ISERT-2023
International Conference on Interdisciplinary Studies in Education, Research and Technology
In Association with International Journal of Scientific Research in Science and Technology
Volume 10 | Issue 14 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

A Hybrid GAN-ANN-Based Model for Diabetes Prediction

Djalila Boughareb *[1], Hazem Bensalah[2], Hamid Seridi[1]
[1*]Labstic laboratory, Computer Science Department, University of 8 May 1945, Guelma-Algeria
[2]Liap laboratory, Computer Science Department, Faculty of Exact Sciences University of El Oued, El Oued-Algeria
[1]Labstic laboratory, Computer Science Department, University of 8 May 1945, Guelma-Algeria

ABSTRACT

Millions of people around the globe have diabetes, which is a widespread chronic condition. It occurs when the body cannot properly process glucose, the primary energy source for the body's cells. This can lead to various health complications, including cardiovascular diseases, kidney damage, and blindness. Therefore, a predictive tool is urgently needed to help physicians detect the disease early and suggest the necessary lifestyle changes to halt its progression. Artificial intelligence technologies, such as machine and deep learning, have emerged as a means to reduce human effort and increase automation with minimal errors. This paper proposes a diabetes detection and prediction system based on deep learning techniques. Our approach employs Generative Adversarial Networks (GAN) to augment and impute data, while Artificial Neural Networks (ANN) are used for data training and decision-making. Experimental results on the Pima dataset demonstrate promising prediction outcomes compared to eight other machine learning techniques.

Keywords: Artificial Intelligence, Deep Learning, GAN, ANN, Machine Learning, Diabetes Detection, Diabetes Prediction.

I. INTRODUCTION

The number of chronic patients worldwide, particularly those with diabetes, has significantly risen. Unfortunately, underdeveloped countries struggle with insufficient infrastructure to handle the growing demand, causing private health services to become increasingly costly. While diabetes is not a fatal disease, severe issues in organs like the kidneys, eyes, and peripheral organs might result from it.

Diabetes can be divided into two main types: (i) Type 1 diabetes, commonly known as insulin-dependent diabetes, is frequently discovered in adolescents and young children. It happens due to the immune system attacking and destroying insulin-producing cells, which stops all insulin synthesis completely. As a result, patients require daily insulin injections or an insulin pump. Failure to manage Type 1 diabetes correctly can result in complications such as neuropathy, nephropathy, retinopathy, and cardiovascular disease.

(ii) Diabetes type 2 is also called non-insulin-dependent diabetes. This type of diabetes is the most prevalent. It is usually diagnosed in older adults but can also occur in younger individuals. This type occurs when the body still produces insulin, but it is either not enough or the body is unable to use it effectively. The majority of the time, oral medications and lifestyle modifications (such as diet and exercise) help treat type 2 diabetes, although insulin may occasionally be required as well. If it is not adequately treated, it can have the same consequences as type 1 diabetes.

There is a variation of diabetes that occurs during pregnancy named gestational diabetes mellitus (GDM). Although it usually goes away after birth, both the mother and the child are at an increased risk of having type 2 diabetes in the future. During pregnancy, the body becomes less sensitive to insulin in GDM, which raises blood sugar levels.

As they say, "Prevention is better than cure." It is important to note that proper management, including regular monitoring and treatment, can help prevent or delay the onset of complications in all forms of diabetes. The prediction of diabetes has been examined in several works [4], [8], [13], [17], [20], [32]. These studies have considered various patient data elements, such as molecular trait variability, environmental factors, electronic health records (EHRs), and way of life. The developed models are based on historical experience and claims using health conditions and vital signs. The most commonly used dataset is the Pima Indians Diabetes Database [19], which consists of 768 samples, 268 individuals who are diagnosed with the condition, and eight independent variables that are used to determine if the patient has diabetes or not. Classes in this dataset are underrepresented, and when analyzing a medical dataset where the proportion of healthy patients is much higher than that of affected patients, classification algorithms may struggle to identify the minority classes because even if they classify every minority class incorrectly, the program will still have a low error rate [10]. Data augmentation is a possible solution to tackle this problem, which consists of increasing the representation of the minority class and helping to avoid overfitting.

With considerable data flows, it is now possible to apply algorithms that give automatic and more accurate answers to different medical problems. Despite the promising results obtained using machine or/and deep learning (ML/DL) approaches. In order to improve patient outcomes and reduce the burden of diabetes on healthcare systems, this work proposes an accurate and helpful system allowing from a set of clinical data to detect whether the person has diabetes or to predict whether he/she is in the pre-diabetes phase.

