G-Tech : Jurnal Teknologi Terapan
Volume 8, No. 1, Januari 2023, hlmn. 196-208
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Machine Learning Approach of Obesity Level Classification: A Systematic Literature Review of Methods and Factors

Steven Marcelino Tandiono[1] , **Samuel Ady Sanjaya**[2]✉

[1, 2] Information Systems Department, Faculty of Engineering and Informatics, Multimedia Nusantara University, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 02-12-2023
Direvisi : 08-12-2023
Diterima : 12-12-2023

ABSTRAK

Meningkatnya prevalensi obesitas, yang memperbesar risiko penyakit seperti diabetes, penyakit jantung, dan kanker, mendorong kekhawatiran global. Otoritas kesehatan, peneliti, dan masyarakat menuntut solusi komprehensif. Machine learning menjanjikan dalam memproyeksikan risiko obesitas dengan mengidentifikasi penyebabnya secara akurat. Penelitian ini mereview literatur dari 2018-2023, fokus pada faktor-faktor yang berpengaruh terhadap obesitas dan aplikasi machine learning dalam mendeteksi risikonya, menggunakan pendekatan SPAR-4-SLR untuk panduan sistematis. Temuan ini menyoroti faktor risiko utama dan metode machine learning untuk prediksi, serta strategi optimalisasi model prediksi. Dengan memahami hubungan antara faktor risiko, obesitas dan pendekatan penyelesaian dengan machine learning, penelitian ini bertujuan untuk mengidentifikasi, mengevaluasi, dan mensintesis literatur yang relevan dalam suatu bidang penelitian untuk menyediakan pemahaman komprehensif tentang topik tersebut.

Kata Kunci:

Faktor Risiko, Obesitas, Machine Learning, Penyakit, Tinjauan Literatur Sistematis

ABSTRACT

The increasing prevalence of obesity, which amplifies the risk of diseases such as diabetes, heart disease, and cancer, has raised global concerns. Health authorities, researchers, and the community demand comprehensive solutions. Machine learning holds promise in projecting obesity risks by accurately identifying its causes. This study reviews literature from 2018-2023, focusing on factors influencing obesity and the application of machine learning in risk detection, using the SPAR-4-SLR approach for systematic guidance. The findings highlight key risk factors and machine learning methods for prediction, as well as optimization strategies for predictive models. By understanding the relationship between risk factors, obesity, and machine learning-based solutions, this research aims to identify, evaluate, and synthesize relevant literature in a specific research field to provide a comprehensive understanding of the topic.

***Keywords* :**

Risk Factors, Obesity, Machine Learning, Disease, Systematic Literature Review

***Corresponding Author* :**
Samuel Ady Sanjaya
Information Systems Department, Faculty of Engineering of Informatics, Multimedia Nusantara University
Jl. Scientia Boulevard, Curug Sanereng, Kec. Klp. Dua, Kabupaten Tangerang, Banten 15811
Email: samuel.ady@umn.ac.id

INTRODUCTION

Obesity is a complex health condition that not only reduces overall body quality, but also increases susceptibility to life-threatening chronic diseases. The World Health Organization

(WHO), in its official website, defines obesity as the excessive accumulation of fat, which presents a dangerous risk to one's health. This excessive accumulation of fat is not just aesthetic, but also a significant precursor to various chronic diseases. The implications of obesity go far beyond the physical aspect, permeating the intricate web of metabolic, cardiovascular, and respiratory systems, thus increasing the urgency for comprehensive understanding, prevention, and management of this pervasive health problem (WHO, 2023). Furthermore, WHO also states that the underlying cause of obesity is an energy imbalance between calories consumed and calories expended; as a result, the body stores unused calories in the form of fat. This health problem can be experienced by individuals of all ages, regardless of age, gender, or socioeconomic status (Solomon et al., 2023). Therefore, understanding and effectively managing obesity is crucial to raising health awareness.

The high prevalence of obesity worldwide over the past few decades is a health challenge that is not easy to overcome. Overall, the prevalence of obesity has increased from 3.2% in 1975 to 10.8% in 2014 for men and 6.4% to 14.9% for women. If the current trend continues, it is expected to reach 18% in men and exceed 21% in women by 2025 (Zhang et al., 2018). These percentages indicate a very alarming increase in the global obesity problem. Numerous studies have consistently concluded that obesity is a complex and multifactorial problem (Safaei et al., 2021). The combination of factors that influence a person's likelihood of being obese can be classified into several categories, such as demographic, socioeconomic, lifestyle, and genetic factors (Thamrin, Arsyad, et al., 2021). Researchers also consider that obesity is an acquired condition due to personal choices, such as lack of physical activity and overeating, regardless of possible genetic factors. In a broader context, gaining insights into the root causes of obesity and understanding the various influencing factors are important milestones in formulating impactful prevention strategies to combat this global health problem. Utilizing advanced research methodologies, particularly Machine Learning (ML), is emerging as a powerful avenue to improve our predictive capabilities in the area of individual obesity prognosis. ML, with its ability to independently discern causative factors, stands out as a valuable tool in this complicated landscape.

