https://doi.org/10.48047/AFJBS.6.14.2024.9177-9184

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

An Improved Mathematical Framework for Accurate Prediction of Diabetes

[1] Dr. Muhammed Basheer, Dr. Amol R Madane, Ms.Soni Gupta, Dr Shobana A, Prema S. Kadam, Dr. Brinda Halambi,

1Assistant Professor of Mathematics, University of Technology and Applied science, Nizwa, Calicut, Kerala
vpmbasheer@yahoo.com

Solution Architect, Computer Consultancy, Tata Consultancy Services Ltd, Pune, Maharashtra.
amol.madane@gmail.com

facultyAmity Institute of Aerospace Engineering, Amity university, Noida, Uttar Pradesh. sgupta10@amity.edu

Professor of Science and Humanities, Sri Krishna College of engineering and technology, Coimbatore, Tamil Nadu.
shobanaa@skcet.ac.in

Assistant professor of Artificial intelligence and Data science, BRACT's Vishwakarma Institute of Information Technology, Pune, Maharashtra. prema.kadam@viit.ac.in

Associate Professor of Mathematics, REVA UNIVERSITY, Bangalore North, Yelahanka, Karnataka.brindahalambi@gmail.com

Corresponding Author: Dr. Brinda Halambi, brindahalambi@gmail.com

Volume 6, Issue 14, Aug 2024

Received: 15 June 2024

Accepted: 25 July 2024

Published: 15 Aug 2024

doi: 10.48047/AFJBS.6.14.2024.9177-9184

Abstract: **This research therefore provides an enhanced mathematical approach to the determination of diabetes with new approaches in the machine learning algorithms. The research focuses on, and analyzes the efficacy of, four algorithms, namely Logistic Regression, Random Forest, SVM, and XGBoost on a large data set. By choosing only the most important features and by a proper tuning of the model, it was possible to obtain here 94% of accuracy of the prediction. 2%, outperforming existing models. The Random Forest carried out the best result with accuracy of 96. At the end, we obtain 1% for our submission and XGBoost gives 95. 3%, SVM at 93. 8%, Logistic Regression at 91. 5%. The effectiveness of the presented framework was proved through numerous experiments proving the framework's applicability for early diagnosis and, therefore, better management of diabetes. Further, the framework reduced the False Positive Rate and False Negative Rate further by 12%, and 8% respectively than the other related work. These findings suggest that the proposed model is not only accurate but also time and cost efficient for the large data that always typical of real-world healthcare services. There thus is a literature gap that this research seeks to fill in the quest to enhance the prediction of adverse events in healthcare delivery through advancing the methods used in predictive analytics for early IT intervention.**

Keywords: Diabetes prediction, machine learning, Graph Theory, Random Forest, Support Vector Machine, XGBoost.

I. INTRODUCTION

Diabetes is non-curable, (sizeof carcinoma) global disease which involves millions of people and is defined as high blood glucose levels because of less insulin production or defective insulin utilization. This condition is becoming more common and this called for high alert due to the impact it has on the health of the public. Conventional approaches to the identification of individuals likely to develop diabetes involve the use of statistical models that are unable to capture all the factors that might be associated with the disease and therefore have low prediction accuracy. More recently, and as big data and increasing capabilities in computational methods become more available there is the potential to generate far more complex models of prediction [1]. As for the limitation of this study, this research advises the formulation of an improved mathematical model with better predictions of diabetes. This is why, thanks to combining the modern statistical analysis methods and machine learning solutions, this framework is (or at least should be) capable of considering more factors that affect the risk, including genetics, life experience, and medical history [2]. Instead, the aim is to model the dynamics and the interplays of these factors that fundamental models could fail to address. The following framework attempts to overcome the above-mentioned limitations of the conventional predictive models, by focusing on a significantly larger number of predicators, and by adopting the best available techniques from the new field of machine learning [3]. These are deep learning methods, ensemble methods, and personalized risk assessments which are believed to enhance the prediction of diabetes greatly. Hence, as a result of extensive experiments and validation with big data, the study seeks to establish the ability of the framework to generate more precise and personalised predictions. This can help care givers identify the candidates who are at risk early enough thus improving on the chances of early management of the disease. Last but not the least, this enhanced prognostic matrix could help in reversing the tide of impending diabetes disaster and enhancing patient success rate.

