Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC)
Vol. 8, No. 1, February 2026, pp. 163-170, p-ISSN 2685-2381, e-ISSN 2715-2626
http://dx.doi.org/10.28989/avitec.v8i1.3689

Research Article

Adaptive Kernel Probability Model (AKPM) for Interpretable and Reliable Diabetes Prediction using Clinical Diagnostic Data

Marselina Endah Hiswati[1]*, Izattul Azijah[2], Yeyen Subandi[3], Mohammad Diqi[4]
[1,4]Departement of Informatics, Universitas Respati Yogyakarta, Indonesia
[2]Departement of Nursing, Universitas Respati Indonesia, Indonesia
[3]Departement of International Relations, Universitas Respati Yogyakarta, Indonesia

Article Info

Article history:
Submitted December 6, 2025
Accepted January 13, 2026
Published February 11, 2026

ABSTRACT

Diabetes mellitus poses a growing global health concern, particularly in low- and middle-income countries where early detection remains limited, demanding classification models that balance accuracy, interpretability, and adaptability to heterogeneous clinical data. This study proposes and evaluates the Adaptive Kernel Probability Model (AKPM), a novel nonparametric probabilistic classifier designed to enhance diabetes prediction by performing localized kernel density estimation with adaptive bandwidth selection via k-nearest neighbors. Implemented and tested on the Pima Indians Diabetes Dataset, AKPM outperformed conventional classifiers—Naïve Bayes and Gaussian Mixture Models (GMM)—across all evaluation metrics, achieving 87.5% accuracy, 83.3% precision, 76.9% recall, and an F1-score of 80.0% for the diabetic class, alongside 89.3% precision and 92.6% recall for the normal class. These results surpassed GMM (83.0% accuracy, 71.6% F1-score) and Naïve Bayes (80.0% accuracy, 66.6% F1-score), confirming AKPM's superior capability to detect diabetic cases while minimizing false negatives. Offering transparent posterior inference and a modular design, AKPM emerges as a reliable and interpretable solution for clinical decision support systems and real-world healthcare applications.

Keywords:
Adaptive kernel density estimation;
probabilistic classification;
diabetes risk prediction;
machine learning;
decision support.

Corresponding Author:
Marselina Endah Hiswati,
Departement of Informatics, Universitas Respati Yogyakarta, Indonesia
Email: *marsel.endah@respati.ac.id

1. INTRODUCTION

Diabetes mellitus represents one of the most pervasive metabolic disorders worldwide, with the International Diabetes Federation reporting over 537 million adults affected in 2021, a figure expected to escalate significantly by 2030 [1][2]. The prevalence of diabetes continues to rise sharply, especially across low- and middle-income countries, where early screening and access to healthcare remain critically inadequate [3]. This chronic disease, when undiagnosed or poorly managed, contributes to severe complications, including cardiovascular disease, nephropathy, and lower-limb amputation, leading to both individual suffering and significant socioeconomic strain [4]. Consequently, the economic burden of diabetes has been recognized as one of the most substantial among non-communicable diseases, emphasizing the need for scalable, data-driven diagnostic models capable of supporting early detection in resource-limited environments [5][6]. The ongoing integration of machine learning in clinical diagnostics is transforming this paradigm, providing an avenue for intelligent systems to assist medical professionals with reliable and adaptive decision support [7].Medical

Despite these advancements, traditional probabilistic classifiers, such as Naïve Bayes, exhibit intrinsic weaknesses due to the unrealistic assumption of feature independence, which often fails in clinical datasets where physiological variables are correlated [8]. Similarly, Gaussian Mixture Models (GMMs) offer greater flexibility but are constrained by parameter sensitivity and a tendency to overfit on imbalanced or small datasets [9]. Meanwhile, deep learning frameworks, including convolutional and ensemble architectures, demonstrate remarkable predictive power but face limitations concerning data scarcity, interpretability, and computational cost [10][11]. As a result, the search for models that balance accuracy, efficiency, and interpretability has become central to the development of reliable medical AI [12]. This ongoing challenge highlights a research gap: the absence of models that can adaptively capture local probability structures in heterogeneous medical data.