The impetus behind this systematic review is to summarize the findings of recent studies from 2018-2023. It aims to contribute to the ongoing discourse in uncovering potential triggers of obesity and shed light on various Machine Learning (ML) techniques used in obesity predictive research. Through this synthesis, it is hoped that a more comprehensive understanding of the nature of obesity will emerge, paving the way for more appropriate interventions and preventive measures. As for the research method, it is done by designing relevant research questions, defining important keywords, and finally compiling the collected research in an accessible and systematic way. This required a review protocol, Scientific Procedures and Rationales for Systematic Literature Reviews (SPAR-4-SLR), designed to guide and justify decisions in this systematic literature review. Through this method, the authors detail the process of searching, selecting studies, extracting data, and assessing the quality of relevant studies. Ultimately, the approach helped in answering 4 research questions that could provide in-depth insights into the significant risk factors, understand the different machine learning methods/algorithms that can be used for obesity prediction and which perform the best, and know the efforts that need to be made to maximize the performance of the built models. By understanding the correlation between risk factors and obesity, this research is expected to provide a strong foundation for the development of more effective prevention and intervention strategies in addressing the obesity problem globally.

RESEARCH METHODS

To achieve the objectives of this systematic review, the Systematic Literature Review (SLR) utilizes the Scientific Procedures and Rationales for Systematic Literature Reviews (SPAR-4-SLR) protocol. According to (Paul et al., 2021), the purpose of SPAR-4-SLR is to provide a reliable review protocol that researchers can depend on to guide and justify the accuracy of SLR outcomes. Furthermore, this protocol enables the inclusion of all necessary data to document the SLR comprehensively (Sauer & Seuring, 2023). The SPAR-4-SLR process comprises three main phases: assembling, arranging, and assessing. Within these three phases, there are six sub-derivatives, namely identification, acquisition, organization, purification, evaluation, and reporting. (Tushar & Sooraksa, 2023). As depicted in Figure 1, illustrating the entire systematic review process using the SPAR-4-SLR protocol, there are three main phases. In the assembling phase, this study identified various scientific literature with the main topic being the classification of obesity levels using ML techniques. Research questions were formulated to assess the relevance to the research topic. Three research questions were defined as benchmarks to achieve the research objectives. Table 1 provides a detailed explanation of the research questions to attain the overall research goals systematically. The review was conducted using a comprehensive and fundamental search strategy through the Publish or Perish (POP) platform. The scholarly works obtained were sourced from the Scopus online database spanning from 2018 to 2023. To systematically explore the landscape of research at the intersection of Machine Learning (ML) and obesity, a targeted search strategy was employed using specific keywords. The chosen search criteria included the combination of ("Machine Learning") AND ("Obesity" OR "Overweight"). This deliberate selection aimed to narrow down the focus on literature directly addressing ML classification techniques and their application in the context of obesity. The search encompassed titles, abstracts, and keywords, ensuring a comprehensive scan of relevant information. This initial phase of the review was conducted with a pragmatic lens, critically assessing whether the identified literature met the predefined prerequisites. The outcome of this meticulous integrated search process resulted in the identification and extraction of a substantial corpus, comprising a total of 500 scholarly works from the Scopus database. This robust selection laid the foundation for the subsequent stages of the systematic literature review, facilitating a thorough investigation into the intricate relationship between ML methodologies and obesity-related insights.

Table 1. Research Questions

Indeks	***Research Questions***	**Tujuan**
RQ1	What potential factors or parameters contribute to obesity?	To identify potential factors or parameters that may cause obesity.
RQ2	Which techniques or algorithms are commonly used for predicting classifications?	To understand the variation in ML algorithms used to predict obesity.
RQ3	Which techniques or algorithms yield the best performance in predicting obesity?	To identify algorithms that achieve the best performance in predicting obesity.
RQ4	What are the best efforts to maximize model performance?	What are the best efforts to maximize model performance?