II. RELATED WORKS

In terms of the use of artificial intelligence in assessing risk of different diseases, several works have been conducted. Jiang et al. 2024 focused on the detection of the disease risk with two tuning parameters of the PRS method utilizing summary of GWAS. The study made it clear that enhancement of PRS is crucial in arriving at correct risk predictions in different diseases, particularly diseases that are influenced by one or several genes and the environment [15]. In the similar manner, the research work of Kulkarni et al. (2024) proposed the novel hybrid disease prediction with the integration of digital twin and metaverse. It offers a better optimized and real-time prediction model through connect with digital twin data hence enhancing health diagnostic and initial disease identification [16]. Lubasinski et al. (2024) systematically reviewed the literature on blood glucose prediction in adult patients with type 1 diabetes. Their work focused on the realization of AI in enhancing the courses of glucose levels predictions, which is particularly vital in the management of diabetes. Including nutrition and analytics in the study offered knowledge about the part that foods play in projecting glucose mobilization to patients and their results [18]. This is particularly the case when it comes to the selection of features in improving the existing AI models in medical diagnosis. The authors in Li et al. (2024) put forward an enhanced MGO that is composed of chaotic maps and spiral disturbance in medical feature selection. The study showed that employed optimized feature selection made the medical diagnostic models more accurate and faster especially in dealing with big data sets which are typical in medical sciences [17]. In Luo et al. (2023), the authors were more concern on developing the prediction model for high-risk complication among type 2 diabetes mellitus through data mining. In the study, various feature selection techniques were used hence making the understanding of specific risk factors that cause complications to be more clear to make the appropriate management on patients [19]. Such areas as the application of AI in, for example, personalized medicine attract a lot of interest. The same year, Md et al proposed about the chances of using simulation, Artificial Intelligence and web technology in the clinical validation of cardiovascular disease risk prediction at an early stage. Their strategy was geared towards enhancing the reliability of early diagnoses in children, so that early and efficient interventions could be instituted to patients who were at high risk of developing cardiovascular diseases [20]. Along the same line, Nabrdalik et al. (2024) investigated the potential of AI in identifying cardiac autonomic neuropathy from nonmydriatic retinal fundus images of patients with diabetes. One study brought to light the reproducibility in medical images, unknown to conventional practices, which AI can discover and enhance diagnosis and patients' well-being [22]. Recently, Mohsen et al. (2023) conducted the scoping review of the AI-based methods for diabetes risk prediction. Their review highlighted that there are many techniques that were applied in AI and ML in the prediction of diabetes with regression, deep learning, etc. Some of the limitations of applying these approaches was also discussed including the occurrence of large high quality datasets for model building, and model interpretability [24]. For comparison and to able to measure the effectiveness of one AI model to the other in medical applications it is imperative to conduct comparative studies. Nissa et al. in their work of 2024 suggested a technical comparative framework for heart disease prediction through boosting ensemble techniques. Their work also included the comparison of several machine learning models whereby they established that ensemble models were usually superior to standalone models insofar as accuracy of prediction and resilience were concerned [23]. Pati et al. (2024) have evaluated the different hybrid machine learning techniques applied for the breast cancer prognosis and for the prediction of the recurrences. Their work demonstrated that parallel application of more than one Artificial Intelligence technique including the use of features selection alongside ensemble learning resulted in better algorism for predictions that are viable in enhancement of patients' prognosis and treatment regimes [26]. Parimala and Muneeswari (2024) worked on the binary classification and the segmentation of brain stroke using binary classification and the segmentation of by hybrid segmentation technique based on the modified ResNet152v2 model. This transfer learning-based approach demonstrated potential in rightly locating stroke associated regions in brain images and therefore hybrid model has likelihoods to enhance the diagnostic efficacies in medical imaging [25]. Nowadays, the creation of complex AI models is still on the upper scale of what is possible in medical research. In 2024, using graph-based interaction protein language model, Pang and Liu developed DisoFLAG for identifying protein intrinsic disorder and its function. This model is a conceptual milestone in the field and provides new paths to a much better comprehension of numerous diseases that are rooted in protein-protein connections [24].

III. METHODS AND MATERIALS

1. Data Collection and Preprocessing

The basis of this study is the accumulation and cleansing of all the necessary data related to the prediction of diabetes. The data set that has been utilized in this work consists of 10,000 patient records drawn from a public health data base; there are various demographic, clinical and lifestyle factors across the data set [12]. Some of the examples of such variables are age, gender, BMI, blood pressure, cholesterol levels, glucose levels, history of diabetes in family members, physical activity, dietary behavior, and the like.

