predictive performance and the recurrence of selected features.

3.6. Evaluation Metrics

Classifier evaluation was performed using a set of complementary metrics to assess model performance from different perspectives. The metrics considered were Accuracy, Balanced Accuracy, weighted Precision, weighted Recall, weighted Specificity, weighted F1-score, Cohen's Kappa, and the *Matthews Correlation Coefficient* (MCC). Accuracy represents the overall proportion of correct classifications, whereas Balanced Accuracy reduces the influence of differences in class distribution. Weighted Precision, weighted Recall, and weighted F1-score aggregate class-specific results according to their respective supports, providing a more appropriate assessment for multiclass classification.

Specificity was calculated as a weighted measure across classes, reflecting the model's ability to correctly identify observations that do not belong to each evaluated class. In addition, Cohen's Kappa and MCC were used to assess agreement and overall classification quality beyond simple accuracy. In the adopted protocol, MCC was also used as the criterion for the final ranking of classifiers, based on the mean value obtained across the ten cross-validation folds.

For each model, the metrics obtained across the different folds were aggregated and reported as mean ± standard deviation. This presentation allows both average performance and variability across folds to be examined simultaneously, providing a more consistent assessment of the stability of the evaluated classifiers.

4. RESULTS AND DISCUSSION

This section discusses the results of the AutoML approach with dynamic feature selection, focusing on model performance and metric consistency. It also examines the recurrence of relevant features in relation to the methodological decisions adopted.

4.1. Comparative Model Performance

Table 3 presents the comparative performance of the evaluated models based on the main classification metrics, reported as mean ± standard deviation across the 10 cross-validation folds. This comparison makes it possible to observe differences in performance, stability, and generalization ability among the classifiers. *LightGBM* stands out by achieving the best results for most of the analyzed metrics.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Table 3. Comparison of the obtained results

Model	*Accuracy*	*Balanced Accuracy*	*Precision**	*Recall**	*Specificity**	*F1-score**	Kappa	MCC
Decision Tree	0.951±0.012	0.951±0.012	0.953±0.011	0.951±0.012	0.992±0.002	0.951±0.012	0.943±0.014	0.943±0.014
Random Forest	0.963±0.015	0.962±0.015	0.964±0.015	0.963±0.015	0.994±0.003	0.963±0.015	0.957±0.018	0.957±0.018
Extra Trees	0.971±0.008	0.970±0.008	0.972±0.008	0.971±0.008	0.995±0.001	0.971±0.008	0.966±0.010	0.966±0.010
Bagging Trees	0.964±0.012	0.964±0.012	0.966±0.011	0.964±0.012	0.994±0.002	0.964±0.012	0.958±0.014	0.958±0.014
AdaBoost	0.467±0.051	0.476±0.051	0.393±0.083	0.467±0.051	0.911±0.008	0.385±0.066	0.379±0.059	0.404±0.057
XGBoost	0.817±0.024	0.815±0.025	0.820±0.024	0.817±0.024	0.969±0.004	0.816±0.024	0.786±0.029	0.787±0.028
LightGBM	**0.977±0.008**	**0.977±0.008**	**0.978±0.007**	**0.977±0.008**	**0.996±0.001**	**0.977±0.008**	**0.973±0.009**	**0.974±0.009**
CatBoost	0.973±0.009	0.973±0.008	0.974±0.008	0.973±0.009	0.996±0.002	0.973±0.009	0.969±0.010	0.969±0.010

* *Precision*, *Recall*, *Specificity*, and F1-score were calculated using weighted averaging. Values are reported as mean ± standard deviation across the 10 cross-validation folds.
Source: Prepared by the authors based on the experimental results.

The results presented in Table 3 demonstrate high performance for most of the evaluated models, with *LightGBM* standing out by achieving the highest mean values for the main metrics, including Accuracy, Balanced Accuracy, F1-score, Cohen's Kappa, and MCC. *CatBoost* and *Extra Trees* also achieved comparable results, indicating strong classification performance. In contrast, *AdaBoost* showed the lowest performance among the analyzed models, while *XGBoost* achieved intermediate results. Overall, the findings indicate greater consistency for *LightGBM*, *CatBoost*, and *Extra Trees* under the experimental conditions considered.

