Jurnal ELTIKOM :
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 8, No. 2, December 2024, page. 192-200
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v8i2.1347
Available online at http://eltikom.poliban.ac.id

RADIAL BASIS FUNCTION MODEL FOR OBESITY CLASSIFICATION BASED ON LIFESTYLE AND PHYSICAL CONDITION

Farhan Radhiansyah Razak*, Muhammad Kunta Biddinika, Herman Yuliasnyah
Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
e-mail: farhanrrazak@gmail.com, muhammad.kunta@mti.uad.ac.id, herman.yuliansyah@tif.uad.ac.id

Received: 21 October 2024 – Revised: 3 December 2024 – Accepted: 4 December 2024

ABSTRACT

Obesity is a chronic condition affecting millions worldwide, influenced by genetic predispositions, environmental factors, lifestyle habits, and excessive caloric intake surpassing energy expenditure. widespread prevalence, existing studies lack a comprehensive exploration of classification models that effectively address the complex interplay between lifestyle and physical attributes. This study tackles the absence of an optimal machine learning model for accurately classifying obesity based on these multifaceted factors. To address this gap, the study evaluates the performance of three machine learning algorithms: Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, Naïve Bayes, and K-Nearest Neighbor (KNN). The primary objectives are to identify the most accurate classification approach, analyze the strengths of these algorithms, and highlight the importance of lifestyle and physical attributes in obesity prediction. Experimental findings show that SVM with RBF kernel achieves the highest accuracy at 89%, surpassing the performance of the other models. This study advances the field of obesity classification by offering a detailed comparative analysis of machine learning algorithms and underscoring the critical role of integrating lifestyle and physical factors into predictive modeling.

Keywords*: classification, KNN, naïve bayes, obesity, SVM.*

I. INTRODUCTION

OBESITY is a chronic health condition [1] resulting from a combination of genetic, environmental, and lifestyle factors [2], [3], [4]. It occurs when calorie intake exceeds calorie expenditure[5], [6], leading to excessive body fat accumulation [7], [8]. This condition is associated with several health complications, including cardiovascular diseases, diabetes, and even mortality [9]. Globally, obesity has become a serious epidemic. According to the World Health Organization (WHO), more than 2.5 billion adults aged 18 years and older are overweight, with 890 million classified as obese [10]. Furthermore, obesity and being overweight rank as the fifth leading cause of mortality worldwide, contributing to at least 2.8 million deaths annually due to related comorbidities [11]. Thus, accurate identification and classification of obesity are crucial for preventing and managing associated diseases.

In recent years, machine learning methods have been widely applied to obesity classification, offering valuable insights into anticipating and addressing weight-related issues. These methods can assist in determining an individual's body weight category [12]. Researchers have explored various classification algorithms, such as Naïve Bayes [13], Support Vector Machine (SVM) [14], Neural Network, Decision Tree [15], and K-Nearest Neighbor (KNN) [16].

For example, a prior study comparing KNN, Naïve Bayes, and SVM algorithms found that Decision Trees achieved the highest accuracy at 84.98% for specific datasets [17]. Another study demonstrated that Naïve Bayes could predict obesity levels with moderate accuracy of 84.15% when applied to male obesity data using a 60:40 data split [5]. Similarly, KNN with k=1 yielded an accuracy of 79.96%, focusing on attributes such as weight and age [18]. Despite these advancements, existing studies have not fully explored the interactions between obesity characteristics, lifestyle factors, and physical conditions within a comprehensive classification framework.