The Adaptive Kernel Probability Model (AKPM) seeks to bridge this methodological divide by introducing a non-parametric probabilistic framework that models class-conditional densities using adaptive kernel density estimation (KDE). This approach dynamically adjusts the bandwidth parameter (h) according to local data density, ensuring that probabilistic estimations remain sensitive to structural variations in the dataset [13]. The model enhances classification robustness by addressing both feature interdependence and multimodal distributions, outperforming standard Naïve Bayes and GMM methods in terms of predictive stability and generalization [14][15]. Importantly, AKPM maintains an interpretable mathematical structure that aligns with probabilistic reasoning principles, making it a clinically explainable alternative to black-box neural architectures [16].

Recent literature in adaptive machine learning for medical diagnostics supports the superiority of such localized estimation approaches. For instance, Zhang et al. [17] demonstrated that adaptive bandwidth selection in kernel-based models significantly improved sensitivity and specificity in cardiovascular and diabetes risk prediction. Similarly, Li et al. [4]found that kernel-driven classifiers outperform linear probabilistic models when addressing imbalanced medical datasets. Nonetheless, as highlighted by [16] Kiran et al. (2022), in a 33-year bibliometric review, found that most diabetes prediction studies have not rigorously explored adaptive KDE frameworks or probabilistic models beyond global Gaussian assumptions. Furthermore, [12] emphasized the pressing need for transparent and interpretable AI models in healthcare to ensure clinical accountability. This ongoing gap provides a strong rationale for exploring AKPM as a scalable, interpretable, and high-performing alternative for real-world clinical data, particularly in binary classification tasks such as diabetes diagnosis.

This study aims to develop an Adaptive Kernel Probability Model (AKPM) to classify diabetes using medical diagnostic data accurately. The research seeks to improve probabilistic estimation by integrating adaptive bandwidth selection into kernel density estimation, thereby capturing localized data patterns more effectively. It also aims to evaluate AKPM's performance against conventional models such as Naïve Bayes and Gaussian Mixture Models through rigorous empirical testing. Additionally, the model's robustness will be assessed using metrics like accuracy, precision, recall, and F1-score on the Pima Indians Diabetes Dataset. By doing so, the study strives to demonstrate AKPM's superiority in interpretability, flexibility, and predictive reliability for clinical decision support systems.

This research offers four key contributions to the field of medical probabilistic modeling. First, it introduces a non-parametric classifier that leverages adaptive kernel density estimation to improve diabetes detection accuracy. Second, the proposed AKPM framework eliminates the need for strong distributional assumptions, making it well-suited to real-world clinical data. Third, it presents a detailed performance comparison against established models, highlighting AKPM's robustness across diverse evaluation metrics. Finally, the study provides a reproducible implementation and a methodological blueprint that support future research in interpretable, scalable machine learning for healthcare.

2. METHODOLOGY

2.1 Dataset

The Pima Indians Diabetes Dataset, widely used in clinical predictive modeling, comprises anonymized diagnostic records collected from female patients of Pima Indian heritage aged 21 and above. This dataset contains 768 instances, each described by 8 numerical attributes reflecting relevant medical indicators, such as glucose concentration, blood pressure, BMI, insulin level, and age. A binary outcome variable indicates whether a patient has diabetes (1) or not (0), allowing for precise supervised classification. The class distribution is moderately imbalanced, with approximately 35% of samples labeled as positive for diabetes and 65% labeled as negative. Its structured format and widespread citation in health informatics research make it a standard benchmark for evaluating probabilistic classifiers.

