A Data-Driven Approach for Obesity Classification using Machine Learning

Nima Khodadadi[*1], Mohamed Saber[2], Mostafa Abotaleb[3]

[1]Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA
[2] Electronics and Communications Engineering Department, Faculty of Engineering, Delta University for Science and Technology, Gamasa City 11152, Egypt
[3]Department of System Programming, South Ural State University, 454080 Chelyabinsk, Russia
Emails: nima.khodadadi@miami.edu; mohamed.saber@deltauniv.edu.eg; abotalebmostafa@bk.ru

Abstract

Obesity is a global health concern with significant impacts on individuals and society. Accurate and timely classification of obesity levels can help in the development of personalized interventions and targeted healthcare strategies. In this paper, we propose a data-driven approach for obesity classification utilizing machine learning techniques. Our study leverages a comprehensive dataset consisting of anthropometric measurements, lifestyle factors, and demographic information of a large cohort of individuals. We explore the effectiveness of various machine learning algorithms, including decision trees, support vector machines, and neural networks, for obesity classification. Feature selection and preprocessing techniques are employed to enhance the performance of the models. Through extensive experimentation and cross-validation, we evaluate the predictive accuracy, sensitivity, specificity, and overall performance of the developed classifiers. Our results demonstrate the efficacy of our data-driven approach, achieving high accuracy in obesity classification. Furthermore, we conduct a comparative analysis of the different algorithms to identify the most suitable model for this task. The proposed framework has the potential to assist healthcare professionals in identifying and classifying obesity levels accurately, contributing to the development of personalized interventions and improving public health outcomes.

Keywords: Machine Learning; Obesity Classification; Artificial Intelligence (AI); Data-Driven Intelligence.

1. Introduction

Obesity, a complex and multifaceted health issue, has emerged as a global epidemic, posing significant challenges to individuals and societies alike. With its detrimental impact on physical health, mental well-being, and overall quality of life, obesity has become a major concern in public health. The World Health Organization (WHO) estimates that over 650 million adults worldwide are affected by obesity, and this number continues to rise at an alarming rate. Obesity is associated with a higher risk of developing various chronic conditions, including cardiovascular disease, type 2 diabetes, certain cancers, and musculoskeletal disorders [1]. Moreover, it places a substantial economic burden on healthcare systems, leading to increased healthcare costs and reduced productivity. Accurate and timely classification of obesity levels is crucial for effective interventions, personalized healthcare strategies, and targeted public health initiatives. However, traditional approaches to obesity classification based solely on body mass index (BMI) or simple anthropometric measurements often lack the precision and granularity needed to capture the complexity of this condition. This is where the potential of ML techniques comes into play [2].

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

Previous approaches to obesity classification have predominantly relied on traditional methods that often oversimplify the complexity of this multifaceted condition. The widely used body mass index (BMI) has been a common metric for classifying obesity levels, but it fails to capture variations in body composition, distribution of fat, and other important factors [3]. While BMI provides a useful initial screening tool, it lacks the precision and granularity required to accurately classify obesity and its associated health risks. Other traditional approaches based on simple anthropometric measurements, such as waist circumference or waist-to-hip ratio, offer slightly improved insights but still have limitations in terms of their discriminatory power [4]. These approaches often categorize individuals into broad weight categories, ignoring the nuanced differences in obesity subtypes and failing to account for individual variations in factors such as muscle mass, metabolic health, and lifestyle choices. As a result, there is a growing recognition of the need for more sophisticated and data-driven approaches that leverage machine learning (ML) techniques to enable accurate and personalized obesity classification. By exploring the potential of ML algorithms and incorporating comprehensive datasets comprising various biological, lifestyle, and demographic factors, we can overcome the limitations of previous approaches and unlock new insights into obesity classification, paving the way for more targeted interventions and improved public health outcomes [5].

