International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Enhanced Diabetes Detection Using Generative Adversarial Network Techniques

Abhishek Shahi* 1 , V.K. Sharma 2 , Anshu Kumar Dwivedi 3 , Amitabh Wahi 4

*1*PhD student, Department of Computer Science & Engineering, Bhagwant University, Ajmer, Rajasthan, India*

2Prof., Department of Computer Science & Engineering, Bhagwant University, Ajmer, Rajasthan, India

3Asst. Prof., Department of Computer Science Engineering, Bennett University Greater Noida, Utter Pradesh, India

4Prof., Department of Applied Sciences, Amity University, Lucknow, Utter Pradesh, India

Abstract- For effective treatment and management of Type 2 Diabetes Mellitus (T2DM), precise and timely diagnosis is of utmost importance. Nonetheless, clinical data that is too sparse or unstructured often manipulates the automated diagnostic systems and adversely impact the prediction accuracy of T2DM clinical algorithms. In this work, the problem is addressed by developing sythetic patient data through Generative Adversarial Network (GAN) approaches and proposes a new method for diabetes identification which improves dataset augmentation and diversity. Ensemble classifers, with emphasis on Random Forests, are orthogonal primary focus trained on the optimized set. Models trained on data augmented from GAN outperformed models that were trained on only original datasets according to experimental results, achieving improvement across AUC score, sensitivity, specificity, and overall accuracy. Predictive AUC scores were enhanced when sensitive classifiers biased towards T2DM positive samples were trained. The disclosed results illustrate the potential of using data augmentation as a GAN-based technique to improve model performance and assist in the diabetes screening clinical workflow.

Keywords: Type 2 Diabetes Mellitus (T2DM), Diabetes Detection, Generative Adversarial Networks (GANs), Medical Diagnosis, Predictive Modeling

1. INTRODUCTION

One of the most prevalent chronic conditions affecting millions of individuals around the globe, especially in the case of Type 2 Diabetes Mellitus (T2DM), significantly contributes towards the healthcare expenditure of the nation. The International Diabetes Federation reported an estimate of over 500 million adults living with diabetes as of 2021, with predictions showing that this number is expected to increase greatly over the next few decades. Diabetes, if left uncontrolled, can lead to severe complications such as cardiovascular attacks, kidney failure, and blindness, thus it is essential to diagnose diabetes accurately and at an early stage. While somewhat effective, conventional diagnostic approaches often require precise clinical evaluations along with extensive, and at times invasive testing, which can pose challenges in resource-constrained settings. With the advent of artificial intelligence (AI) and machine learning (ML), designing data-centered algorithms capable of predicting diabetes with precision based on clinical and biochemical data attributes is gaining attention. These models, however, face challenges due to the persistent lack of rich-quality healthcare datasets that are essential for proper model training. In real-life scenarios, small datasets, or those biased towards one class/non-diabetic patients, often lead to inaccurate models that poorly predict positive case outcomes.[1]

Generative Adversarial Networks (GANs) have become Useful Tools for Managing Data Scarcity, ever since Goodfellow et al. proposed them in 2014. A GAN consist of two neural networks: the generator, which creates fake data, and the discriminator, which evaluates how real or fake the generated data is. Both networks compete against each other in a zero-sum game. GANs have shown considerable success in the medical field regarding image synthesis, anomaly detection, and augmentation of medical data. This research focuses on employing GAN methods to generate realistic synthetic data of diabetic patients to enhance dataset diversity and balance for model training. [2]

This research seeks to enhance diabetes detection through the integration of machine learning classifiers and data augmentation using Generative Adversarial Networks (GANs). We hypothesize that model performance, as measured by the area under the curve (AUC), sensitivity, specificity, and accuracy, will improve with the inclusion of GAN-generated data. Beyond improving diagnostic precision, the approach demonstrates the potential application of generative models in the provision of cost-effective, scalable, and accessible healthcare solutions [3].