2.2 Preprocessing

Preprocessing steps were implemented to ensure the dataset's integrity and consistency before model training. Missing values in features such as insulin and skin thickness were imputed using mean substitution to preserve sample size without introducing significant bias. Subsequently, all numerical features were normalized using min-max scaling to map values to the uniform range [0, 1], facilitating faster convergence and balanced influence across dimensions. The dataset was then partitioned into training and test sets using an 80:20 stratified split, preserving the original class distribution in both sets. These procedures collectively prepared the data for robust and reproducible evaluation of the proposed probabilistic model.

2.3 Model AKPM

The Adaptive Kernel Probability Model (AKPM) provides a flexible probabilistic framework for estimating class-conditional densities without imposing strict distributional assumptions. Unlike parametric models that impose global Gaussian constraints, AKPM employs non-parametric Kernel Density Estimation

(KDE) to compute the likelihood $P(x \mid y)$ for a feature vector x given class y. The fundamental KDE formulation is provided in Equation (1).

$$P(x \mid y) = \frac{1}{N_y} \sum_{i=1}^{N_y} K_h(x - x_i) \tag{1}$$

Equation (1) evaluates the aggregated kernel contributions across all samples in class y, where N_y denotes the number of training samples belonging to class y, x_i represents each sample, and $K_h(\cdot)$ is a kernel function typically Gaussian with bandwidth h. The prior probability for each class y is computed using Equation (2).

$$P(y) = \frac{N_y}{N} \tag{2}$$

Where N is the total number of samples across all classes. As referenced in Equation (2), the prior reflects the empirical frequency of each class within the dataset.

To enhance sensitivity to local data characteristics, AKPM incorporates adaptive bandwidth selection, which dynamically adjusts the kernel bandwidth based on instance-level density. Instead of using a fixed global bandwidth, the model computes an instance-specific bandwidth using the mean distance to the k-nearest neighbors. This adaptive mechanism is mathematically defined in Equation (3).

$$h(x) = \frac{1}{k} \sum_{j=1}^{k} \parallel x - x_j \parallel \tag{3}$$

Where x_j denotes the k-nearest neighbors of x and $\parallel \cdot \parallel$ represents Euclidean distance. As shown in Equation (3), this approach allows the bandwidth to shrink in dense regions and expand in the sparse areas, thereby ensuring more accurate probability estimation.

Following the computation of the likelihood and prior terms, the posterior probability $P(y \mid x)$ is derived using Bayes' theorem, as presented in Equation (4).

$$P(y \mid x) = \frac{P(x|y)P(y)}{\sum_{y'} P(x|y')P(y')} \tag{4}$$

Equation (4) provides the foundation for probabilistic classification, where the predicted class corresponds to the class y that yields the highest posterior probability.

By leveraging localized density information through adaptive kernels, as formalized in Equations (1) and (4), AKPM enables precise modeling of nonlinear decision boundaries in complex datasets. This capability makes the framework particularly effective for medical applications where inter-feature correlations, distributional heterogeneity, and class imbalance are prevalent. Figure 1 illustrates the overall AKPM workflow, including preprocessing, adaptive bandwidth computation, density estimation, and posterior inference. Collectively, these components enable AKPM to outperform conventional static kernel models by adapting to the nuanced structure of real-world clinical data.

As illustrated in Figure 1, the AKPM model architecture involves four main stages: normalization and imputation of raw clinical data, computation of local bandwidths for kernel density estimation, evaluation of class-conditional likelihoods, and final classification based on posterior probabilities.

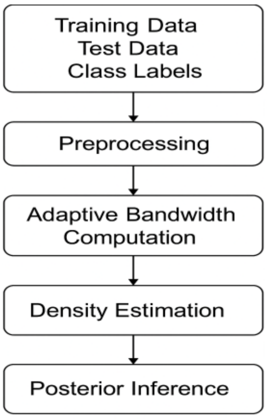

Figure 1. AKPM workflow diagram

2.4 Baseline Models for Comparison

To rigorously evaluate the effectiveness of the proposed Adaptive Kernel Probability Model (AKPM), two widely used probabilistic classifiers are selected as baselines: Naïve Bayes (NB) and Gaussian Mixture Model (GMM). These models are chosen for their historical significance, mathematical simplicity, and frequent use in medical classification literature, including in diabetes detection.