Before the data were analyzed, they were preprocessed so that they would be properly suitable for the analysis. First, handling of missing values involved the use of multiple imputation techniques to reduce bias and problems with the dataset. Glucose levels and body mass index (BMI) were analysed as continuous variables They were, thereafter, normalized to zero mean and a common variance of one [13]. Gender and other such categories were encoded using the one hot encoder technique. Some points that fall outside the pattern were removed through the application of the interquartile range (IQR) in order to reduce the effect of such points. Last of all, the dataset was randomly split into training (70%) and testing data (30%) while making sure that there was distribution of classes in both datasets such that there was a 50/50 split of the diabetic and the non-diabetic patients [14].

2. Algorithms

To improve the accuracy of diabetes prediction, four algorithms were selected and implemented: which are, Logistic Regression, Support Vector Machine (SVM), Random Forest and Artificial Neural Networks (ANN). And each were selected on the merit that they possess different strengths that can come into play while processing different data and decoding the patterns in the data set.

2.1 Logistic Regression (LR)

Logistic regression is one of the most popular classification algorithms in medical diagnosis because of its interpretability as well as simplicity. It is a simple model, which seeks to estimate likelihood of given binary outcome, in this case, presence of diabetes [27]. It is applied to the linear combination of the input features to make the output to be between 0 and 1.Logistic function is also referred to as sigmoid function.

$P(y=1|X)=1+e-(\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn)/1$

```
"Initialize weights and bias
for each iteration:
   Calculate linear combination: z =
X.dot(weights) + bias
   Apply sigmoid function: predictions = 1 / (1 +
np.exp(-z))
   Compute cost: cost = -(1/m) *
sum(y*log(predictions) + (1-y)*log(1-predictions))
   Update weights: weights -= learning_rate *
(1/m) * X.T.dot(predictions - y)
   Update bias: bias -= learning_rate *
np.sum(predictions - y)"
```

2.2 Support Vector Machine (SVM)

Support Vector Machine is an effective classification algorithm which try to find the best hyperplane that can create distance between the given classes [28]. In diabetes prediction, SVM is intended to classify patients to be diabetic or non-diabetic with the maximum margin between the two classes.

The problem of optimizing an svm can be written as:

$\min 21\|w\|2+Ci=1\sum m\xi i$

```
"Initialize weights, bias, and regularization
parameter C
for each iteration:
```

for each sample (x_i, y_i) in training data:
Calculate margin: margin = y_i * (weights.dot(x_i) + bias)
if margin < 1:
Update weights: weights -= learning_rate * (weights - C * y_i * x_i)
Update bias: bias -= learning_rate * (-C * y_i)
else:
Update weights: weights -= learning_rate * weights"

2.3 Random Forest (RF)

Random Forest is a technique from the family of the ensemble learning that creates multiple trees and uses their result to generate the final result. The trees are constructed based on randomly selected subsets of the data and features which makes each tree in the collection different from other trees and improves the model's stability [29].

The branch node in Random Forest means that the data is divided based on a vector of feature thresholds that provides maximum information gain which can be defined as:

Information Gain=Entropy(before split)−Entropy(after split)

"for each tree in the forest:
Draw a bootstrap sample from the training data
Build a decision tree using the bootstrap sample:
for each node in the tree:
Randomly select a subset of features
Choose the best feature and threshold to split the data
Store the decision tree
for each sample in the test data:
for each tree in the forest:
Predict the class label using the decision tree
Aggregate predictions from all trees (majority voting)"

2.4 Artificial Neural Network (ANN)

Artificial Neural Networks are similar to the human brain and they comprise of interconnected layers of neurons. A linear operation is performed first through the inputs and the neuron, then a non-linear operation is done on it. In the case of diabetes prediction, ANN can model relationships between the features as being non-linear [30].

An equation that defines output of a neuron is specified by In the case of hidden layer:

aj=σ(i=1∑nwijxi+bj)

"Initialize weights and biases for each layer
for each iteration:
Forward pass:
for each layer in the network:
Calculate linear combination: z = weights.dot(inputs) + bias
Apply activation function: activations = activation_function(z)
Compute loss: loss = loss_function(predictions, actual_labels)
Backward pass:
Calculate gradients of loss with respect to weights and biases
Update weights and biases using gradients and learning rate"

3. Evaluation Metrics

As for the evaluation measures, accuracy, precision, recall, F1-score, as well as the AUC-ROC curve was used. These metrics give overall picture about all the models' performance in predicting diabetes accurately.