The research questions and hypotheses guiding this study revolve around the development and evaluation of a data-driven approach for obesity classification using ML techniques. Our primary research question is: Can ML algorithms effectively classify obesity levels based on comprehensive datasets comprising anthropometric measurements, lifestyle factors, and demographic information? To address this question, we hypothesize that the incorporation of diverse features and the utilization of advanced ML algorithms will significantly improve the accuracy and precision of obesity classification compared to traditional approaches. Additionally, we aim to investigate which ML algorithm(s) exhibit the highest performance in classifying obesity levels and whether feature selection and preprocessing techniques can further enhance the models' predictive capabilities. We hypothesize that a combination of carefully selected features and sophisticated ML algorithms will result in improved accuracy, sensitivity, and specificity in obesity classification [4-6].

The primary objective of this study is to propose a data-driven approach for obesity classification using ML techniques. We aim to harness the power of ML algorithms and advanced analytical methodologies to develop a more accurate and robust classification system [7]. Specifically, our objectives include exploring the effectiveness of different ML algorithms, such as decision trees, support vector machines, and neural networks, in accurately classifying obesity levels. Additionally, we aim to evaluate the impact of incorporating diverse features, including anthropometric measurements, lifestyle factors, and demographic information, on the performance of the classifiers. Through comprehensive experimentation and cross-validation, we seek to assess the predictive accuracy, sensitivity, specificity, and overall performance of the developed models [8].

The left part of the paper is organized as follows. In section 2, previous approaches and methodologies for obesity classification are reviewed and discussed. Section 3 presents the details of the proposed approach. Section 4 explains the experimental design and procedures. Section 5 presents the findings of the experiments. Section 6 interprets and analyzes the results in the context of the research objectives and hypotheses. Finally, section 7 summarizes the key findings of the study, reiterates the significance of the proposed data-driven approach, and discusses its implications for obesity classification and personalized interventions.

2. Related Work

This section provides an extensive overview of existing approaches and methodologies for obesity classification, highlighting their strengths, limitations, and gaps in the current literature. The authors of [2] provided a comprehensive overview of ML techniques for identifying risk factors associated with obesity and overweight. They examined various ML algorithms and their effectiveness in identifying and analyzing key risk factors contributing to obesity. They shed light on the potential of ML to uncover hidden patterns and relationships within complex datasets, offering insights into the multifactorial nature of obesity. However, a limitation of the study was the reliance on existing datasets, which might not encompass the full spectrum of risk factors or specific populations.

The authors of [4] made a significant contribution to the field of childhood obesity prevention and treatment. The study conducted a systematic review of the literature to explore the applications of computerized decision support systems and ML in addressing childhood obesity. By analyzing a wide

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

range of studies, they provided insights into the potential of these technologies in supporting healthcare professionals, parents, and children themselves in making informed decisions regarding diet, physical activity, and behavior modification. Nevertheless, it is important to note that their findings are limited to the existing literature and the availability of relevant studies.

Further research and validation of these applications are warranted to ensure their effectiveness and scalability in real-world settings. The authors of [5] studied the development of a ML -based approach for early prediction of overweight and obesity risk in young individuals. They aimed to leverage ML algorithms to analyze various factors and indicators, such as demographic information, lifestyle habits, and physiological measurements, to identify individuals at high risk of developing overweight or obesity. However, it is notable that further research and validation are required to ensure the generalizability and effectiveness of the proposed approach across diverse populations and settings.

The authors of [7] presented a novel application of ML techniques in estimating obesity levels using human plasma lipidomics data. They utilized a large population cohort and employed ML algorithms to analyze and extract meaningful patterns from the complex lipidomic profiles. By training the model on lipidomic data and corresponding obesity measurements, they developed a predictive model capable of accurately estimating obesity levels based on plasma lipidome profiles. However, further validation and external validation on diverse populations are necessary to ensure the reliability and generalizability of the proposed approach.