In the arranging phase, the second step of this systematic review, an exhaustive selection process was undertaken to identify scholarly works that aligned with the initial criteria. Subsequently, these chosen works underwent meticulous organization, utilizing a structured spreadsheet format that encompassed crucial details such as journal specifics, author information, title, issue, research objectives, methods/techniques employed, findings, and gap analyses. Throughout the purification process, certain literature pieces were identified as unrelated or irrelevant to the research focus. This discernment was achieved by scrutinizing the titles and abstracts of each piece individually. The rigorous screening process led to the exclusion of 480 articles, leaving a curated selection of 20 journal articles. Ultimately, the current review culminated in the extraction of 20 pertinent journal articles that delve into the intersection of Machine Learning classification techniques and the study of obesity.

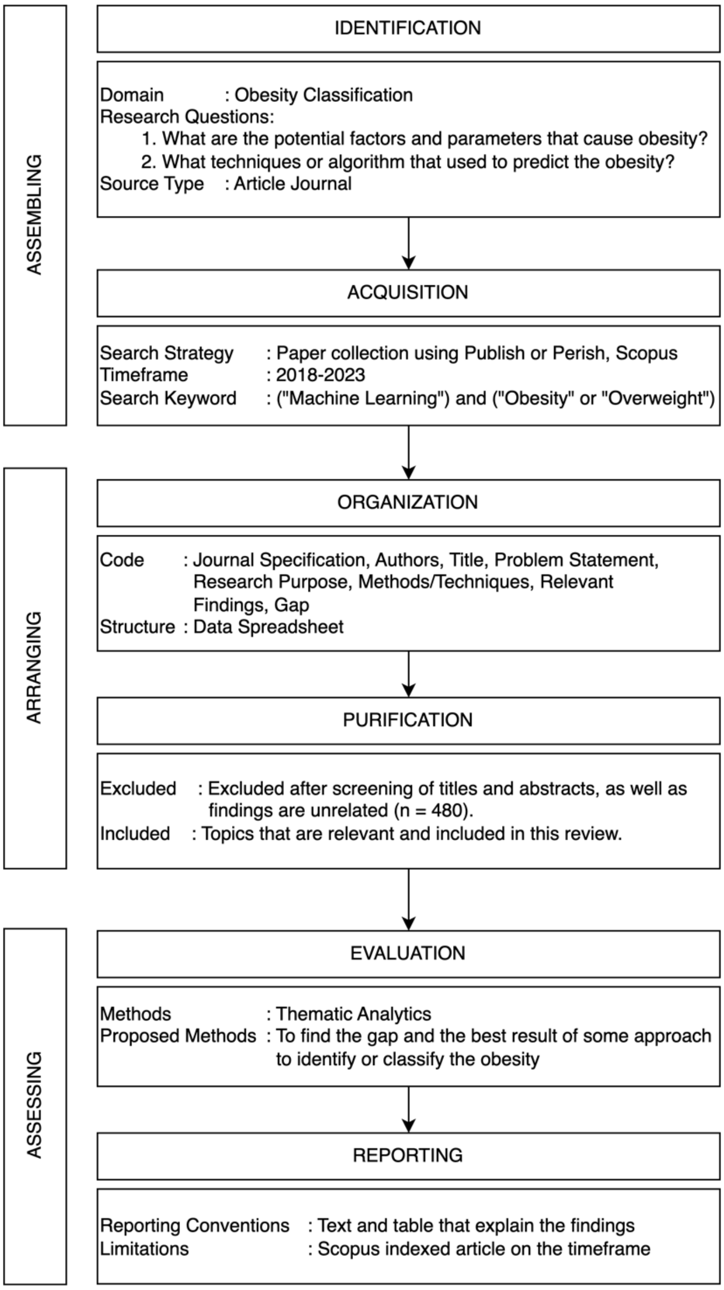

Figure 1. Process Flow of SPAR-4-SLR

In the final stage, the assessment of the suitability of evaluating journal articles utilized a thematic analysis approach. Various ML algorithm techniques and categorizations of predictors causing obesity were identified in the 20 journal articles. This study also aims to obtain the best ML method results in classifying obesity levels and analyze gaps between scientific studies. The reporting phase of presenting SLR findings is done in the form of writing and tables, intended to clarify the obtained results. Additionally, this research evaluates the limitations that may arise.

RESULTS AND DISCUSSIONS

From the screening process of findings to evaluation, a total of 20 selected scholarly works have been deemed relevant and met the criteria set by the research questions and objectives. These works were retrieved from the Scopus journal database covering the last five years (2018 - 2023). Regarding the selected journal articles, the following is a summary presented using a tabular structure. Presenting the summary in table form is useful for providing more structured information, facilitating analysis, and supporting further research development.

