Artificial Intelligence in Educational Decision Sciences
E-ISSN: 3123-8823
Vol. 1, No. 1, Month 2026

Data-Driven Obesity Classification Integrating Genetic and Lifestyle Determinants Using Naive Bayes

Yusion Gandjang✉
Universitas Negeri Makassar, Makassar, Indonesia
Amaliah Safitri K
Universitas Negeri Makassar, Makassar, Indonesia
Nabila Dwi Anugra
Universitas Negeri Makassar, Makassar, Indonesia
Iyang Yuyung S
Universitas Negeri Makassar, Makassar, Indonesia
Akhmad Affandi
Dresden University, Germany

ABSTRACT

Purpose – This study aims to develop a data-driven obesity classification framework that integrates genetic predisposition and lifestyle determinants using the Naive Bayes algorithm, while empirically evaluating optimal training–testing data proportions for health decision support systems.

Methods – A systematic computational workflow was applied to a public obesity dataset comprising 2,112 records, which was refined to 1,259 valid instances after preprocessing. Genetic indicators and lifestyle-related variables were encoded and classified into four obesity categories: normal weight, obesity type I, obesity type II, and obesity type III. The Naive Bayes model was evaluated using three training–testing data partition ratios (75:25, 80:20, and 85:15). Model performance was assessed using six metrics: Area Under the Curve (AUC), classification accuracy, F1-score, precision, recall, and Matthews Correlation Coefficient.

Findings – The results demonstrate that the 80:20 and 85:15 data partitions achieved the highest performance, with an accuracy of 0.878 and an AUC of 0.979. The model showed excellent sensitivity in identifying severe obesity cases, while moderate misclassification occurred between obesity type I and type II due to phenotypic overlap in lifestyle patterns.

Research limitations – This study relies on a single public dataset and lacks population-specific genetic calibration, which may limit generalizability to diverse regional contexts.

Originality – This study provides empirical validation of a probabilistic obesity classification framework that integrates genetic and lifestyle factors, offering an interpretable and computationally efficient approach to support data-driven health decision making.

OPEN ACCESS

ARTICLE HISTORY
Received: 20-05-2026
Revised: 25-07-2026
Accepted: 10-08-2026

KEYWORDS
Classification;
Obesity;
Genetic factors;
Lifestyle;
Naive bayes.

Correspondence Author: ✉yusiongandjang@mail.com
To cite this article : Yusion Gandjang, Amaliah Safitri K, Nabila Dwi Anugra, & Iyang Yuyung S, Akhmad Affandi (2026). Data-Driven Obesity Classification Integrating Genetic and Lifestyle Determinants Using Naive Bayes. Artificial Intelligence in Educational Decision Sciences, 1(2), 49–59.

INTRODUCTION

Obesity represents one of the most pressing global health challenges of the twenty-first century, with its prevalence escalating at an alarming rate across both developed and developing nations alike. According to the most recent epidemiological assessments, the global obesity prevalence has nearly tripled since 1975. By 2022, over 890 million adults worldwide were classified as obese, while more than 2.5 billion adults were overweight (Mulder et al., 2025). Current projections from the World Obesity Atlas 2025 indicate this figure may reach 1.13 billion adults with obesity by 2030 if current trends persist (Parums, 2025). This escalating epidemic serves as a primary etiological factor for numerous chronic conditions, including type 2 diabetes, hypertension, and cardiovascular diseases, which collectively contribute to millions of premature deaths annually (Bhutta, 2025; Zhao et al., 2021). The situation is particularly concerning in developing nations like Indonesia, where rapid urbanization, economic transition, and cultural shifts have precipitated dramatic changes in nutritional patterns and physical activity levels. Recent national health survey data reveals that Indonesia's adult obesity prevalence continues to rise significantly, representing one of the fastest growing rates in Southeast Asia and imposing substantial economic burdens on the national healthcare system (Muharram et al., 2025; Tee & Voon, 2024). This phenomenon is driven by a complex interplay of environmental factors and biological predispositions, necessitating a more profound understanding of how lifestyle choices and genetic architecture converge to determine an individual's metabolic fate.

Contemporary lifestyle transformations have significantly exacerbated Indonesia's obesity epidemic through interconnected pathways that disrupt metabolic homeostasis (Alchamdani & Anas, 2024; Muharram et al., 2025). Rapid urbanization has catalyzed a dual burden: dietary patterns increasingly dominated by ultra-processed, energy-dense foods while physical activity levels diminish through technology-mediated sedentary behaviors and motorized transportation systems (Colozza et al., 2023; Stappers et al., 2023). These environmental shifts collectively create an obesogenic ecosystem that challenges traditional metabolic adaptations, particularly among vulnerable urban populations experiencing accelerated nutrition transitions(Hosseinpour-Niazi et al., 2024; Verde et al., 2024).