Our method stands out by offering a complete solution that not only increases the accuracy of classification but also tackles the common problems of unequal class distribution in healthcare data. It provides a more effective approach for use in real-life healthcare situations. In this work, our contributions are:

- We used Generative Adversarial Networks (GANs) for data imputation to minimize the impact of missed data.
- We used Artificial Neural Networks (ANNs) for data training and decision-making.
- We evaluated the suggested method's effectiveness compared to eight machine learning techniques, including Naive Bayes, artificial neural networks, support vector machines, and random forests.

The remaining sections of the paper are structured as follows: Section 2 evaluates several comparable publications, Section 3 provides the study methods, Section 4 specifies the assessment, Section 5 discusses the collected results, and Section 6 ends the paper and offers potential areas for future work.

II. Related Work

The identification of diabetes using machine learning and deep learning has been the subject of numerous relevant publications in the literature.

A. Machine Learning Approaches

Using machine learning algorithms to predict diabetes has been an important subject in research for more than a decade. Numerous machine learning techniques have been employed to forecast diabetes, such as decision trees [6], [7], [25], Random Forest [3], [21], [24], [27], k-Nearest Neighbors (k-NN) [29], neural networks [9],

[30], and Support Vector Machine (SVM) [31], [37]. For instance, [31] explored the use of SVM models to predict the onset of type 2 diabetes based on various clinical and demographic features. The outcomes demonstrated that the SVM models had a high degree of accuracy in predicting the onset of type 2 diabetes. The authors reported an overall accuracy of 94.5% over PIMA for their SVM model.

For predicting diabetes, hybrid approaches that incorporate several machine learning algorithms have also been explored. By combining these methods, the effectiveness of individual models is increased while taking advantage of the advantages of various algorithms. For instance, in [36] a hybrid model of k-means and decision trees has been proposed. Also, [38] used a combination of random forest and support vector machine (SVM) algorithms to improve the prediction performance of diabetes.

Additionally, the study of [18] examined the performance of different machine learning algorithms on the Pima Indian dataset, including decision trees, k-NN, logistic regression, and SVM. They discovered that the SVM method has the highest level of precision. While [16] employed mutual information as a feature selection strategy and machine learning classification techniques, such as decision tree, SVM, Random Forest, Logistic Regression, K-NN, and various ensemble techniques, to ascertain which algorithm delivers the best prediction results. With 81% accuracy, the suggested system delivered the best result in the XGBoost classifier.

More recently, [33] using Decision Tree, SVM, and Naive Bayes, they created a model that successfully predicts diabetes in a patient group. Additionally, [2] combined PIMA with logistic regression, XGBoost, gradient boosting, decision trees, ExtraTrees, random forest, and the light gradient boosting machine (LGBM). According to the results of these classifiers, the LGBM classifier has the highest accuracy (95.20%), when compared to the other algorithms. An overview of the state of the art in using machine learning to predict diabetes mellitus can be found in [5], [40] and [41].

B. Deep Learning Approach

Artificial neural networks are the foundation of the machine learning subfield known as deep learning. It has been increasingly used for diabetes prediction in recent years. These models have been shown to be effective in handling high-dimensional and non-linear data and have achieved good performance on various datasets. Some works that have used Convolutional Neural Networks (CNNs) [11], Recurrent Neural Networks (RNNs) [39], Generative Adversarial Networks (GANs) [22], and Deep Belief Networks (DBNs) [28] have shown their effectiveness in identifying signs of diabetic retinopathy, analyzing time series data and generating synthetic data, and analyzing various features such as age, body mass index, and blood pressure. For instance, [12] used CNNs for diabetes prediction by analyzing images of the retina to identify signs of diabetic retinopathy. Additionally, [35] used a modified version of the generative adversarial network, where the generator computed Blood glucose (BG) predictions using a recurrent neural network with gated recurrent units and the discriminator used a one-dimensional convolutional neural network to differentiate between the predictive and real BG values.

More recently, DBNs have been used by [1] for solving regression problems in detecting diabetes by analyzing various features such as blood pressure, body mass index, and age. Furthermore, [26] used vote ensemble feature selection and DBNs for diabetes early detection in a Bangladeshi online library of prediagnosed patients' answers. Ref. [14], [23], and [34] provide an overview of the state of the art in applying deep learning to predict diabetes mellitus.