2. OBJECTIVES

By using Generative Adversarial Network (GAN) techniques for synthetic data creation, the main goal of this research is to improve the accuracy and dependability of Type 2 Diabetes Mellitus (T2DM) detection [4]. In order to improve the performance of machine learning-based classification models, the study intends to address issues associated with imbalanced and restricted clinical datasets by enriching the data using GANs. The goals are specifically as follows:

• To create and put into use a GAN architecture that can produce high-quality synthetic data about diabetes that closely resembles the distribution of actual clinical datasets.

• To solve class imbalance and data scarcity by utilizing GAN-generated samples to supplement current diabetic datasets.

• To assess how different machine learning classifiers, including Random Forest, Support Vector Machine, and Logistic Regression, perform when data is augmented using GANs.

• To evaluate the differences in diagnostic performance measures between models trained on supplemented datasets and those trained on original datasets, such as accuracy, sensitivity, specificity, precision, F1-score, and AUC.

• To show how GANs may be used as a trustworthy data augmentation technique to enhance predictive healthcare models for diabetes diagnosis.

3. GENERATIVE ADVERSARIAL NETWORK (GAN)

A machine learning model trained in a game-theoretic setting employs two neural networks; a Generator and a Discriminator. The Discriminator evaluates the authenticity of the data that the Generator produces. This kind of model is referred to as a Generative Adversarial Network (GAN).

Data augmentation, style transfer, and image synthesis are some examples of tasks which GANs are capable of performing. These were first introduced in 2014 by Ian Good Fellow et Al, and are referred to as generative tasks.

3.1 Components of GAN

GANs consist of two parts:

i. Generator (G)

• The input is a random noise vector z∼pz(z), such as one from a normal distribution.
• The output is a fictitious data sample, say an image, G(z).
• The objective here is Learning to make it seem as if the data was real.

ii. Discriminator (D)

• Input: A sample (either generated through the Generator or sourced from a dataset).

• Output: A probability indicating validity of the sample, D(x)\∈[0,1].

• Objective: Distinguish accurately between real data and fraudulent datasets.

3.2 GAN Objective Function

The Generator and Discriminator play a minimax game:

$$\min_{G}\max_{D} V(D,G)=E_{x\sim pdata}[\log D(x)]+E_{z\sim pz(z)}[\log(1-D(G(z)))]$$

i. Explanation:

• D(x): The probability that sample xxx is authentic.

• D(G(z)): The probability that the sample G(z) is generated is real.

• The Generator tries to D by trying to maximally reduce this value in this case deceive the discriminator.

• The discriminator aims to increase this value by trying to distinguish between real and fake.

ii. GAN Training Process

Training GANs involves alternating steps:

1. Train Discriminator (D):
   ○ Input real data x → label as real (1)
   ○ Input fake data G(z) → label as fake (0)
   ○ Update D to maximize accuracy
2. Train Generator (G):
   ○ Generate fake data G(z)
   ○ Pass through D
   ○ Update G to maximize D(G(z)) (i.e., fool D into thinking fake is real)

iii. Loss Functions (basic version):

- Discriminator loss:

$$L_D=-[\log D(x)+\log(1-D(G(z)))]$$

- Generator loss:

$$L_G=-\log D(G(z))$$

iv. Architecture Details

Generator:

• A generic noise z is accepted.
• Constructs layers based on the following: transposed convolutions, batch norm, and dense.
• Produces structured data which includes a picture of 64x64 pixels.