Crucially, modern research now demonstrates that genetic factors account for a significant portion of the phenotypic variance in body mass index, with genome-wide association studies identifying numerous genetic loci associated with obesity susceptibility (Downie et al., 2025; Loos, 2025). Recent advances in polygenic risk scoring have enabled more precise quantification of genetic predisposition, showing that genetic variation is capable of increasing the prediction of obesity risk within the Indonesian population (Siswanto et al., 2025). These studies demonstrate that incorporating population-specific genetic variants enhances obesity risk prediction accuracy across diverse ancestral groups, while emphasizing that optimal predictive frameworks require integration of polygenic susceptibility with modifiable behavioral and environmental factors rather than substituting one for the other (Chen et al., 2025; Jansen et al., 2024).

The application of machine learning algorithms in obesity classification has gained substantial traction over the past decade, with Naive Bayes emerging as a particularly promising approach due to its computational efficiency and probabilistic interpretability (Thamrin et al., 2021). Naive Bayes is a statistical classification method used to predict the probability of class membership with high accuracy and speed, making it suitable for clinical environments where rapid decision support is required (Wang, 2025). Recent comparative analyses demonstrate that while complex ensemble methods like Random Forest often achieve marginally superior accuracy in controlled settings, Naive Bayes maintains remarkable competitiveness when considering the practical constraints of real-world healthcare implementation (Hadi et al., 2025; Phatcharathada & Srisuradetchai, 2025). Indonesian researchers have conducted studies applying machine learning algorithms such as Naive Bayes for diabetes risk classification based on lifestyle factors, demonstrating moderate to high predictive performance in computational models (Awalia et al., 2025; Kurniawan et al., 2024). Similarly, analyses of datasets containing lifestyle attributes have consistently shown Naive Bayes

achieving classification accuracy that is highly effective for many probabilistic classification cases (Airlangga, 2025; Awalia et al., 2025). This algorithm remains competitive in terms of efficiency, especially when preprocessing is performed correctly to handle the nuances of medical data.

International research continues to affirm the clinical value of Naive Bayes classifiers, particularly their interpretability and computational efficiency in healthcare applications where transparent decision-making remains essential for clinician adoption. Recent analysis of Indonesian Basic Health Research data by Thamrin et al. (2021) demonstrated how probabilistic models incorporating lifestyle variables can effectively predict obesity risk in resource-limited settings, though such approaches typically remain confined to single-domain predictors. Current literature increasingly recognizes that obesity prediction models relying solely on lifestyle or anthropometric factors face significant limitations, as evidenced by emerging studies integrating polygenic risk scores to capture individual genetic predisposition alongside environmental exposures (Hüls et al., 2021; Jansen et al., 2024). This methodological gap is particularly consequential given that obesity fundamentally emerges from complex gene-environment interactions that shape long-term risk trajectories, suggesting that future research must prioritize multimodal frameworks combining genetic markers with behavioral and clinical data.

This research gap is evidenced by three specific methodological limitations that our study directly addresses through empirical validation. First, despite growing recognition of gene-environment interactions in obesity pathogenesis, few computational frameworks systematically integrate polygenic susceptibility markers with lifestyle determinants into unified classification systems, particularly for Southeast Asian populations where such integration remains underexplored (Bineid et al., 2025; Dashti et al., 2022). Second, existing machine learning studies predominantly employ fixed training-testing ratios without empirical justification, overlooking how data partitioning strategies critically influence model stability in obesity prediction tasks (Sivakumar et al., 2024). Third, conventional evaluation protocols often rely on singular accuracy metrics despite clinical requirements for multidimensional performance assessment across sensitivity, specificity, and clinical interpretability domains (Andersen et al., 2024; Kocak et al., 2025). Our methodology directly addresses these constraints by empirically validating model stability across three distinct data partition schemes while employing a comprehensive six-metric evaluation framework, creating foundational protocols for obesity prediction systems that account for Indonesia's distinctive demographic and genetic context.

Building on this methodological foundation, Our research directly confronts these methodological challenges by developing an integrated predictive framework that combines genetic predisposition and lifestyle factors within an optimized Naive Bayes classification system. Through systematic evaluation of model performance across multiple training-testing configurations and strategic data preprocessing approaches, we establish an optimized prediction protocol specifically designed for Indonesia's unique demographic and genetic profile. This work contributes a clinically actionable decision support tool that enables early identification of obesity risk, facilitating timely public health interventions while addressing critical gaps in personalized obesity prevention strategies for Southeast Asian populations.

METHOD

Research Design

This study employs a systematic computational framework to classify obesity status by integrating genetic predispositions and lifestyle variables into a unified predictive model. The comprehensive workflow, illustrated in Figure 1, follows five sequential phases from data acquisition through evaluation. This methodological approach reflects current best practices for integrating multidimensional health predictors in machine learning applications, ensuring the model effectively captures complex interactions between biological and behavioral determinants of obesity. The design specifically addresses methodological gaps identified in recent literature regarding the combined analysis of genetic and lifestyle factors in obesity classification systems.