As far as we are aware, no research has combined artificial neural networks (ANNs) and generative adversarial networks (GANs) to predict diabetes.

III. Research Methodology

This section delineates the research methodology employed in this study.

A. *Dataset*

The Pima Indians diabetes database is a collection of medical data on Pima Indian women who live close to Phoenix, Arizona and who underwent World Health Organization-required diabetes testing. The dataset includes 8 features and 768 instances, including:

- Pregnancies: The total number of pregnancies;
- Blood Pressure: Diastolic blood pressure (mm Hg);
- Plasma Glucose Concentration: 2-hour oral glucose tolerance test;
- Triceps skin fold thickness in millimetres;
- Body mass index is calculated as follows: weight in kg/(height in m)2;
- Insulin: 2-hour serum insulin (mu U/ml);
- BMI: body mass index;
- DiabetesPedigreeFunction: diabetes pedigree function;
- Age: Age (years); Class variable: 0 or 1;
- Outcome: Age (years);

This dataset, which has 268 samples classified as diabetes and 500 as non-diabetic, is frequently used for binary classification issues.

The Pima Indian dataset is often regarded as biased due to its imbalanced class distribution, focusing on a specific ethnic group (Pima Indians). Also, the dataset is often criticized for its imbalanced class distribution, meaning that there are disproportionately more instances of individuals with diabetes compared to those without. This imbalance can pose challenges for machine learning models, as they might exhibit a bias towards predicting the majority class, potentially leading to less accurate predictions for the minority class. Addressing the imbalance in the dataset and ensuring accurate model evaluation will be discussed in the following sections, employing Generative Adversarial Networks (GAN) for enhanced data generation and improved representation of minority classes.

B. *Data imputation with Generative Adversarial Networks*

A generative modeling strategy called generative adversarial networks, or GANs [15], was suggested by Goodfellow et al. in 2014. The foundation of GANs is a game theoretic scenario where the generating network faces off against an opponent.

A generative model is trained in this stage to deceive a feedforward classifier. The feedforward classifier should recognize all samples from the generative model as false, whereas all samples from the training set

should be recognized as real. Any organized pattern the feedforward network may identify is highly prominent in this approach. The discriminator and the generator are the two primary parts of a GAN.

The generator is a neural network that uses random noise input to generate fresh data samples that resemble training data; its network directly generates samples using equation (1), which maps noise variable z to data space x and defines a stochastic process to simulate data x.

$$x=g(z;\theta(g)) \quad (1)$$

Our model has three layers, with the 'ReLU' function activating two of them. Then, the linear function will trigger the output layer, and its dimension will be the same as the dimension of the dataset (9 columns).

The discriminator is also a neural network that takes in a sample and produces a probability that the sample is real. It then compares this probability to a threshold value (usually 0.5) to make its decision. It tries to distinguish between the generated samples and the real samples emitting a probability value given by $d(x;\theta(d))$. This indicates the likelihood that x represents an authentic training sample, as opposed to a synthetic sample generated by the model. Our discriminator is a sequential model with three dense layers. The function activates the first two layers, while the 'sigmoid' function activates the output layer since it will determine if the input samples are true or false. We produced 1602 lines, comprising 602 lines of real data and 1000 lines of generated data.

C. *Diabetes prediction using ANN*

We employed neural networks and a sequential model with three layers for data training. The rectified linear unit (ReLU) function operates on the first two levels to activate them. 1000 neurons make up the first layer, while 500 neutrons make up the second. The sigmoid function activates the output layer because we have a 0 or 1 output. A mathematical function called the sigmoid function converts any input value to a number between 0 and 1. Equation 2 defines it.

$$F(x) = 1 / (1 + e^{(-x)}) \quad (2)$$

The output of the sigmoid function is also known as the "squashing" function, as it compresses the output of a neuron to the range between 0 and 1. The sigmoid function is often used as the activation function in the output layer of a binary classification neural network, as it can be interpreted as the probability of the positive class. It was employed to choose the neural network's output. It converts the outcomes to a number between 0 and 1 or -1 and 1.

