Optimizing Predictive Performance: Hyperparameter Tuning in Stacked Multi-Kernel Support Vector Machine Random Forest Models for Diabetes Identification

Dimas Chaerul Ekty Saputra [1], Alfian Ma'arif [2*], Khamron Sunat [3]

[1] Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand
[2] Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3] Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand
Email: [1] dimaschaerulekty.s@kkumail.com, [2] alfian.maarif@te.uad.ac.id, [3] skhamron@kku.ac.th
*Corresponding Author

Abstract—This study addresses the necessity for more advanced diagnostic tools in managing diabetes, a chronic metabolic disorder that leads to disruptions in glucose, lipid, and protein metabolism caused by insufficient insulin activity. The research investigates the innovative application of machine learning models, specifically Stacked Multi-Kernel Support Vector Machines Random Forest (SMKSVM-RF), to determine their effectiveness in identifying complex patterns in medical data. The innovative ensemble learning method SMKSVM-RF combines the strengths of Support Vector Machines (SVMs) and Random Forests (RFs) to leverage their diversity and complementary features. The SVM component implements multiple kernels to identify unique data patterns, while the RF component consists of an ensemble of decision trees to ensure reliable predictions. Integrating these models into a stacked architecture allows SMKSVM-RF to enhance the overall predictive performance for classification or regression tasks by optimizing their strengths. A significant finding of this study is the introduction of SMKSVM-RF, which displays an impressive 73.37% accuracy rate in the confusion matrix. Additionally, its recall is 71.62%, its precision is 70.13%, and it has a noteworthy F1-Score of 71.34%. This innovative technique shows potential for enhancing current methods and developing into an ideal healthcare system, signifying a noteworthy step forward in diabetes detection. The results emphasize the importance of sophisticated machine learning methods, highlighting how SMKSVM-RF can improve diagnostic precision and aid in the continual advancement of healthcare systems for more effective diabetes management.

Keywords—*Diabetes; Machine Learning Model; SMKSVM-RF; Hyperparameter Tuning; Predictive Performance Optimization.*

I. INTRODUCTION

Diabetes, a chronic metabolic disorder, results in abnormalities in glucose, lipid, and protein metabolism due to insufficient insulin activity [1]. This health crisis is marked by high blood sugar levels, caused by disruptions in insulin production or reduced cellular responsiveness. Dysfunction in the pancreas, the organ responsible for insulin generation, leads to glucose buildup in the bloodstream. The severe organ dysfunctions and life-threatening complications resulting from diabetes underline the crucial requirement for strong diagnostic tools and effective disease management [2], [3].

To address this urgent need, the implementation of machine learning models has become a revolutionary technique for identifying diabetes. Stacked Multi-Kernel Support Vector Machines Random Forest (MKSVM-RF) models are highly regarded for their ability to identify complex patterns within medical data. Researchers have focused on these models in the field of diabetes research. However, the effectiveness of these models is reliant on careful hyperparameter tuning, which is crucial for achieving the best predictive performance tailored to the nuances of diabetes datasets [4].

The need for effective identification of diabetes is highlighted by the increasing prevalence of the disease, demonstrated by projections from the International Diabetes Federation (IDF) [5]. For instance, Indonesia is expected to experience a rise 47% rise in its diabetic population by 2045, resulting in 28.57 million people affected. Overall, the IDF estimates that the global diabetic population will surge to 783.7 million by 2045, necessitating novel strategies for prompt detection and control [6]–[8].

This paper explores the intersection of healthcare and machine learning with an emphasis on the essential role of Stacked MKSVM with Random Forest models in improving diabetes diagnostics. Optimization of hyperparameters emerges as a crucial factor in the ever-changing healthcare landscape, ensuring the accuracy and effectiveness of these sophisticated models [9]. The following research involves an in-depth analysis of predictive analytics, incorporating machine learning models and deep learning techniques to enhance the forecasting performance of diabetes patients [10]. We aim to build simple and understandable models, identify important attributes, and utilize ensemble methods that combine the advantages of Machine Learning.

As we begin this exploration, it is crucial to acknowledge the multifaceted nature of diabetes. With two distinct types, Type 1 and Type 2, each presenting unique challenges, it is essential to understand the complexities of their etiology [11], [12]. While Type 1 diabetes is frequently autoimmune, resulting from immune system attacks on pancreatic cells that generate insulin, Type 2 diabetes is primarily due to insulin

resistance in cells [13]. A comprehensive approach to identifying diabetes should cover both types while taking into account the diverse contributing factors and tailoring diagnostic strategies accordingly [14].

This study goes beyond traditional diagnostic methods by exploring predictive analytics that utilize a combination of machine learning algorithms and statistical methodologies. The goal is to utilize current and past data to provide objective evaluations and projections that advance patient care, streamline resources, and enhance clinical outcomes. We employ advanced data preparation methods, such as the MinMax Scaler, together with machine learning algorithms, to develop models that yield swift, accurate, and practical diabetes evaluations.

In summary, this comprehensive investigation is driven by the imperative to address the growing challenges posed by diabetes through cutting-edge technological solutions [15]. By emphasizing the importance of integrating machine learning, fine-tuning hyperparameters, and utilizing predictive analytics, this study aims to refine diabetes identification methods and ultimately improve the quality of care and outcomes for individuals affected by this prevalent metabolic disorder.

Inadequate diabetes management not only affects individual health outcomes but also has a broader impact on the societal and economic spheres [16,17]. Diabetes complications pose a significant economic burden, resulting in increased healthcare costs and productivity losses. Effective strategies for identification and intervention can improve individual health and mitigate the economic burden associated with diabetes [18], [19].

Efforts to address diabetes globally extend beyond healthcare institutions and involve policymakers, researchers, and communities. Initiated to prevent and manage diabetes, emphasis is placed on a multi-faceted approach that includes education [20], lifestyle modifications [21], and innovative technologies [22]. Machine learning models can analyze broad datasets and detect subtle patterns. This aligns with a holistic approach by delivering individualized insights and predictions for those who may be at risk [23].

In the context of diabetes diagnostics, the interpretability and explainability of machine learning models are of utmost importance [24], [25]. As we examine the amalgamation of MKSVM and Random Forest models, it is vital to guarantee that the resultant prognostications are both precise and intelligible to medical experts and most significantly, to patients undergoing diagnosis. This element is becoming increasingly crucial in healthcare contexts where conclusions driven by machine learning models must be transparent and reliable.

The importance of hyperparameter tuning in machine learning models cannot be overstated [26]. By optimizing hyperparameters, models can be fine-tuned to perform better on specific datasets. When working with diabetes data, a nuanced approach to hyperparameter tuning is required to strike a balance between model complexity and generalizability [27], [28].