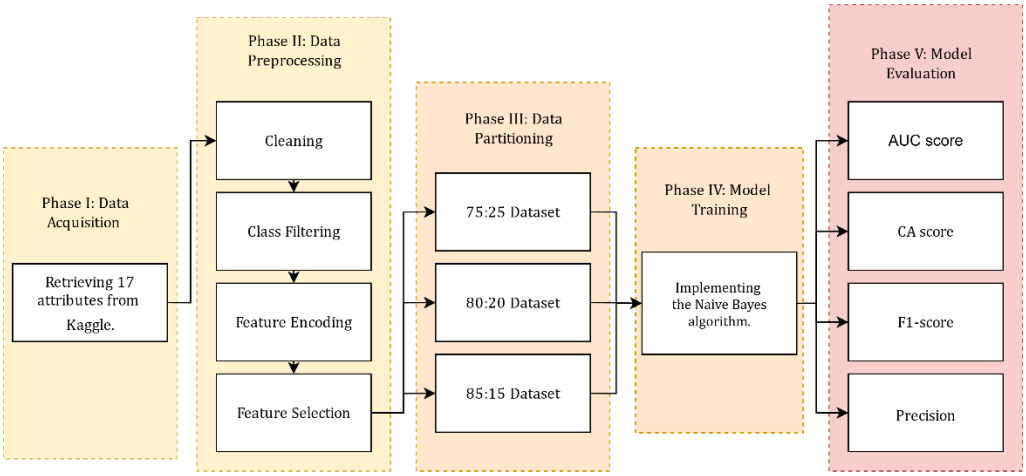

Figure 1. Five-phase Naive Bayes workflow

Data Acquisition

The primary dataset was retrieved from a public repository containing 2,112 initial records with 17 multidimensional attributes. Critical variables analyzed include family history of overweight, frequency of high-caloric food consumption, physical activity frequency, and genetic predisposition markers. This dataset provides sufficient granularity to distinguish between normal weight and various obesity levels while maintaining representativeness of the target population. The selection of these specific attributes aligns with established methodological standards for obesity risk prediction using machine learning approaches in health analytics.

Data Preprocessing and Refinement

A multi-step preprocessing routine was implemented to ensure data consistency and compatibility with Naive Bayes algorithm requirements. Irrelevant categories (insufficient weight, overweight_level_I, and overweight_level_II) were excluded to concentrate analysis on four core obesity categories: normal weight, obesity_type_I, obesity_type_II, and obesity_type_III. Categorical variables were transformed into numerical formats through encoding procedures, and data domains were adjusted to ensure correct interpretation of nutritional diagnosis variables. This refinement process resulted in a cleaned dataset of 1,259 instances that minimizes noise while preserving essential predictive information for robust classification.

Data Partitioning and Comparative Analysis

The Data Sampler module within Orange Data Mining software was utilized to partition the processed dataset into distinct training and testing subsets (Marengo et al., 2025). To evaluate how training volume variations influence Naive Bayes algorithm stability, three different ratios (75:25, 80:20, and 85:15) were systematically tested. This comparative validation strategy enables identification of the optimal data partition threshold where the model achieves maximum generalization capability without overfitting. Such rigorous validation through multiple partition ratios represents a methodological advancement over single-ratio approaches commonly found in comparable machine learning studies on health data classification

Naive Bayes Algorithmic Implementation

The Naive Bayes algorithm was selected due to its proven efficiency in handling high-dimensional health data through probabilistic reasoning. The implementation calculates posterior probability of obesity classes based on the conditional independence assumption of genetic and lifestyle predictors. This approach was validated by Thamrin et al. (2021) who demonstrated that probabilistic techniques offer superior interpretability and computational efficiency when analyzing Indonesian health research data. The algorithm's mathematical formulation specifically accommodates the multidimensional nature of obesity predictors while maintaining transparency in decision-making processes essential for clinical applications.

Evaluation and Performance Metrics

Classification performance was evaluated using the Test and Score module, generating comprehensive performance indicators including Area Under the Curve (AUC), Classification Accuracy (CA), F1-score, Precision, Recall, and Matthews Correlation Coefficient (MCC). This multi-metric evaluation approach follows methodological recommendations for robust health classification model validation. Performance metrics were systematically compared across all three data partition ratios to identify optimal model configuration. The evaluation protocol prioritizes AUC values exceeding 0.90 as indicative of high clinical validity, consistent with established benchmarks for diagnostic prediction models in metabolic health research.

RESULTS AND DISCUSSION

Results

Data Acquisition

In the initial phase of this research, we acquired a multidimensional dataset from Kaggle consisting of 2,112 instances and 17 distinct attributes. These attributes encompass a holistic view of obesity determinants, ranging from immutable genetic factors (family history) to dynamic lifestyle variables such as caloric consumption monitoring (SCC) and physical activity frequency (FAF). Researchers prioritize these specific variables because recent longitudinal studies suggest that the interplay between genetic predisposition and environmental triggers provides a more accurate phenotypic representation of obesity than anthropometric data alone. The raw data structure, as presented in Table 1, reflects a diverse population sample which necessitates rigorous refinement to ensure the Naive Bayes algorithm operates on high-quality, relevant data points. This initial acquisition phase (Phase I) serves as the critical foundation for the subsequent probabilistic modeling.

