Journal of Robotics and Control (JRC)
Volume 4, Issue 4, 2023
ISSN: 2715-5072, DOI: 10.18196/jrc.v4i4.18412

Early Prediction of Gestational Diabetes with Parameter-Tuned K-Nearest Neighbor Classifier

Tsehay Admassu Assegie [1*], Tamilarasi Suresh [2], Raguraman Purushothaman [3], Sangeetha Ganesan [4], Napa Komal Kumar [5]
[1] Department of Computer Science, College of Engineering and Technology, Injibara University, Injibara, Ethiopia
[2] Department of Information Technology, St. Peter's Institute of Higher Education and Research, Chennai, India
[3] Department of Artificial Intelligence and Data Science, Madanapalle Institute of Technology & Science, Angallu, Madanapalle, Andhra Pradesh, India
[4] Department of Artificial Intelligence and Data Science, RMK College of Engineering Technology, Thiruvallur District, India
[5] Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Tamil Nadu, India
Email: [1] tsehayadmassu2006@gmail.com
*Corresponding Author

Abstract—**Diabetes is one of the quickly spreading chronic diseases causing health complications, such as diabetes retinopathy, kidney failure, and cardiovascular disease. Recently, machine-learning techniques have been widely applied to develop a model for the early prediction of diabetes. Due to its simplicity and generalization capability, K-nearest neighbor (KNN) has been one of the widely employed machine learning techniques for diabetes prediction. Early diabetes prediction has a significant role in managing and preventing complications associated with diabetes, such as retinopathy, kidney failure, and cardiovascular disease. However, the prediction of diabetes in the early stage has remained challenging due to the accuracy and reliability of the KNN model. Thus, gird search hyperparameter optimization is employed to tune the K values of the KNN model to improve its effectiveness in predicting diabetes. The developed hyperparameter-tuned KNN model was tested on the diabetes dataset collected from the UCI machine learning data repository. The dataset contains 768 instances and 8 features. The study applied Min-max scaling to scale the data before fitting it to the KNN model. The result revealed KNN model performance improves when the hyperparameter is tuned. With hyperparameter tuning, the accuracy of KNN improves by 5.29% accuracy achieving 82.5% overall accuracy for predicting diabetes in the early stage. Therefore, the developed KNN model applied to clinical decision-making in predicting diabetes at an early stage. The early identification of diabetes could aid in early intervention, personalized treatment plans, or reducing healthcare costs reducing associated risks such as retinopathy, kidney disease, and cardiovascular disease.**

Keywords—Optimization; Machine Learning; Automated Diagnosis; Parameter Tuning; Classification.

I. INTRODUCTION

Diabetes is a chronic disease characterized by high glucose levels in the human blood [2]. Recently it is among one of the most rapidly spreading diseases in the world. Diabetes can be categorized into three types. The first type is Type 1 diabetes occurs when the immune system is weakened, and the cells are unable to produce enough insulin required to regulate the blood glucose level. Type 2 diabetes occurs when the body cells are unable to generate enough insulin or the body fails to utilize the insulin properly.

Gestational diabetes occurs when pregnant women acquire high blood sugar [3]. It occurs at any stage of pregnancy and causes problems for the woman and the baby during and after birth. Gestational diabetes develops in some women when they are pregnant. Most of the time, this type of diabetes goes away after the baby is born. However, women with gestational diabetes have a greater chance of developing type 2 diabetes later in life. Sometimes diabetes diagnosed during pregnancy is type 2 diabetes.

With the advancement in computing and the availability of labeled diabetes datasets in the healthcare sector, machine learning (ML) has improved the diagnosis of diabetes [4]. A pre-processing method such as feature selection is found effective in improving the performance of ML techniques for the accurate prediction of diabetes in the early stages. With feature selection as preprocessing technique, the KNN model achieves an accuracy of 76.25% on early-stage diabetes prediction.