In the following sections of this article, we will discuss the particulars of Stacked Multi-Kernel SVM and Random Forest models, highlighting their unique advantages and reasoning behind choosing them for diabetes identification. We will focus on exploring hyperparameter tuning techniques that illuminate the methodologies used to optimize the efficiency of these models [29]. Through this comprehensive analysis, we aim to provide valuable insights that extend beyond theoretical considerations and have practical applications in the field of diabetes diagnostics.

This elaborate introduction sets the foundation for a thorough examination of diabetes identification with the help of machine learning. By combining the disease's significance, the potential for machine learning models to transform diagnostics, the critical role of hyperparameter tuning, and the broader societal and economic implications, this research aims to provide a comprehensive understanding of the complex diabetes diagnostic landscape. As the following sections unfold, we will delve into the intricacies of Stacked Multi-Kernel SVM with Random Forest models, highlighting their potential as effective tools in combating the global diabetes epidemic. This research presents significant contributions via its incorporation of these pivotal components: (a) the creation of a model embodying simplicity and comprehensibility, (b) the identification of highly significant attributes to augment interpretability, and (c) the utilization of a variety of machine learning algorithms to build a stacked model. This integration of techniques entails a complex procedure of hyperparameter tuning, designed to optimize the overall performance of the model.

II. RELATED WORK

Diabetes is a complex metabolic disorder that can occur due to inadequate insulin production or an impaired response to insulin, a critical hormone that regulates blood sugar levels [30]. Symptoms can include frequent urination, excessive thirst, sudden weight loss, vision impairment, and fatigue [31]. These symptoms can lead to severe health complications such as abortions, blindness, strokes, and organ failure [32]. Medical recommendations emphasize early detection of at-risk patients and proactive lifestyle monitoring to mitigate risk factors [33]. It is imperative to utilize precise vocabulary and adhere to grammatical correctness while maintaining formal and objective language. Subjective evaluations are excluded unless labeled as such. Conventional techniques like rule engines or score estimates have been found inadequate in identifying potential health risks [34]. However, recent advances in machine learning and feature engineering offer promising options for early detection and improved treatment outcomes in chronic diabetes [35].

Ramesh conducted a significant study using the PIMA Indian Diabetes Dataset which applied four machine learning techniques: KNN, LR, Gaussian NB, and SVM. The study aimed to collect patient information through personal healthcare equipment and wearable devices, to facilitate early detection of diabetes risk and keep medical professionals informed of developments. The SVM RBF kernel demonstrated an accuracy rating of 83.20% and the highest precision. Hassan conducted an important study that used 250

data points featuring 16 unique attributes and employed LR, SVM, and random tree models. Through ten-fold cross-validation, the random tree model demonstrated the highest accuracy, achieving an impressive 97.50% [36].

Sivaranjani's research concentrated on diabetes-related disorders, using support vector machine (SVM) and random forest (RF) machine learning methods. Principal component analysis (PCA) was utilized to decrease the number of factors considered, leading to a reduction in dimensionality. Random Forest (RF) demonstrated an accuracy rate of 83% while processing data obtained from the PIMA diabetes database. In a thorough study by Khanam using the PIDD data collection accessible via the UCI Machine Learning Repository, seven machine learning algorithms were trained. Among these methods, the Neural Network (NN) with 2 hidden layers was found to be the most accurate, achieving a remarkable accuracy rate of 88.6%. These studies collectively demonstrate the potential of machine learning techniques to advance diabetes diagnostics. They offer nuanced insights and robust predictions that can enhance patient care [37].

The high prevalence of thyroid diseases worldwide constitutes a major health concern, particularly in India, where an estimated 42 million people suffer from this condition [38]. The complexities of diagnosing thyroid-related health issues drive medical scientists to explore robust classification models. As such, our proposed model optimizes feature selection and leverages kernel-based classifiers to classify thyroid data [39]. Employing multi-kernel support vector machines, our model excels by integrating an enhanced gray wolf optimization technique for feature selection. This methodology strategically boosts the model's performance by removing insignificant features from the dataset, a crucial step in refining the classification process. The thyroid classification model proposed by [40] shows some results from another paper, with an impressive accuracy rate of 97.49%, sensitivity of 99.05%, and specificity of 94.5%. These results surpass those of previously established models. These metrics were obtained from analyzing a confusion matrix, which demonstrates the superiority of our proposed model for thyroid disease classification compared to individual approaches, existing classifiers, and optimization techniques. Thus, our model's efficacy is validated.

In the field of medical informatics, the classification of medical data is a crucial process, and machine learning (ML) algorithms have greatly enhanced the landscape of medical diagnosis. This paper presents a new ML-based medical data classification model, WOA-MKELM, that utilizes the Whale Optimization Algorithm (WOA) for feature selection in combination with the Multi-kernel Extreme Learning Machine (MKELM) [41]. The WOA-MKELM model has two distinct stages. First, the WOA-FS algorithm performs feature selection on medical data to generate a reduced feature subset [42]. Second, the MKELM algorithm assigns appropriate class labels to the medical data. The integration of the WOA-FS model, as a precursor to classification, significantly improves the detection rate. The effectiveness of the proposed WOA-MKELM model is evaluated using three distinct medical datasets, namely hepatitis, UCI-Indian Liver

Patient (UCI-ILD), and thyroid. The experimental results confirm its efficacy by achieving maximum accuracies of 98.36%, 98.72%, and 98.93% for the hepatitis, UCI-ILD, and thyroid datasets, correspondingly. This innovative approach demonstrates superior classification accuracy and robustness across various medical datasets, showcasing promising advancements in the field [41].

III. MATERIAL AND METHODS

This section outlines the materials and methods utilized in our study. In subsection A, we specify the dataset used and detail any preprocessing steps taken to ensure data quality and relevance. In subsection B, we provide a thorough explanation of the Stacked Multi-Kernel SVM and Random Forest models chosen for diabetes identification, including the reasoning behind their selection. Lastly, our study focuses on hyperparameter tuning (section C), which provides an analysis of the specific hyperparameters considered for optimization in both models. Furthermore, we discuss the methodology employed in tuning the models for optimal predictive performance, including whether a grid search, random search, or another method was used.

A. Data Collection

The study utilizes the Pima Indian Diabetes Dataset sourced from Kaggle, comprising 768 data entries with nine criteria. The dataset is categorized into normal (500 entries) and diabetes (268 entries) groups, with 80% assigned to training data and the remaining 20% to test data. Originating from the National Institute of Diabetes and Digestive and Kidney Diseases, the dataset aims to identify diabetes through various diagnostic tests. The selection criteria for participants involved being at least 21 years old and of Pima Indian descent. This dataset focuses on 768 women from Phoenix, Arizona, with 258 testing positive for diabetes and 500 testing negatives. The dataset features eight variables, including pregnancies, Oral Glucose Tolerance Test (OGTT), blood pressure, skin thickness, insulin, Body Mass Index (BMI), age, and pedigree diabetes function, to identify familial diabetes trends. Conducted biennially since 1965, the demographic study by the National Institute of Diabetes and Digestive and Kidney Diseases provides valuable information on traits associated with diabetes onset and potential consequences. The study employs various machine learning models, including support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), and a proposed model (Hyperparameter Tuning with MKSVM-RF).