Table 2. Paper Summary: Factors and Parameters caused Obesity

Authors	Factors Influencing Obesity	Authors	Factors Influencing Obesity
(Blüher, 2019)	**Food Intake (Overeating)** - Uncontrolled eating - Emotional eating - Lack of knowledge about healthy eating **Metabolism and Energy Expenditure (Low Energy Burn)** - Aging - Genetics and epigenetics **Physical Activity (Physical Inactivity)** - Physical challenges - Low fitness levels - Chronic fatigue	(Kadouh & Acosta, 2017)	**Biological Factors** - Genetics - Gut microbiome - Prenatal determinants - Pregnancy - Menopause - Medication use - Viral infections **Environmental Factors** - Food abundance - Societal and cultural influences - Environmental chemicals **Individual Behavioral Factors** - Excessive calorie intake - Unhealthy eating patterns - Sedentary lifestyle - Lack of sleep
(S. Sun et al., 2021)	**Biological or Genetic Factors** - Physiological **Personal Lifestyle Factors** - Unhealthy eating patterns - Lack of physical activity - Family reasons - Work-related stress - Irregular lifestyle **Social Factors** - Prevalence of fast food - New transportation and/or technology - Lack of facilities - Misleading health and consumption-related advertisements	(Omer, 2020)	**Genetic Vulnerability** - Polygenic - Monogenic or syndromic **Human Behavior** - Specific endocrinological conditions - Medication consumption - Depression, anxiety, stress - Excessive intake of high-energy foods - Lack of physical activity **Environmental Factors** - Workplace environment and long working hours - Difficulty accessing healthy food - Social networks
(Weihrauch-Blüher & Wiegand, 2018)	Physical inactivity Low physical activity behavior Socioeconomic factors Duration and intensity of sleep Obesogenic environment Excessive energy intake	(Masood & Moorthy, 2023)	Physical Activity Environment Food Consumption Food Production Individual and Social Psychology
(Y. Sun et al., 2020)	**Socioeconomic Status** - Poverty - Unemployment **Behavioral Factors** - Consumption of sugary drinks - Consumption of fruits and vegetables	(Lecube et al., 2020)	Physical activity Nutritional habits Consumption of unhealthy beverages, both soft drinks and alcohol Sleep patterns Difficulty in adhering to a healthy diet

	Environmental Factors - Prevalence of physical activity - Supermarket accessibility

RQ1: Factors causing obesity

Several studies have indicated that obesity is not solely caused by a single factor; rather, it results from a combination of interactions among various factors (Longo et al., 2019). From the findings presented in various literature, as depicted in Table 2, it can be concluded that the causes of obesity are multifactorial and complex. These factors can interact with each other, creating an energy imbalance in the body that ultimately leads to weight gain. The gathered studies reveal that determinants of obesity can generally be grouped into several main categories, such as biological, environmental, and individual behavioral factors.

Biological factors, such as genetics, can play a significant role in causing obesity. Genetic vulnerability can increase the risk of obesity in individuals by influencing metabolism, the body's response to food, and the body's ability to burn calories. From an environmental perspective, factors such as easy access to high-calorie and sugary fast food and the influence of the surrounding culture can affect individuals' eating habits and physical activity levels. An environment that does not support a healthy lifestyle can be a cause of obesity. Additionally, individual behavior is also a crucial factor in causing obesity. Unhealthy eating patterns, including overeating, uncontrolled eating, and emotional eating, can contribute to obesity. Physical inactivity, such as a lack of physical activity, physical challenges, low fitness levels, and chronic fatigue, can impact body weight.

The explanation of the factors causing obesity above demonstrates a high level of complexity and interaction within them. A thorough understanding of these factors is valuable for identifying highly relevant potential parameters in the context of testing ML classification models. Thus, information about various categories of obesity determinants can serve as variables to build predictive models that assist in identifying, preventing, and managing obesity more effectively.

Table 3. Summary of Selected Scientific Work Related to the Use of Classification Techniques

Authors	Research Objective	Comparison of Algorithms	Best Algorithm and Results	Support Techniques
(Thamrin, Arsyad, et al., 2021)	Predicting obesity in adults	Logistic Regression, Classification and Regression Trees, Naïve Bayes	Logistic Regression Accuracy: 72%, AUC: 79%	SMOTE to resolve data imbalance
(Cheng et al., 2021)	Prediction of the effect of physical activity on obesity	Naïve Bayes, Radial Basis Function (RBF), Local K-Nearest Neighbors (KNN), Classification Via Regression (CVR), Random Subspace, Decision Table, Multiobjective Evolutionary Fuzzy Classifier, Random Tree, J48, And Multilayer Perceptron Classification	Random Subspace *Accuracy*: 67% AUC: 64%	-