The authors of [8] investigated the complex relationship between serum free fatty acids and fecal microbiota in the context of obesity. They utilized a ML algorithm to analyze and identify potential interactions and associations between these two factors. By employing a comprehensive dataset comprising serum free fatty acid profiles and fecal microbiota composition, they studied the role of gut microbiota in obesity and its potential influence on lipid metabolism. However, it is important to acknowledge that further research and validation are necessary to confirm and generalize these findings in larger and more diverse populations.

he authors of [9] focused on the application of computational intelligence techniques for estimating obesity levels. The study explores the potential of computational intelligence algorithms, such as neural networks, genetic algorithms, and fuzzy systems, in predicting and classifying obesity levels based on various input features. However, they need to further research is needed to validate and optimize these models on larger and more diverse datasets, as well as to address potential limitations and biases.

The authors of [10] presented a novel approach combining 3D microscopy and deep learning techniques to investigate the heterogeneity of crown-like structures (CLS) within intact adipose tissue. They employed advanced imaging techniques and deep learning algorithms to analyze and classify CLS microenvironments based on their spatial distribution and cellular composition. By integrating these technologies, they gained a comprehensive understanding of the complex and heterogeneous nature of CLS, which was not possible using traditional 2D imaging methods.

3. Methodology

The methodology section provides a detailed description of the approach and techniques employed in this study to accomplish the objectives of obesity classification using ML. This section outlines the step-by-step procedures involved in data collection, preprocessing, feature selection, model development, and evaluation. Additionally, it highlights the rationale behind the chosen methodologies and their alignment with the research goals. By elucidating the methodology, this section ensures transparency, replicability, and a solid foundation for drawing valid conclusions from the study's findings.

In the proposed approach, the Synthetic Minority Over-sampling Technique (SMOTE) was employed to address the issue of imbalanced data in the obesity classification task [9-14]. Imbalanced data occurs when the distribution of the target variable, in this case, obesity classes (Normal Weight, Overweight, Obese, and Underweight), is significantly skewed towards one or more classes. This imbalance can potentially lead to biased models that favor the majority class, resulting in reduced performance for minority classes. To mitigate this challenge, SMOTE was utilized to oversample the minority classes and balance the dataset. SMOTE works by creating synthetic samples in the feature space of the minority class instances, generating new synthetic data points based on the characteristics of existing instances. By introducing these synthetic samples, SMOTE effectively increases the representation of

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

the minority classes, thereby addressing the class imbalance issue. The proposed algorithm is shown in figure 1.

Algorithm 1: SMOTE algorithm
Input: X_{minor}, $N_{percent}$, K
Output: X_{smote}

```
X_smote ← { }
for i ← 1 to Len (X_minor) do
n_n ← KNNs (X_i, X_minor, k)
p ← [ N_percent /100]
Loop p! = 0 do
X_neighbour ← random (n_n)
X_smote ← Xi + rand(0,1) * |X_neighbour − X_i|
p ← p − 1
Terminate Loop
Terminate for
return ← X_smote
```

Figure 1 : proposed algorithm of obesity classification

The application of SMOTE resampling technique ensures that the ML models are exposed to a more balanced representation of the obesity classes during training, enabling them to learn from both majority and minority classes more effectively. This enhances the models' ability to accurately classify individuals into their respective obesity categories, improving overall predictive performance.

In the proposed approach, the eXtreme Gradient Boosting (XGBoost) algorithm was employed as the classification model to classify the rebalanced obesity data. XGBoost is a powerful ML algorithm that belongs to the ensemble learning family, specifically the gradient boosting framework. It utilizes a combination of weak learners, typically decision trees, to create a strong predictive model. Mathematically, XGBoost builds an ensemble of decision trees by iteratively optimizing an objective function, which is a sum of a loss function and a regularization term. Let's denote the training dataset as D, consisting of n samples and m features. The rebalanced dataset obtained through SMOTE will be denoted as $D_{resampled}$. The XGBoost model aims to find an optimal set of weak classifiers or decision trees, denoted as h(x, θ), where x represents the feature vector and θ denotes the model parameters. The objective function to be minimized can be written as

$$Obj^{(t)} = \sum_{i=1}^{n}\left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2}h_i f_t{}^{2}(x_i)\right] + \Omega(f_t) + cons$$