For diabetes prediction, A.H. Osman et al. [5] investigated the performance of KNN on the Pima Indian Diabetes dataset. The study revealed that the SVM model achieves 80.39% accuracy on the test dataset. The result demonstrated higher precision for ML techniques in predicting diabetes. However, pre-processing techniques such as missing value analysis, and cross-validation technique is not investigated to validate the accuracy achieved by the model.

Another research conducted in [6] compared and tested the effectiveness of ensemble learning techniques, namely random forest (RF), extreme boosting (XGB), decision tree (DT), and light extreme boosting (LGB) for diabetes prediction. The result revealed that the highest accuracy of scored 73.5% accuracy with the ensemble learning methods. The grid search method is employed to tune the hyperparameter of the ensemble technique for diabetes prediction for improving their accuracy.

The diagnosis of and prediction of diabetes in the early stage significantly reduces the health complications due to diabetes [7]. Studies on the application of machine learning have proliferated in recent years for the improvement of diabetes treatment. The association between different test symptoms and test results to develop classification models that generalize and classify a given sample into the diabetic or non-diabetic class.

Research article [8] further investigated the performance of an ensemble model for gestational diabetes prediction. The investigation reveals gradient booting model achieved a receiver operating characteristic curve of 0.71. Different health-related issues affect the performance of machine learning predictors. A study [9] investigated the effect of health-related issues on the performance machine-learning model for diabetes prediction.

Similar research [10] evaluated the performance of deep learning model for diabetes prediction. The study suggested that early prediction of diabetes significantly helps the patient to change life styles and improve their health condition. The study also revealed that deep learning model outperforms other supervised learning model.

In addition, another research article [11], [12], [13] analyzed the effectiveness of several machine learning models such as KNN, DT, RF, Naïve Bayes (NB), support vector machine (SVM), and logistic regression. The comparison of the performance of these machine-learning models reveals that the logistic regression model outperforms other models with an accuracy score of 75.32%.

A research article [14], [15], [16] predicted diabetes by developing a KNN model. The KNN model performance improves when the model is trained on pre-processed PIMA Indian diabetes dataset (PIDD). The study suggested that pre-processing with feature scaling increases the performance of KNN with an accuracy of 8.48% on PIDD.

Diabetes has several consequences such as an increased risk of retinopathy, hypertension or high blood pressure, renal damage, and cardiovascular disease [17], [18], [19], [20], [21]. Recently, diabetes is becoming one of the most prevalent diseases affecting numerous people all over the world. The advancement in machine learning has become significant in the reduction of the ever-growing risk of diabetes through early prediction of diabetes to avoid latter associated consequences such as retinopathy, and cardiovascular disease.

For the early diabetes prediction systems, the existing studies employ different machine learning models, trained on diverse amounts of data, to predict the presence of gestational diabetes. The KNN model fails to categorize the diversity of sample instances with a relatively small set of data. To address this issue, this study proposes a KNN classification model based on distance measurement to predict the presence of gestational diabetes. Furthermore, existing prediction models employ a common set of factors for constructing the model. According to the healthcare professional principle in the diagnosing process, a wide swath of health conditions results in different disease diagnoses and treatment decisions [22], [23], [24], [25], [26]. However, the existing KNN-based gestational diabetes prediction has scope for improvement for accurate prediction of gestational diabetes. For instance, a study [27] reveals 71.4% accuracy on gestational diabetes prediction with the KNN model even though the KNN model outperforms the SVM model [28]. Similarly, another research article [29] compared the performance of supervised learning algorithms. The result revealed that the KNN model achieved 74.89% accuracy in gestational diabetes prediction. The findings revealed improving the performance of the KNN model is recommended for further research. Thus, this study proposes a novel KNN model for gestational diabetes prediction. Overall, the objective of this study is summarized as follows:

- To develop a KNN model that can learn the pattern from the diabetes dataset and perform automated analysis.
- To apply the grid search technique and improve the performance of the KNN classifier.
- To apply correlation analysis and extract underlying patterns from the diabetes dataset, which will assist in determining the diabetic tendency of a new set of patients in the early stage.