Authors	Research Objective	Comparison of Algorithms	Best Algorithm and Results	Support Techniques
(Solomon et al., 2023)	Developing effective prediction and classification models for obesity	Gradient boosting classifier, XGB classifier, Multilayer perceptron, K-nearest-neighbor classifier, Logistic regression, Naïve Bayes classifier, Random forest classifier, Decision tree	*Hybrid* Model (Gradient boosting classifier, XGB classifier, dan multilayer perceptron classifier) *Accuracy*: 97.16%	Application of a voting-based ensemble learning approach to create a hybrid model
(Ferdowsy et al., 2021)	Predicting obesity risk using ML algorithms	K-Nearest Neighbor (K-NN), Random Forest, Logistic Regression, Multilayer Perceptron (MLP), Support Vector Machine (SVM), Naïve Bayes, Adaptive Boosting (ADA Boosting), Decision Tree, dan Gradient Boosting Classifier	Logistic Regression *Accuracy*: 97.09%	-
(Santisteban Quiroz, 2022)	Identifying obesity levels based on lifestyle through ML techniques	Light Gradient Boosting Machine (Light GBM) classifier, random forest (RF), decision tree (DT), Extremely Randomized Trees (ET), dan logistic regression (LR)	Light GBM *Accuracy*: 97.45% AUC: 99.90%	-
(Musa et al., 2022)	Developing an ML model to predict obesity status	Gboost Classifier, Random Forest Classifier, Decision Tree Classifier, K-Nearest Neighbor, dan Support Vector Machine	Gboost Classifier *Accuracy*: 99.05%	-
(Jeon et al., 2023)	Identifying risk factors that predict obesity using ML classifiers	Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Light Gradient Boosting (LGBM), dan Extreme Gradient Boosting (XGB)	Multi-Layer Perceptron (MLP) AUC: 78%	-
(Rashmi et al., 2021)	Use of ML to evaluate obesity	Support Vector Machine (SVM),	SVM *Accuracy*: 98%	*Feature fusion, Feature selection,* dan *feature*

Authors	Research Objective	Comparison of Algorithms	Best Algorithm and Results	Support Techniques
	through thermal imaging	Naïve Bayes, dan Random Forest		*dimension reduction* (PCA)
(Li et al., 2023)	Application of ML models in classifying stroke levels	Multiple logistic regression, Multiple Classification Support Vector Machine, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine	XGBoost *Accuracy*: 99.27% AUC: 99.99%	SHAP used to assess the impact of one feature on different stroke rates
(Jian et al., 2021)	Development of an ML model to predict a person's obesity category	Random Forest, Support Vector Machine, Decision Tree, dan K-Nearest Neighbor	Random Forest *Accuracy*: 98,48%	-
(Maria et al., 2023)	Predict various complications associated with diabetes using ML techniques	Logistic Regression (LR), support vector machine (SVM), decision tree (DT CART), random forest (RF), AdaBoost, and XGBoost	XGBoost *Accuracy*: 97.8%	The use of SMOTE for oversampling the minority class and cluster centroids for undersampling the majority class is used.
(Alsareii et al., 2022)	Application of ML techniques to create an IoT system for diagnosing obesity	Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), dan Naïve Bayes (NB)	SVM *Accuracy*: 96%	-
(Rodríguez et al., 2021)	Application of ML techniques to build a model that can identify a person's obesity	Decision Tree,Support Vector Machines, K-Nearest Neighbors, Gaussian Naive Bayes, Multilayer Perceptron, Random Forest, Gradient Boosting, dan Extreme Gradient Boosting	Decision Tree *Accuracy*: 78%	-
(Singh & Tawfik, 2019)	Use of appropriate ML techniques to predict BMI values	Multivariate Linear Regression, SVM, dan Multi-Layer Perceptron Feed Forward Artificial Neural Networks (MLPFFANN)	MLPFFANN *Accuracy*: 93.04%	-

RQ2: Techniques/Algorithms used for obesity classification

Based on the findings of various scientific works that have been collected, a number of machine learning methods that are commonly used in the context of classification can be identified. The use of ML to diagnose disease has promising potential (Zou et al., 2018). Table 3 provides a comprehensive overview of algorithms that are often used in previous research. Several algorithms that have been regularly tested to solve classification problems include Logistic Regression, K-Nearest Neighbor (KNN), Naïve Bayes, Support Vector Machine (SVM), Decision Tree, and Random Forest. Most of this previous research has had a main focus on discussing comparisons and evaluating the performance of these algorithms. Methods such as Adaptive Boosting (ADA Boosting), Gradient Boosting Classifier, and Extreme Gradient Boosting (XGBoost) have also received significant attention as effective choices in classification modeling.