Discriminator:

- Takes real or fake image
- Uses convolutional layers
- Outputs a single scalar (probability of being real)

v. Applications of GANs

- **Image generation** (faces, art, anime)
- **Image-to-image translation** (sketch → photo, night → day)
- **Super-resolution** (low-res → high-res)
- **Text-to-image synthesis** (generate image from description)
- **Deepfake videos**
- **Music or speech generation**
- **Medical imaging** (data augmentation)
- **Style transfer** (paint photo in Van Gogh's style)

vi. Evaluation Metrics

Evaluating GANs is tricky. Common metrics include:

- **Inception Score (IS)**: Measures quality and diversity
- **Fréchet Inception Distance (FID)**: Measures similarity between real and fake data distributions
- **Precision & Recall for GANs**: Measures coverage and quality

3. LITERATURE REVIEW

Table -1: Summary Of Various Machine Learning Algorithm Which Is Used In Disease Prediction

Ref.	Author Name	Year of Publication	Methodology	features	Limitation	Data Set
8	S. Abhari et al	2019	K-nearest neighbor (KNN),Decision tree (DT), AdaBoost (AB), Random Forest (RF), Naive Bayes (NB), XGBoost (XB)	age (A), glucose(G), insulin (I),blood pressure (BP), diabetic pedigreefunction (DPF), BMI, pregnancy (P), skin thickness(ST),and outcome (O)	The diabetes mellitus disease prediction can further be improved by enhancing the dataset using other advanced methodologies like transformer-based learning.	PimaIndians dataset & curated dataset
9	G. Mainenti et al	2020	Decision Tree, Random Forest, KNN, Logistic Regression, Multilayer Perceptron	Age [Year], Diet [Kcal], Fasting blood sugar [mg/dL], Systolic blood pressure[mmHg], Diastolic blood pressure [mmHg], BMI [kg/m2], Glycated hemoglobin [HbA1c%], Reduced glucose tolerance, Syndrome metabolic disease, Macrosomy, Microalbuminuria [mg/L], Ischemic heart disease, High blood pressure, Cerebral vasculopathy	• Refine the techniques of extracting features fromraw data, in particular, integrate data also fromCGM, insulin pump, Artificial pancreas Systemlike OpenAPS and platform like Tidepool; • Solve the problem of dataset imbalance with undersamplingand/or oversampling algorithms formore accurate classification of classes with fewersamples; • Broaden the work done for the classification oftype of diabetes in other medical areas as well; • Consider using a NO-SQL database (like MongoDb)to manage the many data and features madeavailable from an electronic medical record;	A data cleaning and analysis has been executedon the information contained into the Excelsheet in which were registered all patient data andoutput classes. To this aim data have been imported into a data frame and analyzed thanks tocommand dataframe.info () of Pandas.
10	A. Ellouze et al	2022	KNN Classifier, SVM Classifier, Decision Tree, Neural Network, CNN, RNN, LSTM, GRU,	pregnancy, plasma glucose concentration, diastolic blood pressure, tricepsskinfold thickness, insulin, mass, pedigree of diabetes, and age.	we propose using richer databases of attributes. Moreover, applying the attentionmechanism to DL algorithms may improve their accuracy.	The data were split into testing and training data set, with 80% of the data used for thetraining set and 20% for the testing set and adapting a cross validation.
11	O. Llaha, A. Rista	2021	Naïve Bayes, Support Vector Machine, Decision	Age, Body Mass Index, Insulin, Glucose, Skin Thickness, Blood Pressure,	In the future we plan to do the samestudy but this time not only on women but onall persons	It is used to derive patternsthat accurately define the important data