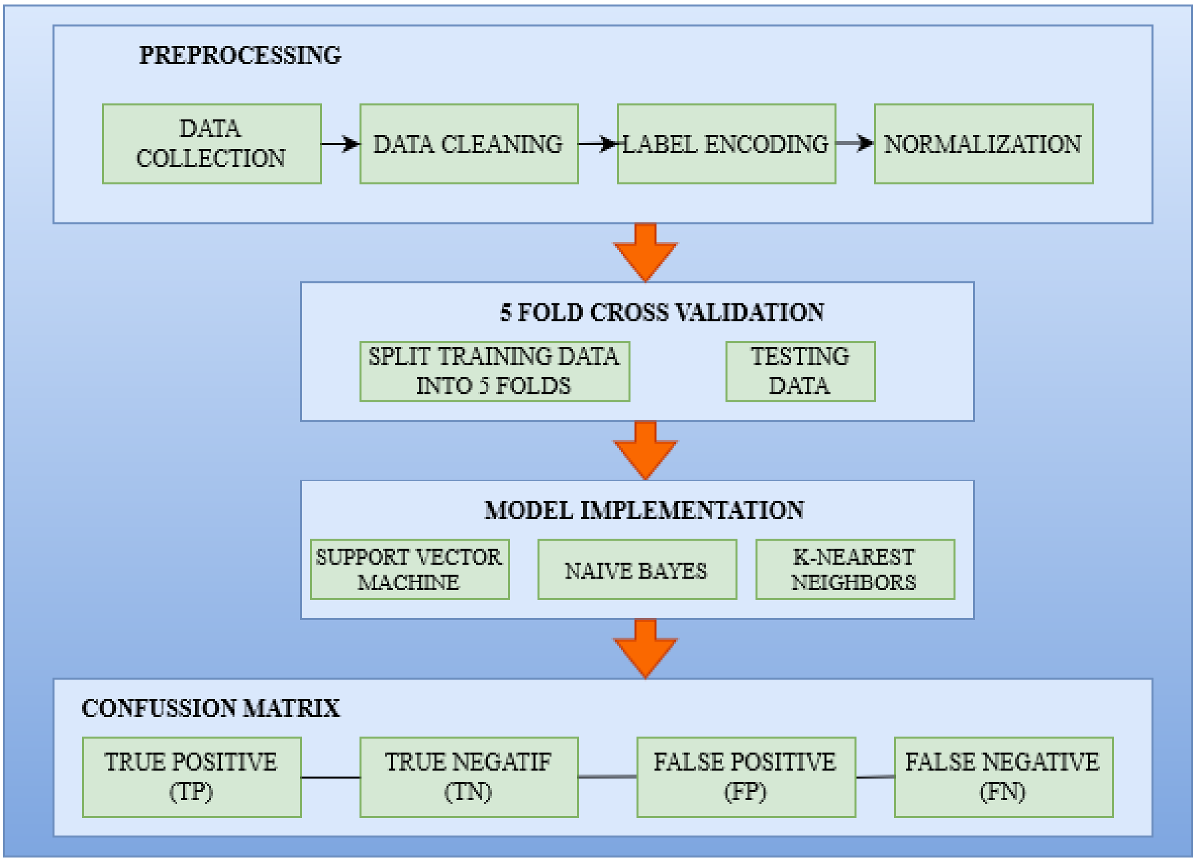

Figure 1. Research Implementation

The primary issue addressed in this study is the limited understanding of how different machine learning algorithms (SVM, Naïve Bayes, and KNN) perform in classifying obesity levels based on a combination of lifestyle factors (e.g., high-calorie food consumption, physical activity, alcohol use) and physical conditions (e.g., age, height, family history of obesity). This study seeks to fill this gap by investigating which model best captures the complex relationships within obesity-related data.

This study evaluates and compares the effectiveness of three machine learning classification methods—Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbor (KNN)—in predicting obesity based on comprehensive datasets encompassing lifestyle and physical condition factors. The objective is to determine which algorithm offers the highest accuracy and robustness for obesity classification. By providing a comparative analysis of multiple classification methods, this study contributes to the field by emphasizing the significance of integrating both lifestyle-related and physical condition attributes, offering a more holistic approach to obesity classification compared to prior research.

The novelty of this study lies in its focus on the combined impact of lifestyle and physical attributes on obesity classification. It introduces a comprehensive comparison of SVM with a Radial Basis Function (RBF) kernel, Naïve Bayes, and KNN, addressing the existing lack of consensus on optimal models for complex obesity datasets. To address the research problem, this study assesses the classification accuracy of SVM, Naïve Bayes, and KNN to identify the most suitable model for obesity classification using a balanced dataset. This analysis provides insights into the performance of each model in handling obesity-related factors, contributing valuable knowledge to academia and healthcare for more effective obesity classification.

II. Research Method

Implementing an effective machine learning model involves several critical steps. The process begins with understanding the problem, followed by data preprocessing, 5-fold cross-validation, model implementation, and evaluation using a confusion matrix. Figure 1 illustrates the stages of research implementation, detailing key steps from data collection to model evaluation.

