2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) | 979-8-3503-7647-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/INNOCOMP63224.2024.00093

An In-Depth Exploration of Machine Learning Algorithms and Performance Evaluation Approaches for Personalized Diabetes Prediction

Inderdeep Kaur
Dept. of Computer Science and Engineering
Chandigarh University
Kharar, India
inderdeep.sgtb@gmail.com

Aleem Ali
Dept. of Computer Science and Engineering
Chandigarh University
Kharar, India
aleem.e12948@cumail.in

***Abstract*—Diabetes is now recognised as a major worldwide health problem, demanding effective early detection and prediction measures to limit its repercussions. Machine learning algorithms have shown to be excellent tools in healthcare for creating prediction models. This paper delves into an extensive investigation of machine learning (ML) algorithms and methodologies for assessing their performance in the context of personalized diabetes prediction. Different machine learning models are rigorously explored and evaluated for how well they can customise forecasts to the unique characteristics of individual patients. The study also critically assesses several performance evaluation techniques in an effort to improve the accuracy and reliability of diabetes prediction algorithms. Furthermore, interpretability problems and class imbalance are addressed as challenges and limitations to the application of machine learning in the prediction of diabetes. This study aims to motivate researchers, facilitating a deeper understanding of disease prediction algorithms and conduct related research.**

***Index Terms*—Diabetes, Risk Prediction, ML Classifier, Interpretable AI**

I. Introduction

Diabetes has emerged as a major global health problem, affecting millions worldwide, and posing significant challenges to individuals, health systems, and nations [1].A long-term metabolic illness, diabetes is marked by increased quantities of glucose in blood which is caused due to insulin resistance, insufficient insulin, or both.

Diabetes has progressively grown in frequency during the last few decades, and reached epidemic proportions. The International Diabetes Federation (IDF) claims that, Diabetes affected roughly 463 million adults (aged 20-79) in 2019, and This number is anticipated to rise to 700 million by 2045 if present trends continue, making it one of the top causes of mortality globally. [2].

The rate of development of diabetes can be attributed to many factors. Lifestyle changes, such as a sedentary lifestyle, an unhealthy diet, and obesity play a crucial part in the development of diabetes [3]. In addition, an aging population, urbanization, globalization, and dietary changes contribute to the burden of diabetes. Genetic predisposition also plays a role, although the interactions between genes and environment are complex and poorly understood.

Addressing the global diabetes epidemic requires a multi-pronged approach. Earlier detection and prevention are important to reduce the risk of diabetes [4]. Lifestyle interventions that promote physical activity, healthy eating, and weight management are important preventive measures. Timely diagnosis and effective management of diabetes is essential to maintain levels of glucose in blood and reduce the risk of complications [4]. Affordable and quality health care including diabetes education, medication and routine monitoring is essential for individuals with diabetes [5].

Machine learning has received considerable attention in diabetes research and has shown promise in diabetes prediction and management. Using large datasets and advanced algorithms, machine learning models can help with early detection, risk stratification, personalized treatment and outcome prediction for individuals with diabetes [2].

A. Types of Diabetes

It has four significant kinds: Type 1 diabetes, Type 2 diabetes, Gestational diabetes and Pregestational diabetes. Every type has different causes, risk factors and management strategies. A quick summary of each category is as follows:

- **Type-1 diabetes:** Type-1 diabetes, also called as juvenile diabetes or insulin-dependent diabetes, basically is an auto-immune disorder in which the body's immune system assaults insulin-producing cells in the pancreas, preventing the body from manufacturing insulin, a hormone that regulates blood sugar levels. Type-1 diabetes is more common in kids or adolescence, although it can occur at any age. [1]. Insulin treatment is required for life for those with type 1 diabetes.
- **Type-2 diabetes**: The most prevalent kind of diabetes,from a large proportion of instances, is type 2 diabetes. This happens if the body becomes insulin resistance or lacks the ability to generate sufficient insulin to control blood sugar levels. [4]. Type-2 diabetes is frequently associated with lifestyle factors such as obesity, lack of physical activity, unhealthy diet and genetic predisposition. Initially, lifestyle changes including dietary changes, exercise and weight loss are