The remainder of the study is structured as follows: Section 2 describes the method, explaining the dataset, preprocessing steps, and the proposed KNN model. Section 3 presents the results and discussion. The result section compares the performance of the KNN model on the original and feature-scaled dataset with the help of an accuracy metric. Finally, Section 4 concludes the study. The conclusion presents the results obtained implication of the study, limitations, and recommendations for feature work.

II. Method

The study considered four procedures for developing a predictive model for gestational diabetes. The procedures involve data acquisition, data preprocessing, and development of the KNN model and analysis of the performance of the developed model on the test set. Grid search and hyperparameter tuning have been widely employed for improving the performance of the KNN algorithm. For instance, research article [30], [31], [32]. The study employed feature scaling due to its significance in improving the performance of the KNN model for diabetes diagnosis [33]. Each step is discussed in Section A, and Section B. Fig. 1 demonstrates the flow diagram for the procedure followed to develop the proposed KNN model.

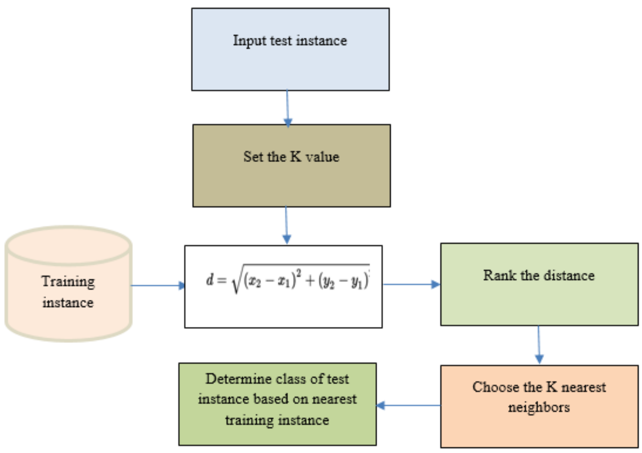

Fig. 1. The flow chart for the developed KNN model

A. Data Acquisition

The PIDD is employed for the development of the KNN model for gestational diabetes prediction. The PIDD dataset is one of the standard datasets previously employed in several studies [34], [35], [36], [37] for the development of gestational diabetes prediction, prognosis, and diagnosis with machine learning algorithms. The dataset is collected from the online Kaggle repository available at the following link which if previously employed by the study [38], [39], [40] can be downloaded at https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database. There were 768 instances each with 8 features. The dataset consists of several medical predictor variables and one target variable, outcome. The predictor variables include the number of pregnancies the patient has had, their body mass index (BMI), insulin level, age, blood pressure (BP) glucose level, skin thickness, and diabetes pedigree function (DPF). The dataset is analyzed for missing values and the features are scaled with a standard scaler. In the analysis and dataset feature, exploration panda's data frame is employed.

B. Data Preprocessing

In data preprocessing, feature scaling with a standard scaler is performed. The standard scaler is one of the most commonly employed types of feature scaling methods for distance-based learners such as KNN, and SVM [41], [42], [43], [44]. Standard scaler improves the performance of distance based learning algorithm [45], [46], [47], [48], [49].

Feature scaling through standardization (or Z-score normalization) can be an important preprocessing step for many machine-learning algorithms [50], [51], [52], [53]. Standardization involves rescaling the features such that they have the properties of a standard normal distribution with a mean of zero and a standard deviation of one. The standard scaler is defined by the mathematical equation [54], [55], [56], [57], [58], [59], [60] given in equation (1).

$$Zscaled = x - \mu / \alpha \tag{1}$$

Where μ is the mean and σ is the standard deviation.

III. Results And Discussion

This section presents the result of the constructed KNN model on gestational diabetes prediction. Firstly, the performance of the KNN model is evaluated on the test dataset with random K values. Secondly, the grid search technique is employed to evaluate corss-vlaidated accuracy on K values in the range 1 to 25 and the optimal value of K is determined. Then, the performance of the KNN model is compared on random K values and the optimized K value.