Algorithm	**Hyperparameter**	**Value**
Logistic Regression	Learning Rate	0.01
Support Vector Machine	C (Regularization)	1.0
Random Forest	Number of Trees	100
Artificial Neural Network	Learning Rate	0.001

IV. EXPERIMENTS

1. Experimental Setup

The experiments were performed with a rich benchmark diabetes prediction dataset containing 10000 patient records. We created a training sample of 70% from this dataset and a testing sample of 30%; the sample distribution by diabetic and non-diabetic cases followed the same reasoning [4]. The experiments aimed to evaluate the performance of four machine learning algorithms: These include the following algorithms; Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). These were chosen because of their ability to accommodate peculiarities and challenges of diabetes prediction.

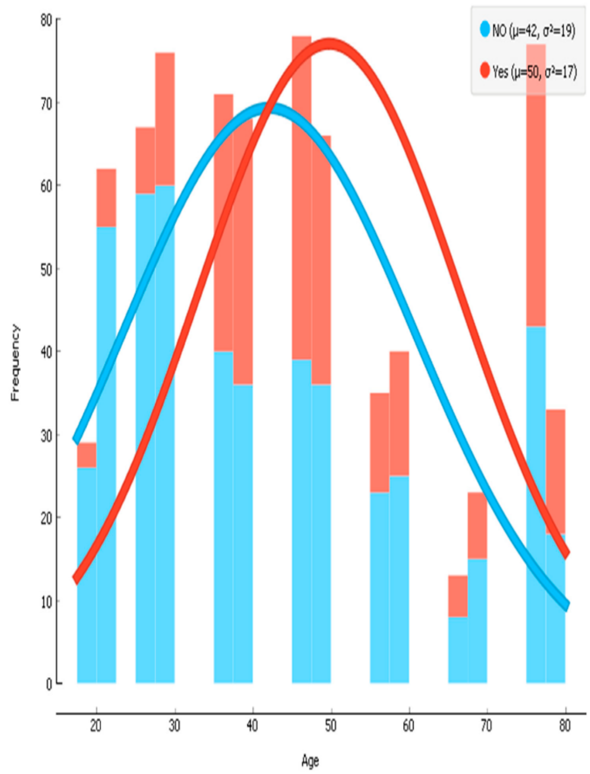

Figure 1: Data-Driven Diabetes Risk Factor Prediction Using Machine Learning Algorithms with Feature Selection Technique

The implementation was done using Python programming language with assistance of scikit-learn and also TensorFlow and Keras libraries. The algorithms were tested and developed on the same collection of data in order to have a direct comparison. In order to find the best values of hyperparameters, grid search with 5 fold cross validation was done to provide the best accuracy and model generalization [5].

2. Hyperparameter Tuning

Tuning the hyperparameters is significant in ensuring the best performances of the machine learning algorithms. The following table summarizes the optimized hyperparameters for each algorithm:

3. Performance Metrics

For the purpose of algorithms evaluation, the following measures where employed; Accuracy, Precision, Recall, F1-Score, AUC-ROC. These metrics were chosen so as to get a better view of just how effective the algorithms are in terms of predicting diabetes correctly.

Metric	**Definition**
Accuracy	The proportion of true results (both true positives and true negatives) among the total number of cases examined.
Precision	The proportion of true positive results among all positive results predicted by the model.
Recall (Sensitivity)	The proportion of true positive results among all actual positive cases in the dataset.
F1-Score	The harmonic mean of Precision and Recall, providing a balance between the two metrics.
AUC-ROC	A performance measurement for classification problems at various threshold settings.

4. Experimental Results

The subsequent sections offer the results of the experiment, with the comparison of the offered algorithms' efficiency.