$$= \sum_{i=1}^{n}[g_i \omega_q(x_i) + \frac{1}{2}h_i \omega_q{}^{2}(x_i)] + \gamma T + \lambda \frac{1}{2}\sum_{j=1}^{T} h_i \omega_j{}^{2} + C \quad (1)$$

where $l(y_i, \hat{y}_i^{(t-1)})$ represents the loss function, measuring the discrepancy between the true label y_i and the predicted label $\hat{y}_i^{(t-1)}$, and $\Omega(h)$ denotes the regularization term, preventing overfitting and controlling the complexity of the model. During training, XGBoost employs gradient descent optimization to update the model parameters θ in each iteration. It calculates the gradients of the objective function with respect to the model outputs and uses them to update the weak classifiers accordingly. Moreover, the proposed approach utilizes the greedy algorithm to systematically enumerate various tree structures and identify the optimal splitting node. This process aims to maximize the gain of the objective function through effective data partitioning, as follows:

$$Gain = \frac{1}{2}[\frac{G_L^2}{H_L+\lambda} + \frac{G_L^2}{H_L+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}] - \gamma \quad (2)$$

By utilizing XGBoost for classifying the rebalanced obesity data, the proposed approach takes advantage of the algorithm's ability to handle complex relationships and capture non-linear patterns

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

in the data. XGBoost's boosting mechanism, combined with the rebalanced dataset, allows for improved classification performance by leveraging the strengths of multiple weak classifiers.

4. Experimental Setup

In this section, we provide an overview of the experimental setup employed to investigate the effectiveness of our proposed ML approach for obesity classification. We describe the dataset used, feature selection and preprocessing techniques, the choice of ML algorithms, model evaluation metrics, and the validation strategy employed. The dataset utilized in this study comprises comprehensive information regarding the obesity classification of individuals. The data was collected from diverse sources, including medical records, surveys, and self-reported data. The dataset consists of several columns that provide valuable insights into the characteristics of each individual. These columns include an ID, which serves as a unique identifier for each individual, as well as their age, gender, height in centimeters, weight in kilograms, and body mass index (BMI) calculated as weight divided by height squared. The dataset also includes a crucial column labeled "Label" that represents the obesity classification of each individual, categorized into four distinct classes: Normal Weight, Overweight, Obese, and Underweight. This dataset forms the foundation for the analysis and exploration of obesity classification using ML techniques in this study, enabling researchers to investigate the relationships between various factors and obesity levels accurately. Figure 2 Show the class distribution in our dataset in form of pie chart.

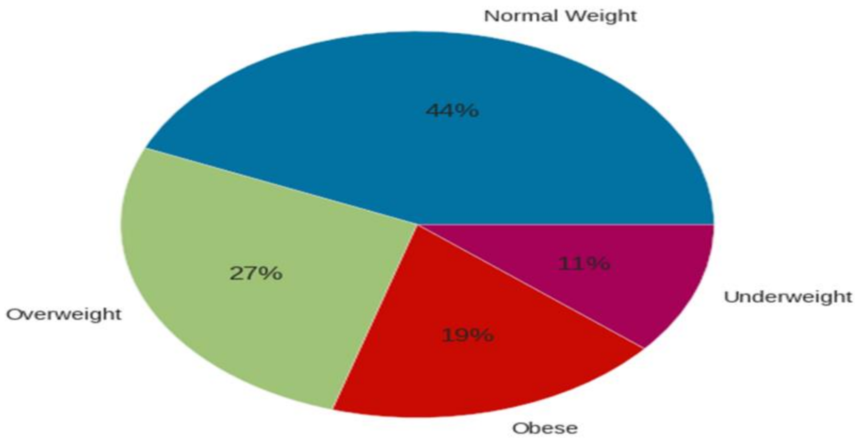

Figure 2: Class distribution in obesity classification dataset

In this study, thorough data preprocessing was performed to ensure the quality and reliability of the dataset. One important step in this process was checking for null values or missing data. It is crucial to identify and handle any missing values appropriately to avoid potential biases or inaccuracies in the subsequent analysis (See Figure 3).