TABLE 1
LIFESTYLE DATASET ATTRIBUTES

No	Attributes	Detail
1	FAVC	Frequent Consumption of High-Calorie Food
2	FCVC	Frequency of Consumption of Vegetables
3	NCP	Number of Meals per Day
4	CAEC	Consumption of Food between Meals
5	SMOKE	Smoking Habit
6	CH20	Water Consumption
7	SCC	Daily Calorie Intake
8	FAF	Physical Activity Frequency
9	TUE	Time Using Electronic Devices
10	CALC	Alcohol Consumption
11	MTRANS	Mode of Transportations

TABLE 2
PHYSICAL CONDITION DATASET ATTRIBUTES

No	Attributes	Detail
1	Age	Age of the person
2	Gender	Gender of the person
3	Height	Height of the person
4	Weight	Weight of the person
5	Family History	Family History of being Overweight

TABLE 3
OBESITY DATASET

Gender	Age	Height	Weight	Family History Overweight	FAFC	FCVC	NCP	CAEC	SMOKE	CH20	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
0	21	1.62	64.00	1	0	2.0	3.0	2	0	2.00	0	0.00	1.00	3	3	1
0	21	1.52	56.00	1	0	3.0	3.0	2	1	2.00	1	3.00	0.00	2	3	1
1	23	1.80	77.00	0	0	2.0	3.0	2	0	2.00	0	2.00	1.00	1	3	1
1	27	1.80	87.00	0	0	3.0	3.0	2	0	2.00	0	2.00	0.00	1	4	5
1	22	1.80	89.00	0	0	2.0	1.0	2	0	2.00	0	0.00	0.00	2	3	6
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0	20	1.71	131.40	1	1	3.0	3.0	2	0	1.72	0	1.67	0.90	2	3	4
0	21	1.74	133.74	1	1	3.0	3.0	2	0	2.00	0	1.34	0.59	2	3	4
0	22	1.75	133.68	1	1	3.0	3.0	2	0	2.05	0	1.41	0.64	2	3	4
0	24	1.73	133.34	1	1	3.0	3.0	2	0	2.85	0	1.13	0.58	2	3	4
0	23	1.73	133.47	1	1	3.0	3.0	2	0	2.86	0	1.02	0.71	2	3	4

TABLE 4
BMI CLASSIFICATION

Classification	BMI (Kg/m$^{2)}$
Infusion Weight	<18.5
Normal Weight	18.5 – 24.9
Overweight	25.0 – 29.9
Obesity Type 1	30.0 – 34.9
Obesity Type 2	35 – 39.9
Obesity Type 3	≥ 40

A. Data Collections

The dataset used for this study includes attributes categorized into two groups: Lifestyle (Table 1) and Physical Condition (Table 2). The target variable, NObeyesdad, indicates an individual's obesity level. These attributes collectively assess a person's risk of obesity. A Kaggle dataset was employed, containing information to estimate obesity levels. The dataset includes seven classes for classification: Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III. It consists of 2,111 records and 17 attributes. The class variable, NObeyesdad (Obesity Level), is detailed in Table 3.

Obesity is defined as an excessive accumulation of body fat or a body fat percentage significantly higher than average, resulting in abnormal weight relative to a person's height and age [19]. According to WHO, obesity is classified into levels based on Body Mass Index (BMI). BMI is calculated using a person's weight and height, as shown in Table 4 [20]. BMI is calculated using (1).

$$BMI = \frac{(Weight\ (kg))}{(Height\ (m)^2}$$

$$BMI = \frac{85}{(1.63)^2} = 31.99 \text{ Based on this result, the individual is classified as Obesity Level I)} \quad (1)$$

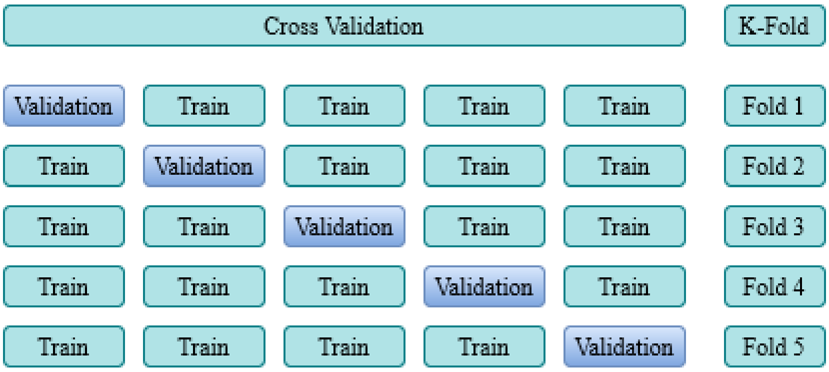

Figure 2. 5-Fold Cross Validation

B. Preprocessing

As shown in Table 3, the preprocessing analysis stage includes label encoding and data normalization. This step is crucial for preparing the data by converting categorical variables into numerical formats and scaling features to ensure consistent ranges. Proper preprocessing enables machine learning algorithms to interpret the data efficiently, reducing the risk of bias caused by varying scales and improving overall model performance. Without appropriate preprocessing, the model may struggle to learn from the data effectively, which can negatively impact its ability to generalize to unseen data.