A. Optimization of KNN with K vlaue

The performance of the developed KNN model is optimized by tuning the K values, which produces the highest possible accuracy for the KNN model. After selecting the combination of hyperparameters with the highest accuracy for the KNN model, the K value that produces optimal test accuracy is determined. The variability of the training and test accuracy for different values of K is revealed in Fig. 2. As indicated in Fig. 2, the KNN model score higher test accuracy of 77.21% with K value=13.

Fig. 2. The optimization of K value for standard KNN model

Fig. 3 indicates the training and testing accuracy for the KNN model for K values ranging from 1 to 25 on the scaled dataset. As indicated in Fig. 3, the KNN model's training and testing accuracy varies for the varying value of K (the number of test instances considered for comparison with the test instance) to predict the class of the new instance. Overall, the KNN model when trained on scaled data improves the KNN model accuracy by 5.29% with K value=21. Thus, feature scaling, and parameter tuning with grid search significantly improve the performance of KNN to predict diabetes.

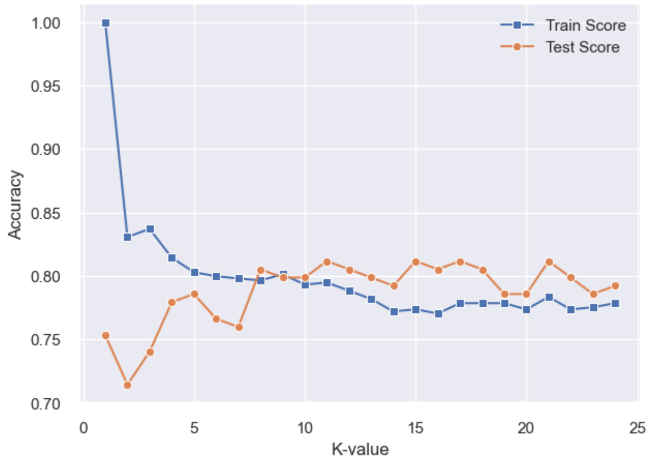

Fig. 3. The optimization of K value for scaled dataset

In addition to the feature scaling, and grid search for improving the performance of KNN in predicting gestational diabetes, the impact of each feature on test instance is analyzed with the help of Shapley Additive Explanation (SHAP). Fig. 4 indicates that for test instances with feature values (number of times pregnant=6, age=50, glucose 148, blood pressure-72, skin thickness=35, body mass index=33.6, and diabetes pedigree function=0.627) with the outcome of diabetes, the age of the patient significantly impacts the prediction of the KNN model.

In Fig. 5 indicates the variability of the cross-validation score for varying values of K. five-fold cross-validation score of the KNN model indicated in Fig. 5 reveals that the variability of the training score tends to decrease with an increase in the value of K. It is evident that with K values greater than 10 the testing, and the cross-validation score tends to converge revealing better performance on diabetes prediction.