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

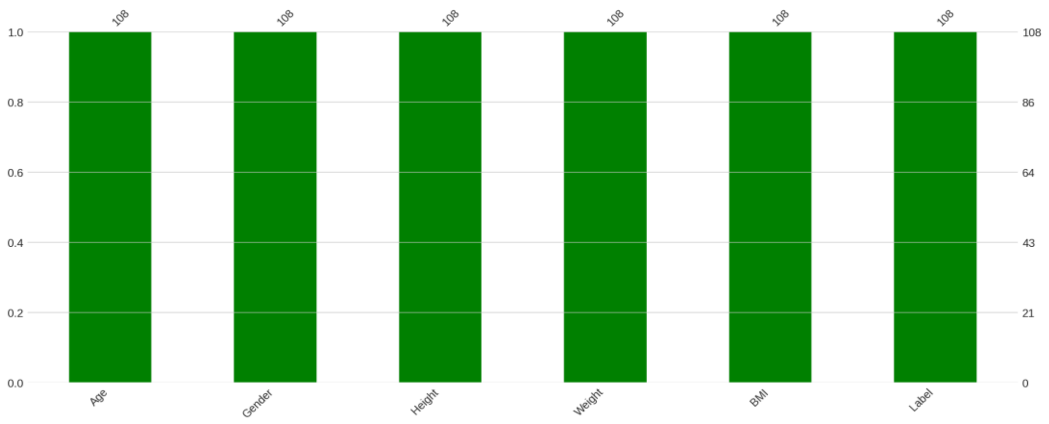

Figure 3: Visualization of null count plot for obesity classification dataset

Upon examining the dataset, it was found that there were no missing values or null entries across the variables of interest, including age, gender, height, weight, BMI, and obesity classification. This indicates that the dataset was complete and free from any substantial data gaps, which enhances the reliability of the subsequent analyses and results.

5. Results

The results section begins by presenting a pair plot among the features of the dataset to visually explore their relationships and identify potential patterns or trends. The pair plot provides a comprehensive overview of the pairwise interactions between variables, allowing for a preliminary assessment of their correlations and distributions. As shown in figure 4, the pair plot displays scatter plots along the diagonal and scatterplots between each pair of variables on the off-diagonal. Each scatter plot represents the relationship between two variables, with the x-axis corresponding to one variable and the y-axis corresponding to another. The scatter plots are accompanied by histograms along the diagonal, illustrating the distribution of each individual variable.