Apart from that, researchers not only use a single algorithm to predict but also use hybrid methods to improve model performance. Research (Solomon et al., 2023), have applied a hybrid method through an ensemble learning approach to obtain satisfactory model performance. From the summary of several scientific works, it can be concluded that the use of various algorithms illustrates the diversity of approaches taken by researchers to predict classification. Although each method certainly has its advantages and disadvantages, the choice depends on the specific needs of the research and the characteristics of the data used. Thus, this literature review provides an overview of the various methods and algorithms commonly used in predicting classification. It is also hoped that this can be an important reference in selecting ML models in the context of research related to classification.

RQ3: Technique/Algorithm that provides the best performance in completing classification

The development of the application of ML algorithms in the health sector has been used to predict the development of certain health conditions based on predetermined characteristics (Singh & Tawfik, 2019). According to the findings of scientific work on a number of algorithms applied in various studies for classification in the health sector, the highest performance in predicting classification categories is generally obtained through the use of several superior models. Research (Solomon et al., 2023) applies a hybrid model by combining the Gradient Boosting Classifier, XGB Classifier, and Multilayer Perceptron Classifier to obtain very high accuracy, reaching 97.16%. These advantages can be attributed to the ability of hybrid models to capture complex patterns and non-linear relationships in obesity data. In addition, the XGBoost algorithm in research (Li et al., 2023) and (Maria et al., 2023) shows excellent results with accuracy ranging from 97% to 99%. Light GBM also provides satisfactory results with an accuracy of 97.45% and AUC reaching 99.90%, proving that this model is efficient and effective in the context of obesity prediction.

Other techniques such as Logistic Regression, Gboost Classifier, Random Forest, MLPFFANN, and SVM also provide good performance in solving classification problems. These performance results prove that the use of ML techniques may be a reliable solution in creating predictive models for obesity classification, although in the end it all depends on the characteristics of the data and analysis needs. In other previous research, there were several cases showing lower model accuracy results, but this research still made an important contribution to the development of predictive approaches. Thus, the decision to choose the best algorithm must be based on the characteristics and complexity of the data faced. It must also be in accordance with the objectives, limitations and needs of the research context of obesity classification research.

RQ4: Efforts are made to maximize model performance

In various previous studies, several approaches and supporting techniques were identified as an effort to maximize model performance effectively. One effort to overcome data imbalance is the use of the Synthetic Minority Over-sampling Technique (SMOTE). The easiest way to correct imbalanced data is to balance it (Thamrin, Sidik, et al., 2021). SMOTE helps the model to better identify less representative minority class variations in the dataset. Apart from that, in research (Maria et al., 2023) also uses Cluster Centroids for undersampling the majority class. The goal of this technique is to achieve a balanced class distribution to improve the model's capabilities. Another approach used to improve model performance results is the application of ensemble learning. This strategy works by creating a hybrid model that can combine the advantages of various algorithms. Many studies reveal that the results of ensemble algorithms show better performance than single algorithms (Karthikeyan et al., 2020).

Feature engineering techniques, such as feature fusion, feature selection, and feature dimension reduction, especially using Principal Component Analysis (PCA), are also strategies in an effort to improve model performance. This approach aims to improve feature representation, reduce irrelevant dimensions, and select the most informative features. In the case of other research, efforts to assess the impact of features at the level of model results can be quantified using Shapley Additive exPlanations (SHAP). SHAP provides a deep understanding of the importance of features and the impact of each feature which helps researchers to make more targeted decisions in improving models (Ekanayake et al., 2022).

Overall, in an effort to maximize model performance, previous research involved a combination of data balancing strategies, the application of ensemble learning, as well as special attention to the role and impact of the features used in the model. These techniques represent the best efforts to improve the model's ability to handle the complexity of the dataset and improve the quality of predictions. The selection and combination of techniques in an effort to maximize model performance should be based on specific data characteristics and research objectives to achieve optimal results.

CONCLUSION

This systematic literature review (SLR) has undertaken a comprehensive examination of diverse prior studies, delving into the intricate realm of predictors for obesity causes and the judicious application of Machine Learning (ML) classification techniques. The synthesis of literature analyzing factors contributing to obesity illuminates the complex and multifaceted nature of this health condition, intricately interwoven with biological, lifestyle, and environmental variables. Exploring the realm of ML techniques for obesity prediction reveals a consensus among various studies on the efficacy of classification algorithms such as Logistic Regression, Gradient Boosting, and XGBoost, showcasing commendable performance. Despite nuanced differences in results across studies, the majority underscores the reliability of these algorithms in achieving classification with satisfactory accuracy.