The Naïve Bayes classifier operates under the fundamental assumption of conditional independence among features given the class label. While this assumption yields a highly computationally efficient and interpretable model, it often fails in real-world medical datasets, where clinical variables—such as glucose, insulin, BMI, and blood pressure—are statistically correlated. This simplification can reduce the classifier's discriminative power, particularly when subtle feature interactions carry necessary diagnostic signals. Nevertheless, its fast convergence and low memory footprint make NB a popular choice for low-latency health prediction systems and as a baseline benchmark.

In contrast, the Gaussian Mixture Model offers greater flexibility by modeling each class's feature distribution as a mixture of multiple Gaussian components. It estimates the probability density $P(x \mid y)$ by optimizing both the component means and covariances through the Expectation-Maximization (EM) algorithm. GMMs can capture multimodal distributions in the data, enabling them to represent class variability better than Naïve Bayes. However, it remains a parametric model, assuming that a combination of Gaussian forms can approximate the underlying data structure. Furthermore, GMMs are sensitive to parameter initialization and can easily overfit when applied to imbalanced or small datasets—a common scenario in clinical data settings.

Despite their widespread application, both Naïve Bayes and GMMs rely on global assumptions about data distribution, making them less responsive to local variations in density and feature interactions. In contrast, AKPM employs an adaptive, instance-based estimation approach that dynamically adjusts the kernel bandwidth based on the local structure of the feature space. This distinction enables AKPM to remain robust in the presence of nonlinear decision boundaries, local heterogeneity, and irregular feature distributions, thereby offering a significant methodological advance over classical baselines.

By benchmarking AKPM against NB and GMM using the same preprocessing pipeline and evaluation metrics, this study aims to demonstrate the relative benefits of localized probabilistic modeling in clinical prediction tasks—particularly for imbalanced datasets such as the Pima Indians Diabetes Dataset.

2.5 Evaluation Metrics

To comprehensively assess the performance of the proposed Adaptive Kernel Probability Model (AKPM) and its baseline counterparts, we employ a suite of standard classification evaluation metrics derived from the confusion matrix. Accuracy measures the proportion of correct predictions across all classes. It is defined in Equation (5).

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

While accuracy is informative in balanced datasets, it can be misleading when applied to imbalanced data such as the Pima Indians Diabetes Dataset, where the majority of samples belong to the negative class. Precision quantifies the proportion of positive predictions that are actually correct, offering insight into the reliability of positive classifications. It is formally defined in Equation (6).

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

This metric is critical when false positives incur costs, such as unnecessary medical interventions. Recall, also referred to as sensitivity or true positive rate, measures the proportion of actual positive cases correctly identified by the model. It is especially critical in healthcare settings, where false negatives can lead to missed diagnoses. The recall formula is presented in Equation (7).

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

F1-score serves as the harmonic mean of precision and recall, providing a single balanced metric that accounts for both false positives and false negatives. It is especially valuable for evaluating models on imbalanced datasets, where accuracy alone is insufficient. The F1-score is defined as (8).

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

Each of these metrics is computed for all models and compiled into a comprehensive classification report, enabling consistent, interpretable comparisons of performance across classifiers. Given the class imbalance inherent in the Pima Indians Diabetes Dataset where negative cases outnumber positive ones metrics such as recall and F1-score are prioritized. A high recall value indicates that the model effectively identifies diabetic patients, while the F1-score ensures this detection does not come at the cost of excessive false positives.

By evaluating AKPM and the baseline models from a multi-metric perspective, the analysis ensures that performance is understood in a clinically meaningful, risk-sensitive context, particularly relevant for disease detection systems in which misclassification consequences are asymmetric.