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

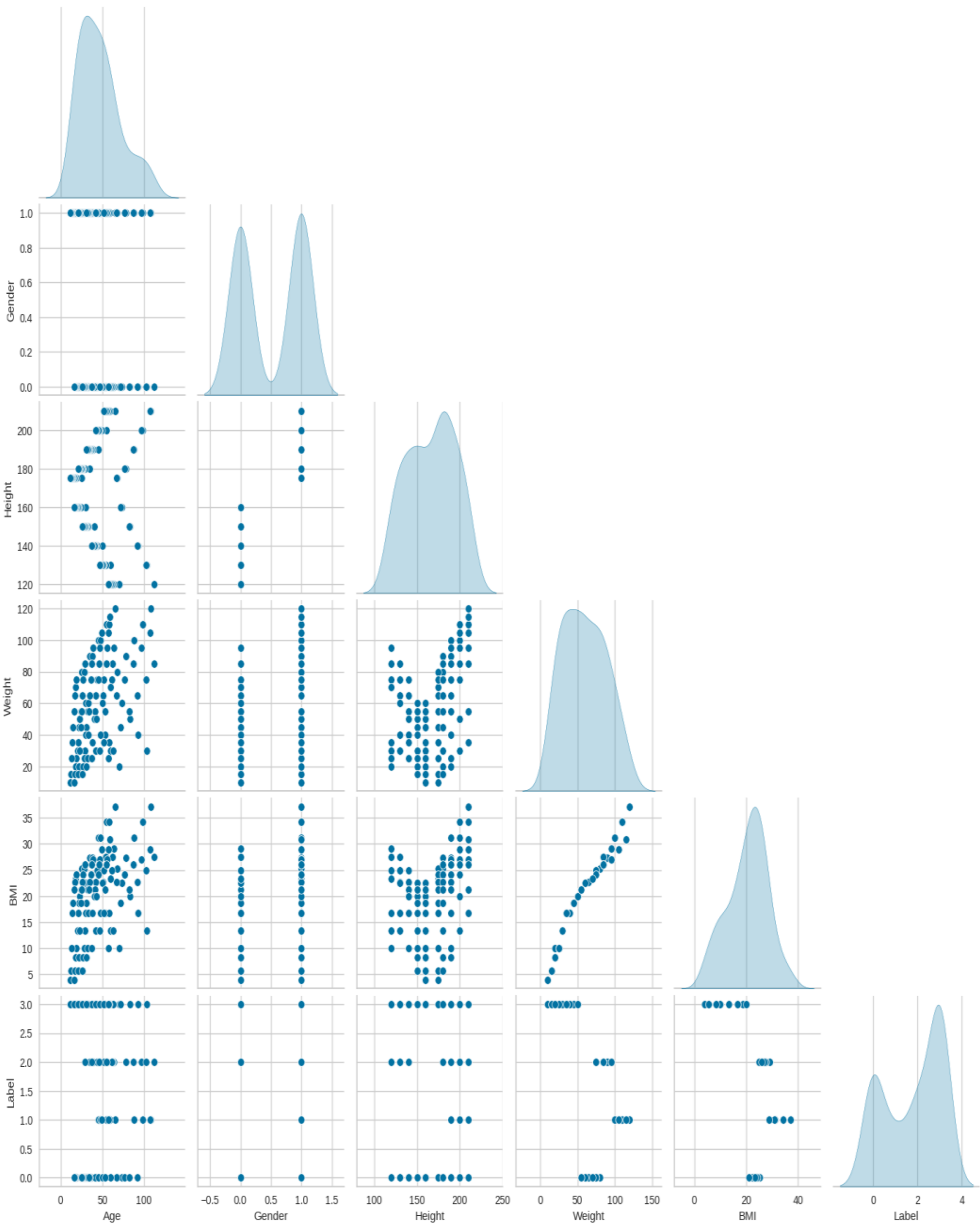

Figure 4: visualization of pair plot for obesity classification dataset

In addition, a comprehensive statistical description of the dataset was conducted to provide valuable insights into its characteristics. Table 1 presents a summary of key statistical measures, including mean, standard deviation, minimum, maximum, and quartile values for relevant variables such as age, gender, height, weight, BMI, and obesity classification.

Table 1: summary of statistical information of obesity classification dataset

	Count	Mean	Std	Min	25%	50%	75%	Max
ID	108	56.0463	31.91794	1	28.75	56.5	83.25	110
Age	108	46.55556	24.72062	11	27	42.5	59.25	112
Hight	108	166.5741	27.8763	120	140	175	190	210

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

Weight	108	59.49074	28.85623	10	35	55	85	120
BMI	108	20.54907	7.583818	3.9	16.7	21.2	26.1	37.2

Moreover, a table presenting the Pearson correlation coefficients for the variables in the dataset was generated to examine the relationships between different attributes. Table 2 displays the correlation values between variables such as age, gender, height, weight, BMI, and obesity classification. The correlation coefficients range from -1 to 1, with values closer to -1 indicating a strong negative correlation, values closer to 1 indicating a strong positive correlation, and values close to 0 indicating a weak or no correlation.