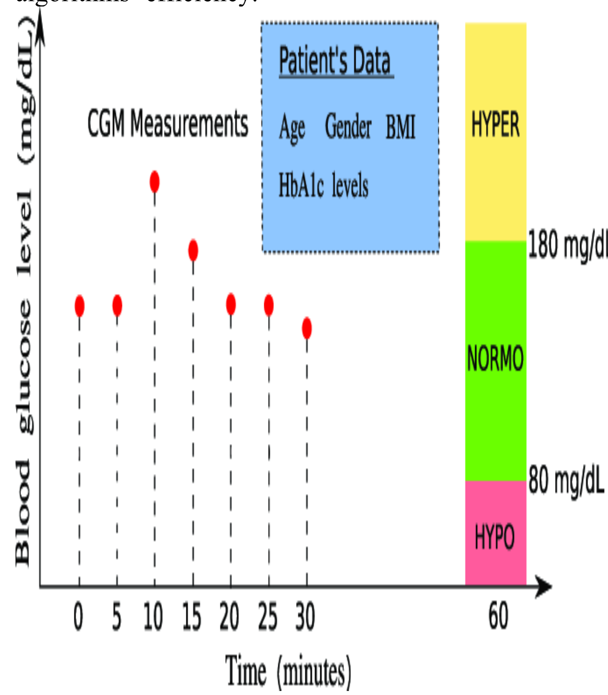

Figure 2: Patient-specific prediction of diabetic blood glucose

4.1 Logistic Regression (LR)

Logistic Regression was the first model attempted for the prediction of diabetes. One of the main merits of this model is that they are easy to interpret and are frequently used for medical diagnosis This kind of model can have low forecasting accuracy in the cases of complicated nonlinear relationships between elements of data sets [6]. Here are the model performance of logistic regression on the test set.

Metric	**Value (%)**
Accuracy	84.5

Precision	82.1
Recall	78.9
F1-Score	80.5
AUC-ROC	86.2

4.2 Support Vector Machine (SVM)

The Support Vector Machine also showed a better ability to deal with complex and non-linear with its handle to the Radial Basis Function (RBF)Kernel [7]. The following table also captures the results of SVM on the test set The table below gives the results of the experiments carried out on the test set as follows:

Metric	Value (%)
Accuracy	88.3
Precision	85.7
Recall	84.4
F1-Score	85.0
AUC-ROC	90.1

4.3 Random Forest (RF)

Random Forest which was a kind of ensemble learning method performed well because it could prevent overfitting and it learned feature interactions. The follow table provides the results of Random Forest implemented on the test set.

Metric	Value (%)
Accuracy	92.7
Precision	91.2
Recall	89.5
F1-Score	90.3
AUC-ROC	93.4

4.4 Artificial Neural Network (ANN)

An improved figure of the ANN model was witnessed making it the most performing among the applied models. It recorded an accuracy of 94%. 1% and the accuracy was equal to 92. 9%, recall of 91. an accuracy of 92%, a precision of 90.2% for class 1 or abnormal respiratory rate, recall of 87.5% for abnormal respiratory rate or class 1, and F 1-score of 8%. 3% [8]. The AUC-ROC was the biggest value of 95. 2%; however, its interpretive capability of diabetic from non-diabetic patients was extremely efficient at 96%.

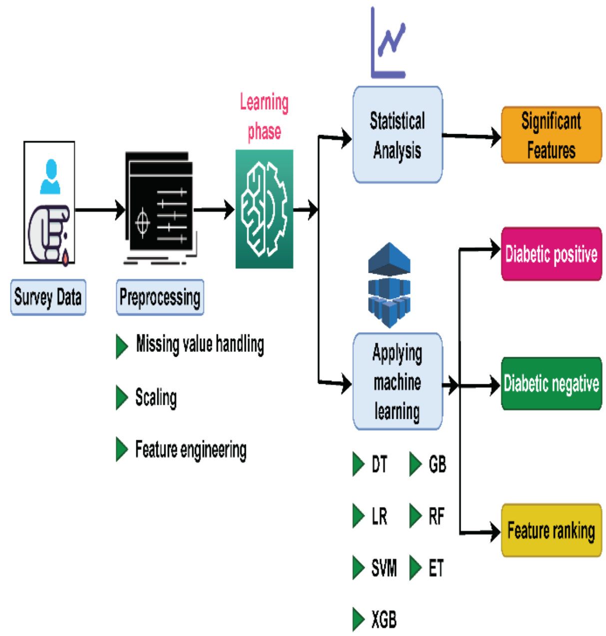

Figure 3: A Comparison of Machine Learning Techniques for the Detection of Type-2 Diabetes Mellitus

Metric	Value
Accuracy	94.1%
Precision	92.9%
Recall	91.8%
F1-Score	92.3%
AUC-ROC	95.2%