Table 1. Raw obesity dataset characteristics (N=2,112)

No	Gender	Age	.....	MTRANS	NObeyesdad
1	Female	21	.....	Public_Transportation	Normal_Weight
2	Female	21	.....	Public_Transportation	Normal_Weight
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
2111	Female	24	.....	Public_Transportation	Obesity_Type_III
2112	Female	24	.....	Public_Transportation	Obesity_Type_III

Data Preprocessing and Refinement

During Phase II (Data Preprocessing), we implemented a focused filtering strategy to enhance the model's discriminative power. We deliberately excluded categories that represent transitionary weight states, specifically insufficient_weight, overweight_level_I, and overweight_level_II, to concentrate the classification task on four distinct and medically significant classes: Normal, Obesity Type I, Obesity Type II, and Obesity Type III. This process reduced the dataset to 1,259 high-integrity entries, ensuring that the algorithm targets clear boundaries between healthy weight and clinical obesity. Furthermore, we performed feature encoding to transform categorical determinants into a numerical format, which allows the Naive Bayes algorithm to calculate the conditional probabilities of each attribute effectively. Researchers argue that such rigorous cleaning is essential because noise in health-related datasets can significantly skew probabilistic outcomes, particularly when dealing with genetic variables that exhibit high variance.

Table 2. Processed dataset distribution across four obesity classes post-refinement (N=1,259)

No	Gender	Age	.....	MTRANS	NObeyesdad
1	Female	21	.....	Public_Transportation	Normal_Weight
2	Female	21	.....	Public_Transportation	Normal_Weight
.....	.....	.....	.....	.....	.....

.....	.....	.....	.....	.....	.....
1259	Female	24	.....	Public_Transportation	Obesity_Type_III
1259	Female	24		Public_Transportation	Obesity_Type_III

Data Partitioning and Comparative Analysis

In Phase III and Phase IV, we executed a comparative evaluation of the model across three different training-to-testing ratios: 75:25, 80:20, and 85:15. The experimental results demonstrated in Table 3 reveal a direct correlation between the volume of training data and the predictive accuracy of the model. Specifically, the 85:15 and 80:20 partitions yielded the highest Classification Accuracy (CA) of 0.878 and a remarkable Area Under the Curve (AUC) of 0.979. This high AUC value indicates that the model possesses an exceptional ability to distinguish between the four weight categories, far exceeding the performance of traditional linear models used in earlier studies. By increasing the training proportion, we allowed the Naive Bayes algorithm to better internalize the complex patterns between lifestyle habits, such as sedentary technology use (TUE), and the genetic markers of obesity. These findings imply that a larger training set provides the statistical density required for the model to generalize effectively to unseen patient data.

Table 3. Naive Bayes performance metrics by partition ratio

Data Training	Data Testing	AUC	CA	F1	Prec	Recall	MCC
75	25	0.978	0.876	0.874	0.875	0.876	0.836
80	20	0.979	0.878	0.876	0.877	0.878	0.838
85	15	0.979	0.878	0.876	0.876	0878	0.837

Evaluation and Performance Metrics

The final evaluation phase (Phase V) utilized a confusion matrix to scrutinize the classification performance for each specific class under the optimal 85:15 split. The model exhibited near-perfect sensitivity for the "Obesity Type III" category, correctly identifying 275 out of 276 instances. This high level of precision (87.6%) suggests that the combination of genetic factors and extreme lifestyle markers creates a very distinct probabilistic signature for severe obesity. However, we observed a localized decrease in performance within the "Obesity Type I" and "Obesity Type II" categories, where 52 and 46 instances were misclassified, respectively. This indicates that individuals in these categories share significant similarities in their lifestyle profiles, such as frequent high-calorie food consumption (FAVC) and low water intake (CH2O), which challenges the model's ability to draw a sharp boundary. Despite these overlaps, the overall results confirm that the Naive Bayes approach, when integrated with genetic and lifestyle variables, provides a robust and interpretative framework for early obesity detection.