On the other hand, the ReLU function is a straightforward mathematical formula that converts every input value to the maximum of that value and zero. Eq. (3) provides a definition for it.

$$F(x) = \max(0, x) \quad (3)$$

The ReLU function is particularly useful in deep neural networks as it can improve the training time and the model's performance by solving the vanishing gradient problem. The ReLU function is often used as the activation function in the hidden layers of neural networks, as it is computationally efficient and does not saturate for positive input values. Figure 1 illustrates a flowchart detailing the proposed system.

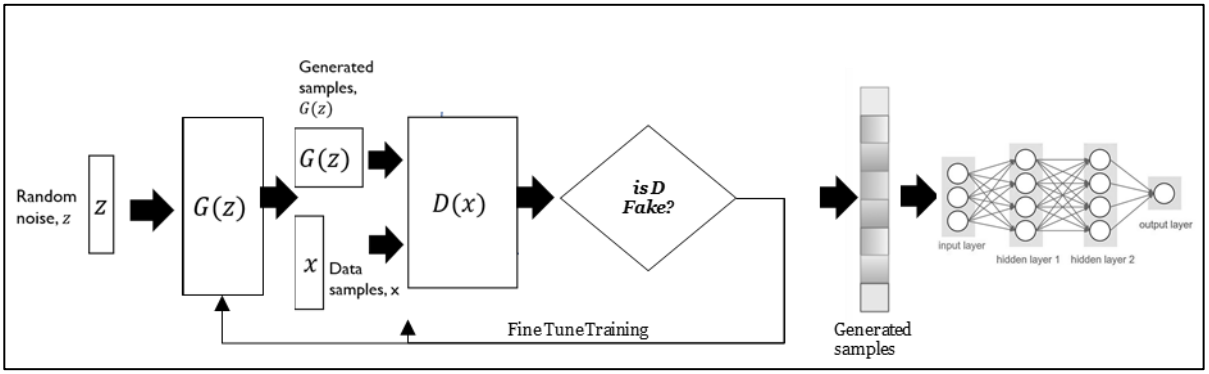

Fig. 1 The proposed GAN-ANN Approach

IV. Results and Discussion

The proposed approach sought to address the issue of imbalanced datasets within the biomedical domain. In this phase, we compared the GAN-ANN model with eight alternative machine learning methods, encompassing Naive Bayes, AdaBoost, and ANN, along with Decision Trees, K Nearest Neighbours (KNN), Quadratic Discriminant Analysis (QDA), Random Forest, and Support Vector Machine (SVM). Indeed, these algorithms are commonly used and generally robust for binary classification tasks. However, their performance may vary based on the dataset's characteristics and the specific requirements of the problem at hand, which the achieved results can prove.

The assessment of precision, recall, f1-score, and accuracy in Equations (4–7) was employed to evaluate performance. Table 1 and Figure 2 compare eight machine learning algorithms and the GAN-ANN-based technique.

$$\text{Precision}=TP/(TP+FP) \quad (4)$$

$$\text{Recall} =TP/(TP+FN) \quad (5)$$

$$\text{F1-score}=TP/(TP+1/2(FP+FN)) \quad (6)$$

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \quad (7)$$

Such as:

- TP stands for true positives.
- FP stands for the number of false positives.
- TN stands for the number of true negatives.
- FN stands for false negatives.

TABLE I. COMPARISON OF THE GAN-ANN-BASED APPROACH AND EIGHT MACHINE LEARNING ALGORITHMS

Approaches	+/-	Precision	Recall	F1-score	Accuracy
GAN-ANN	0	0.94	0.96	0.95	0.94
	1	0.95	0.93	0.94	
Decision tree	0	0.84	0.77	0.80	0.75
	1	0.63	0.73	0.68	
KNN	0	0.77	0.82	0.79	0.73

	1	0.63	0.56	0.60	
QDA	0	0.84	0.82	0.83	0.78
	1	0.68	0.71	0.70	
Random Forest	0	0.81	0.81	0.81	0.75
	1	0.65	0.65	0.65	
SVM	0	0.78	0.88	0.83	0.77
	1	0.72	0.56	0.63	
Naïve bayes	0	0.83	0.80	0.81	0.77
	1	0.66	0.71	0.68	
ANN	0	0.70	0.92	0.79	0.69
	1	0.67	0.29	0.41	
AdaBoost	0	0.80	0.79	0.79	0.73
	1	0.62	0.64	0.63	