C. 5-Fold Cross Validation

K-fold cross-validation is a widely used statistical method for evaluating the performance of machine learning models or algorithms. At this stage, the dataset is split into training and validation subsets. The model is trained on the training data and validated on the validation data across *k* iterations. This study employs 5-fold cross-validation, one of the most recommended techniques for assessing a model's generalization capability. In this process, the dataset is divided into five subsets (folds). The model training occurs five times, with a different combination of training and validation data for each iteration [21]. This method ensures that each subset is used as validation data exactly once, while the remaining subsets are used for training. Figure 2 illustrates the 5-fold cross-validation process.

D. Model Implementation

1) Support Vector Machine

The accuracy of the SVM model depends on the kernel function and parameters used. The RBF kernel is a powerful tool that transforms data into a higher-dimensional space, making it easier to classify [22]. It is particularly advantageous when dealing with non-linear data, as it can effectively capture complex relationships between features that may not be linearly separable in their original space [23]. The RBF kernel calculates the similarity between data points based on their distance in feature space. It uses the (2).

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{2}$$

Here, *K* represents the kernel function, x_ix_i and x_jx_j are data points, and $\gamma\gamma$ controls the influence of a single data point. A small $\gamma\gamma$ value results in a wider decision boundary, while a large $\gamma\gamma$ value creates a tighter boundary that may fit the training data more closely.

To achieve optimal performance with the RBF kernel, it is crucial to tune two key parameters: CC and $\gamma\gamma$. The CC parameter balances the trade-off between minimizing training errors and improving the model's ability to generalize to new data. Cross-validation is commonly employed to determine the best values for these parameters.

Figure 3 illustrates the stages of the SVM model development workflow for obesity classification. The workflow begins with Kernel Selection, where the RBF kernel is applied to map data into higher dimensions, enabling the identification of complex patterns. Next, Parameter Tuning optimizes the CC and γ\gamma parameters to balance model complexity and accuracy. During the Model Training stage, the SVM is trained on the dataset to identify optimal decision boundaries. In the Model Validation phase, the model's performance is evaluated using a validation dataset to mitigate overfitting. Finally, in the

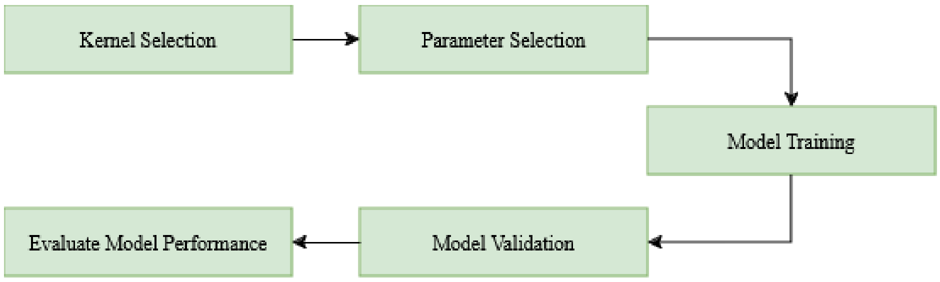

Figure 3. Support Vector Machine Workflow

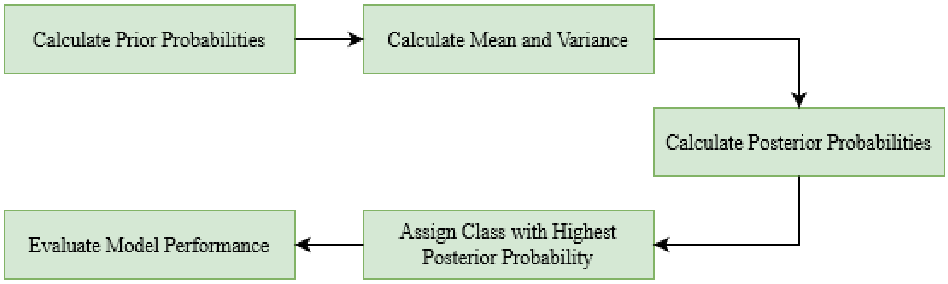

Figure 4. Naïve Bayes Workflow

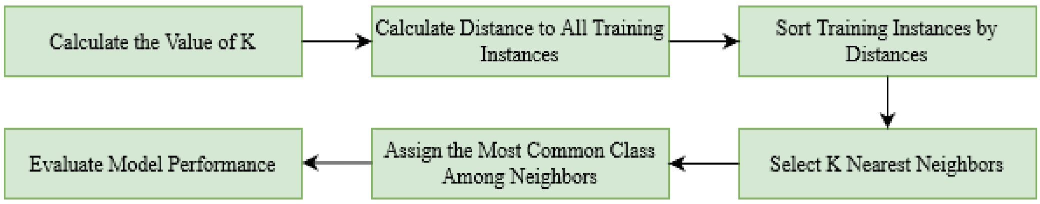

Figure 5. K-Nearest Neighbors Workflow

Evaluate Model Performance stage, metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model's effectiveness.