979-8-3503-7647-0/24/$31.00 ©2024 IEEE
DOI 10.1109/INNOCOMP63224.2024.00093

often recommended. In certain instances, oral medication or insulin therapy may be necessary to control blood sugar levels [6].
- **Gestational diabetes**: Gestational diabetes occurs during pregnancy in women who did not have diabetes before pregnancy. Hormonal changes during pregnancy can cause insulin resistance, causing blood sugar levels to rise. [3] Gestational diabetes usually resolves after delivery, but managing blood sugar levels during pregnancy through a combination of diet, exercise, and sometimes medication is essential to prevent complications that raise the chance of type 2 diabetes later in the life of the mother and child [1].
- **Pregestational Diabetes**: Pregestational diabetes is the diabetes that develops before the beginning of pregnancy-related insulin-dependent diabetes. [3].

There are other rare forms of diabetes [4], such as monogenetic diabetes, caused by specific gene mutations, and secondary diabetes, caused by other conditions or medications, but T1 diabetes, T2 and Gestational diabetes are the most common and widely accepted forms.

II. Machine Learning Algorithms

Machine learning, an essential component of AI (Artificial Intelligence), is a research field of algorithms as well as statistical models that allow systems such as computers to automatically learn from data without being explicitly programmed.

Early diagnosis and prediction in diabetes management using machine learning can greatly improve patient care [7]. Machine learning is capable of analysing massive volumes of patient data, identifying trends, and predicting the risk of diabetes development or progression. [9]. This allows health professionals to intervene early, delivering targeted interventions and personalized treatment plans. Several widely used methods for predicting diabetes beforehand include:

- **Logistic regression**: LR is an algorithm used for binary classification, which can also be used to estimate the likelihood of diabetes based on input characteristics. [9] The approach is based on a logistic function, which converts a linear equation into a probability value ranging from 0 to 1.

$$log[\frac{Y}{1-Y}] = b0 + b1x + b2x + b3x + ... \quad (1)$$

  In LR, y can be between 0 and 1 only. log(Y/1-Y) is the link function. A logarithmic adjustment on the result variable allows a non-linear connection to be modelled linearly. This is the Logistic Regression equation.
- **Decision trees**: DT are multifunctional algorithms that make tree-like models of decisions and possible outcomes [14]. It can be used to predict diabetes by classifying data based on various and limited parameters and classifying individuals as diabetic or non-diabetic [14].
- **Random forest**: RF is a cluster approach to learning that blends several decision trees together. [14]. Each tree in the forest is built on a subset of features and samples, and predictions of all trees are combined into a final prediction Random forests are effective in using datasets with large features and reduce overreduction.
- **Support Vector Machines**: For applications involving regression and classification, SVM is a very reliable method. Its goal is to select the optimal hyperplane for grouping the data. [12]. SVMs work well with large-dimensional data and can handle nonlinear relationships through kernel functions.
- **Neural networks**: NN, particularly deep learning models, are very popular due to their ability to learn complex patterns from big data [11]. It consists of multiple interconnected networks (neurons) that process information [16]. It can automatically extract relevant features from the data and make predictions based on them.
- **Naive Bayes**: A probabilistic algorithm based on the Bayes theorem is called Naive Bayes. It assume that the features are conditionally not dependent [23]. Based on input characteristics, Naive Bayes estimates the probability of each category and predicts the most likely category.