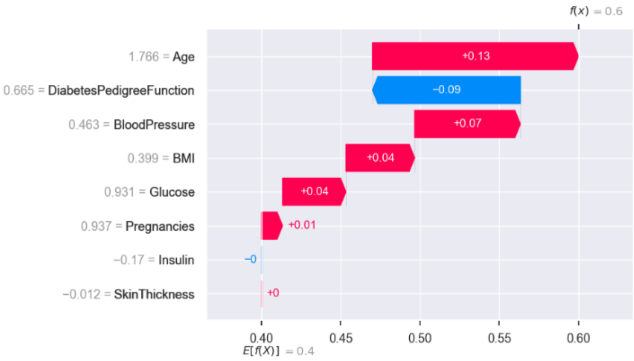

Fig. 4. The relative impact of features on test instance using KNN model

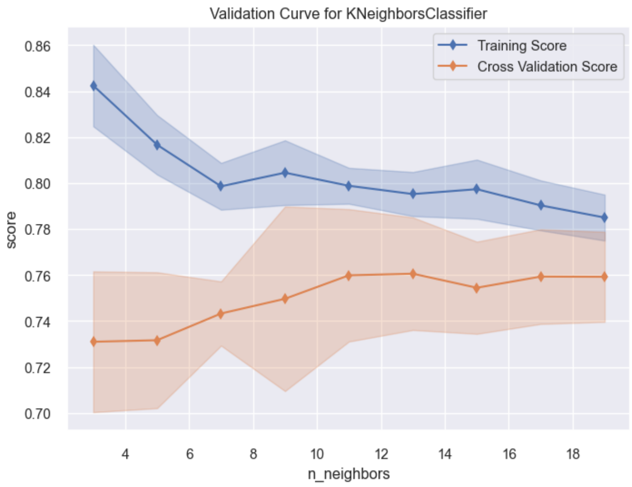

Fig. 5. The validation score for KNN model

IV. Conclusion

In this study, hyperactive parameter tuning, and standardization of input features are employed for pre-processing methods. With these pre-processing methods, the KNN model performance improves for diabetes prediction. The result of the experiment revealed that the constructed KNN model outperforms other existing KNN models. Thus, it is suggested that feature scaling significantly improves the performance of the KNN model for the prediction of diabetes. The study improves the existing KNN model with pre-processing and grid-searching techniques. The developed model significantly improves the accuracy of the previous studies. However, the performance of the KNN model and other pre-processing methods with other datasets seek further studies to validate the findings of the current study.

The limitations of this study include the study lack of generalizability of the findings across various data sources to confirm the reliability of the KNN model performance improvement with preprocessing methods such as feature scaling and hyperparameter tuning. The researchers recommend further investigations using different datasets and validate the effectiveness of the KNN model and pre-processing methods. Additionally, a comparative analysis of the KNN model with other machine learning algorithms commonly used for diabetes prediction such as SVM, and RF is recommended for future work.

Acknowledgment

The authors would like to acknowledge Injibara University of the provision of facilities to complete this work.

References

[1] R. N. Patil. S. Rawandale, N. Rawandale, U. Rawandale, and S. Patil, "An efficient stacking based NSGA-II approach for predicting type 2 diabetes," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 1015-1023, 2023, doi: 10.11591/ijece.v13i1.pp1015-1023.

[2] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, H, "Predicting Diabetes Mellitus with Machine Learning Techniques," *Frontiers in genetics*, vol. 9, no. 515, 2018, doi: 10.3389/fgene.2018.00515.

[3] R. Jader and S. Aminifar, "Predictive Model for Diagnosis of Gestational Diabetes in the Kurdistan Region by a Combination of Clustering and Classification Algorithms: An Ensemble Approach," *Applied Computational Intelligence and Soft Computing*, vol. 2022, 2022, doi: 10.1155/2022/9749579.

[4] E. Sabitha and M. Durgadevi, "Improving the Diabetes Diagnosis Prediction Rate Using Data Preprocessing, Data Augmentation and Recursive Feature Elimination Method," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, 2022.

[5] A. H. Osman and H. M. Aljahdali, "Diabetes Disease Diagnosis Method based on Feature Extraction using K-SVM," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 1, 2017.

[6] A. Dutta *et al.*, "Early Prediction of Diabetes Using an Ensemble of Machine Learning Models," *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, p. 12378, 2022, doi: 10.3390/ijerph191912378.

[7] L. Shrinivasan, R. Verma, and M. D. Nandeesh, "Early prediction of diabetes diagnosis using hybrid classification techniques," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1139-1148, 2023, doi: 10.11591/ijai.v12.i3.pp1139-1148.

[8] Y. N. Chan *et al.*, "A machine learning approach for early prediction of gestational diabetes mellitus using elemental contents in fingernails," *Scientific Reports*, vol. 13, no. 1, p. 4184, 2023, doi: 10.1038/s41598-023-31270-y.

[9] H. F. Ahmad, H. Mukhtar, H. Alaqail, M. Seliaman, and A. Alhumam, "Investigating Health-Related Features and Their Impact on the Prediction of Diabetes Using Machine Learning," *Applied Sciences*, vol. 11, no. 3, p. 1173, 2021, doi: 10.3390/app11031173.