2) Naïve Bayes

Naïve Bayes is a probabilistic classification method that estimates probabilities in a straightforward manner by using known probabilities to calculate unknown ones. The approach is based on Bayes' Theorem, expressed in (3) [5] where X is the observed data or evidence, H is the hypothesis that X belongs to a specific class, $P(H|X)$ is the posterior probability of hypothesis H given condition X, $P(H)$ is the prior probability of hypothesis H, $P(X|H)$ is the likelihood of X given condition H, and $P(X)$ is the marginal probability of X.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{3}$$

Figure 4 illustrates the Naïve Bayes model workflow in machine learning. In Figure 4, the Gaussian Naïve Bayes algorithm classifies data by first calculating the prior probabilities for each class using the training data. It then computes the mean and variance for each feature within each class. For each test instance, the likelihood of each feature is calculated using the Gaussian distribution. Finally, Bayes' Theorem is applied to compute the posterior probability for each test instance. Afterward, the model performance is evaluated using accuracy, precision, recall, and F1-score.

3) K-Nearest Neighbors

The KNN method is a data classification technique that estimates the class of a data point by analyzing its proximity to other points. The distance between points is computed using (4) [24].

$$\mathrm{d(X,Y)} = \sqrt{\sum \mathrm{(pi - qi)\ 2\ n\ i = 1}} \tag{4}$$

Figure 5 illustrates the KNN model workflow in machine learning. In Figure 5, the KNN classification process begins by selecting an appropriate value of KK (the number of nearest neighbors). For each test instance, the algorithm iterates over the training set and computes the distance between the test instance and all training instances. The training instances are then sorted by their calculated distances, and the KK nearest neighbors are selected. The test instance is classified based on the most common class among these KK nearest neighbors. Model performance is evaluated using accuracy, precision, recall, and F1-score.

E. Confusion Matrix

The outcomes of the analytical models evaluated in this section using assessment metrics, including accuracy, precision, recall, and the F1-score. A rationale is also provided for selecting the best model for predicting obesity levels based on the comparison of results. The performance of the SVM, Naïve Bayes (NB), and K-Nearest Neighbor (KNN) models is assessed using four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [25].

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{5}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP}} \tag{6}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{TP}} \tag{7}$$

$$F1 - Score = \frac{2 \text{ x precision x recall}}{\text{precision} + \text{recall}} \tag{8}$$

Accuracy measures how often the model makes correct predictions. It is calculated as the ratio of correct predictions (TP and TN) to the total number of predictions. Equation (5) expresses accuracy [25]. Precision indicates how often the model's positive predictions are correct. It is defined as the ratio of true positives to the total predicted positives, as shown in (6) [25]. Recall (also known as sensitivity or true positive rate) measures the proportion of actual positives correctly identified by the model. It is calculated using (7) [25]. The F1-Score combines precision and recall into a single metric to provide a balanced measure, especially when both false positives and false negatives are significant. It is the harmonic mean of precision and recall, calculated using (8) [25]:

III. RESULT AND DISCUSSION

The application of the radial basis function (RBF) for obesity classification is based on physical and lifestyle conditions. The results include the model's performance and evaluation based on key metrics.

A. Result

1) Performance Metrics of SVM

Table 5 summarizes the classification metrics for the SVM model, demonstrating its superior accuracy and balance across all key metrics. The model achieves an accuracy of 89%, meaning that 89% of all predictions are correct. Its precision is also 89%, indicating that 89% of the predictions classified as positive are indeed positive. With a recall of 88%, the model successfully identifies 88% of the actual positive data. The harmonic mean of precision and recall, represented by the F1-score, is 89%, reflecting a strong balance between the two. In conclusion, the SVM model with the RBF kernel delivers the best performance in terms of accuracy, precision, recall, and F1-score, effectively handling data complexity and providing accurate and consistent predictions.