B. Stacked Multi-Kernel Support Vector Machine

Stacked Multi-Kernel SVM with Random Forest (SMKSVM-RF) is an innovative ensemble learning method that combines the capabilities of two influential machine learning models, leveraging the diversity and complementary features of Support Vector Machines (SVM) and Random Forest (RF). Within this ensemble, the SVM component utilizes multiple kernels, each capturing unique patterns within the data [43]. The SMKSVM-RF model combines various types of kernels, ranging from linear to highly complex, to enable the SVM to model intricate relationships effectively [44]. Meanwhile, the Random Forest component comprises an ensemble of decision trees, which together

contribute to the model's robust and accurate prediction [45]. By integrating these models into a stacked architecture, SMKSVM-RF optimizes their strengths, improves overall predictive performance, and can be applied to tasks such as classification or regression.

SMKSVM-RF combines the strengths of SVM and RF in an ensemble learning framework [46], [47]. The SVM component utilizes multiple kernels, which are designated as

$$K_1, K_2, \ldots, K_p, \quad (1)$$

where p is the number of kernels. The decision function for each SVM in the ensemble is given by

$$f_i(X) = \sum_{j=1}^{m} \alpha_{ij} K_{ij} + b_i, \quad (2)$$

where α_{ij} represents the SVM coefficients, b_i is the bias term, and i is the index of the SVM. The overall of the SVM ensemble is expressed as

$$F(X) = \sum_{i=1}^{p} w_i f_i(X), \quad (3)$$

where w_i represents the weights assigned to each SVM.

In parallel, the RF component comprises N decision trees, each providing an independent prediction denoted as

$$\hat{Y}_j = T_j(X), \quad (4)$$

where j in ranges from 1 to N. The overall output of the RF is then obtained by averaging the predictions of individuals trees, given by

$$\hat{Y}_{RF}(X) = \frac{1}{N} \sum_{j=1}^{N} \hat{Y}_j. \quad (5)$$

The ensemble process involves stacking the outputs of the SVM ($F(X)$) and RF ($\hat{Y}_{RF}(X)$) components. The final prediction $\hat{Y}_{Ensemble}(X)$ is generated through a meta-learner, often using a weighted combination, given by

$$\hat{Y}_{Ensemble}(X) = \sigma\left(W\left([F(X), \hat{Y}_{RF}(X)]\right) + b\right), \quad (6)$$

where W is the weight matrix, b is the bias term, and σ is the activation function (e.g., softmax for classification). This comprehensive ensemble approach aims to leverage the diversity of SVMs and RF for enhanced predictive performance across various machine learning tasks. Table I show the pseudo-code for SMKSVM-RF.

C. Hyperparameter Tuning

In the process of tuning hyperparameters for the SMKSVM-RF ensemble, our focus is on optimizing crucial parameters to improve the performance of both the Stacked Multi-Kernel SVM and Random Forest elements [48]–[52]. Essential hyperparameters for the Stacked Multi-Kernel SVM comprise the selection of SVM kernels, particularly the Radial Basis Function (RBF) kernel (K_{rbf}) and the Linear kernel (K_{linear}), the SVM regularization parameter C, regulating the balance between a smooth decision boundary and precise classification of training points, and the weights w_i allocated to individual SVMs in the ensemble. The SVM decision function is denoted in Eq. (2).

The decision function of each SVM in the ensemble is represented by $f_i(X)$, where α_{ij} denotes the SVM coefficients, K_{ij} is the selected kernel function (either K_{rbf} or K_{linear}) and b_i is the bias term. The overall output of the Stacked Multi-Kernel SVM is then expressed in Eq. (5), where p is the number of SVMs.

Concurrently, the Random Forest component has critical hyperparameters that include N, the number of trees, the maximum depth of each decision tree, the minimum samples required to split an internal node, and the minimum samples needed at a leaf node. It is imperative to carefully consider these hyperparameters. The formula output from the random forest is denoted by Eq. (7). The Random Forest's output is acquired by averaging the forecasts from each decision tree, identified as Eq. (5). This collective output signifies the agreement of the Random Forest ensemble.

In the following stages, the SVM and Random Forest predictions are merged to create the input for the meta-learner, which might be another Random Forest instructed on this combined result. For Meta-learner input denoted as

$$Input_{MetaLearner} = [F(X), RF_{Output}] \quad (7)$$

TABLE I. PSEUDO-CODE SMKSVM-RF

1	**Input:**
2	X: Input features ($m \times n$ matrix)
3	Y: Target labels ($m \times 1$ vector)
4	N: Number of decision trees in Random Forest
5	C: SVM regularization parameter
6	Meta-Learner Trees: Number of decision trees in the meta-learner Random Forest
7	Meta-Learner Depth: Maximum depth of the meta-learner Random Forest
8	**Process:**
10	for each kernel in Kernels: Train SVM with kernelized decision function using Eq. (2) Combine SVM outputs using Eq. (3)
12	for each tree in range (N): Train Decision Tree using Eq. (4) RF output using Eq. (5)
14	Input_MetaLearner using Eq. (7)
16	for each tree in range(Meta-Learner Trees): Train Decision Tree using Eq. (6)
17	**Output:**
18	Return $Y_{Ensemble_j}$

IV. RESULT AND DISCUSSION

In this section, we provide an overview of the experimental setup and methodology employed in our study. For the experimental configuration (Section A), we outline the hardware and software specifications, ensuring transparency in our approach. We also elucidate how we meticulously divided the dataset into training and testing sets for robust model evaluation. Shifting to the presentation of experimental results (Section B), we start by showcasing the baseline performance of the Stacked Multi-Kernel SVM and Random Forest models. Subsequently, in the same Section, we delve into the outcomes of hyperparameter tuning, discussing the impact on model performance and presenting the optimal hyperparameter configurations. Concluding our discussion in Section C, we interpret the findings in the context of diabetes identification, exploring potential implications for future research and clinical applications.