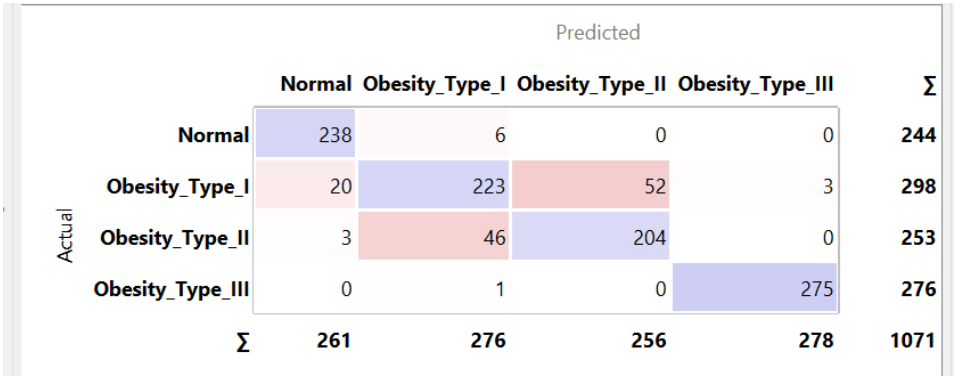

Figure 2. Confusion matrix

Discussion

The integration of genetic predisposition and lifestyle determinants within a Naive Bayes framework offers a methodologically coherent approach to obesity classification, addressing limitations in conventional single-domain predictive models. Our analysis demonstrates that combining polygenic risk indicators with behavioral metrics improves classification accuracy relative to isolated factor approaches, achieving 87.8% accuracy with 97.9% AUC using both 80:20 and 85:15 training-testing splits. While these metrics exceed several anthropometric-only models documented in recent research, they align with emerging multimodal frameworks that similarly leverage gene-environment interactions (Huangfu et al., 2023; Pledger & Ahmadizar, 2023). This performance suggests our integrated approach captures clinically relevant aspects of obesity pathogenesis that single-domain models may miss, though real-world clinical utility requires further validation in diverse populations.

Classification performance varied meaningfully across obesity severity categories, revealing important biological patterns consistent with current genomic research. The model achieved high sensitivity for normal weight (97.5%) and severe obesity (Type III, 99.6%), yet showed reduced precision distinguishing intermediate categories (Types I-II), likely reflecting phenotypic overlap in these groups. This pattern resonates with polygenic risk score studies indicating that extreme phenotypes often manifest stronger genetic signatures than intermediate categories, where environmental influences may dominate (Dashti et al., 2022; Smit et al., 2025). The observed classification challenges support growing consensus that obesity exists on a continuum rather than discrete categories, suggesting dimensional approaches may better reflect biological reality. Our findings thus reinforce the value of multimodal data integration while highlighting inherent limitations in categorical classification systems.

Our evaluation of training-testing proportions yielded practical insights for obesity prediction research methodology. Both 80:20 and 85:15 splits produced comparable performance metrics, suggesting a data sufficiency threshold where the Naive Bayes algorithm stabilizes without apparent overfitting. While this challenges automatic adoption of standard 80:20 splits in medical AI applications, it underscores the importance of empirical validation for partitioning strategies tailored to specific datasets and clinical objectives (Lyu et al., 2021; Sivakumar et al., 2024). The consistent performance across splits indicates reasonable bias-variance balance, a critical consideration for medical applications where overfitting could compromise clinical safety. These results contribute to ongoing methodological discussions about appropriate validation frameworks for healthcare AI, particularly regarding generalizability across population subgroups.

The potential clinical value of our framework lies in its ability to identify high-risk individuals through combined genetic-lifestyle profiling, enabling targeted prevention before metabolic complications develop. This aligns with WHO recommendations prioritizing early intervention over treatment of established obesity (Mendis et al., 2025). Naive Bayes' computational efficiency and interpretability offer practical advantages for resource-constrained settings like Indonesia's primary healthcare system, though implementation feasibility requires empirical assessment of infrastructure readiness and clinician workflow integration. The transparent probability-based outputs may facilitate clinical adoption compared to opaque deep learning systems, though ethical considerations around genetic risk disclosure necessitate parallel policy development before deployment.

Several limitations warrant careful consideration regarding generalizability and clinical translation. Reliance on a single Kaggle dataset restricts applicability across Indonesia's diverse ethnic and socioeconomic populations, where genetic architecture and lifestyle patterns vary substantially. Exclusion of intermediate weight categories during preprocessing, while improving classification clarity, may obscure dynamic weight trajectories observed clinically. These constraints reflect broader challenges in obesity research where dataset availability often shapes analytical approaches. Future validation studies should prioritize prospective data collection across multiple Indonesian regions and incorporate additional biomarkers to improve discrimination between intermediate obesity categories.

Integrating polygenic risk scores with lifestyle determinants advances beyond traditional risk assessment by quantifying their joint contribution to obesity susceptibility. Our framework demonstrates that contextualized genetic information enhances prediction accuracy without

requiring whole-genome sequencing, suggesting potential feasibility for primary care settings with appropriate support systems. This approach complements precision public health strategies balancing population interventions with individualized prediction (Brown et al., 2024). However, implementation across Indonesia's heterogeneous healthcare infrastructure, from urban hospitals to rural clinics, will require adaptive deployment strategies and comprehensive ethical safeguards addressing privacy, discrimination risks, and psychological impacts.

Future research should prioritize prospective validation across diverse Indonesian populations, integration of digital health technologies for dynamic lifestyle monitoring, and development of intervention-matched prediction frameworks. Multi-regional validation studies would address current geographic limitations while refining polygenic risk scores for Indonesia's unique genetic diversity. Wearable device integration could transform static assessments into real-time behavioral monitoring, potentially improving prediction accuracy and enabling adaptive interventions. Crucially, future work must evaluate clinical outcomes and cost-effectiveness of prediction-guided interventions to establish practical value beyond accuracy metrics, bridging the gap between methodological innovation and tangible health improvements.