[10] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," *Journal of Diabetes & Metabolic Disorders*, vol. 19, pp. 391-403, 2020, doi: 10.1007/s40200-020-00520-5.

[11] A. Bansal and A. Singhrova, "Performance Analysis of Supervised Machine Learning Algorithms for Diabetes and Breast Cancer Dataset," *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 137-143, 2021, doi: 10.1109/ICAIS50930.2021.9396043.

[12] R. Saxena, "Role of K-nearest neighbour in detection of Diabetes Mellitus," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 10, pp. 373-376, 2021.

[13] M. T. Alasaady, T. N. M. Aris, N. M. Sharef, and H. Hamdan, "A proposed approach for diabetes diagnosis using neuro-fuzzy technique," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3590-3597, 2022, doi: 10.11591/eei.v11i6.4269.

[14] N. AlRefaai and S. Z AlRashid, "Classification of gene expression dataset for type 1 diabetes using machine learning methods," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, 2023, pp. 2986-2992, doi: 10.11591/eei.v12i5.4322.

[15] P. A. Rajendra and S. Latif, "Prediction of diabetes using logistic regression and ensemble techniques," *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100032, 2021, doi: 10.1016/j.cmpbup.2021.100032.

[16] S. Khairunnisa, S. Suyanto, and P. Eko Yunanto, "Removing Noise, Reducing dimension, and Weighting Distance to Enhance kk-Nearest Neighbors for Diabetes Classification," *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 471-475, 2020, doi: 10.1109/ISRITI51436.2020.9315515.

[17] M. Ahsan, P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," *Technologies*, vol. 9, no. 52, 2021, doi: 10.3390/ technologies9030052.

[18] T. A. Assegie, V. Elanangai, J. S. Paulraj, M. Velmurugan, and D. F. Devesan, "Evaluation of feature scaling for improving the performance of supervised learning methods*," Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1833-1838, 2023, doi: 10.11591/eei.v12i3.5170.

[19] P. Ferreira, D. C. Le, and N. Zincir-Heywood, "Exploring Feature Normalization and Temporal Information for Machine Learning Based Insider Threat Detection," *2019 15th International Conference on Network and Service Management (CNSM)*, pp. 1-7, 2019, doi: 10.23919/CNSM46954.2019.9012708.

[20] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 729-735, 2020, doi: 10.1109/ICSSIT48917.2020.9214160.

[21] M. J. Pendekal and S. Gupta, "An Ensemble Classifier Based on Individual Features for Detecting Microaneurysms in Diabetic Retinopathy," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 60-71, 2022, doi: 10.52549/ijeei.v10i1.3522.

[22] N. Ahmed *et al*., "Machine learning based diabetes prediction and development of smart web application," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229-241, 2021, doi: 10.1016/j.ijcce.2021.12.001.

[23] T. A. Assegie and P. S. Nair, "The Performance of Different Machine Learning Models on Diabetes Prediction," *International Journal of Scientific & Technology Research*, vol. 9, no. 1, 2020.

[24] M. F. Faruque, A. Asaduzzaman, and I. H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-4, 2019, doi: 10.1109/ECACE.2019.8679365.

[25] H. Butt, I. Khosa, and M. A. Iftikhar, "Feature Transformation for Efficient Blood Glucose Prediction in Type 1 Diabetes Mellitus Patients," *Diagnostics*, vol. 13, no. 3, p. 340, 2023, doi: 10.3390/ diagnostics13030340.

[26] N. P. Miriyala *et al*., "Diagnostic Analysis of Diabetes Mellitus Using Machine Learning Approach," *Revue d'Intelligence Artificielle*, vol. 36, no. 3, pp. 347-352, 2022, doi: 10.18280/ria.360301.

[27] B. C. Kusumaatmajaa and H. Thamrin, "Web-Based Prediction of Potential Diabetes Outbreaks Using Django with the KNN Algorithm," *AIP Conference Proceedings*, vol. 2727, no. 1, p. 040004, 2023, doi: 10.1063/5.0141932.