Moreover, to augment model performance, a range of strategies have been explored, including the utilization of techniques like Synthetic Minority Over-sampling Technique (SMOTE), ensemble learning, and feature evaluation through SHAP. These supporting methodologies not only enhance the model's efficiency but also offer profound insights into the intricate factors contributing to obesity. In conclusion, the outcomes derived from this SLR aspire to serve as a foundational reference for addressing the challenge of classifying obesity levels. Acknowledging the variability in results and the evolving nature of research, it is essential to tailor decisions based on the distinctive characteristics of the encountered data and the specific objectives and constraints inherent in each research endeavor. This synthesis, therefore, serves not only as a current snapshot of the field but also as a dynamic resource guiding future investigations into the complex landscape of obesity classification.

ACKNOWLEDGEMENT

This research received support from the Institution of Research and Community Services at Universitas Multimedia Nusantara. We extend our appreciation to our colleagues at the Big Data Laboratory within the Information Systems Department at Universitas Multimedia Nusantara, whose valuable input and expertise greatly enriched the research.

REFERENCES

Alsareii, S. A., Shaf, A., Ali, T., Zafar, M., Alamri, A. M., AlAsmari, M. Y., Irfan, M., & Awais, M. (2022). IoT Framework for a Decision-Making System of Obesity and Overweight Extrapolation among Children, Youths, and Adults. *Life*, *12*(9). https://doi.org/10.3390/life12091414

Blüher, M. (2019). Obesity: global epidemiology and pathogenesis. In *Nature Reviews Endocrinology* (Vol. 15, Issue 5). https://doi.org/10.1038/s41574-019-0176-8

Cheng, X., Lin, S. Y., Liu, J., Liu, S., Zhang, J., Nie, P., Fuemmeler, B. F., Wang, Y., & Xue, H. (2021). Does physical activity predict obesity—a machine learning and statistical method-based analysis. *International Journal of Environmental Research and Public Health*, *18*(8). https://doi.org/10.3390/ijerph18083966

Ekanayake, I. U., Meddage, D. P. P., & Rathnayake, U. (2022). A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Studies in Construction Materials*, *16*. https://doi.org/10.1016/j.cscm.2022.e01059

Ferdowsy, F., Rahi, K. S. A., Jabiullah, M. I., & Habib, M. T. (2021). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences*, *2*. https://doi.org/10.1016/j.crbeha.2021.100053

Jeon, J., Lee, S., & Oh, C. (2023). Age-specific risk factors for the prediction of obesity using a machine learning approach. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.998782

Jian, Y., Pasquier, M., Sagahyroon, A., & Aloul, F. (2021). A machine learning approach to predicting diabetes complications. *Healthcare (Switzerland)*, *9*(12). https://doi.org/10.3390/healthcare9121712

Kadouh, H. C., & Acosta, A. (2017). Current paradigms in the etiology of obesity. In *Techniques in Gastrointestinal Endoscopy* (Vol. 19, Issue 1). https://doi.org/10.1016/j.tgie.2016.12.001

Karthikeyan, R., Geetha, P., & Ramaraj, E. (2020). Prediction of diabetes and cholesterol diseases based on ensemble learning techniques. *International Journal of Scientific and Technology Research*, *9*(2).

Lecube, A., Sánchez, E., Monereo, S., Medina-Gomez, G., Bellido, D., Garcia-Almeida, J. M., Martinez De Icaya, P., Malagon, M. M., Goday, A., & Tinahones, F. J. (2020). Factors Accounting for Obesity and Its Perception among the Adult Spanish Population: Data from 1,000 Computer-Assisted Telephone Interviews. In *Obesity Facts* (Vol. 13, Issue 4). https://doi.org/10.1159/000508111

Li, J., Luo, Y., Dong, M., Liang, Y., Zhao, X., Zhang, Y., & Ge, Z. (2023). Tree-Based Risk Factor Identification and Stroke Level Prediction in Stroke Cohort Study. *BioMed Research International*, *2023*. https://doi.org/10.1155/2023/7352191

Longo, M., Zatterale, F., Naderi, J., Parrillo, L., Formisano, P., Raciti, G. A., Beguinot, F., & Miele, C. (2019). Adipose tissue dysfunction as determinant of obesity-associated metabolic complications. *International Journal of Molecular Sciences*, *20*(9). https://doi.org/10.3390/ijms20092358