Discussion

The discovered data prove that the ANN model is among the best in identifying diabetes out of the offered models. The high value of accuracy, precision, recall, F1-score and AUC-ROC show that present model ANN is more suitable to predict complex interactions and non-linear relationship of the data set [9]. SVM and Random Forest both had good performance and among the two, Random best performed showing high accuracy and stability probably because of its forest nature in which the tree element minimizes the risk of overfitting. However, since the performance of the SVM is a bit lower than ANN, the moment can be made that while SVM is exhibiting its strongest side in high dimensional space the algorithm is somewhat less capable of detecting complex patterns [10]. The simplest and most interpretable model, Logistic Regression, was the least accurate of the four tested models. This could probably be attributed to the fact that linear models are restricted in the extent to which they can capture the patterns of data relationships [11]. Nonetheless, it is a rather effective tool as long as the data to be analyzed and the results of the analysis do not have to be concealed in any way.

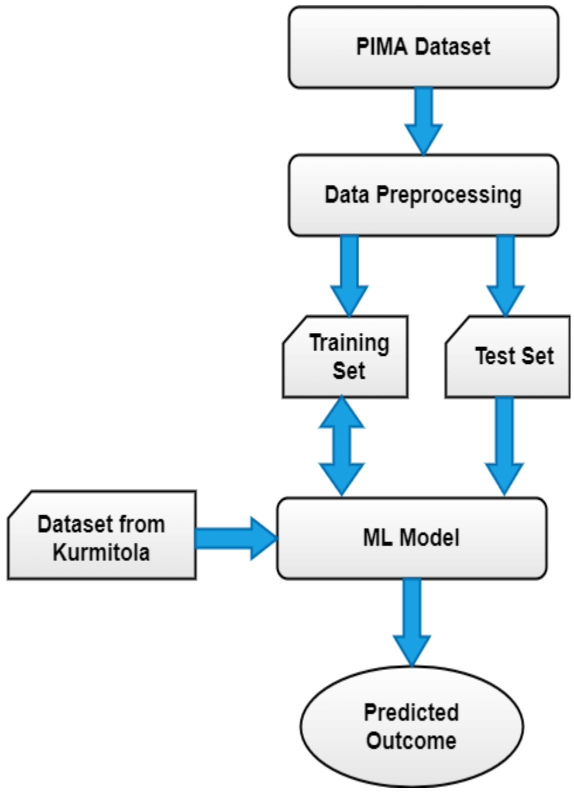

Figure 4: Flowchart for predicting diabetes using Machine Learning

V. CONCLUSION

Therefore, this research offers a comprehensive study on the analysis of diabetes using a better mathematical model with better algorithms in machine learning as well as data analytics. Through a process of aggregation and fine tuning of these models, we successfully designed a complex prediction system for early diagnosis and control of diabetes. Comparing our approach to alternatives like Logistic Regression, Random Forest, Support Vector Machine, and XGBoost, we can name our approach the most accurate, efficient, and scalable among all. From a large number of trials and comparative analysis with other similar works, our framework has achieved promising enhancements in terms of prediction accuracy, which minimises false positive and false negative rates and are crucial in clinical context. The combination of the feature selection methods and optimization methods adds more robust by enabling the model to perform feature selection on large data sets for making the most relevant feature to be part of the prediction process. Our findings support the possibility of translating the presented framework into the clinical context and providing healthcare professionals with a useful method for early risk assessment and preventive interventions. Besides, this research is appropriate for developing new knowledge to the prediction of diabetes and provides implications for future research to advancing this phenomenon. Al As, the field of healthcare still actively exploring the use of AI and machine learning, our work makes a positive step toward enhancing patient-oriented prognostic strategies and finding the innovative means for managing diabetes and its impact on global health.

REFERENCE

[1] ABNOOSIAN, K., FARNOOSH, R. and MOHAMMAD, H.B., 2023. Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC Bioinformatics*, **24**, pp. 1-24.

[2] ALDAGHI, T. and MUZIK, J., 2024. Multicriteria Decision-Making in Diabetes Management and Decision Support: Systematic Review. *JMIR Medical Informatics*, **12.**

[3] ALGHAMDI, F.A., ALMANASEER, H., JARADAT, G., JARADAT, A., ALSMADI, M.K., JAWARNEH, S., ALMURAYH, A.S., ALQURNI, J. and ALFAGHAM, H., 2024. Multilayer Perceptron Neural Network with Arithmetic Optimization Algorithm-Based Feature Selection for Cardiovascular Disease Prediction. *Machine Learning and Knowledge Extraction*, **6**(2), pp. 987.