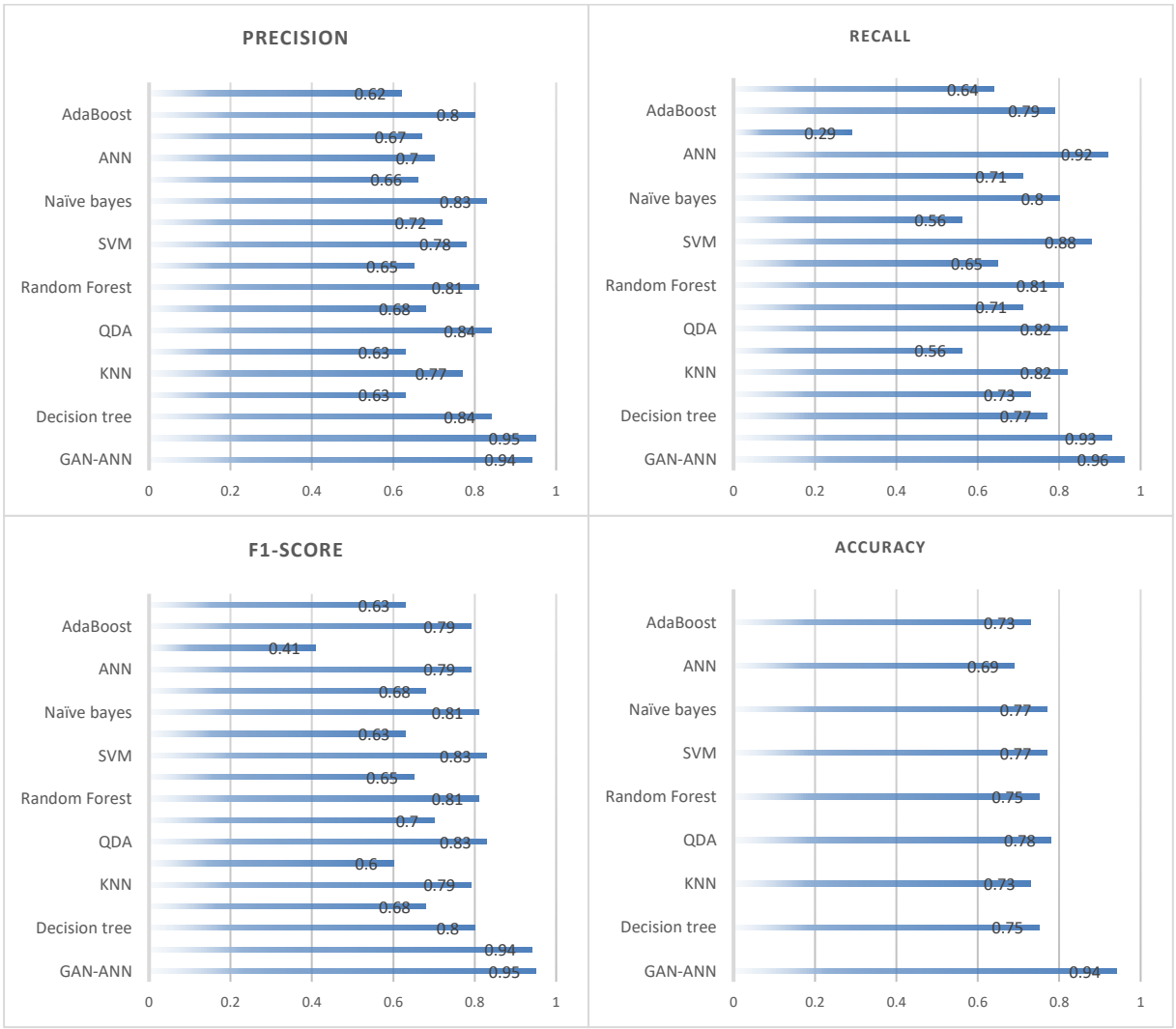

Fig. 2 Evaluating the GAN-ANN-based Approach Against Eight Machine Learning Algorithms

Based on the results, the developed GAN-ANN model exhibited notable performance metrics, including an accuracy of 94%, precision of 95%, recall of 93%, and an f1-score of 95% for predicting diabetes. Significantly, the model outperformed the eight alternative machine learning techniques regarding precision, recall, f1-score, and overall accuracy.