2) Performance Metrics of Naive Bayes

As shown in Table 6, the Gaussian Naive Bayes model demonstrates the lowest performance among the models. It achieves an accuracy and precision of 57%, meaning that only 57% of the predictions classified as positive are truly positive. The recall value is 55%, indicating that the model correctly

TABLE 5
CLASSIFICATION REPORT FOR SVM

Method	Accuracy	Precision	Recall	F1-Score
SVM with Radial Basis Function	0.89	0.89	0.88	0.89

TABLE 6
CLASSIFICATION REPORT FOR NAÏVE BAYES

Method	Accuracy	Precision	Recall	F1-Score
Gaussian Naïve Bayes	0.57	0.57	0.55	0.48

TABLE 7
CLASSIFICATION REPORT FOR KNN

Method	Accuracy	Precision	Recall	F1-Score
KNN with KneighborsClassifier	0.81	0.80	0.80	0.80

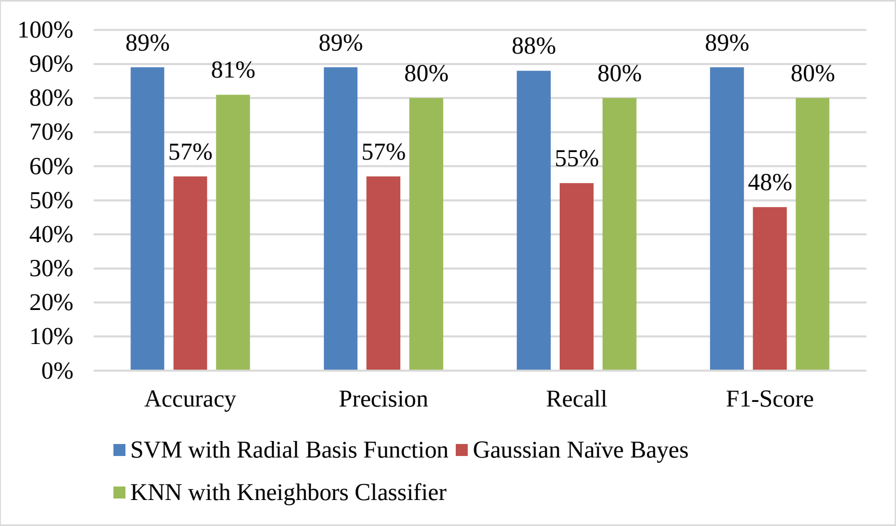

Figure 6. Comparison of Performance Report

detects 55% of the actual positive data. The F1-score is 48%, highlighting an imbalance between precision and recall. In conclusion, the Gaussian Naive Bayes model struggles to handle data complexity, resulting in less accurate and inconsistent predictions.

3) Performance Metrics of K-Nearest Neighbors

Table 7 presents the performance of the KNN method using the KNeighborsClassifier. The model achieves an accuracy of 81%, meaning that 81% of all predictions are correct. Its precision and recall are both 80%, signifying that 80% of the predictions classified as positive are truly positive and that the model successfully identifies 80% of the actual positive data. The F1-score is also 80%, reflecting a good balance between precision and recall. In conclusion, the KNN model ranks second, delivering fairly good performance. However, it is not as robust as the SVM model with the RBF kernel in handling data complexity, although it still provides acceptable results.

B. Discussion

1) Comparative Analysis

Figure 6 presents a bar chart comparing the performance metrics (accuracy, precision, recall, and F1-score) of SVM, Naïve Bayes, and KNN. The visualization highlights the superior performance of SVM with the RBF kernel across all metrics, demonstrating its strong capability to handle complex data and provide accurate predictions. The advantage of SVM with the RBF kernel in this study lies in its ability to address non-linear relationships between data attributes. The RBF kernel effectively maps features into a higher-dimensional space, capturing intricate interactions between lifestyle and physical conditions that other models, such as Naïve Bayes and KNN, may struggle to identify.

Although KNN does not perform as strongly as SVM with the RBF kernel, it still demonstrates fairly good performance. KNN remains suitable for classification tasks involving simpler data or when model interpretability is a priority.