A. Experimental Configuration

When configuring our experiments, the processing power of the MacBook Pro M1 2020 was essential for efficiently handling the data set. Python, a versatile and commonly used programming language in the field of machine learning, functioned as the backbone for our implementation, providing a rich ecosystem of libraries and tools. The data set analyzed contained 768 PIMA data points, each with nine distinct features vital for identifying diabetes (Fig. 1).
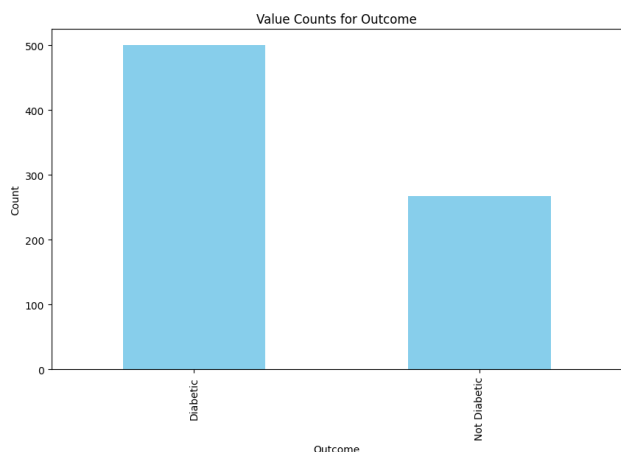

Fig. 1. Value counts for each class

To ensure a thorough evaluation of our models, we carefully partitioned the dataset, allocating 80% (615 data points) for model training and reserving the remaining 20% (154 data points) for meticulous testing. This deliberate split enabled us to assess the models' performance on previously unseen data and ensured their reliability and adaptability in real-world scenarios. The collaboration of hardware capabilities, chosen programming language, and meticulous dataset management established the groundwork for a methodical and insightful inquiry into the detection of diabetes via machine learning models.

B. Experimental Results

Building on these experimental results, the increase in accuracy from 66.88% to 73.37% demonstrates the effectiveness of hyperparameter tuning in improving the predictive power of the models. In addition to accuracy, other performance metrics such as precision, recall, and F1 score were evaluated to provide a comprehensive assessment of the models' capabilities. The precise examination of the models is essential for comprehending their accuracy and ability to identify diabetic cases accurately while minimizing the number of false positives and negatives. The nuanced perspective that refined performance metrics provide sheds light on the models' strengths and areas for improvement, thereby contributing to a comprehensive and nuanced evaluation. Table II shown the performance results of our proposed model.

TABLE II. PERFORMANCE RESULTS OF SMKSVM-RF MODEL

Model	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
SMKSVM-RF	73.37	71.62	70.13	71.34

Furthermore, the hyperparameter tuning process entailed a thorough exploration of the parameter space, testing various combinations to identify the optimal settings [53]–[56]. This rigorous methodology guarantees that the models are finely tuned to apprehend the complex relationships embedded within the diabetes dataset. The iterative refinement process successfully demonstrates the models' adaptability and responsiveness to specific configurations, emphasizing the necessity of tailored adjustments for optimal performance in medical diagnostics.

Furthermore, these findings underscore the potential of advanced machine learning models, particularly the Stacked Multi-Kernel SVM and Random Forest, in yielding accurate and dependable predictions within the realm of diabetes identification [57]–[61]. The iterative process of refining models through hyperparameter tuning indicates a dedication toward continuous enhancement, ultimately bolstering the development of diagnostic tools for diabetes and related medical conditions.

The study aims to identify the relative strengths and weaknesses of the proposed SMKSVM-RF by conducting a thorough comparison with established alternative methods such as SVM, KNN, and RF. This comparative analysis enhances the understanding of the performance of the combined SVM and RF methodology in comparison to other commonly used techniques in diabetes identification.

Our study innovates through the creation of a new approach, the fusion of SVM with RF. This combination aims to optimize on the respective strengths of SVMs and RF, utilizing the discriminative power of one and the ensemble learning capabilities of the other. By combining these two distinct methodologies, our goal is to attain a synergistic effect resulting in an innovative and ideal solution for detecting diabetes. This creative approach signifies a conscious attempt to navigate unfamiliar territories in machine learning, striving for advancements that surpass the individual capabilities of traditional models. Table III describe about the index of performance results for each class in each method.

TABLE III. INDEX OF PERFORMANCE RESULTS FOR EACH CLASS IN EACH METHOD

Model	Classes	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
SVM	Diabetes	76.62	62.22	59.57	60.87
	Non-Diabetes		85.57	84.11	83.33
RF	Diabetes	81.82	65.96	72.09	68.89
	Non-Diabetes		88.78	85.58	87.15
KNN	Diabetes	79.22	59.57	68.29	63.63
	Non-Diabetes		87.85	83.18	85.45
SMKSVM-RF	Diabetes	73.37	77.78	80.20	78.97
	Non-Diabetes		65.45	62.06	63.71

The incorporation of comparative analyses and the introduction of a pioneering approach demonstrate our dedication to advancing current methodologies. By doing so, we endeavor to provide valuable insights into the field of

medical diagnostics, encompassing not only a nuanced comprehension of model performance but also potentially paving the way for more effective and precise tools in diabetes identification.

Based on Fig. 2 about the heatmap correlation depicts the strength and direction of the linear relationships among variables in the dataset, measured by correlation coefficients. Notably, the highest correlation coefficient of 0.54 (age with pregnancies) indicates a moderate positive correlation, meaning that as one variable increases, the other tends to increase as well, although the relationship is not perfectly linear. While there is a noticeable correlation, its strength may not be adequate to suggest a direct linear correlation between the variables. Conversely, the weakest correlation, measured by a coefficient of -0.11 (age with skin thickness), shows a feeble negative correlation. In this instance, as one variable increases, the other tends to decrease slightly. The negative sign indicates an inverse relationship. However, the strength of the correlation is relatively weak. Moreover, the magnitude being close to zero implies that the variables are not strongly related. This in-depth comprehension of correlation coefficients yields insights into the nature and extent of relationships in the dataset, thereby guiding further analysis and interpretation.
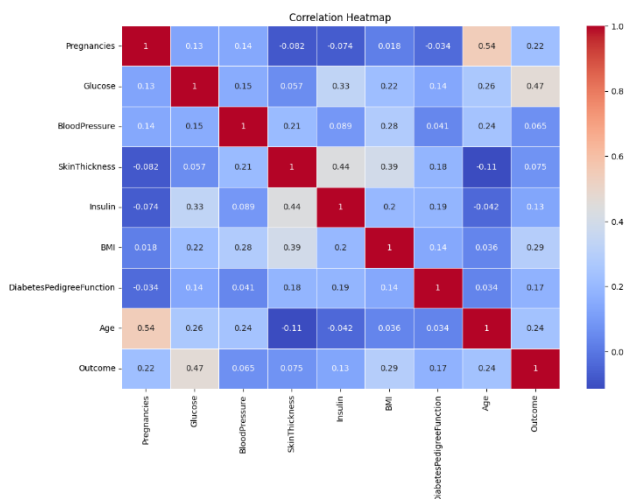

Fig. 2. Correlation heatmap

C. Discussions

In the realm of medical diagnosis, misidentifying healthy individuals as unwell and vice versa can have severe repercussions [62]. The increasing use of data mining technologies in healthcare emphasizes the need for accurate diagnosis [63]. This study aims to advance diabetes classification with the utilization of an SMKSVM-RF model. The primary objective is to develop a decision support system that enhances medical professionals' decision-making capabilities and offers valuable insights for more accurate diagnoses. We aim to create a system that can assist clinicians in making more precise diagnoses for diabetes patients [64].