All machine learning achieved accuracy values are very close to each other. This suggests that the performance of the compared models is very similar and that there is little difference in their accuracies. Generally, SVM is often adequate for binary classification tasks; in our case, it achieved an accuracy 0.77. Adaboost and Naïve Bayes rely on probabilities to make predictions. In this case, Naïve Bayes outperformed AdaBoost with an accuracy of 0.77 and 0.73, respectively.

Meanwhile, QDA is a generative model that uses Bayes' theorem to compute the posterior probability of each class given an input. The algorithm gave the best accuracy value among the eight comparison algorithms; it achieved 0.78. KNN and decision trees gave low accuracy values, 0.73 and 0.75, respectively. Decision trees can be sensitive to even minor changes in the data, resulting in various tree architectures and forecasts. The tree may become weaker and more challenging to understand as a result. At the same time, KNN can be sensitive to the number of neighbors (k) selected. While a significant value of k can result in underfitting and subpar performance, a small value of k can result in overfitting. When the training set is small, artificial neural networks (ANNs) are prone to overfitting the training data. This can lead to poor performance when applied to new, untested data, as reflected in its lowest accuracy. However, In the GAN-ANN model, GANs could generate realistic data even when there was noise or missing information. This capability improved the ability of the ANN classifier to handle input data that is noisy or incomplete. By combining the strengths of both networks, we enhanced the accuracy and robustness of the classification task, leading to improved overall performance.

Our approach has practical implications that span from advancing early-stage diagnostics and treatment to influencing policies and ultimately elevating the overall quality of patient care while mitigating complications.

V. CONCLUSION

The study utilized a hybrid GAN-ANN approach to predict diabetes in an Indian population, demonstrating superior performance compared to eight other machine learning algorithms with a 94% accuracy. Despite this achievement, the study's limitations include a relatively small sample size, even after data imputation, potentially limiting the generalizability of the findings to other populations. Furthermore, the study overlooked additional diabetes-associated variables, such as primary predictor characteristics, family history, and lifestyle. To enhance future research, there will be a concerted effort to gather more comprehensive and representative datasets spanning a wider range of ethnicities, socio-economic backgrounds, and health conditions. Additionally, addressing biases in biomedical datasets will be a crucial focus, emphasizing developing and implementing robust strategies, including using fairness-aware machine learning algorithms, to ensure equitable representation across diverse demographic groups.

VI. REFERENCES

[1] M. Wijayaa, D. S. Ikawahyunia, R. Geaa, F. Maedjaja, "Role Comparison between Deep Belief Neural Network and NeuroEvolution of Augmenting Topologies to Detect Diabetes," International Journal On Informatics Visualization, vol. 5, no. 2, pp. 156–161, 2021.

[2] A.Shamreen, A. M. Sumeet, N. V. A. Osvin, "Prediction of Type-2 Diabetes Mellitus Disease Using Machine Learning Classifiers and Techniques," Frontiers in Computer Science, vol. 4, 2022. doi: 10.3389/fcomp.2022.835242.

[3] K. Shrivastava, V. Karthikeyan, S. Kaushik, M. Sudagar, "Early Diabetes Prediction using Random Forest," in Proceedings of the 3rd International Conference on Electronics and Sustainable Communication Systems, Coimbatore, India, August 17-19, 2022, pp. 1154–1159.

[4] K. Srivastava, Y. Kumar, P. K. Singh, "Hybrid diabetes disease prediction framework based on data imputation and outlier detection techniques," Expert Systems, vol. 39, no. 3, 2021. doi: 10.1111/exsy.12785.

[5] Patil, M. Parhi, B. K. Pattanayak, "A review on prediction of diabetes using machine learning and data mining classification techniques," International Journal of Biomedical Engineering and Technology, vol. 41, no. 1, pp 83-109, 2023. doi: 10.1504/IJBET.2023.128514.

[6] S. Abdullah, "Assessment of the risk factors of type II diabetes using ACO with self-regulative update function and decision trees by evaluation from Fisher's Z-transformation," Medical and Biological Engineering and Computing, vol. 60, pp. 1391–1415, 2022. doi: 10.1007/s11517-022-02530-2.

[7] U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, A. Ali, "Intelligent Machine Learning Approach for Effective Recognition of Diabetes in E-Healthcare Using Clinical Data," Sensors, vol. 20, no. 9, 2020.

[8] Viloriaa, Y. Herazo-Beltranb, D. Cabrerac, O. B. Pineda, "Diabetes Diagnostic Prediction Using Vector Support Machines," in Proceedings of the 11th International Conference on Ambient Systems, Networks and Technologies, Warsaw, Poland, April 6-9, 2020.