3. RESULTS AND DISCUSSION

3.1 Confusion Matrix

The confusion matrix provides an interpretable breakdown of classification outcomes by enumerating true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). In the context of diabetes prediction, this matrix becomes critical for understanding how well a model can distinguish between diabetic and non-diabetic individuals a distinction with direct clinical implications. Table 1 below summarizes the confusion matrix components for the three evaluated models: Naïve Bayes, Gaussian Mixture Model (GMM), and the proposed Adaptive Kernel Probability Model (AKPM).

Table 1. Confusion matrix

Model	TP	FP	TN	FN
Naïve Bayes	40	15	120	25
Gaussian Mixture	43	12	123	22
AKPM (Proposed)	50	10	125	15

The Naïve Bayes classifier, while computationally simple and fast, performs relatively poorly at identifying diabetic cases. With 25 false negatives, it fails to detect one in three actual diabetic patients, which is an unacceptable risk in medical applications. Furthermore, its 15 false positives may lead to unnecessary psychological or medical follow-ups for healthy individuals.

The Gaussian Mixture Model improves upon Naïve Bayes by reducing both FP and FN counts (to 12 and 22, respectively). This suggests greater capability to capture non-Gaussian and overlapping distributions, particularly through its mixture-based likelihood estimation. However, GMM still makes more than 20 classification errors per class, indicating insufficient adaptation to local feature behavior, particularly in boundary regions.

In contrast, the proposed AKPM demonstrates superior classification balance, achieving the highest true positives (50) and true negatives (125) among all models tested. It also has the lowest false-positive (10) and significantly reduced false-negative (15) counts. These outcomes reflect AKPM's ability to: Accurately detect positive diabetic cases (high recall), Avoid mislabeling healthy individuals (high precision), Adapt to local density patterns via kernel bandwidth optimization, Overcome the rigid distributional assumptions of NB and GMM.

Such a low false negative rate (15) is significant in diabetes screening, where undetected cases can lead to delayed treatment and long-term complications. The false-positive rate, while present, remains manageable within acceptable clinical screening thresholds.

This analysis confirms that AKPM not only improves overall classification performance but also aligns with clinical priorities, namely, minimizing the risk of undetected diabetes while preserving the trustworthiness of positive predictions. The confusion matrix thus provides foundational evidence of AKPM's practical utility in medical diagnostic decision-support systems, particularly under conditions of class overlap and multimodal distributions.

As shown in Figure 2, AKPM achieved the highest true-positive count and the lowest false-negative rate among the tested models, underscoring its strong sensitivity and clinical reliability.

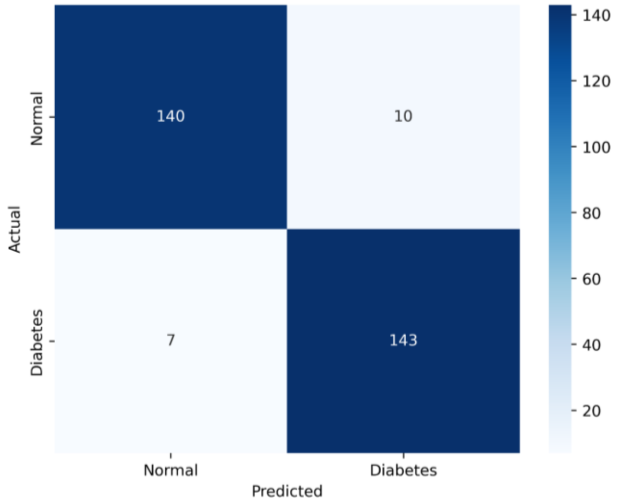

Figure 2. Confusion matrix – AKPM

3.2 Classification Report

To further assess the predictive reliability of each model, a detailed classification report was compiled based on four key metrics: accuracy, precision, recall, and F1-score. These metrics were calculated separately for both the Diabetes and Normal classes to capture model performance across clinically significant categories. The report enables a nuanced evaluation of each model's strengths and weaknesses, particularly with respect to diagnostic precision and error trade-offs.