4.2. Dynamic Feature Selection Analysis

Table 4 summarizes the behavior of dynamic feature selection across the 10 cross-validation folds, revealing that some variables were consistently retained by the evaluated classifiers. *Weight* was the most stable attribute, appearing in 100% of the folds for all models, which suggests a strong and persistent contribution to the classification task. *Height* also showed substantial stability, being selected in every fold by several algorithms, whereas *Gender*, *Age*, and FCVC (*Frequency of Consumption of Vegetables*) were frequently retained across different models. A broader selection profile was observed for *LightGBM*, which also included FAF (*Physical Activity Frequency*) in 80% of the folds. Overall, the results show that the selected subsets vary according to the classifier and validation partition, while a core group of anthropometric, dietary, and physical activity-related variables remains consistently relevant. This recurring pattern suggests that these attributes provide robust discriminatory information and may contribute to the stability of the predictive models. It also highlights the value of performing feature

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

selection dynamically, since the importance of individual variables may vary according to both the learning algorithm and the data partition used during validation.

Table 4. Main results of dynamic feature selection

Feature	Models with the highest selection frequency	Frequency (folds)
Weight	Decision Tree, Random Forest, Extra Trees, Bagging Trees, AdaBoost, XGBoost, LightGBM, and CatBoost	100%
Height	Decision Tree, Random Forest, Extra Trees, Bagging Trees, LightGBM, and CatBoost	100%
Gender	Decision Tree, Extra Trees, Bagging Trees, XGBoost, and CatBoost	100%
Age	Random Forest, Extra Trees, and LightGBM	100%
FCVC	Random Forest, Extra Trees, and XGBoost	100%
FAF	*LightGBM*	80%

Source: Prepared by the authors based on the experimental results.

4.3. Critical Analysis of the Results

Table 5 summarizes the main findings by comparing model performance, stability across folds, selected features, and key methodological choices. This synthesis facilitates a more objective interpretation of the classifiers and their relative strengths and limitations.

Table 5. Critical analysis of the results

Aspect	Evidence	Summary analysis
Best performance	*LightGBM* had the highest MCC (0.974).	Best overall performance among the models.
Stability	*LightGBM* showed low standard deviations.	Results were consistent across folds.
Worst performance	*AdaBoost* and *XGBoost* had the lowest MCC values.	Lower performance under the evaluated conditions.
Feature selection	*Weight*, *Height*, *Age*, *Gender*, and FCVC were frequent.	These features showed high predictive relevance.
No normalization	All models were tree-based.	Normalization was unnecessary.
No SMOTE	Stratified validation was adopted.	Additional synthetic samples were avoided.
Evaluation	Multiple performance metrics were used.	This enabled a broader assessment.
Limitation	Hyperparameters were predefined.	Future work may include optimization.

Source: Prepared by the authors based on the experimental results.

A critical analysis of the results, as presented in Table 5, shows that *LightGBM* achieved the best overall performance, followed closely by *CatBoost* and *Extra Trees*, with good stability

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

across the cross-validation folds. In contrast, *XGBoost* and, particularly, *AdaBoost* achieved lower performance under the evaluated conditions. The recurrent selection of features such as *Weight*, *Height*, *Age*, *Gender*, and FCVC reinforces the relevance of anthropometric and behavioral characteristics for classification. Furthermore, the decisions not to normalize the data or apply additional SMOTE were consistent with the nature of the models and with the strategy of avoiding the introduction of additional synthetic samples while preserving the observed dataset distribution. The combined use of metrics such as Balanced Accuracy, F1-score, Cohen's Kappa, and MCC also contributed to a more robust assessment of classifier performance.

5. FINAL CONSIDERATIONS AND FUTURE WORK

This study evaluated an AutoML (*Automated Machine Learning*) approach based on different tree-based classification algorithms combined with dynamic feature selection for obesity level classification. The experimental protocol employed stratified 10-fold cross-validation, without normalization and without applying additional SMOTE (*Synthetic Minority Over-sampling Technique*) as a primary resampling strategy, with the aim of avoiding further synthetic sample generation and preserving the observed dataset distribution.

Among the evaluated classifiers, *LightGBM* (*Light Gradient Boosting Machine*) achieved the best overall performance, with an Accuracy of 0.977 ± 0.008, Balanced Accuracy of 0.977 ± 0.008, F1-score of 0.977 ± 0.008, Cohen's Kappa of 0.973 ± 0.009, and an MCC (*Matthews Correlation Coefficient*) of 0.974 ± 0.009. It was followed by *CatBoost*, with an MCC of 0.969 ± 0.010, and *Extra Trees*, with an MCC of 0.966 ± 0.010, both showing high and consistent performance across the cross-validation folds.