[9] Ossai, N. Wickramasinghe, "Sentiments prediction and thematic analysis for diabetes mobile apps using Embedded Deep Neural Networks and Latent Dirichlet Allocation," Artificial Intelligence in Medicine, volume 138, 2023.

[10] F. H. K. dos Santos Tanaka, C. Aranha, "Data Augmentation Using GANs," in 2020 IEEE International Conference on Big Data, Atlanta, GA, USA, December 10-13, 2020, pp. 5048-5053.

[11] G. Swapna, Kp. Soman, R. Vinayakumar, "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals," Procedia Computer Science, vol. 132, pp. 1253–1262, 2018.

[12] G. Thomas, A. Sampaul, R. Y. Harold, J. E. Golden, S. Vimal, R. Seungmin, N. Yunyoung, "Intelligent Prediction Approach for Diabetic Retinopathy Using Deep Learning Based Convolutional Neural Networks Algorithm by Means of Retina Photographs," Computers Materials and Continua, vol. 66, no. 2, pp. 1613–1629, 2021.

[13] H. Aktün, M. Acet, "Gestational diabetes mellitus screening and outcomes," Journal of the Turkish German Gynecology Association, vol. 16, no. 1, pp. 25–29, 2015.

[14] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, O. P. Verma, "Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction," Complex and Intelligent Systems, vol. 8, pp. 3073–3087, 2022.

[15] J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial networks, in Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, December 8-13, 2014, Z. Ghahramani, M. Welling, C. Cortes, Eds. MA: MIT Press, pp. 2672–2680.

[16] Tasin, T. U. Nabil, S. Islam, R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," Healthcare Technology Letters, pp. 1–10, 2022.

[17] J. Li, P. Yuan, X. Hu, J. Huang, L. Cui, J. Cui, X. Ma, T. Jiang, X. Yao, J. Li, Y. Shi, Z. Bi, Y. Wang, H. Fu, J. Wang, Y. Lin, C. Pai, X. Guo, C. Zhou, L. Tu, J. Xu, "A tongue features fusion approach to predicting prediabetes and diabetes with machine learning," Journal of Biomedical Informatics, vol. 115, 2021. doi: 10.1016/j.jbi.2021.103693.

[18] J. Revathya, D. Selvanayagi, "Comparative Analysis of Predicting the Diabetic Disease Using Machine Learning Techniques," Advances in Parallel Computing Algorithms, Tools and Paradigms, vol. 41, D. J. Hemanth, T. N. Nguyen, J. Indumathi, Advances in Parallel Computing, IOS Press, pp. 155–160, 2022.

[19] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, R. S. Johannes, "Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus," in Proceedings of the annual symposium on computer application in medical care, pp. 261–265, 1988.

[20] K. J. Wang, A. M. Adrian, K. H. Chen, K. M. Wang, "An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus," Journal of Biomedical Informatics, vol. 54, pp. 220–229, 2015. doi: 10.1016/j.jbi.2015.02.001.

[21] K. VijiyaKumar, B. Lavanya, I. Nirmala, S. S. Caroline, "Random Forest Algorithm for the Prediction of Diabetes," in 2019 IEEE International Conference on System, Computation, Automation and Networking, Pondicherry, India, March 29-30, 2019, pp. 1–5.

[22] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, D. Lee, "DETERRENT: Knowledge Guided Generative Adversarial Networks for Detecting Healthcare Misinformation," in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, USA, ACM, NY, USA, pp. 492–502, 2020.

[23] L. Fregoso-Aparicio, J. Noguez, L. Montesinos, J. A. García-García, "Machine learning and deep learning predictive models for type 2 diabetes: a systematic review," Diabetology and Metabolic Syndrome, vol. 13, 2021.

[24] M. Butwall, S. Kumar, "A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier," International Journal of Computer Applications, vol 120, no. 8, 2015.

[25] M. Saberi-Karimian, A. Mansoori, M. M. Bajgiran, Z. S. Hosseini, A. Kiyoumarsioskouei, E. S. Rad, M. M. Zo, N. Y. Khorasani, M. Poudineh, S. Ghazizadeh, G. Ferns, H. Esmaily, M. Ghayour-Mobarhan, "Data mining approaches for type 2 diabetes mellitus prediction using anthropometric measurements," Journal of Clinical Laboratory Analysis, vol. 37, no. 1, 2023.