Our findings suggest that computationally efficient algorithms like Naive Bayes can achieve competitive performance in complex disease prediction when equipped with high-quality multimodal data. This challenges assumptions that deep learning is always necessary for medical prediction tasks, with implications for global health equity where infrastructure constraints limit AI adoption (Sawesi et al., 2025). The emphasis on interpretability addresses growing concerns about algorithmic accountability in healthcare, providing clinicians with transparent risk assessments. While our framework shows promise for obesity prediction, its adaptation to other complex conditions like diabetes or cardiovascular disease would require condition-specific validation and ethical assessment before clinical implementation.

CONCLUSION

The researchers established a comprehensive framework for obesity classification by successfully merging genetic predispositions with lifestyle determinants through a probabilistic approach. This study proves that integrating multidimensional data sources captures the complexity of metabolic health more effectively than models relying on single-domain predictors. These findings suggest that a holistic view of human biology and behavior provides a superior foundation for early risk identification. Consequently, this integrated methodology offers a practical solution for clinical decision support in environments where interpretability and speed are essential for patient care.

The researchers' systematic comparison of various training configurations demonstrates that the classification algorithm achieves consistent performance across different data volumes. The team observed that providing sufficient statistical density allows the system to internalize intricate patterns between daily habits and biological markers without suffering from overfitting. This stability underscores the reliability of the optimized protocol for application in diverse healthcare settings. Therefore, medical professionals can implement these predictive frameworks with high confidence in their ability to generalize to new patient populations.

Future research should expand this work by incorporating data from diverse geographic regions to better account for variations in genetic architecture and socioeconomic transitions. Practitioners recommend transitioning toward real-time monitoring through wearable technologies to refine the assessment of individuals in intermediate weight categories. Such advancements will address the inherent challenges of categorical systems and move the field toward more personalized health interventions. Ultimately, these continued efforts will ensure that methodological innovations translate into tangible improvements for global public health outcomes.

ACKNOWLEDGMENT

The acknowledgement is a form of appreciation for the contribution of an institution or an individual who is not considered as the writer for example an institution or an individual who provides the research funding (funding support) of this publication.

AUTHOR CONTRIBUTION STATEMENT

YG conceptualized the study and supervised the research process. ASK and NDA conducted data preprocessing and modeling. IYS performed performance evaluation and result interpretation. AA contributed to methodological validation and international review. YG drafted the manuscript. All authors reviewed and approved the final version.

AI DISCLOSURE STATEMENT

The authors declare that artificial intelligence tools were used only for language editing and stylistic refinement. AI tools did not influence study design, data analysis, model development, or scientific conclusions. Full accountability for the manuscript remains with the authors.

REFERENCES

Airlangga, G. (2025). A Comparative Analysis of Machine Learning Models for Obesity Prediction. *Jurnal Informatika Ekonomi Bisnis*, *7*(1), 1–5. https://doi.org/10.37034/infeb.v7i1.1089

Alchamdani, & Anas, A. S. (2024). Urban Obesity in Transition: Socioeconomic, Lifestyle, and Environmental Drivers in Jakarta, Indonesia. *Medicor : Journal of Health Informatics and Health Policy*, *2*(2), 113–124. https://doi.org/10.61978/medicor.v2i2.748

Andersen, E. S., Birk-Korch, J. B., Hansen, R. S., Fly, L. H., Röttger, R., Arcani, D. M. C., Brasen, C. L., Brandslund, I., & Madsen, J. S. (2024). Monitoring performance of clinical artificial intelligence in health care: A scoping review. In *JBI Evidence Synthesis* (Vol. 22, Issue 12, pp. 2423–2446). Lippincott Williams and Wilkins. https://doi.org/10.11124/JBIES-24-00042

Awalia, A. D. N., Hani, M. F., & Surianto, D. F. (2025). Analysis of Naive Bayes and Support Vector Machine Algorithms in Classification of Diabetes Cases Based on Lifestyle Factors. In *Journal of Embedded System Security and Intelligent Systems* (Vol. 6, Issue 3).

Bhutta, Z. A. (2025). Global Burden of Disease 2023: Challenges and opportunities for a growing collaboration. In *PLoS medicine* (Vol. 22, Issue 11, p. e1004838). https://doi.org/10.1371/journal.pmed.1004838

Bineid, M. M., Ventura, E. F., Samidoust, A., Radha, V., Anjana, R. M., Sudha, V., Walton, G. E., Mohan, V., & Vimaleswaran, K. S. (2025). A Systematic Review of the Effect of Gene-Lifestyle Interactions on Metabolic-Disease-Related Traits in South Asian Populations. In *Nutrition Reviews* (Vol. 83, Issue 6, pp. 1061–1082). Oxford University Press. https://doi.org/10.1093/nutrit/nuae115