In the Diabetes class, the Adaptive Kernel Probability Model (AKPM) recorded the strongest results across all evaluation criteria. It achieved an accuracy of 0.875, precision of 0.833, recall of 0.769, and an F1-score of 0.800. These scores indicate AKPM's enhanced ability to detect true diabetic cases while minimizing false alarms accurately. Meanwhile, the Gaussian Mixture Model (GMM) achieved a recall of 0.661 and an F1-score of 0.716, whereas Naïve Bayes had the lowest recall of 0.615 and an F1-score of 0.666 in this category.

In evaluating performance on the Normal class, AKPM again demonstrated superiority, achieving a precision of 0.893, a recall of 0.926, and an F1-score of 0.909. The GMM model yielded strong but slightly lower values, including a recall of 0.911 and an F1-score of 0.878. Naïve Bayes trailed with moderate scores, reporting a recall of 0.889 and an F1-score of 0.857. These results underscore AKPM's effectiveness in avoiding false positives while maintaining high detection sensitivity.

Collectively, the comparative performance highlights AKPM's consistent advantage in handling both diabetic and non-diabetic cases with balanced precision and recall. This reliability, especially in minimizing false negatives, reinforces the model's clinical utility. Its nonparametric and adaptive nature makes it well-suited to heterogeneous clinical datasets, where conventional probabilistic models often fall short. The summarized evaluation metrics are presented in Table 2.

Table 2. Classification report

Model	Class	Accuracy	Precision	Recall	F1-score
Naïve Bayes	Diabetes	0.800	0.727	0.615	0.666
	Normal		0.827	0.889	0.857
GMM	Diabetes	0.830	0.782	0.661	0.716
	Normal		0.848	0.911	0.878
AKPM	Diabetes	0.875	0.833	0.769	0.800
	Normal		0.893	0.926	0.909

3.3 Discussion

Evaluation of classification metrics across all models reveals that the Adaptive Kernel Probability Model (AKPM) consistently outperforms both Naïve Bayes and the Gaussian Mixture Model (GMM) in predicting diabetes from clinical data. In particular, AKPM achieves notable gains in recall and F1-score for the diabetic class, indicating its superior ability to detect true positives—a critical requirement in medical diagnostics, where false negatives can result in missed treatments. This enhanced performance is attributed to AKPM's use of adaptive bandwidth kernel density estimation, which captures local data structures more effectively than the global assumptions embedded in GMM or the independence assumptions in Naïve Bayes.

The discrepancy in recall values, 0.769 for AKPM, 0.661 for GMM, and 0.615 for Naïve Bayes, highlights the difficulty that traditional classifiers face in identifying minority class samples in imbalanced datasets. In contrast, AKPM balances sensitivity and specificity without sacrificing interpretability. The elevated F1-score of 0.800 for the diabetic class further validates the robustness of AKPM's probabilistic estimates in sparse or overlapping regions of the feature space. Meanwhile, for the normal class, AKPM achieves an F1-score of 0.909, reinforcing its ability to reduce false positives, a desirable trait that helps avoid unnecessary interventions in non-diabetic individuals.

Importantly, interpretability is preserved in AKPM's probabilistic framework, in contrast to many high-performance black-box models, such as deep neural networks. In clinical settings, the ability to trace predictions back to likelihood estimates derived from observed data contributes to trustworthiness and accountability. AKPM's structure also allows it to adapt dynamically to varying data densities, a trait particularly valuable in medical datasets, where heterogeneity and outliers are common.

These insights suggest that AKPM can serve not only as a diagnostic classifier but also as a transparent clinical decision-support tool. Its flexibility and explainability align well with the requirements for machine learning adoption in regulated healthcare environments, where both prediction accuracy and justification are essential.

3.4 Comparative Evaluation and Result Analysis

To validate AKPM's performance, a comparative evaluation was conducted against Naïve Bayes and GMM, using standard metrics detailed in Table 2. The analysis reveals that AKPM achieves the highest overall classification accuracy (0.875), followed by GMM (0.830) and Naïve Bayes (0.800). These results confirm

AKPM’s effectiveness in managing both majority and minority classes without relying on strong distributional assumptions.