[26] O. Olabanjo, A. Wusu, M. Mazzara, "Deep Unsupervised Machine Learning for Early Diabetes Risk Prediction using Ensemble Feature Selection and Deep Belief Neural Networks," Preprints 2023. doi: 10.20944/preprints202301.0208.v1.

[27] P. Palimkar, R. N. Shaw, A. Ghosh, "Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach," in Proceedings of the International Symposium on Engineering Accreditation, Novomber 4-5, 2021 M. Bianchini, V. Piuri, S. Das, R. N. Shaw, Eds. Advanced Computing and Intelligent Technologies

[28] P. Prabhu, and S. Selvabharathi, "Deep Belief Neural Network Model for Prediction of Diabetes Mellitus," in: 2019 3rd International Conference on Imaging, Signal Processing and Communication, Singapore, July, 27-29, 2019, pp. 138–142, 2019.

[29] R. Garcia-Carretero, L. Vigil-Medina, and I. Mora-Jimenez, "Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population", Medical & Biological Engineering & Computing, vol. 58, pp. 991–1002, 2020.

[30] R. Jader, and S. Aminifar, "Fast and Accurate Artificial Neural Network Model for Diabetes Recognition," NeuroQuantology, vol 20(10), pp. 2187–2196, 2022.

[31] R. Patil, S. Tamane, S. A.Rawandale, and K. Patil, "A modified mayfly-SVM approach for early detection of type 2 diabetes mellitus," International Journal of Electrical and Computer Engineering, vol. 12(1), pp. 524–533, 2022.

[32] S. Tarumi, W. Takeuchi, R. Qi, X. Ning, L. Ruppert, H. Ban, D. H Robertson, T. Schleyer, and K. Kawamoto, "Predicting pharmacotherapeutic outcomes for type 2 diabetes: An evaluation of three approaches to leveraging electronic health record data from multiple sources," Journal of Biomedical Informatics, vol. 129, 2022. https://doi.org/10.1016/j.jbi.2022.104001.

[33] S. Thaiyalnayaki, G. Kalaiarasi, M. Nithya, and R. Padmavathy, "Classification system on diabetes prediction using deep learning approach," in: Proceedings of the 2023 AIP Conference, 2023. https://doi.org/10.1063/5.0110266.

[34] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep Learning for Diabetes: A Systematic Review," IEEE Journal of Biomedical and Health Informatics, vol. 25(5), pp. 2744–2757, 2021.

[35] T. Zhu, X. Yao, K. Li, P. Herrero, and Georgiou. P, "Blood glucose prediction for type 1 diabetes using generative adversarial networks," in: Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data co-located with 24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, August 29-30, 2020, K. Bach, R. Bunescu, C. Marling, N. Wiratunga, Eds.NY: ACM , pp. 90-94, 2020.

[36] W. Chen, S. Chen, H. Zhang, and T. Wu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree," in: 8th IEEE International Conference on Software Engineering and Service Science, Beijing, China, November 24-26, 2017, pp. 386–390, 2017.

[37] W. Yu, T. Liu, and R. Valdez, "Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes," BMC Medical Informatics and Decision Making vol. 10(16), 2010. https://doi.org/10.1186/1472-6947-10-16.

[38] X. Wang, M. Zhai, Z. Renet, H. Ren, M. Li, D. Quan, L. Chen, and L. Qiu, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," BMC Medical Informatics and Decision Making, vol. 21, 2021.

[39] Y. Dong, R. Wen, Z. Li, K. Zhang and L. Zhang, "Clu-RNN: A New RNN Based Approach to Diabetic Blood Glucose Prediction," in: 2019 IEEE 7th International Conference on Bioinformatics and Computational Biology, Hangzhou, China, March 21-23, 2019, pp. 50–55, 2019.

[40] Puneeth N. Thotad, Geeta R. Bharamagoudar, Basavaraj S. Anami,Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods,Diabetes & Metabolic Syndrome: Clinical Research & Reviews,17(1),2023,102690, ISSN 1871-4021, https://doi.org/10.1016/j.dsx.2022.102690.

[41] R. Dewan, R. Polishetty, K. K. Ravulakollu, N. Jagadam, M. K. Goyal and B. Sharan, "Diabetes Prediction using Data Mining Techniques: A state-of-the-art Survey," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1140-1144.