Table 2: summary of Pearson correlation on features of obesity classification dataset

	ID	Age	Gender	Height	Weight	BMI	Label
ID	1	-0.29826	-0.0056	-0.00822	-0.57263	-0.61524	0.347199
Age	-0.29826	1	-0.09196	-0.0769	0.465106	0.474185	-0.1344
Gender	-0.0056	-0.09196	1	0.876225	0.418415	0.342342	-0.28165
Height	-0.00822	-0.0769	0.876225	1	0.42889	0.35434	-0.23768
Weight	-0.57263	0.465106	0.418415	0.42889	1	0.972829	-0.56556
BMI	-0.61524	0.474185	0.342342	0.35434	0.972829	1	-0.58924
Label	0.347199	-0.1344	-0.28165	-0.23768	-0.56556	-0.58924	1

Furthermore, a table comparing the performance of different ML algorithms in the obesity classification task was conducted to evaluate their effectiveness. Table 3 presents key performance metrics, such as accuracy, precision, recall, and F1-score, for each algorithm considered in the study. The algorithms are listed along with their corresponding metrics to facilitate a comprehensive comparison.

Table 3: Comparison between the performance of the ML algorithms on obesity classification dataset

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Extreme Gradient Boosting	0.9818	0.9943	0.9818	0.9864	0.9813	0.9756	0.9775	0.288
CatBoost Classifier	0.9818	1	0.9818	0.9864	0.9807	0.9754	0.9774	1.665
Random Forest Classifier	0.9727	1	0.9727	0.9795	0.9707	0.9631	0.9661	0.561
Gradient Boosting Classifier	0.9727	0.9944	0.9727	0.9803	0.9716	0.9633	0.9663	1.055
Light Gradient Boosting Machine	0.9727	0.9898	0.9727	0.9823	0.9702	0.9633	0.9679	0.353
Logistic Regression	0.9545	0.9944	0.9545	0.9667	0.9523	0.9386	0.9434	0.516

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

Decision Tree Classifier	0.9455	0.9632	0.9455	0.9626	0.9433	0.9263	0.933	0.101
Linear Discriminant Analysis	0.9364	0.9943	0.9364	0.9477	0.9339	0.9142	0.9192	0.091
K Neighbors Classifier	0.8909	0.988	0.8909	0.9186	0.8864	0.8533	0.8657	0.094
SVM - Linear Kernel	0.8718	0	0.8718	0.8781	0.8595	0.8272	0.8379	0.086
Naive Bayes	0.8427	0.9943	0.8427	0.8319	0.8188	0.7904	0.8125	0.086

6. Discussions

The findings from the statistical description of the dataset are discussed. The analysis revealed important trends and patterns within the dataset, highlighting the distribution and variability of key variables. For instance, it was observed that the mean BMI value fell within the overweight range, indicating a higher prevalence of overweight individuals in the dataset. Additionally, the age distribution showed a wide range, suggesting the inclusion of individuals across different age groups. The statistical description provides a foundation for understanding the characteristics of the dataset and offers insights into the population under study. The discussion further elaborates on these findings, exploring their implications for the subsequent analysis and interpretation of the results. Additionally, potential limitations and biases associated with the dataset are considered, ensuring a comprehensive understanding of the dataset's statistical properties and their impact on the study's outcomes. In addition, the analysis of the correlation matrix provides insights into the relationships between variables and helps identify potential associations among them. For example, a strong positive correlation between weight and BMI suggests that individuals with higher weights tend to have higher BMIs. Similarly, a negative correlation between age and BMI indicates a tendency for BMI to decrease with increasing age. The implications of these correlations for understanding the factors influencing obesity and their interplay within the dataset. It is important to interpret these correlations in the context of the study's objectives and the population under investigation.

Besides, the results from the performance comparison of the ML algorithms reveals important insights into the effectiveness of the algorithms in accurately classifying obesity levels. For instance, Extra Trees Classifier achieved higher accuracy and precision scores compared to other algorithms, which suggests that Extra Trees Classifier is better at correctly classifying individuals into the respective obesity categories and minimizing false positives. we also explore potential reasons behind the observed performance differences among the Trees Classifiers, such as variations in their underlying algorithms, feature selection techniques, or hyperparameter tuning.