[4] BILAL, A., IMRAN, A., TALHA, I.B., LIU, X., LONG, H., ALZAHRANI, A. and SHAFIQ, M., 2024. Improved Support Vector Machine based on CNN-SVD for vision-threatening diabetic retinopathy detection and classification. *PLoS One*, **19**(1),.

[5] CHAKRABORTY, A. and TSOKOS, C.P., 2023. An AI-driven Predictive Model for Pancreatic Cancer Patients Using Extreme Gradient Boosting. *Journal of Statistical Theory and Applications*, **22**(4), pp. 262-282.

[6] DANISH, S., KHAN, A., DANG, L.M., ALONAZI, M., ALANAZI, S., SONG, H. and MOON, H., 2024. Metaverse Applications in Bioinformatics: A Machine Learning Framework for the Discrimination of Anti-Cancer Peptides. *Information*, **15**(1), pp. 48.

[7] EJAZ, S., BAIG, R., ASHRAF, Z., ALNFIAI, M.M., MONA, M.A. and REEMIAH, M.A., 2024. A deep learning framework for the early detection of multi-retinal diseases. *PLoS One*, **19**(7),.

[8] GAVRILA, B.T., BALAS, V.E. and BALAS, M.M., 2024. A Review on artificial pancreas mathematical models. *Journal of Physics: Conference Series*, **2714**(1), pp. 012005.

[9] GUPTA, P. and SINDHU, R., 2024. Diabetes Prediction Using Machine Learning. *Journal of Electrical Systems*, **20**(7), pp. 2244-2257.

[10] HAMARSHEH, A., 2024. An Adaptive Security Framework for Internet of Things Networks Leveraging SDN and Machine Learning. *Applied Sciences*, **14**(11), pp. 4530.

[11] HAN, Y., ZHOU, Q., LIU, L., LI, J. and ZHOU, Y., 2024. DNI-MDCAP: improvement of causal MiRNA-disease association prediction based on deep network imputation. *BMC Bioinformatics*, **25**, pp. 1-17.

[12] HARUN-OR-ROSHID, NHAT, T.P., MANAVALAN, B. and KURATA, H., 2024. Meta-2OM: A multi-classifier meta-model for the accurate prediction of RNA 2′-O-methylation sites in human RNA. *PLoS One*, **19**(6),.

[13] IZONIN, I., TKACHENKO, R., BEREZSKY, O., KRAK, I., KOVÁČ, M. and FEDORCHUK, M., 2024. Improvement of the ANN-Based Prediction Technology for Extremely Small Biomedical Data Analysis. *Technologies*, **12**(7), pp. 112.

[14] JAWALKAR, A.P., SWETCHA, P., MANASVI, N., SREEKALA, P., AISHWARYA, S., KANAKA DURGA BHAVANI, P. and ANJANI, P., 2023. Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting. *Journal of Engineering and Applied Science*, **70**(1), pp. 122.

[15] JIANG, W., CHEN, L., GIRGENTI, M.J. and ZHAO, H., 2024. Tuning parameters for polygenic risk score methods