Diabetes, or diabetes mellitus, is indicated by high glucose levels in the bloodstream and occurs when the body produces an insufficient amount of insulin or is unable to efficiently use it [65]. These deficiencies lead to a range of diabetic symptoms. Prolonged high levels of glucose carry

the risk of causing damage to vascular and neural systems, which in turn can lead to complications in cardiovascular, renal, ophthalmic, and podiatric health [66]. Diabetes remains a widely prevalent global health concern, impacting 415 million people, or one out of every eleven adults. Alarmingly, underdiagnosis of diabetes reaches 46%, underscoring the critical need for timely and precise detection [67]. Certain populations, including Pakistani and Indian women, demonstrate elevated susceptibility to diabetes, potentially arising from a combination of genetic, economic, and lifestyle factors [68]. In light of these complexities, researchers have leveraged artificial intelligence, particularly machine learning, to forecast diabetes and improve early intervention tactics [69]–[74].

The incorporation of machine learning and deep learning techniques has greatly impacted research related to diabetes, where datasets like the PIMA Diabetes Dataset have played a crucial role. PIMA Diabetes Dataset obtained from the Phoenix Pregnancy Incidence Database (PPID), contains data on 768 women living in the Phoenix region. Its substantial usage in multiple studies underscores its importance. The computational capacity of these approaches allows for the prediction of diabetes in its early stages, promoting proactive medical care. A variety of machine learning-based algorithmic methods have emerged throughout the medical field, including SVM [75], ANN [76], ELM [77], AdaBoost [78], RF [79], Bagging [80], KNN [81], and DNN [53][82]. This landscape of methodologies reflects the ongoing commitment to advancing diagnostic precision through innovative computational techniques (Table IV).

TABLE IV. Competitive Results From Another Method

Authors	Year	Dataset	Method	Accuracy (%)
Joshi et al. [83]	2021	PIDD	**LR** and ML	78.26
Gupta et al. [10]	2022	PIDD	QML and DL	95.00
Krishnamoorthi et al. [84]	2022	PIDD	LR, KNN, SVM, RF, **Proposed Framework LR**	83.00
Kibria et al. [85]	2022	PIDD	ANN, RF, SVM, LR, AdaBoost, XGBoost, and **Ensemble (RF, XGB)**	90.00
Proposed Model	**2023**	**PIDD**	**SMKSVM-RF**	**73.37**

V. Conclusion

Diabetes, a common health concern, is affected by considerations such as genetic predisposition and lifestyle decisions. The ongoing dialogue regarding diabetes between affected individuals and non-diabetics remains a topic of interest, as there is a growing awareness of the various risk factors that are linked to the condition. This study examines the domain of artificial intelligence, showcasing its effectiveness in identifying complex patterns and achieving high levels of accuracy. Furthermore, the combination of machine learning and deep learning techniques highlights the potential for synergy in creating a third strategy that is both efficient and precise. This study makes notable contributions, including the introduction of a novel approach called SMKSVM-RF, which resulted in an increased accuracy rate of 73.37%, recall is 71.62%, precision is 70.13%, and it has

a noteworthy F1-Score of 71.34% at confusion matrix. Anticipating that this innovative technique will enhance existing methods, the formula applied exhibits potential for developing into an optimal healthcare system, marking a substantial progression in the realm of diabetes detection.

ACKNOWLEDGMENT

REFERENCES

[1] Z. Mushtaq, M. F. Ramzan, S. Ali, S. Baseer, A. Samad, and M. Husnain, "Voting Classification-Based Diabetes Mellitus Prediction Using Hypertuned Machine-Learning Techniques," *Mobile Information Systems*, vol. 2022, pp. 1–16, 2022, doi: 10.1155/2022/6521532.

[2] L. Jiang *et al.*, "A global view of hypertensive disorders and diabetes mellitus during pregnancy," *Nat. Rev. Endocrinol.*, vol. 18, no. 12, pp. 760–775, 2022, doi: 10.1038/s41574-022-00734-y.

[3] F. Hill-Briggs *et al.*, "Social Determinants of Health, Race, and Diabetes Population Health Improvement: Black/African Americans as a Population Exemplar," *Curr. Diab. Rep.*, vol. 22, no. 3, pp. 117–128, 2022, doi: 10.1007/s11892-022-01454-3.

[4] M. Nagassou, R. W. Mwangi, and E. Nyarige, "A Hybrid Ensemble Learning Approach Utilizing Light Gradient Boosting Machine and Category Boosting Model for Lifestyle-Based Prediction of Type-II Diabetes Mellitus," *JDAIP*, vol. 11, no. 4, pp. 480–511, 2023, doi: 10.4236/jdaip.2023.114025.

[5] K. Ogurtsova *et al.*, "IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021," *Diabetes Research and Clinical Practice*, vol. 183, p. 109118, 2022, doi: 10.1016/j.diabres.2021.109118.

[6] H. Sun *et al.*, "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Research and Clinical Practice*, vol. 183, p. 109119, 2022, doi: 10.1016/j.diabres.2021.109119.

[7] R. L. Thomas, S. Halim, S. Gurudas, S. Sivaprasad, and D. R. Owens, "IDF Diabetes Atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018," *Diabetes Research and Clinical Practice*, vol. 157, p. 107840, 2019, doi: 10.1016/j.diabres.2019.107840.

[8] W. Perng, R. Conway, E. Mayer-Davis, and D. Dabelea, "Youth-Onset Type 2 Diabetes: The Epidemiology of an Awakening Epidemic," *Diabetes Care*, vol. 46, no. 3, pp. 490–499, 2023, doi: 10.2337/dci22-0046.

[9] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nat. Med.*, vol. 29, no. 8, pp. 1930–1940, 2023, doi: 10.1038/s41591-023-02448-8.

[10] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, "Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction," *Complex Intell. Syst.*, vol. 8, no. 4, pp. 3073–3087, 2022, doi: 10.1007/s40747-021-00398-7.

[11] A. Farooqi *et al.*, "A systematic review and meta-analysis to compare the prevalence of depression between people with and without Type 1 and Type 2 diabetes," *Primary Care Diabetes*, vol. 16, no. 1, pp. 1–10, 2022, doi: 10.1016/j.pcd.2021.11.001.