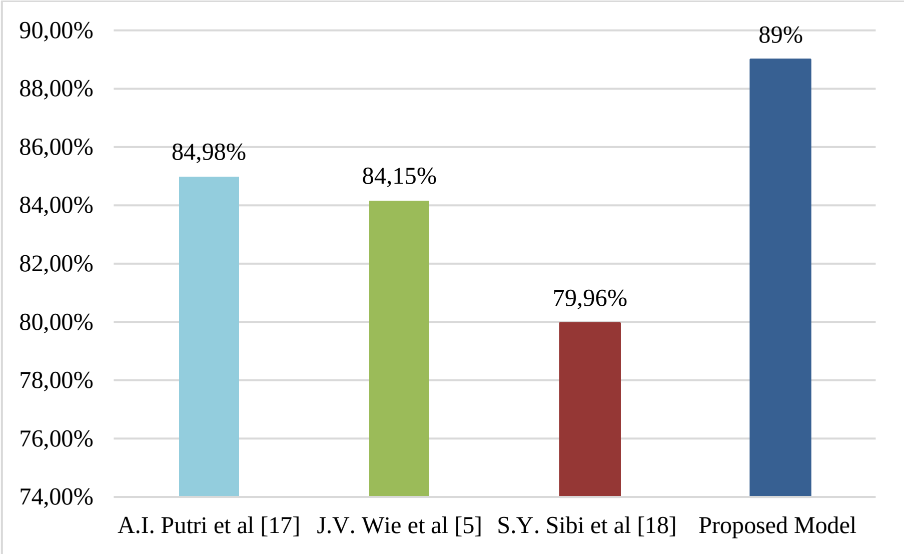

Figure 7. Comparison of the Proposed Model with Other Studies

In contrast, Gaussian Naïve Bayes shows the lowest performance, indicating its unsuitability for complex data or data that does not follow a normal distribution. However, this method remains fast and effective for problems involving simple and well-structured data.

2) Best Model Selection

A novel model has been developed to enhance prediction quality and compares favorably with existing models. This study evaluates the predictive accuracy of the Radial Basis Function (RBF), emphasizing its superior performance. As shown in Figure 7, the findings reveal that the proposed model achieves an accuracy of 89%, outperforming previous models.

The dataset for this study was obtained from Kaggle and is limited to a specific demographic, which may restrict the generalizability of the model's findings. To enhance robustness, future studies should test the model on more diverse datasets representing broader populations. This would provide a more comprehensive assessment of its applicability across different groups and help address potential demographic bias.

To further refine the model, conducting a thorough bias analysis on the dataset is essential. Techniques such as SMOTE or other oversampling methods could be employed to balance class distributions, ensuring equitable performance across all categories. Additionally, challenges such as overfitting in the SVM model with the RBF kernel, caused by hyperparameter tuning, need to be addressed. Another key concern is the scalability of the model when handling larger datasets. Future research should focus on regularization methods and evaluate the model's performance on larger and more diverse datasets to improve its generalizability. Addressing these issues will ultimately enhance the model’s robustness and expand its real-world applicability.

3) Potential Implementation in Healthcare Systems

The proposed model can be implemented in clinical or public health settings where individuals input their data based on predefined attributes such as age, weight, height, and lifestyle factors. The system would then process the input and classify the individual's obesity status. This approach can assist healthcare providers in making early and accurate diagnoses, enabling them to deliver tailored health interventions. Furthermore, integrating this model into user-friendly platforms, such as mobile applications or self-service kiosks in healthcare facilities, could enhance accessibility and user engagement.

IV. Conclusion

This study evaluated three machine learning methods—Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel, K-Nearest Neighbor (KNN), and Naïve Bayes—for classifying obesity based on lifestyle and physical conditions. The SVM model with the RBF kernel achieved an accuracy, precision, and F1-score of 89%, with a recall of 88%, demonstrating its superior performance. The SVM model holds significant potential for clinical applications, such as early obesity detection and personalized health interventions. Integrating this model with wearable devices could enable real-time health monitoring, allowing healthcare providers to design tailored health programs more effectively. Despite these promising results, the study is limited by the demographic specificity of the dataset. To improve the model's generalizability and performance, future studies should focus on testing it with larger and more diverse datasets that represent a broader range of populations, thereby enhancing its applicability.

ACKNOWLEDGMENT

This study was supported by *Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)* Universitas Ahmad Dahlan (UAD) under the *Hibah Penelitian Internal Skema Penelitian Dasar* for the year 2024. The authors gratefully acknowledge the financial assistance and support provided by LPPM UAD in facilitating this study.