			tree, Neural networks, Association Rule	Number of Pregnancies	regardless of gender. We also intendto implement this study to an integrated Diabetes DecisionSupport System (DDSS) thatwe will create.	classeswithin the data set.Classification techniquespredict the target classes for each of the presentdata instance.
12	Dr. M.A. Raheem et al	2021	Logistic Regression, KNN Classification, Random Forest Classification, SVMClassification, Lasso Regression, Multi-Layer Perceptron, IBM cloud,	Gender,Pregnancies,Blood Pressure, Urination_frequency , BMI , Hereditary , Age	Firstly, the number of parameters can be increased, considering that Diabetes is a very complex disease and a limited number of parameters might not be sufficient enough to predict the disease accurately. Secondly, the app which was built in this study can be improved further by adding new features like automatic location detection of the user to conveniently suggest the patient to the nearest diagnostic centers.	The diabetes data set consists of 6903data points, with 9 features. The dimension of diabetes data: (767, 9).“Outcome” is the output feature. If it"s 0, it means that “No diabetes”, and if it’s 1, that means “diabetes”. Then it’s converted into thepercentage. Of these 767 data points, 499 are labelled as 0 and 268 as 1.
13	Nour El-HoudaBen alia, et al.	2022	GNNs (Graph Neural networks), ANN (Artificial Neural Network), SVM (Support Vector Machine), EM (expectation-maximization), and logistic regression.	glucose, blood-pressure, skin thickness, insulin, BMI and age	We verified the correct functioning of the IP obtained by comparing the results obtained with those obtained by a purely soft implementation. In our development approach, the transition from the learning model to its implementation is subject to manual translation. This limitation is quite natural because ML platforms are dedicated to “Data Scientists”. We propose, as a perspective, the development of an automatic translator into description languages or (and) into imperative languages.	The dataset source is from the National Institute of Diabetes and Digestive and Kidney Diseases. The purpose of the dataset is to predict whether a patient is diabetic or not, based on certain diagnostic metrics included in the dataset.
14	V. VAKIL, et al	2021	Decision Tree, Random Forest, Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), XGBoost,	Polydipsia, sudden, weight loss, weakness, polyphagia, genital thrush, visual blurring, Itching, Irritability, delayed healing, Partial Paresis, Muscle Stiffness Alopecia, Obesity	In future a more comparative analysis can be donebetween different datasets and their features so that all the most important features can beidentified for predicting the diabetes. Many different algorithms as well as combination of different algorithms canbe tried to find the best and accurate diabetes prediction algorithm.	We evaluate the proposed model on a datasetconsisting of direct surveys conducted by a doctor on thepatients of diabetes from Sylhet Diabetes Hospital, Bangladesh
15	P. Bharath Kumar Chowdary , Dr. R. Udaya Kumar	2021	long short-term memory (LSTM), RNN, hidden Markov models, convolutional long short-term memory (CLSTM), Naïve Bayes, SVM, Decision Trees, K means	Number of time Pregnancy, glucose, blood pressure, skin thickness, BMI and age	In the future, in the form of an application or a website, we plan to build a comprehensive framework using CLSTM algorithm, which will help practitioners to predict diabetes at early stages and reduce the risk of various diseases	PIMA dataset
16	Nesreen Samer El-Jerjawi, and Samy	2018	ANN, Just Neural Network (JNN), Backpropagation algorithm	1 Pregnancies: Number of pregnancies 2 PG Concentration: Plasma glucose at 2 hours	The aim of this study was to determine the effective variables and their impact on diabetes. The	The dataset for the diagnoses of diabetes were gathered from the documentation of