[12] M. Akbarizadeh, M. Naderi Far, and F. Ghaljaei, "Prevalence of depression and anxiety among children with type 1 and type 2 diabetes: a systematic review and meta-analysis," *World J. Pediatr.*, vol. 18, no. 1, pp. 16–26, 2022, doi: 10.1007/s12519-021-00485-2.

[13] B. A. Al-Ghamdi, J. M. Al-Shamrani, A. M. El-Shehawi, I. Al-Johani, and B. G. Al-Otaibi, "Role of mitochondrial DNA in diabetes Mellitus Type I and Type II," *Saudi Journal of Biological Sciences*, vol. 29, no. 12, p. 103434, 2022, doi: 10.1016/j.sjbs.2022.103434.

[14] G. T. Russo, V. Manicardi, M. C. Rossi, E. Orsi, and A. Solini, "Sex-and gender-differences in chronic long-term complications of type 1 and type 2 diabetes mellitus in Italy," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 32, no. 10, pp. 2297–2309, 2022, doi: 10.1016/j.numecd.2022.08.011.

[15] B. M. Schmidt, "Emerging Diabetic Foot Ulcer Microbiome Analysis Using Cutting Edge Technologies," *J. Diabetes Sci. Technol.*, vol. 16, no. 2, pp. 353–363, 2022, doi: 10.1177/1932296821990097.

[16] D. Costa *et al.*, "Social Aspects of Diabetic Foot: A Scoping Review," *Social Sciences*, vol. 11, no. 4, p. 149, 2022, doi: 10.3390/socsci11040149.

[17] L. E. Glenn, C. B. Thurlow, and M. Enriquez, "The "Ups and Downs" of Living With Type 2 Diabetes Among Working Adults in the Rural South," *J. Prim. Care Community Health*, vol. 13, 2022, doi: 10.1177/21501319221143715.

[18] S. Vidolov, S. Geiger, and E. Stendahl, "Affective Resonance and Durability in Political Organizing: The case of patients who hack," *Organization Studies*, vol. 44, no. 9, pp. 1413–1438, 2023, doi: 10.1177/01708406231162002.

[19] L. E. Egede *et al.*, "Nonmedical Interventions For Type 2 Diabetes: Evidence, Actionable Strategies, And Policy Opportunities: Review article examines type 2 diabetes nonmedical interventions," *Health Affairs*, vol. 41, no. 7, pp. 963–970, 2022, doi: 10.1377/hlthaff.2022.00236.

[20] S. Ciężki, E. Kurpiewska, A. Bossowski, and B. Głowińska-Olszewska, "Multi-Faceted Influence of Obesity on Type 1 Diabetes in Children – From Disease Pathogenesis to Complications," *Front. Endocrinol.*, vol. 13, p. 890833, 2022, doi: 10.3389/fendo.2022.890833.

[21] M. Levy, S. Israel, and M. Pauzner, "Multifaceted Requirements Engineering: Developing A MESH (Municipal-Environmental-Social-Health) Platform," in *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, pp. 24–29, 2022, doi: 10.1109/REW56159.2022.00014.

[22] C. Papoutsi *et al.*, "Implementation and delivery of group consultations for young people with diabetes in socioeconomically deprived, ethnically diverse settings," *BMC Med.*, vol. 20, no. 1, p. 459, 2022, doi: 10.1186/s12916-022-02654-0.

[23] S. Neethirajan, "Artificial Intelligence and Sensor Innovations: Enhancing Livestock Welfare with a Human-Centric Approach," *Human-Centric Intelligent Systems*, pp. 1-16, 2023, doi: 10.1007/s44230-023-00050-2.

[24] R. Hendawi, J. Li, and S. Roy, "A Mobile App That Addresses Interpretability Challenges in Machine Learning–Based Diabetes Predictions: Survey-Based User Study," *JMIR Form. Res.*, vol. 7, p. e50328, 2023, doi: 10.2196/50328.

[25] F. Curia, "Explainable and transparency machine learning approach to predict diabetes develop," *Health Technol.*, vol. 13, no. 5, pp. 769–780, 2023, doi: 10.1007/s12553-023-00781-z.

[26] N. Bacanin, C. Stoean, M. Zivkovic, M. Rakic, R. Strulak-Wójcikiewicz, and R. Stoean, "On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting," *Energies*, vol. 16, no. 3, p. 1434, 2023, doi: 10.3390/en16031434.

[27] Y. Ali, E. Awwad, M. Al-Razgan, and A. Maarouf, "Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity," *Processes*, vol. 11, no. 2, p. 349, 2023, doi: 10.3390/pr11020349.

[28] M. Misdram, E. Noersasongko, and F. Y. Pamuji, "Gaussian Based-SMOTE Method for Handling Imbalanced Small Datasets," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 4, 2023, doi: http://dx.doi.org/10.26555/jiteki.v9i4.26881.

[29] F. Zhang, M. Petersen, L. Johnson, J. Hall, and S. E. O'Bryant, "Hyperparameter Tuning with High Performance Computing Machine Learning for Imbalanced Alzheimer's Disease Data," *Applied Sciences*, vol. 12, no. 13, p. 6670, 2022, doi: 10.3390/app12136670.

[30] D. Rotariu *et al.*, "Oxidative stress – Complex pathological issues concerning the hallmark of cardiovascular and metabolic disorders," *Biomedicine & Pharmacotherapy*, vol. 152, p. 113238, 2022, doi: 10.1016/j.biopha.2022.113238.

[31] C. Xin, H. Fan, J. Xie, J. Hu, X. Sun, and Q. Liu, "Impact of Diabetes Mellitus on Lower Urinary Tract Symptoms in Benign Prostatic Hyperplasia Patients: A Meta-Analysis," *Front. Endocrinol.*, vol. 12, p. 741748, 2022, doi: 10.3389/fendo.2021.741748.

[32] A. Khanam, G. Hithamani, J. Naveen, S. R. Pradeep, S. Barman, and K. Srinivasan, "Management of Invasive Infections in Diabetes Mellitus: A Comprehensive Review," *Biologics*, vol. 3, no. 1, pp. 40–71, 2023, doi: 10.3390/biologics3010004.

[33] K. Sękowski, J. Grudziąż-Sękowska, J. Pinkas, and M. Jankowski, "Public knowledge and awareness of diabetes mellitus, its risk factors, complications, and prevention methods among adults in Poland—A 2022 nationwide cross-sectional survey," *Front. Public Health*, vol. 10, p. 1029358, 2022, doi: 10.3389/fpubh.2022.1029358.

[34] E. Vlachou, A. Ntikoudi, D. A. Owens, M. Nikolakopoulou, T. Chalimourdas, and O. Cauli, "Effectiveness of cognitive behavioral therapy-based interventions on psychological symptoms in adults with type 2 diabetes mellitus: An update review of randomized controlled trials," *Journal of Diabetes and its Complications*, vol. 36, no. 5, p. 108185, 2022, doi: 10.1016/j.jdiacomp.2022.108185.