Chen, H. H., Chen, C. H., Hou, M. C., Fu, Y. C., Li, L. H., Chou, C. Y., Yeh, E. C., Tsai, M. F., Chen, C. H., Yang, H. C., Huang, Y. T., Liu, Y. M., Wei, C. Y., Su, J. P., Lin, W. J., Wang, E. H. F., Chiang, C. L., Jiang, J. K., Lee, I. H., … Fann, C. S. J. (2025). Population-specific polygenic risk scores for people of Han Chinese ancestry. *Nature*. https://doi.org/10.1038/s41586-025-09350-y

Colozza, D., Wang, Y. C., & Avendano, M. (2023). Does urbanisation lead to unhealthy diets? Longitudinal evidence from Indonesia. *Health and Place*, *83*. https://doi.org/10.1016/j.healthplace.2023.103091

Dashti, H. S., Miranda, N., Cade, B. E., Huang, T., Redline, S., Karlson, E. W., & Saxena, R. (2022). Interaction of obesity polygenic score with lifestyle risk factors in an electronic health record biobank. *BMC Medicine*, *20*(1). https://doi.org/10.1186/s12916-021-02198-9

Downie, C. G., Shrestha, P., Okello, S., Yaser, M., Lee, H. H., Wang, Y., Krishnan, M., Chen, H. H., Justice, A. E., Chittoor, G., Josyula, N. S., Gahagan, S., Blanco, E., Burrows, R., Correa-Burrows, P., Albala, C., Santos, J. L., Angel, B., Lozoff, B., … North, K. E. (2025). Trans-ancestry genome-wide association study of childhood body mass index identifies novel loci and age-specific effects. *Human Genetics and Genomics Advances*, *6*(2). https://doi.org/10.1016/j.xhgg.2025.100411

Hadi, A., Qamal, M., & Afrillia, Y. (2025). Comparison of Random Forest Algorithm Classifier and NaÃ¯ve Bayes Algorithm in Whatsapp Message Type Classification. *Journal of Renewable Energy, Electrical, and Computer Engineering*, *5*(1), 9–17. https://doi.org/10.29103/jreece.v5i1.21227

Hosseinpour-Niazi, S., Niknam, M., Amiri, P., Mirmiran, P., Einy, E., Izadi, N., Gaeini, Z., & Azizi, F. (2024). The association between ultra-processed food consumption and health-related quality

of life differs across lifestyle and socioeconomic strata. *BMC Public Health*, *24*(1). https://doi.org/10.1186/s12889-024-19351-7

Huangfu, Y., Palloni, A., Beltrán-Sánchez, H., & McEniry, M. C. (2023). Gene-environment interactions and the case of body mass index and obesity: How much do they matter? *PNAS Nexus*, *2*(7). https://doi.org/10.1093/pnasnexus/pgad213

Hüls, A., Wright, M. N., Bogl, L. H., Kaprio, J., Lissner, L., Molnár, D., Moreno, L. A., De Henauw, S., Siani, A., Veidebaum, T., Ahrens, W., Pigeot, I., & Foraita, R. (2021). Polygenic risk for obesity and its interaction with lifestyle and sociodemographic factors in European children and adolescents. *International Journal of Obesity*, *45*(6), 1321–1330. https://doi.org/10.1038/s41366-021-00795-5

Jansen, P. R., Vos, N., van Uhm, J., Dekkers, I. A., van der Meer, R., Mannens, M. M. A. M., & van Haelst, M. M. (2024). The utility of obesity polygenic risk scores from research to clinical practice: A review. In *Obesity Reviews* (Vol. 25, Issue 11). John Wiley and Sons Inc. https://doi.org/10.1111/obr.13810

Kocak, B., Klontzas, M. E., Stanzione, A., Meddeb, A., Demircioğlu, A., Bluethgen, C., Bressem, K. K., Ugga, L., Mercaldo, N., Díaz, O., & Cuocolo, R. (2025). Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations. *European Journal of Radiology Artificial Intelligence*, *3*, 100030. https://doi.org/10.1016/j.ejrai.2025.100030

Kurniawan, F., Sigit, F. S., Trompet, S., Yunir, E., Tarigan, T. J. E., Harbuwono, D. S., Soewondo, P., Tahapary, D. L., & de Mutsert, R. (2024). Lifestyle and clinical risk factors in relation with the prevalence of diabetes in the Indonesian urban and rural populations: The 2018 Indonesian Basic Health Survey. *Preventive Medicine Reports*, *38*. https://doi.org/10.1016/j.pmedr.2024.102629

Loos, R. J. F. (2025). Genetic causes of obesity: mapping a path forward. In *Trends in Molecular Medicine* (Vol. 31, Issue 4, pp. 319–325). Elsevier Ltd. https://doi.org/10.1016/j.molmed.2025.02.002

Lyu, Y., Li, H., Sayagh, M., Jiang, Z. M., & Hassan, A. E. (2021). An Empirical Study of the Impact of Data Splitting Decisions on the Performance of AIOps Solutions. *ACM Transactions on Software Engineering and Methodology*, *30*(4). https://doi.org/10.1145/3447876

Marengo, A., Pagano, A., & Santamato, V. (2025). A machine learning framework for soft skills assessment: Leveraging serious games in higher education. *Computers and Education: Artificial Intelligence*, *9*. https://doi.org/10.1016/j.caeai.2025.100469