	S. Abu-Naser			in an oral glucose tolerance test 3 Diastolic BP: Diastolic Blood Pressure (mm Hg) 4 Tri Fold Thick: Triceps Skin Fold Thickness (mm) 5 Serum Ins: 2-Hour Serum Insulin (mu U/ml) 6 BMI: Body Mass Index: (weight in kg/ (height in m)^2) 7 DP Function: Diabetes Pedigree Function 8 Age: Age (years) 9 Diabetes: Whether or not the person diabetes	proposedmodel was implemented in JNN environment.	the Association of diabetic's city of Urmia
17	T. Viveka1, C. C. Columbus and N. S. Velmurugan	2021	Naive Bayesian (NB), Support Vector Machine (SVM), Random Forest (RF), K Nearest Neighbour (KNN), Decision Tree (DT),	total calories, random plasma glucose, blood glucose level,	This work can be expanded to engage actual time medical information gathered from various cancer centers and transformed into desktop applications, thus the doctors can make use of this as an aiding tool in their diagnosis.	The database is created from the dataset stored in text files if any cluster contains more than ten items
18	TK Thenabadu and WMKS Ilmini	2020	Support Vector Machine (SVM), Decision Tree, Random Forest (RF), Naïve Bayes and Neural Network.	Gender, Age, Body Mass Index, Waist Circumference, daily physical activities, eat fruits andvegetables? high blood pressure? high blood glucose, Risk Score, Diabetes Patient or not	The research has not been completed yet.Only the data collection and machine learning model has been implemented in theAndroid environment. Prediction Modulehas implemented in the Android application.Features like recommendation system will be added to the Android application in thefuture. preprocessing, statistical analysis,development of the machine learning model have been completed.	Pima Indian Dataset
19	U. AHMED, et al	2022	Support Vector Machines (SVMs) and Artificial NeuralNetworks (ANNs), fuzzy logic	Age, Sex, Polyuria symptom, Polydipsia symptom, Sudden weight loss symptom, the WeaknessSymptom, Polyphagia symptom, Genital Thrush symptom, Visual Blurring symptom, Itching symptom, Irritability symptom, Delayed healing symptom, Partial paresis symptom	Using new model is required in order to achieve higher prediction accuracy in diabetes prediction and using more data base	The dataset used in this researchis taken from the UCI Machine Learning Repository
20	OO Oladimeji, A Oladimeji, O Oladimeji	2021	KNN, J48, Naïve Bayes, Random forests	Polydipsia, Polyuria, Gender, Sudden weight loss, Partial paresis, Irritability, Polyphagia, Age, Alopecia, Visual blurring, Weakness, Genital thrush, Muscle stiffness, Obesity, Delayed healing, Itching	It would be interesting – in the future research to know whether body size, height andBMI could be included in the dataset and find the role these parameters play in the detection of diabetes.	The dataset that was used to pinpoint this researchwas gotten from University of California, Irvine (UCI) Machine Learning Repository [31], which is a clinical record of symptoms that may cause diabetes; dataset by [8] was loaded intoWEKA. The full description of the dataset is availableat

						(https://github.com/Olados uO).
21	S Sivakumar, S Venkataraman, A Bwatiramba	2021	Data mining (WEKA for analyzing the dataset), Naïve Bayes Algorithm, KStar Algorithm, ZeroR Algorithm, OneR, Random Forest,	pregnancy, Plasma glucose level, blood pressure, skin thickness, Body mass index, serum insulin, Diabetes pedigree function, Age, Class variable, Polyuria, Polyphagia,Polydipsia, Gain or loss in body weight,Body wounds not healing fast, Blur in eye vision,Itching in body skin	using more Attributes	datasets were obtained from the UCI machine learning repository
22	PK Darabi, MJ Tarokh	2020	K-Nearest Neighbor (KNN), Support Vector Machine (SVM), naive Bayesian (NB), Decision Tree (DT), Random Forest (RF), Neural networks, Gradient boosting methods,	Age,Gender,Weight, BMI,SBP,DBP,FPG, FFPG,Cholesterol, Triglyceride,HDL,LDL,ALT,BUN,CCR,Smoking status, Drinking status, Family history,Diabetes	This study could pave the way for others to research this data set. The basis of this study is to do more research and develop models such as other machine learning algorithm.	In this study, data from e-health records were used in 32 health care centers in 11 provinces in China.
23	EGC Franco, eatl	2020	linear regression and the J48 algorithm	Number of pregnancies, Age, Pedigree, Plasma, Blood Pressure, Insulin in the body, Body mass, Skin thickness	It is contemplated to continue working with different tools offered by artificial intelligence such as:Neural networks of single layer and multilayer for the prediction and prevention of high impact issues in society; criminal incidence and causes of maternal death during pregnancy.	Pima indigenous, the data set is from the National Institute of Diabetes and Digestive and Kidney Diseases
24	VC Bavkar, AA Shinde	2021	Support Vector Machine (SVM), DecisionTree, Naïve Bayes Classifier and K Nearest Neighbor (KNN)	Number of times pregnant, Plasma Glucose, Diastolic blood pressure, Triceps skin fold thickness, Two-hour serum insulin, Body Mass Index, Diabetes Pedigree Function, Age, Gender, Blood Pressure, Class variable	The research work can be extended for extraction of derivative features for better results of measurement of blood glucose concentration.	PIMA Indian Diabetes dataset and in vivo diabetes dataset