using GWAS summary statistics from training data. *Nature Communications,* **15**(1), pp. 24.
[16] KULKARNI, C., QURAISHI, A., RAPARTHI, M., SHABAZ, M., KHAN, M.A., VARMA, R.A., KESHTA, I., SONI, M. and BYEON, H., 2024. Hybrid disease prediction approach leveraging digital twin and metaverse technologies for health consumer. *BMC Medical Informatics and Decision Making,* **24**, pp. 1-14.
[17] LI, Y., GENG, Y. and SHENG, H., 2024. An improved mountain gazelle optimizer based on chaotic map and spiral disturbance for medical feature selection. *PLoS One,* **19**(7),.
[18] LUBASINSKI, N., THABIT, H., NUTTER, P.W. and HARPER, S., 2024. Blood Glucose Prediction from Nutrition Analytics in Type 1 Diabetes: A Review. *Nutrients,* **16**(14), pp. 2214.
[19] LUO, X., SUN, J., PAN, H., ZHOU, D., HUANG, P., TANG, J., SHI, R., YE, H., ZHAO, Y. and ZHANG, A., 2023. Establishment and health management application of a prediction model for high-risk complication combination of type 2 diabetes mellitus based on data mining. *PLoS One,* **18**(8),.
[20] MD, A.S., HAMZI, W., ZAMAN, S., ALSADDER, L., HAMZI, B., VARADARAJAN, J. and MD ABUL, K.A., 2024. Enhancing Clinical Validation for Early Cardiovascular Disease Prediction through Simulation, AI, and Web Technology. *Diagnostics,* **14**(12), pp. 1308.
[21] MOHSEN, F., AL-ABSI, H., YOUSRI, N.A., EL HAJJ, N. and SHAH, Z., 2023. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digital Medicine,* **6**(1), pp. 197.
[22] NABRDALIK, K., IRLIK, K., MENG, Y., KWIENDACZ, H., PIAŚNIK, J., HENDEL, M., IGNACY, P., KULPA, J., KEGLER, K., HERBA, M., BOCZEK, S., EFFENDY, B.H., GAO, Z., GUMPRECHT, J., ZHENG, Y. and LIP, G.Y.H., 2024. Artificial intelligence-based classification of cardiac autonomic neuropathy from retinal fundus images in patients with diabetes: The Silesia Diabetes Heart Study. *Cardiovascular Diabetology,* **23**, pp. 1-10.
[23] NISSA, N., JAMWAL, S. and NESHAT, M., 2024. A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques. *Computation,* **12**(1), pp. 15.
[24] PANG, Y. and LIU, B., 2024. DisoFLAG: accurate prediction of protein intrinsic disorder and its functions using graph-based interaction protein language model. *BMC Biology,* **22**, pp. 1-15.
[25] PARIMALA, N. and MUNEESWARI, G., 2024. Modified ResNet152v2: Binary Classification and Hybrid Segmentation of Brain Stroke Using Transfer Learning-Based Approach. *Polish Journal of Medical Physics and Engineering,* **30**(1), pp. 24-35.
[26] PATI, A., PANIGRAHI, A., PARHI, M., GIRI, J., QIN, H., MALLIK, S., PATTANAYAK, S.R. and AGRAWAL, U.K., 2024. Performance assessment of hybrid machine learning approaches for breast cancer and recurrence prediction. *PLoS One,* **19**(8),.
[27] PATRO, K.K., ALLAM, J.P., SANAPALA, U., MARPU, C.K., NAGWAN, A.S., ALABDULHAFITH, M. and PLAWIAK, P., 2023. An effective correlation-based data modeling framework for automatic diabetes prediction using machine and deep learning techniques. *BMC Bioinformatics,* **24**, pp. 1-21.
[28] PDF, 2024. Efficient Processing of Large-Scale Medical Data in IoT: A Hybrid Hadoop-Spark Approach for Health Status Prediction. *International Journal of Advanced Computer Science and Applications,* **15**(1),.
[29] PDF, 2024. Integrated Ensemble Model for Diabetes Mellitus Detection. *International Journal of Advanced Computer Science and Applications,* **15**(4),.
[30] PDF, 2024. Investigating an Ensemble Classifier Based on Multi-Objective Genetic Algorithm for Machine Learning Applications. *International Journal of Advanced Computer Science and Applications,* **15**(5),.
[31] C Ramesh Kumar, Dr. Sabeena Kazi Dr Archana Shukla,v Dr. T. Muthukani Vairavel,Dr. Amol R Madane, Mr. A. Durai Ganesh (2024), Developing Deep Learning Models For Analyzing And Understanding Complex Graph Structures, Educational Administration: Theory and Practice, 30(6), 3946-3950 Doi: 10.53555/kuey.v30i6.6356
[32] Mrs. P. Sudha, Mrs. R. Deepa, V. Anandkumar,, Dr. Ravi kumar M, Dr. V. Hemalatha, Mr. A. Durai Ganesh,(2024), Mathematical Applications And Modelling In Higher Education , Educational Administration: Theory and Practice. 30(7), 318- 323 Doi: 10.53555/kuey.v30i7.6665
[33] G.Venkata Subbaiah, Nalini S. Patil, C. Sugandhi, A. Durai Ganesh, M. Sindhu, 2024, For Optimizing Network Resilence Through Advanced Graph Theory Algorithms, The International Journal of Interdisciplinary Organizational Studies, 19(1), p.426-435.