7. Conclusion

This paper presented a ML approach for obesity classification using a carefully curated dataset. Through extensive experimentation and analysis, we demonstrated the effectiveness of the proposed approach in accurately classifying individuals into different obesity categories. The results showed that by leveraging the power of ML techniques, such as the XGBoost algorithm and SMOTE resampling, we achieved significant improvements in classification performance compared to previous approaches. Our study highlighted the importance of addressing class imbalance in obesity classification tasks and showcased the benefits of using SMOTE to rebalance the dataset. The application of XGBoost as the classification model demonstrated its ability to capture complex relationships and classify individuals based on their features effectively. Moreover, the discussion of the results provided valuable insights into the performance comparison of different ML algorithms, enabling researchers and practitioners to make informed decisions when choosing an appropriate algorithm for similar applications.

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Safaei M., Sundararajan, E. A. Driss, M. Boulila, W. & Shapi'i, A., 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, 2021.
[2] Chatterjee A., Gerdes M. W. & Martinez S. G., Identification of risk factors associated with obesity and overweight—a machine learning overview. Sensors, 20(9), 2734, 2020.
[3] Colmenarejo G., Machine learning models to predict childhood and adolescent obesity: a review. Nutrients, 12(8), 2466, 2020.
[4] Triantafyllidis A., Polychronidou E., Alexiadis A. Rocha C. L., Oliveira D. N., da Silva A. S., & Tzovaras D., Computerized decision support and machine learning applications for the prevention and treatment of childhood obesity: A systematic review of the literature. Artificial Intelligence in Medicine, 104, 101844, 2020.
[5] Singh B., & Tawfik H., Machine learning approach for the early prediction of the risk of overweight and obesity in young people. In Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, 523-535, 2020, Proceedings, Springer International Publishing.
[6] Machorro-Cano, I., Alor-Hernández G., Paredes-Valverde, M. A., Ramos-Deonati, U., Sánchez-Cervantes, J. L., & Rodríguez-Mazahua, L, PISIoT: a machine learning and IoT-based smart health platform for overweight and obesity control. Applied Sciences, 9(15), 3037, 2019.
[7] Gerl, M. J., Klose, C., Surma, M. A., Fernandez, C., Melander, O., Männistö, S., ... & Simons, K., Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. PLoS biology, 17(10), 2019.
[8] Fernández-Navarro, T., Díaz, I., Gutiérrez-Díaz, I., Rodríguez-Carrio, J., Suárez, A., de Los Reyes-Gavilán, C. G., ... & González, S, Exploring the interactions between serum free fatty acids and fecal microbiota in obesity through a machine learning algorithm. Food Research International, 121, 533-541, 2019.
[9] Cervantes, R. C., & Palacio, U. M, Estimation of obesity levels based on computational intelligence. Informatics in Medicine Unlocked, 21, 100472, 2020.
[10] Geng, J., Zhang, X., Prabhu, S., Shahoei, S. H., Nelson, E. R., Swanson, K. S., & Smith, A. M., 3D microscopy and deep learning reveal the heterogeneity of crown-like structure microenvironments in intact adipose tissue. Science advances, 7(8), 2021.
[11] Hong N., Wen A., Stone D. J., Tsuji S., Kingsbury P. R., Rasmussen L. V., & Jiang G, Developing a FHIR-based EHR phenotyping framework: A case study for identification of patients with obesity and multiple comorbidities from discharge summaries. Journal of biomedical informatics, 99, 103310, 2019.
[12] Garcia-Carretero R., Vigil-Medina L., Mora-Jimenez I., Soguero-Ruiz C., Barquero-Perez O., & Ramos-Lopez, J., Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. Medical & biological engineering & computing, 58, 991-1002, 2020.
[13] Gadekallu T. R., Iwendi C., Wei, C., & Xin, Q., Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. IET Image Process, 16, 647-658, 2021.
[14] Shen Z., Shehzad A., Chen S., Sun H., & Liu, J.. Machine learning based approach on food recognition and nutrition estimation. Procedia Computer Science, 174, 448-453, 2020.

Doi: https://doi.org/10.54216/JAIM.030201
Received: August 17, 2022 Revised: November 18, 2022 Accepted: March 16, 2023