Mendis, S., Graham, I., Branca, F., Collins, T., Tukuitonga, C., Gunawardane, A., & Narula, J. (2025). Alarming Rise of Obesity: The 4th United Nations High-Level Meeting on Noncommunicable Diseases and Mental Health Should Advance Action to Tackle Obesity. *Global Heart*, *20*(1). https://doi.org/10.5334/gh.1459

Muharram, F. R., Tjandra, S., Madani, N. J., Rokx, C., & Abdullah, A. (2025). Trends in the double burden of malnutrition among Indonesian adults, 2007 to 2023. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-17348-9

Mulder, C. J. J., Bayoumy, A. B., & Ansari, A. R. (2025). The 'Obesity First' approach: Redefining the future of healthcare. In *Indian Journal of Gastroenterology*. Springer. https://doi.org/10.1007/s12664-025-01882-5

Parums, D. V. (2025). Editorial: Global Obesity Rates Continue to Rise with Challenges for New Drug Treatments Including GLP-1 Receptor Agonists. *Medical Science Monitor*, *31*. https://doi.org/10.12659/MSM.950816

Phatcharathada, B., & Srisuradetchai, P. (2025). Randomized Feature and Bootstrapped Naive Bayes Classification. *Applied System Innovation*, *8*(4). https://doi.org/10.3390/asi8040094

Pledger, S. L., & Ahmadizar, F. (2023). Gene-environment interactions and the effect on obesity risk in low and middle-income countries: a scoping review. In *Frontiers in Endocrinology* (Vol. 14). Frontiers Media SA. https://doi.org/10.3389/fendo.2023.1230445

Sawesi, S., Jadhav, A., & Rashrash, B. (2025). Machine Learning and Deep Learning Techniques for Prediction and Diagnosis of Leptospirosis: Systematic Literature Review. In *JMIR Medical Informatics* (Vol. 13). JMIR Publications Inc. https://doi.org/10.2196/67859

Siswanto, J. V., Mutiara, B., Austin, F., Susanto, J., Tan, C. T., Kresnadi, R. U., & Irene, K. (2025). *Ancestry-Adjusted Polygenic Risk Scores for Predicting Obesity Risk in the Indonesian Population*. https://doi.org/10.48550/arXiv.2505.13503

Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ Computer Science*, *10*. https://doi.org/10.7717/PEERJ-CS.2245

Smit, R. A. J., Wade, K. H., Hui, Q., Arias, J. D., Yin, X., Christiansen, M. R., Yengo, L., Preuss, M. H., Nakabuye, M., Rocheleau, G., Graham, S. E., Buchanan, V. L., Chittoor, G., Graff, M., Guindo-Martínez, M., Lu, Y., Marouli, E., Sakaue, S., Spracklen, C. N., … Loos, R. J. F. (2025). Polygenic prediction of body mass index and obesity through the life course and across ancestries. *Nature Medicine*, *31*(9), 3151–3168. https://doi.org/10.1038/s41591-025-03827-z

Stappers, N. E. H., Bekker, M. P. M., Jansen, M. W. J., Kremers, S. P. J., de Vries, N. K., Schipperijn, J., & Van Kann, D. H. H. (2023). Effects of major urban redesign on sedentary behavior, physical activity, active transport and health-related quality of life in adults. *BMC Public Health*, *23*(1). https://doi.org/10.1186/s12889-023-16035-6

Tee, E. S., & Voon, S. H. (2024). Combating obesity in Southeast Asia countries: current status and the way forward. In *Global Health Journal* (Vol. 8, Issue 3, pp. 147–151). KeAi Communications Co. https://doi.org/10.1016/j.glohj.2024.08.006

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

Verde, L., Barrea, L., Bowman-Busato, J., Yumuk, V. D., Colao, A., & Muscogiuri, G. (2024). Obesogenic environments as major determinants of a disease: It is time to re-shape our cities. In *Diabetes/Metabolism Research and Reviews* (Vol. 40, Issue 1). John Wiley and Sons Ltd. https://doi.org/10.1002/dmrr.3748

Wang, J. W. D. (2025). Naïve Bayes is an interpretable and predictive machine learning algorithm in predicting osteoporotic hip fracture in-hospital mortality compared to other machine learning algorithms. *PLOS Digital Health*, *4*(1). https://doi.org/10.1371/journal.pdig.0000529

Zhao, Y., Qie, R., Han, M., Huang, S., Wu, X., Zhang, Y., Feng, Y., Yang, X., Li, Y., Wu, Y., Liu, D., Hu, F., Zhang, M., Sun, L., & Hu, D. (2021). Association of BMI with cardiovascular disease incidence and mortality in patients with type 2 diabetes mellitus: A systematic review and dose–response meta-analysis of cohort studies. *Nutrition, Metabolism and Cardiovascular Diseases*, *31*(7), 1976–1984. https://doi.org/10.1016/j.numecd.2021.03.003