[28] O. Altay, M. Ulas, and K. E. Alyamac, "Prediction of the Fresh Performance of Steel Fiber Reinforced Self-Compacting Concrete Using Quadratic SVM and Weighted KNN Models," in *IEEE Access*, vol. 8, pp. 92647-92658, 2020, doi: 10.1109/ACCESS.2020.2994562.

[29] R. Patil and S. Tamane, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Diabetes," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 5, pp. 3966-3975, 2018, doi: https://doi.org/10.11591/ijece.v8i5.pp3966-3975.

[30] Y. N. Fuadah, M. A. Pramudito, and K. M. Lim, "An Optimal Approach for Heart Sound Classification Using Grid Search in Hyperparameter Optimization of Machine Learning," *Bioengineering*, vol. 10, no. 1, p. 45, 2023, doi: 10.3390/bioengineering10010045.

[31] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, pp. 1-6, 2020, doi: 10.1109/B-HTC50970.2020.9297949.

[32] D. M. Belete and M. D. Huchaiah, "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results," *International Journal of Computers and Applications*, vol. 44, no. 9, pp. 875-886, 2021, doi: 10.1080/1206212X.2021.1974663.

[33] D. U. Ozsahin, M. Taiwo Mustapha, A. S. Mubarak, Z. S. Ameen, and B. Uzun, "Impact of feature scaling on machine learning models for the diagnosis of diabetes," *2022 International Conference on Artificial Intelligence in Everything (AIE)*, pp. 87-94, 2022, doi: 10.1109/AIE57029.2022.00024.

[34] S. A. Alalwan *et al*., "Diabetic analytics: proposed conceptual data mining approaches in type 2 diabetes dataset," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, pp.88-95, 2019, doi: 10.11591/ijeecs.v14.i1.pp88-95.

[35] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkut, "Prediction of diabetes disease using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, pp. 284-290, 2022, doi: 10.11591/ijai.v11.i1.pp284-290.

[36] K. Yothapakdee, S. Charoenkhum, and T. Boonnuk, "Improving the efficiency of machine learning models for predicting blood glucose levels and diabetes risk," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 555-562, 2022, doi: 10.11591/ijeecs.v27.i1.pp555-562.

[37] H. R. Ismail and M. M. Hassan, "Bayesian deep learning methods applied to diabetic retinopathy disease: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 2, pp. 1167-1177, 2023, doi: 10.11591/ijeecs.v30.i2.pp1167-1177.

[38] G. Nikhila, T. Bhuvan, and R. Jerrard, "Time series prediction of personalized insulin dosage for type 2 diabetics," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 2, pp. 1080-1087, 2023, doi: 10.11591/ijeecs.v31.i2.pp1080-1087.

[39] Y. Tan, H. Chen, J. Zhang, R. Tang, and P. Liu, "Early Risk Prediction of Diabetes Based on GA-Stacking," *Applied Science*, vol. 12, no. 2, p. 632, 2022, doi: 10.3390/app12020632.

[40] T. Alghamdi, "Prediction of Diabetes Complications Using Computational Intelligence Techniques," *Applied Science*, vol. 13, no. 5, p. 3030 2023, doi: 10.3390/app13053030.

[41] O. AlShorman, B. AlShorman, and F. Alkahtani, "A review of wearable sensors based monitoring with daily physical activity to manage type 2 diabetes," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 646-653, 2021, doi: 10.11591/ijece.v11i1.pp646-653.

[42] A. Yasar, "Data Classification of Early-Stage Diabetes Risk Prediction Datasets and Analysis of Algorithm Performance Using Feature Extraction Methods and Machine Learning Techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 4, pp. 273-281, 2021.

[43] J. J. Sonia, P. Jayachandran, A. Q. Md, S. Mohan, A. K. Sivaraman, and K. F. Tee, "Machine-Learning-Based Diabetes Mellitus Risk Prediction Using Multi-Layer Neural Network No-Prop Algorithm," *Diagnostics*, vol. 13, no. 4, p. 723, 2023, doi: 10.3390/diagnostics13040723.