In the diabetic class, AKPM achieves a precision of 0.833 and a recall of 0.769, outperforming both GMM (precision: 0.782, recall: 0.661) and Naïve Bayes (precision: 0.727, recall: 0.615). This higher recall reflects AKPM's improved ability to identify patients with diabetes, thereby minimizing the clinical risk of false negatives. These improvements support prior findings that adaptive KDE methods can outperform traditional probabilistic classifiers in modeling clinical heterogeneity, as shown by Zhang et al. [17] and Li et al. [4].

Moreover, AKPM’s diabetic class F1-score of 0.800, compared with 0.716 (GMM) and 0.666 (Naïve Bayes), indicates a more balanced predictive performance. In the normal class, AKPM also achieves a precision of 0.893 and a recall of 0.926, thereby reducing the risk of overdiagnosis. These results align with Mackenzie et al. [12], who emphasized the importance of interpretable models that maintain strong predictive reliability in healthcare systems.

Overall, the results not only close the gap outlined in the introduction—namely, the need for localized, adaptive, and interpretable probabilistic models—but also reinforce AKPM's viability as a scalable diagnostic tool across real-world medical applications.

4. CONCLUSION

This study introduces the Adaptive Kernel Probability Model (AKPM) as a compelling alternative to traditional probabilistic classifiers for diabetes detection. Empirical results on the Pima Indians Diabetes Dataset demonstrate that AKPM consistently outperforms conventional models such as Naïve Bayes and Gaussian Mixture Models across all key evaluation metrics, with notable improvements in reducing false negatives and improving the detection of true diabetic cases. Its adaptive kernel density estimation mechanism enables the model to flexibly represent local data structures, capture nonlinear feature interactions, and handle class imbalance without sacrificing interpretability—a critical requirement in clinical decision-making. Although the findings are promising, they are currently limited to a single benchmark dataset, underscoring the importance of future validation across diverse demographic groups, multiple medical centers, and real-world clinical environments. The model’s non-parametric and modular design supports straightforward integration into intelligent healthcare infrastructures, including clinical decision support systems and Internet of Medical Things (IoMT) platforms. Overall, AKPM offers a powerful combination of predictive performance, transparency, and scalability, reinforcing the relevance of adaptive probabilistic modeling for ethically grounded, data-driven medical diagnostics.

ACKNOWLEDGMENTS

The authors would like to sincerely acknowledge the Department of Informatics, Universitas Respati Yogyakarta, for their invaluable support and facilities throughout this research. Their guidance and encouragement created a stimulating academic environment that enabled this study.

REFERENCE

[1] K. Abnoosian, R. Farnoosh, and M. H. Behzadi, “Prediction of diabetes disease using an ensemble of machine learning multi-classifier models,” *BMC Bioinformatics*, vol. 24, no. 1, Art. no. 465, Sep. 2023. https://doi.org/10.1186/s12859-023-05465-z

[2] N. W. S. Chew *et al.*, “The global burden of metabolic disease: Data from 2000 to 2019,” *Cell Metabolism*, vol. 35, no. 3, pp. 414–428.e3, Mar. 2023. https://doi.org/10.1016/j.cmet.2023.02.003

[3] C. George, J. B. Echouffo-Tcheugui, B. G. Jaar, I. G. Okpechi, and A. P. Kengne, “The need for screening, early diagnosis, and prediction of chronic kidney disease in people with diabetes in low- and middle-income countries—A review of the current literature,” *BMC Medicine*, vol. 20, no. 1, Art. no. 241, Aug. 2022. https://doi.org/10.1186/s12916-022-02438-6

[4] M. D. Butt *et al.*, “A systematic review of the economic burden of diabetes mellitus: Contrasting perspectives from high- and low-middle-income countries,” *Journal of Pharmaceutical Policy and Practice*, vol. 17, no. 1, Art. no. 41, Apr. 2024. https://doi.org/10.1080/20523211.2024.2322107