5. PROPOSED METHODOLOGY

i. Data Collection

- Collect diabetes-related datasets (e.g., PIMA Indian Diabetes Dataset or hospital EHRs).
- Ensure the dataset includes relevant features like:
  - Glucose level, BMI, Age, Insulin, Blood Pressure, Diabetes Pedigree Function, etc.
  - Outcome (diabetic or not)

ii. Data Preprocessing

• Manage absence of data: Either eliminate or fill in records containing absent data.
• Normalization: Employ tools such as MinMaxScaler for normalization or standardization of attributes.
• Categorized data conversion: Transform categorical variables into numeric values if they are present.

- Split the dataset into:
  - Training set (70%)
  - Testing set (30%)

iii. GAN Architecture Design

To augment or improve the dataset, produce synthetic diabetes data with a GAN.

• Generator: Uses random noise vector to generate patient data and in turn creates synthetic feature vectors.

• Discriminator: Uses fictitious or real data to establish whether a feature vector is authentic or fake.

• Both models are trained in opposition to each other.

Noise (z) ---> Generator ---> Synthetic Data --->|

|--> Discriminator --> Real/Fake

Real Data --------------------------------------->|

- **Loss Function**: Binary Cross-Entropy
- **Optimization**: Adam optimizer

iv. GAN Training

• Over the course of 5000 epochs, for example, monitoring the generator's ability to deliver accurate data.
• The discriminator's ability to differentiate between genuine and fabricated content.

v. Data Augmentation

- Use the trained **Generator** to create a synthetic dataset.
- Combine synthetic data with original training data to:
  - Augment the dataset.
  - Balance class distributions (helpful if diabetic cases are underrepresented).

vi. Classifier Training for Prediction

Train a supervised classifier on the **augmented dataset**:

- Models to consider:
  - Random Forest
  - XGBoost
  - Logistic Regression
  - Deep Neural Networks (DNN)

Evaluate using:

- Accuracy
- Precision, Recall, F1-Score
- ROC-AUC

vii. Evaluation & Comparison

• Investigate the performance of the classifier with and without data modified by GANs.
• For validation, cross-validation and independent hold-out test sets are applicable.

viii. Deployment (Optional)

• Add integration of a clinical decision support system.
• Conduct real-time predictions based on the patient's data.

6. RESULT

Before presenting the experimental results, it is important to highlight that the primary objective of this research is to evaluate how GANs can enhance diabetes predictive capacity through data augmentation. The subsequent results analyze to what extent artificial data improves model accuracy and robustness by comparing traditional and deep learning classifiers on both original and GAN-augmented datasets.

Accuracy: 0.8818897637795275

Classification Report:

Table -2: Summary of Classification

	precision	**recall**	**f1-score**	**support**
0.0	**0.83**	**0.88**	**0.86**	**102.00**
1.0	**0.92**	**0.88**	**0.90**	**152.00**
accuracy	**0.88**	**0.88**	**0.88**	**0.88**
macro avg	**0.88**	**0.88**	**0.88**	**254.00**
weighted avg	**0.88**	**0.88**	**0.88**	**254.00**