[44] J. R. Raut, Y. Sharma, and V. D. Shinde, "Performance Evaluation of Various Supervised Machine Learning Algorithms for Diabetes Prediction," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 8, 2020.

[45] S. P. Menon *et al*., "An Intelligent Diabetic Patient Tracking System Based on Machine Learning for E-Health Applications," *Sensors*, vol. 23, no. 6, p. 3004, 2023, doi: 10.3390/s23063004.

[46] H. B. Kibria, M. Nahiduzzaman, 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.

[47] D. S. Sisodia and R. Agrawal, "Data Imputation-Based Learning Models for Prediction of Diabetes," *2020 International Conference on Decision Aid Sciences and Application (DASA)*, pp. 966-970, 2020, doi: 10.1109/DASA51403.2020.9317070.

[48] J. Ramesh, R. Aburukba, and A. Sagahyroon, "A remote healthcare monitoring framework for diabetes prediction using machine learning," *Healthcare Technology Letters*, vol. 8, no. 3, pp. 45-57, April 2021, doi: 10.1049/htl2.12010.

[49] H. Lu, J. Hirst, J. Yang, L. Mackillop, and D. Clifton, "Standardizing the assessment of caesarean birth using an oxford caesarean prediction

score for mothers with gestational diabetes," *Healthcare Technology Letters*, vol. 9, no. 1-2, pp. 1-8, 2022, doi: 10.1049/htl2.12022.

[50] I. Tasin, T. U. Nabi, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," *Healthcare Technology Letters*, vol. 10, no. 1-2, pp. 1-10, 2022, doi: 10.1049/htl2.12039.

[51] B. Premamayud, K. Muralikrishna, and K. Pramodh "Diabetes Prediction Using Machine Learning KNN -Algorithm Technique," *International Journal of Innovative Science and Research Technology*, vol. 7, no. 5, May 2022.

[52] A. S. Hassan, I. Malaserene, and A. A. Leema, "Diabetes Mellitus Prediction using Classification Techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 5, March 2020.

[53] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*, vol. 1, pp. 679-689, 2021.

[54] A. Sharma, S. Shinde, I. I. Shaikh, M. Vyas, and S. Rani, "Machine Learning Approach for Detection of Diabetic Retinopathy with Improved Pre-Processing," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 517-522, 2021, doi: 10.1109/ICCCIS51004.2021.9397115.

[55] S. Patikar, P. Saha, S. Neogy, and C. Chowdhury, "An Approach towards prediction of Diabetes using Modified Fuzzy K Nearest Neighbor," *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 73-76, 2020, doi: 10.1109/GUCON48875.2020.9231066.

[56] S. M. A. Huda, I. J. Ila, S. Sarder, M. Shamsujjoha, and M. N. Y. Ali, "An Improved Approach for Detection of Diabetic Retinopathy Using Feature Importance and Machine Learning Algorithms," *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, pp. 1-5, 2019, doi: 10.1109/ICSCC.2019.8843676.

[57] M. Nilashi, S. Samad, E. Yadegaridehkordi, A. Alizadeh, and E. Akbari, "Early Detection of Diabetic Retinopathy Using Ensemble Learning Approach," *Journal of Soft Computing & Decision Support Systems*, vol. 6, no. 2, 2019.

[58] P. S. Kumar, A. K. K, S. Mohapatra, B. Naik, J. Nayak, and M. Mishra, "CatBoost Ensemble Approach for Diabetes Risk Prediction at Early Stages," *2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON)*, pp. 1-6, 2021, doi: 10.1109/ODICON50556.2021.9428943.

[59] M. Hassan, S. Mollick, and F. Yasmin, "An unsupervised cluster-based feature grouping model for early diabetes detection," *Healthcare Analytics*, vol. 2, p. 100112, 2022, doi: 10.1016/j.health.2022.100112.

[60] S. H. Shaker, A. Farah, and A. Mahdi, "Diagnosis of Diabetes Mellitus Based Combined of Feature Selection Methods," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, July 2022.