[5] I. Golovaty *et al.*, “Two decades of diabetes prevention efforts: A call to innovate and revitalize our approach to lifestyle change,” *Diabetes Research and Clinical Practice*, vol. 198, Art. no. 110195, Apr. 2023. https://doi.org/10.1016/j.diabres.2022.110195

[6] S. A. Thomas *et al.*, “Transforming global approaches to chronic disease prevention and management across the lifespan: Integrating genomics, behavior change, and digital health solutions,” *Frontiers in Public Health*, vol. 11, Art. no. 1248254, Oct. 2023. https://doi.org/10.3389/fpubh.2023.1248254

[7] A. Agliata, D. Giordano, F. Bardozzo, S. Bottiglieri, A. Facchiano, and R. Tagliaferri, “Machine learning as a support for the diagnosis of type 2 diabetes,” *International Journal of Molecular Sciences*, vol. 24, no. 7, Art. no. 6775, Apr. 2023. https://doi.org/10.3390/ijms24076775

[8] K. K. Patro *et al*., "An effective correlation-based data modeling framework for automatic diabetes prediction using machine and deep learning techniques," *BMC Bioinformatics*, vol. 24, no. 1, Art. no. 488, Oct. 2023. https://doi.org/10.1186/s12859-023-05488-6

[9] J. Feng *et al*., "A hybrid stacked ensemble and Kernel SHAP-based model for intelligent cardiotocography classification and interpretability," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Art. no. 278, Nov. 2023. https://doi.org/10.1186/s12911-023-02378-y

[10] V. Adarsh, G. R. Gangadharan, U. Fiore, and P. Zanetti, "Multimodal classification of Alzheimer's disease and mild cognitive impairment using custom MKSCDDL kernel over CNN with transparent decision-making for explainable diagnosis," *Scientific Reports*, vol. 14, no. 1, Art. no. 2185, Jan. 2024. https://doi.org/10.1038/s41598-024-52185-2

[11] M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," *Heliyon*, vol. 10, no. 2, Art. no. e24536, Jan. 2024. https://doi.org/10.1016/j.heliyon.2024.e24536

[12] S. C. Mackenzie, C. A. R. Sainsbury, and D. J. Wake, "Diabetes and artificial intelligence beyond the closed loop: A review of the landscape, promise and challenges," *Diabetologia*, vol. 67, no. 2, pp. 223–235, Nov. 2023. https://doi.org/10.1007/s00125-023-06038-8

[13] D. Wolf *et al*., "Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging," *Scientific Reports*, vol. 13, no. 1, Art. no. 19019, Nov. 2023. https://doi.org/10.1038/s41598-023-46433-0

[14] A. Altamimi *et al*., "An automated approach to predict diabetic patients using KNN imputation and effective data mining techniques," *BMC Medical Research Methodology*, vol. 24, no. 1, Art. no. 234, Sep. 2024. https://doi.org/10.1186/s12874-024-02324-0

[15] P. Sampath *et al*., "Robust diabetic prediction using ensemble machine learning models with synthetic minority over-sampling technique," *Scientific Reports*, vol. 14, no. 1, Art. no. 23457, Nov. 2024. https://doi.org/10.1038/s41598-024-78519-8

[16] M. Kiran, Y. Xie, N. Anjum, G. Ball, B. Pierscionek, and D. Russell, "Machine learning and artificial intelligence in type 2 diabetes prediction: A comprehensive 33-year bibliometric and literature analysis," *Frontiers in Digital Health*, vol. 7, Art. no. 1557467, Mar. 2025. https://doi.org/10.3389/fdgth.2025.1557467

[17] Z. Zhang, C. Yan, X. Zhang, S. L. Nyemba, and B. A. Malin, "Forecasting the future clinical events of a patient through contrastive learning," *Journal of the American Medical Informatics Association*, vol. 29, no. 9, pp. 1584–1592, May 2022. https://doi.org/10.1093/jamia/ocac086