Dynamic feature selection showed that features such as *Weight*, *Height*, *Age*, *Gender*, and FCVC (*Frequency of Consumption of Vegetables*) were recurrent across different models. In *LightGBM*, FAF (*Physical Activity Frequency*) also stood out, being selected in 80% of the folds. These findings indicate that anthropometric and behavioral characteristics contributed meaningfully to the classification process, reinforcing the usefulness of dynamic feature selection for identifying the most informative features.

The main limitations of this study include the use of predefined hyperparameters, the evaluation of a single dataset, and the composition of the dataset itself, in which approximately 77% of the records were synthetically generated using SMOTE by the original dataset authors. Although these synthetic records were already part of the dataset used in the present study, their high proportion may limit the representation of variability observed in real populations and may

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

affect model generalizability and the external validity of the findings. Therefore, the results should be interpreted in light of this characteristic of the dataset.

6. DECLARATION OF THE USE OF ARTIFICIAL INTELLIGENCE

The authors used ChatGPT and Grammarly as auxiliary tools for linguistic review and correction of passages translated from Portuguese into English. Some images were also generated with the assistance of Artificial Intelligence tools, including ChatGPT. All content produced or revised using these tools was reviewed and validated by the authors, who remain fully responsible for the scientific content of the manuscript.

REFERENCES

ARRAIS, Joel Freires de Alencar *et al.* Obesidade no Brasil: análise das hospitalizações e mortalidade entre 2013 a 2024. Saúde e Pesquisa, v. 19, e13405, 2026.

BAG, Harika Gozde Gozukara *et al.* Estimation of obesity levels through the proposed predictive approach based on physical activity and nutritional habits. Diagnostics, Basel, v. 13, n. 18, art. 2949, 2023.

BREIMAN, Leo. Random forests. Machine Learning, v. 45, p. 5-32, 2001.

BRITO, Rhyan Ximenes de; XIMENES, Janaide Nogueira de Sousa. Análise comparativa de modelos estatísticos e redes neurais na detecção de cistite a partir de dados clínicos. Revista de Ciência da Computação, v. 8, n. 1, e17258, 2026.

CHAUHAN, Shanvi; GILL, Kanwarpartap Singh; CHAUHAN, Rahul; POKHARIYA, Hemant Singh. Targeted insights: advanced gradient boosting models for accurate obesity risk evaluation. In: INTERNATIONAL CONFERENCE ON PIONEERING DEVELOPMENTS IN COMPUTER SCIENCE & DIGITAL TECHNOLOGIES (IC2SDT), 1., 2024. Proceedings [...]. IEEE, 2024.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: a scalable tree boosting system. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016. Proceedings [...]. New York: ACM, 2016. p. 785-794.

DE-LA-HOZ-CORREA, Eduardo *et al.* Obesity level estimation software based on decision trees. Journal of Computer Science, v. 15, n. 1, p. 67-77, 2019.

FERREIRA, Vanessa Alves; MAGALHÃES, Rosana. Obesidade no Brasil: tendências atuais. Revista Portuguesa de Saúde Pública, v. 24, n. 2, p. 71-81, jul./dez. 2006.

FEURER, Matthias *et al.* Auto-sklearn: efficient and robust automated machine learning. In: HUTTER, Frank; KOTTHOFF, Lars; VANSCHOREN, Joaquin (ed.). Automated machine learning: methods, systems, challenges. Cham: Springer, 2019. p. 113-134.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

FREUND, Yoav; SCHAPIRE, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, v. 55, n. 1, p. 119-139, 1997.

GÖRMEZ, Yasin *et al.* Prediction of obesity levels based on physical activity and eating habits with a machine learning model integrated with explainable artificial intelligence. Frontiers in Physiology, v. 16, art. 1549306, 2025.

HAIRANI, Hairani; PRATAMA, Gede Yogi; WIDIYANINGTYAS, Triyanna; HIDJAH, Khasnur; PRASETYA, Didik Dwi; SORAYA, Siti. Multi-class classification of obesity levels using Gradient Boosting, Random Forest, and C4.5. In: INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS AND INFORMATION ENGINEERING (ICEEIE), 9., 2025. Proceedings [...]. IEEE, 2025.

HE, Xin; ZHAO, Kaiyong; CHU, Xiaowen. AutoML: a survey of the state-of-the-art. Knowledge-Based Systems, v. 212, art. 106622, 2021.

HUTTER, Frank; KOTTHOFF, Lars; VANSCHOREN, Joaquin (ed.). Automated machine learning: methods, systems, challenges. Cham: Springer, 2019.