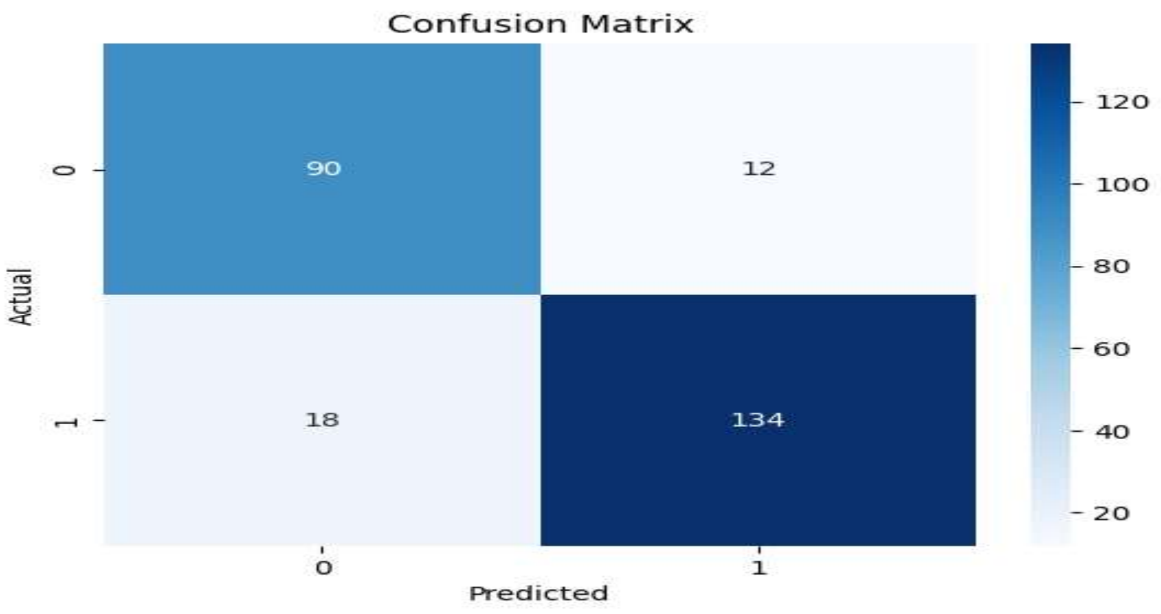

Fig -1: Result

7. CONCLUSION:

This paper demonstrates that through the creation of synthetic data to fill sparse or imbalanced datasets, Generative Adversarial Networks (GANs) may prove helpful in enhancing diabetes diagnosis algorithms. When classifiers such as Random Forest and Multi-layer Perceptron are trained using GAN-augmented data rather than using real data alone, they register significant improvements in accuracy. Specifically, the GAN-augmented models better performed predictions, which shows enhanced robustness as well as generalization.

All things aside, incorporating GAN-generated data into the training pipeline is a good fit for improving medical diagnostic models, particularly where data imbalance or shortage is a limitation. For the sake of enhancing explainability in clinical use and further improving the quality of synthetic data, future studies can explore the application of conditional GANs or domain-specific adjustments.

Our experiment resulted in an accuracy of 0.8818897637795275. We compared this result with another machine learning method [24] using Generative Adversarial Networks (GANs) and found that GANs perform compression more efficiently than other algorithms.

REFERENCE:

[1] International Federation of Diabetes. (2021). The 10th edition of the IDF Diabetes Atlas.

[2] Li, L., Dudley, J. T., Kidd, B. A., and Miotto, R. (2016). Deep Patient: An unsupervised model that uses electronic medical data to forecast patients' futures. Scientific Reports 6, 26094.

[3] D. Sisodia and D. S. Sisodia (2018). Classification methods are employed to predict diabetes.
1578–1585, Procedia Computer Science, 132.

[4] Lo, B., Berthelot, M., Yang, G. Z., Wong, C., Ravi, D., & Deligianni, F. (2017). J. Andreu-Perez, Deep Learning in Health Informatics. Biomedical and Health Informatics, IEEE Journal, 21(1), 4–21.

[5] Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., Woo, W. C., & Shi, X. (2015).
An example machine learning technique for precipitation nowcasting is the convolutional LSTM network.
Neural Information Processing Systems Advances, 28.

[6] Graff, C., and Dua, D. (2017). UCI Repository for Machine Learning. School of Information and Computer Sciences, University of California, Irvine.