Masood, B., & Moorthy, M. (2023). Causes of obesity: a review. *Clinical Medicine, Journal of the Royal College of Physicians of London*, *23*(4), 284–291. https://doi.org/10.7861/clinmed.2023-0168

Musa, F., Basaky, F., & E.O, O. (2022). Obesity prediction using machine learning techniques. *Journal of Applied Artificial Intelligence*, *3*(1). https://doi.org/10.48185/jaai.v3i1.470

Omer, T. (2020). The causes of obesity: an in-depth review. *Advances in Obesity, Weight Management & Control*, *10*(4). https://doi.org/10.15406/aowmc.2020.10.00312

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*. https://doi.org/10.1111/ijcs.12695

Rashmi, R., Umapathy, S., & Krishnan, P. T. (2021). Thermal imaging method to evaluate childhood obesity based on machine learning techniques. *International Journal of Imaging Systems and Technology*, *31*(3), 1752–1768. https://doi.org/10.1002/ima.22572

Rodríguez, E., Rodríguez, E., Nascimento, L., da Silva, A., & Marins, F. (2021). Machine learning techniques to predict overweight or obesity. In *CEUR Workshop Proceedings* (Vol. 3038, pp. 190–204). https://api.elsevier.com/content/abstract/scopus_id/85121261382

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

Santisteban Quiroz, J. P. (2022). Estimation of obesity levels based on dietary habits and condition physical using computational intelligence. *Informatics in Medicine Unlocked*, *29*. https://doi.org/10.1016/j.imu.2022.100901

Sauer, P. C., & Seuring, S. (2023). How to conduct systematic literature reviews in management research: a guide in 6 steps and 14 decisions. In *Review of Managerial Science* (Vol. 17, Issue 5). https://doi.org/10.1007/s11846-023-00668-3

Singh, B., & Tawfik, H. (2019). A Machine Learning Approach for Predicting Weight Gain Risks in Young Adults. *Conference Proceedings of 2019 10th International Conference on Dependable Systems, Services and Technologies, DESSERT 2019*. https://doi.org/10.1109/DESSERT.2019.8770016

Solomon, D. D., Khan, S., Garg, S., Gupta, G., Almjally, A., Alabduallah, B. I., Alsagri, H. S., Ibrahim, M. M., & Abdallah, A. M. A. (2023). Hybrid Majority Voting: Prediction and Classification Model for Obesity. *Diagnostics*, *13*(15). https://doi.org/10.3390/diagnostics13152610

Sun, S., He, J., Shen, B., Fan, X., Chen, Y., & Yang, X. (2021). Obesity as a "self-regulated epidemic": coverage of obesity in Chinese newspapers. *Eating and Weight Disorders*, *26*(2). https://doi.org/10.1007/s40519-020-00886-8

Sun, Y., Wang, S., & Sun, X. (2020). Estimating neighbourhood-level prevalence of adult obesity by socio-economic, behavioural and built environment factors in New York City. *Public Health*, *186*. https://doi.org/10.1016/j.puhe.2020.05.003

Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021). Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research 2018. *Frontiers in Nutrition*, *8*. https://doi.org/10.3389/fnut.2021.669155

Thamrin, S. A., Sidik, D., Kuswanto, H., Lawi, A., & Ansariadi, A. (2021). Exploration of Obesity Status of Indonesia Basic Health Research 2013 With Synthetic Minority Over-Sampling Techniques. *Indonesian Journal of Statistics and Its Applications*, *5*(1). https://doi.org/10.29244/ijsa.v5i1p75-91

Tushar, H., & Sooraksa, N. (2023). Global employability skills in the 21st century workplace: A semi-systematic literature review. In *Heliyon* (Vol. 9, Issue 11). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2023.e21023

Weihrauch-Blüher, S., & Wiegand, S. (2018). Risk Factors and Implications of Childhood Obesity. In *Current obesity reports* (Vol. 7, Issue 4). https://doi.org/10.1007/s13679-018-0320-0

WHO. (2023). *Obesity*. World Health Organization. https://www.who.int/health-topics/obesity#tab=tab_1

Zhang, C., Zhang, J., Liu, Z., & Zhou, Z. (2018). More than an Anti-diabetic Bariatric Surgery, Metabolic Surgery Alleviates Systemic and Local Inflammation in Obesity. In *Obesity Surgery* (Vol. 28, Issue 11). https://doi.org/10.1007/s11695-018-3400-z

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting Diabetes Mellitus With Machine Learning Techniques. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00515