[35] P. Nagaraj, P. Deepalakshmi, V. Muneeswaran, and K. Muthamil Sudar, "Sentiment Analysis on Diabetes Diagnosis Health Care Using Machine Learning Technique," in *Congress on Intelligent Systems*, vol. 114, pp. 491–502, 2022, doi: 10.1007/978-981-16-9416-5_35.

[36] J. Ramesh, R. Aburukba, and A. Sagahyroon, "A remote healthcare monitoring framework for diabetes prediction using machine learning," *Healthc. technol. lett.*, vol. 8, no. 3, pp. 45–57, 2021, doi: 10.1049/htl2.12010.

[37] S. Sivaranjani, S. Ananya, J. Aravinth, and R. Karthika, "Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 141–146, 2021, doi: 10.1109/ICACCS51430.2021.9441935.

[38] D. Das, A. Banerjee, A. B. Jena, A. K. Duttaroy, and S. Pathak, "Essentiality, relevance, and efficacy of adjuvant/combinational therapy in the management of thyroid dysfunctions," *Biomedicine & Pharmacotherapy*, vol. 146, p. 112613, 2022, doi: 10.1016/j.biopha.2022.112613.

[39] M. H. Alshayeji, "Early Thyroid Risk Prediction by Data Mining and Ensemble Classifiers," *MAKE*, vol. 5, no. 3, pp. 1195–1213, 2023, doi: 10.3390/make5030061.

[40] K. H. Priya and K. Valarmathi, "Big Data Analytics on Thyroid: A Review on Algorithms, Attributes and Future," in *2022 1st International Conference on Computational Science and Technology (ICCST)*, pp. 1056–1061, 2022, doi: 10.1109/ICCST55948.2022.10040437.

[41] M. Y. M. Parvees and M. Raja, "Optimal Feature Subset Selection with Multi-Kernel Extreme Learning Machine for Medical Data Classification," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 6, pp. 3612–3623, 2021, doi: https://doi.org/10.17762/turcomat.v12i6.7157.

[42] A. M. Siregar and S. Faisal, "Optimized Machine Learning Performance with Feature Selection for Breast Cancer Disease Classification," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 4, 2023, doi: http://dx.doi.org/10.26555/jiteki.v9i4.27527.

[43] A. Rahman *et al.*, "Robust biometric system using session invariant multimodal EEG and keystroke dynamics by the ensemble of self-ONNs," *Computers in Biology and Medicine*, vol. 142, p. 105238, 2022, doi: 10.1016/j.compbiomed.2022.105238.

[44] F. Melky, S. Sendari, and I. A. Elbaith, "Optimization of Heavy Point Position Measurement on Vehicles Using Support Vector Machine," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, 2023, doi: 10.26555/jiteki.v9i3.26261.

[45] D. C. E. Saputra, Y. Maulana, E. Faristasari, A. Ma'arif, and I. Suwarno, "Machine Learning Performance Analysis for Classification of Medical Specialties," in *Proceeding of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics*, vol. 1008, pp. 513–528, 2023, doi: 10.1007/978-981-99-0248-4_34.

[46] H. Syahputra and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 1, 2023, doi: http://dx.doi.org/10.26555/jiteki.v9i1.25861.

[47] H. Saputra, D. Stiawan, and H. Satria, "Malware Detection in Portable Document Format (PDF) Files with Byte Frequency Distribution (BFD) and Support Vector Machine (SVM)," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 4, 2023, doi: http://dx.doi.org/10.26555/jiteki.v%25vi%25i.27559.

[48] A. Pfob, S.-C. Lu, and C. Sidey-Gibbons, "Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison," *BMC Med. Res. Methodol.*, vol. 22, no. 1, p. 282, 2022, doi: 10.1186/s12874-022-01758-8.

[49] W. Pannakkong, K. Thiwa-Anont, K. Singthong, P. Parthanadee, and J. Buddhakulsomsiri, "Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–17, 2022, doi: 10.1155/2022/8513719.

[50] J. Isabona, A. L. Imoize, and Y. Kim, "Machine Learning-Based Boosted Regression Ensemble Combined with Hyperparameter Tuning for Optimal Adaptive Learning," *Sensors*, vol. 22, no. 10, p. 3776, 2022, doi: 10.3390/s22103776.

[51] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, and N. Aydin, "Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases," *Computational Biology and Chemistry*, vol. 97, p. 107619, 2022, doi: 10.1016/j.compbiolchem.2021.107619.

[52] I. D. Raji, H. Bello-Salau, I. J. Umoh, A. J. Onumanyi, M. A. Adegboye, and A. T. Salawudeen, "Simple Deterministic Selection-Based Genetic Algorithm for Hyperparameter Tuning of Machine Learning Models," *Applied Sciences*, vol. 12, no. 3, p. 1186, 2022, doi: 10.3390/app12031186.

[53] L. Wu, G. Perin, and S. Picek, "I Choose You: Automated Hyperparameter Tuning for Deep Learning-based Side-channel Analysis," *IEEE Trans. Emerg. Topics Comput.*, pp. 1–12, 2022, doi: 10.1109/TETC.2022.3218372.

[54] D. Passos and P. Mishra, "A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks," *Chemometrics and Intelligent Laboratory Systems*, vol. 223, p. 104520, 2022, doi: 10.1016/j.chemolab.2022.104520.

[55] P. P. Ippolito, "Hyperparameter Tuning: The Art of Fine-Tuning Machine and Deep Learning Models to Improve Metric Results," in *Applied Data Science in Tourism*, pp. 231–251, 2022, doi: 10.1007/978-3-030-88389-8_12.

[56] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," *Mobile Information Systems*, vol. 2022, pp. 1–9, 2022, doi: 10.1155/2022/1410169.

[57] Md. S. Reza, U. Hafsha, R. Amin, R. Yasmin, and S. Ruhi, "Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset," *Computer Methods and Programs in Biomedicine Update*, vol. 4, p. 100118, 2023, doi: 10.1016/j.cmpbup.2023.100118.

[58] D. J. Derwin, B. P. Shan, and O. J. Singh, "Hybrid multi-kernel SVM algorithm for detection of microaneurysm in color fundus images," *Med. Biol. Eng. Comput.*, vol. 60, no. 5, pp. 1377–1390, 2022, doi: 10.1007/s11517-022-02534-y.

[59] C. Guo, Z. Jiang, and D. Zhang, "Multi-Feature Complementary Learning for Diabetes Mellitus Detection Using Pulse Signals," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5684–5694, 2022, doi: 10.1109/JBHI.2022.3198792.