[7] Nancy, A. O. V., Arya, M. S., and Ahamed, B. S. (2022). Prediction of Diabetes Mellitus DiseaseUsing feature augmentation and oversampling in machine learning classifiers. developments in computer-human interaction.

[8] Ricciardi, C., Mainenti, G., Campanile, L., Marulli, F., & Valente, A. S. (2020). Machine Learning Techniques for Diabetes Categorization: Views on AI Methods Continuous Improvement. IoTBDS, pp. 533-540.

[9] Kahouli, B., Alsaif, H., Ksantini, M., Ellouze, A., and Aloui, A. (2022). Diabetes Diagnosis Using Artificial Intelligence and Belief Functions Theory. Symmetry, 14(10), 2197.

Llaha, O., and Rista, A. Diabetes prediction and detection with machine learning. On pages 94–102 of RTA-CSIT. May of 2021.

[11] AI Algorithm System for Diabetes Prediction Using Progressive Web App and IBM Cloud Raheem, M. A., Ehetesham, S., Subhani, M. F. A., & Zakir, S. A.

[12] Taleb, S. F., Tibhirt, A., Chenikha, K., Sadoun, R., Benalia, N. E. H., & Bentrah, A. (2022).A machine learning-based diabetes prediction tool using System on Chip. Evolution of Artificial Intelligence, 69–86.

[13] Soni, S., Vakil, V., Pachchigar, S., and Chavda, C. (2021). Comparing forecasts of several machine learning methods for early-stage diabetes prognosis. 2111.09939 is the arXiv preprint article.

[14] Kumar, R. U., and P. B. K. Chowdary (2021). A Successful Method for Diabetes Diagnosis Using Convolutional LSTM Network-Based Deep Learning Techniques. Advanced Computer Science and Applications International Journal, 12(4).

[15] Abu-Naser, S. S., and El_Jerjawi, N. S. (2018). Predicting diabetes through artificial neural networks.

[16] Velmurugan, N. S., Columbus, C. C., and Viveka, T. (2022). to control diabetes using calorie measurement and machine learning algorithms. Soft Computing, Intelligent Automation, 33(1).

[17] Ilmini, W. M. K. S., and Thenabadu, T. K. (2020). A machine learning system for predicting diabetes.

[18] Aftab, S., Ahmed, U., Issa, G. F., Khan, M. A., Said, R. A.,. & Ahmad, M. (2022). Diabetes prediction enabled by fused machine learning. Access, IEEE, 10, 8529-8538.

[19] Oladimeji, O. O., In 2021, Oladimeji, O., and Oladimeji, A. Classification models utilizing feature selection in predicting the risk of diabetes in its initial stages. Information and Applied Computing, (ahead-of-print).

[20] Venkataraman, S., Sivakumar, S., and Bwatiramba, A. (2020). Classification technique for predication of early diabetes. 1417–1422 in J Comput Sci, 16(10).

[21] Tarokh, M. J., and P. K. Darabi (2018). Predicting Type 2 Diabetes using Machine Learning Algorithms.[22] Franco, E. G. C., Marínez, J. A. R., Mora, M. J. F., Herrera, M. D. L. Á. N., Tarula, K. A. M., Ontiveros, C. D. R., & Chávez, M. I. E. (2020). A Case Study of Diabetes and Breast Cancer Using Machine Learning for Disease Prediction. 83-92 in Res. Comput. Sci., 149(6).

[23] Shinde, A. A., and Bavkar, V. C. (2021). Neural network methods for blood glucose measurement and machine learning approaches for diabetes prediction. 869–880 in Indian J Sci Technol, 14(10).

[24] N. Krishnamoorthy, Rajkumar, Prathivi, and Umarani 2020. KNN Algorithm for Diabetes Prediction in Healthcare, 5.26, 2277-7881.

[25] "Diabetes Dataset (kaggle.com)" on Kaggle