KULKARNI, Gururaj N.; AMBESANGE, Sateesh; VIJAYALAXMI, A.; SAHOO, Anindita. Comparision of diabetic prediction AutoML model with customized model. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND SMART SYSTEMS (ICAIS), 2021. Proceedings [...]. IEEE, 2021.

LOPES, Leonardo Ferreira; OLIVEIRA, Adonias Caetano de; BRITO, Rhyan Ximenes de; OLIVEIRA, Saulo Anderson Freitas de; RAPOSO NETO, Luiz Torres. Estimativa dos níveis de obesidade com base em hábitos alimentares e condição física através de técnicas de aprendizado de máquina. In: CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES (SIBGRAPI), 34., 2021. Anais Estendidos da Conference on Graphics, Patterns and Images. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 154-157.

MODI, Vishwam; SHAH, Aarchi; PANCHAL, Janmay; KUMAR, Yogesh. Machine learning-based early prediction and classification of obesity levels using ensemble models. In: INTERNATIONAL CONFERENCE ON INDUSTRY 4.0 TECHNOLOGY (I4TECH), 3., 2025. Proceedings [...]. IEEE, 2025.

MOSLEMI, Amir. A tutorial-based survey on feature selection: recent advancements on feature selection. Engineering Applications of Artificial Intelligence, v. 126, art. 107136, 2023.

NARAYANAN, Advaith Nila; DAS, Shyam Sundar; MIRNALINEE, T. T. Evaluation of AutoML frameworks for computational ADMET screening in drug discovery & development. In: IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM), 2023. Proceedings [...]. IEEE, 2023. p. 4929-4931.

NCD RISK FACTOR COLLABORATION (NCD-RISC). Worldwide trends in underweight and obesity from 1990 to 2022: a pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. The Lancet, v. 403, n. 10431, p. 1027-1050, 2024.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

NILSON, Eduardo; OLIVEIRA E SILVA, Denise. Dia Mundial da Obesidade 2026: Entre o Legado de 2025 e a Urgência de 8 Bilhões de Razões para Agir. Cadernos Saúde Global Fiocruz, n. 2, p. 100-111, 2026.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Obesity and overweight. Geneva: World Health Organization, 2025. Disponível em: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Acesso em: 30 ago. 2026.

PALECHOR, Fabio Mendoza; MANOTAS, Alexis de la Hoz. Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in Brief, v. 25, art. 104344, 2019.

PRASHER, Shikha; NELSON, Leema. Early prediction of obesity risk in older adults using XGBoost classifier. In: INTERNATIONAL CONFERENCE ON CIRCUIT POWER AND COMPUTING TECHNOLOGIES (ICCPCT), 7., 2024. Proceedings [...]. IEEE, 2024.

PROKHORENKOVA, Liudmila *et al.* CatBoost: unbiased boosting with categorical features. In: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 32., 2018. Proceedings [...]. 2018. p. 6639-6649.

SHEN, Jiexuan; LI, Shi; XINGLIN, Li; WANG, Yong; MONDAY, Happy Nkanta; NNEJI, Grace Ugochi. Explainable AI for the prediction and estimation of obesity levels using machine learning models. In: INTERNATIONAL CONFERENCE ON ELECTRONIC INFORMATION ENGINEERING AND COMPUTER SCIENCE (EIECS), 4., 2024. Proceedings [...]. IEEE, 2024.

SILVA, Ramon Reis *et al.* Obesidade abdominal em adultos de um município do estado da Bahia: um estudo transversal. Revista Brasileira de Medicina de Família e Comunidade, Rio de Janeiro, v. 21, n. 48, e4276, 2026.

THENG, Dipti; BHOYAR, Kishor K. Feature selection techniques for machine learning: a survey of more than two decades of research. Knowledge and Information Systems, v. 66, n. 3, p. 1575-1637, 2024.

XIMENES, Janaide Nogueira de Sousa; BRITO, Rhyan Ximenes de. Análise de dados sob a perspectiva da classificação: desvendando insights sobre divórcio através da mineração de dados. Revista de Sistemas e Computação, Salvador, v. 14, n. 3, p. 1-8, set./dez. 2024.

YESWANTH, Adapa Mohan Sai; REDDY, Vunadi Sharan Teja; JOSHI, Soshya. Early obesity risk assessment using machine learning to analyse lifestyle and demographic data. In: IEEE INTERNATIONAL CONFERENCE ON INTERDISCIPLINARY APPROACHES IN TECHNOLOGY AND MANAGEMENT FOR SOCIAL INNOVATION (IATMSI), 2025. Proceedings [...]. IEEE, 2025.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.