REFERENCES

[1] MD. Jonathan Q. Purnell, *Definitions, Classification, and Epidemiology Obesity*. National Library of Medicine, 2023.
[2] N. Jamal *et al.*, "Safety, efficacy, and mechanism of action of herbs used for obesity management: A thematic review," *Obesity Medicine*, vol. 32, Jun. 2022
[3] O. A. Basit, E. Noorhasanah, R. Kirana, A. Rachmadi, and P. Kemenkes Banjarmasin, "Hubungan Aktivitas Fisik dengan status Gizi pada Anak Sekolah selama masa pandemi Covid-19 di SDN Karang Mekar 9 Kota Banjarmasin," *Jurnal Inovasi Penelitian*, vol. 3, no. 1, 2022.
[4] U. R. Wulandari and A. Khalifatunnisak, "Faktor Lingkungan dan Genetik penyebab Kejadian Overweight pada Balita Usia 3-5 Tahun di Kota Kediri," *Journal for Quality in Women's Health*, vol. 2, no. 2, pp. 55–58, Aug. 2019.
[5] J. V. Wie and M. Siddik, "Penerapan Metode Naive Bayes dalam Mengkalsifikasi TIngkaat Obesitas pada Pria," *JOISIE Journal Of Information System And Informatics Engineering*, vol. 6, pp. 69–77, 2022.
[6] J. Aaseth, S. Ellefsen, U. Alehagen, T. M. Sundfør, and J. Alexander, "Diets and drugs for weight loss and health in obesity," *Biomed Pharmacother*, Aug. 2021.
[7] N. Nevita and P. Arum, "The Incidence of Obesity in Adolencents in Antartika High School in Sidoardjo," *Jurnal Keperawatan*, vol. 16, no. 2, Aug. 2022.
[8] A. La Cruz, E. Severeyn, M. Huerta, and S. Wong, "Support vector machine technique as classifier of impaired body fat percentage," *IOS Press*, vol. 340, pp. 175–182, Oct. 2021.
[9] Joseph C. Wong, Sheila O'Neill, Belinda R. Beck, Mark R. Forwood, and Soo Keat Khoo, "Comparison of obesity and metabolic syndrome prevalence using fat mass index, body mass index and percentage body fat.," Jan. 2021.
[10] World Health Organization, "Obesity and Overweight," Mar. 2024.
[11] Sumarni and E. Yane Bangkele, "Persepsi Orang Tua, Guru dan Tenaga Kesehatan tentang Obesitas," *Healthy Tadulako Journal*, vol. 9, no. 1, 2023.
[12] Hanouf Al Hammadi and John J Reilly, "Classification Accuracy of Body Mass Index for Excessive Body Fatness in Kuwaiti Adolescent Girls and Young Adult Women Diabetes, Metabolic Syndrome and Obesity," 2020.
[13] Harliana and Anggraini Dewi, "Penerapan Algoritma Naïve Bayes Pada Klasifikasi Status Gizi Balita di Posyandu Desa Kalitengah (Harliana, Dewi Anggraini)," *Jurnal Informatika Komputer, Bisnis dan Manajemen*, vol. 21, no. 2, pp. 38–45, 2023.
[14] T. Hidayatulloh and L. Yusuf, "Klasifikasi Tipe Berat Tubuh Menggunakan Metode Support Vector Machine," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 71–77, Aug. 2023.
[15] E. De-La-Hoz-Correa, F. E. Mendoza-Palechor, A. De-La-Hoz-Manotas, R. C. Morales-Ortega, and S. H. B. Adriana, "Obesity level estimation software based on decision trees," *Journal of Computer Science*, vol. 15, no. 1, pp. 67–77, 2019.
[16] P. Wirawan, N. Sari, D. A. Lantana, and A. Suryaningtyas, "Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma," *Journal Of Social Science Research*, vol. 3, no. 2, pp. 13880–13892, 2023.
[17] A. I. Putri *et al.*, "Implementation of K-Nearest Neighbors, Naïve Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 1, pp. 26–33, Apr. 2024.
[18] S. Y. Sibi and A. R. Widiarti, "Klasifikasi Tingkat Obesitas Mempergunakan Algoritma KNN," *Seminar Nasional CORISINDO*, Aug. 2022.
[19] S. Septiyanti and S. Seniwati, "Obesity and Central Obesity in Indonesian Urban Communities," *Jurnal Ilmiah Kesehatan (JIKA)*, vol. 2, no. 3, pp. 118–127, Dec. 2020.
[20] C. B. Weir and A. Jan, *BMI Classification Percentile and Cut Off Points*. National Library of Medicine, 2023.
[21] Y. N. Fuadah, I. D. Ubaidullah, N. Ibrahim, F. F. Taliningsing, N. K. SY, and M. A. Pramuditho, "Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 3, p. 728, Jul. 2022.
[22] B. V. Jayadi, T. Handhayani, Manatap, and D. Lauro, "Perbandingan KNN dan SVM untuk Klasifikasi Kualitas Udara di Jakarta," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 11, no. 2, 2023.
[23] B. Sugara and A. Subekti, "Penerapan Support Vector Machine pada Small Dataset untuk Deteksi dini Gangguan Autisme," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 177–182, Sep. 2019.
[24] A. Wulandari, A. Mulya, T. Dermawan, R. R. Haiban, A. Tatamara, and H. D. Khalifah, "Application of Artificial Neural Network, K-Nearest Neighbor and Naive Bayes Algorithms for Classification of Obesity Risk Cardiovascular Disease," *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, vol. 1, no. 1, pp. 9–15, Jan. 2024.
[25] I. Widhi Saputro and B. Wulan Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction," *Citec Journal*, vol. 6, no. 1, 2019.