[60] A. M. Austin *et al.*, "Using a cohort study of diabetes and peripheral artery disease to compare logistic regression and machine learning via random forest modeling," *BMC Med. Res. Methodol.*, vol. 22, no. 1, p. 300, 2022, doi: 10.1186/s12874-022-01774-8.

[61] P. Palimkar, R. N. Shaw, and A. Ghosh, "Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach," in *Advanced Computing and Intelligent Technologies*, vol. 218, pp. 219–244, 2022, doi: 10.1007/978-981-16-2164-2_19.

[62] J. Hendy, T. Vandrevala, A. Ahmed, C. Kelly, L. Gray, and A. Ala, "Feeling misidentified: Understanding migrant's readiness to engage in health care screening," *Social Science & Medicine*, vol. 237, p. 112481, 2019, doi: 10.1016/j.socscimed.2019.112481.

[63] M. Ye, H. Zhang, and L. Li, "Research on Data Mining Application of Orthopedic Rehabilitation Information for Smart Medical," *IEEE Access*, vol. 7, pp. 177137–177147, 2019, doi: 10.1109/ACCESS.2019.2957579.

[64] T. Sharma and M. Shah, "A comprehensive review of machine learning techniques on diabetes detection," *Vis. Comput. Ind. Biomed. Art*, vol. 4, no. 1, p. 30, 2021, doi: 10.1186/s42492-021-00097-7.

[65] G. B. Bolli, F. Porcellati, P. Lucidi, and C. G. Fanelli, "The physiological basis of insulin therapy in people with diabetes mellitus," *Diabetes Research and Clinical Practice*, vol. 175, p. 108839, 2021, doi: 10.1016/j.diabres.2021.108839.

[66] A. Cherkas, S. Holota, T. Mdzinarashvili, R. Gabbianelli, and N. Zarkovic, "Glucose as a Major Antioxidant: When, What for and Why It Fails?," *Antioxidants*, vol. 9, no. 2, p. 140, 2020, doi: 10.3390/antiox9020140.

[67] M. J. C. Zavaleta *et al.*, "Diabetic gastroenteropathy: An underdiagnosed complication," *WJD*, vol. 12, no. 6, pp. 794–809, 2021, doi: 10.4239/wjd.v12.i6.794.

[68] A. Misra *et al.*, "Diabetes in developing countries," *Journal of Diabetes*, vol. 11, no. 7, pp. 522–539, 2019, doi: 10.1111/1753-0407.12913.

[69] S. Kaul and Y. Kumar, "Artificial Intelligence-based Learning Techniques for Diabetes Prediction: Challenges and Systematic Review," *SN Comput. Sci.*, vol. 1, no. 6, p. 322, 2020, doi: 10.1007/s42979-020-00337-2.

[70] E. Dritsas and M. Trigka, "Data-Driven Machine-Learning Methods for Diabetes Risk Prediction," *Sensors*, vol. 22, no. 14, p. 5304, 2022, doi: 10.3390/s22145304.

[71] J. Hao, S. Luo, and L. Pan, "Rule extraction from biased random forest and fuzzy support vector machine for early diagnosis of diabetes," *Sci. Rep.*, vol. 12, no. 1, p. 9858, 2022, doi: 10.1038/s41598-022-14143-8.

[72] P. Nagaraj, V. Muneeswaran, and G. Deshik, "Ensemble Machine Learning (Grid Search & Random Forest) based Enhanced Medical Expert Recommendation System for Diabetes Mellitus Prediction," in *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 757–765, 2022, doi: 10.1109/ICESC54411.2022.9885312.

[73] H. Ohanyan *et al.*, "Associations between the urban exposome and type 2 diabetes: Results from penalised regression by least absolute shrinkage and selection operator and random forest models," *Environment International*, vol. 170, p. 107592, 2022, doi: 10.1016/j.envint.2022.107592.

[74] S. Gündoğdu, "Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique," *Multimed Tools Appl.*, vol. 82, no. 22, pp. 34163–34181, 2023, doi: 10.1007/s11042-023-15165-8.

[75] S. You and M. Kang, "A Study on Methods to Prevent Pima Indians Diabetes using SVM," *Korean Journal of Artificial Intelligence*, vol. 8, no. 2, pp. 7–10, 2020, doi: 10.24225/KJAI.2020.VOL8.NO2.7.

[76] Y. Aggarwal, J. Das, P. M. Mazumder, R. Kumar, and R. K. Sinha, "Heart rate variability features from nonlinear cardiac dynamics in identification of diabetes using artificial neural network and support vector machine," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1002–1009, 2020, doi: 10.1016/j.bbe.2020.05.001.

[77] D. C. E. Saputra, K. Sunat, and T. Ratnaningsih, "A New Artificial Intelligence Approach Using Extreme Learning Machine as the Potentially Effective Model to Predict and Analyze the Diagnosis of Anemia," *Healthcare*, vol. 11, no. 5, p. 697, 2023, doi: 10.3390/healthcare11050697.

[78] P. S. Washburn, Mahendran, Dhanasekharan, Periyasamy, and Murugeswari, "Investigation of severity level of diabetic retinopathy using adaboost classifier algorithm," *Materials Today: Proceedings*, vol. 33, pp. 3037–3042, 2020, doi: 10.1016/j.matpr.2020.03.199.

[79] X. Wang *et al.*, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," *BMC Med Inform Decis Mak*, vol. 21, no. 1, p. 105, 2021, doi: 10.1186/s12911-021-01471-4.

[80] D. C. Yadav and S. Pal, "An Experimental Study of Diversity of Diabetes Disease Features by Bagging and Boosting Ensemble Method with Rule Based Machine Learning Classifier Algorithms," *SN Comput. Sci.*, vol. 2, no. 1, p. 50, 2021, doi: 10.1007/s42979-020-00446-y.

[81] D. C. E. Saputra, A. Azhari, and A. Ma'arif, "K-Nearest Neighbor of Beta Signal Brainwave to Accelerate Detection of Concentration on Student Learning Outcomes," *Engineering Letters*, vol. 30, no. 1, pp. 318–234, 2022.

[82] H. Mzoughi *et al.*, "Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI Gliomas Brain Tumor Classification," *J. Digit Imaging*, vol. 33, no. 4, pp. 903–915, 2020, doi: 10.1007/s10278-020-00347-9.

[83] R. D. Joshi and C. K. Dhakal, "Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches," *IJERPH*, vol. 18, no. 14, p. 7346, 2021, doi: 10.3390/ijerph18147346.

[84] R. Krishnamoorthi *et al.*, "A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–10, 2022, doi: 10.1155/2022/1684017.

[85] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, "An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI," *Sensors*, vol. 22, no. 19, p. 7268, 2022, doi: 10.3390/s22197268.