$$P(S|T) = \frac{P(S|T)P(S)}{P(T)} \quad (2)$$

  Here S and T are two events and, S!T is the conditional probability of S given T. P(S) is the probability of T. P(S) is the probability of T. It functions well with data that is high-dimensional and is computationally efficient. However, in some circumstances, the presumption of independence of the parties may not hold true.
- **K-Nearest Neighbors (K-NN)**: K-NN is a simple and flexible algorithm that classifies new information based on its proximity to the recorded training data [25]. It calculates the distance (e.g., Euclidean distance) between input features and training instances between and assigns K nearest neighbors majority class labels K-NN is nonparametric and flexible, but can be sensitive to telemeter selection and K value.
- **Gradient boosting algorithms**: Gradient boosting algorithms, such as XGBoost and LightGBM, are ensemble methods that sequentially build weak learners to improve prediction accuracy. The loss functions are optimized by adding an iterative model it fixes the errors of the previous models [**?**]. The greatest qualities of gradient boosting algorithms are their excellent accuracy and capacity to manage intricate data interactions. The Advantages and disadvantages of classification algorithms are explained in Table.I

These algorithms can be trained with a variety of parameters such as BMI, age, glucose levels, BP, family history, and more [19]. The algorithm selection relies on variables like the volume and intricacy of the dataset, semantic requirements, statistical features, and desired performance specifications [21].

TABLE I
CLASSIFICATION ALGORITHMS-BENEFITS AND DRAWBACKS

Algorithm	Benefits	Drawbacks
LR [3]	Probabilistic interpretation and interpretability	Assumes linearity and independence
DT [14]	Easy to understand and interpret	Prone to overfitting
RF [22]	Robust to noise and outliers	Computationally expensive
SVM [12]	Effective in highdimensional spaces	Memory intensive for large datasets
NN [11]	Capture complex nonlinear patterns	Requires large amounts of training data
NB [23]	Fast training and prediction	Assumes independence between features
kNN [25]	Simple implementation and nonparametric	Sensitive to irrelevant features and noise
Boosting [15]	Reduces data overfitting and Easy to interpret	Not appropriate for huge datasets and highly susceptible to outliers

III. REVIEW OF LITERATURE

This section provides the review of literature and bibliometric analysis on the ML techniques for prediction of diabetes.There are various databases accessible that include vast amounts of data connected to published publications, including Web Of Science, IEEE, Google Scholar, Scopus, and many others. We largely used Web of Science (WoS) in this work since it includes a vast quantity of data.

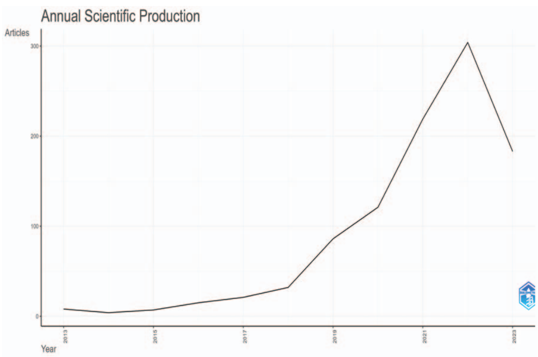

Fig. 1. Publication by Year (WoS)

An initial search using intended keywords yielded 1000 papers from 428 different sources for the WoS databases. Figure 4 depicts the annual scientific production of diabetes prediction from 2013 to 2023 in the WoS database. It can be observed that there is a progressive growth in the number of publications. Furthermore, it can be shown that research interest in the topic expanded quickly between 2020 and 2022. Consistent increase demonstrates research potential and progress in diabetes prediction through machine learning.

N. Ahmed et al. [6] develop a web application for users to monitor their health status. The authors address the gaps in existing studies by integrating trained models into a user's app and using multiple datasets for training. They also employ preprocessing techniques to increase the model's accuracy. The author discusses the performance of different algorithms, such as GB,RF, DT, NB,SVM, LR, and K-NN. The research show that SVM outperforms the other methods. The researchers also analyze the most influential attributes in detecting diabetes and compare the accuracy of their proposal with other works. The proposed model demonstrates better performance and accuracy.

Rastogi R. et. al [3] discussed the several facets of diabetes prediction, such as the different forms of the disease, the effects of diabetes, and the importance of early prediction. The data mining methods employed in the research are described by the author, like classification, clustering, regression, and association. It also describes the implementation strategy and the proposed framework for predicting diabetes outcomes. The authoremphasises the significance of accurate diagnosis and the use of data mining algorithms to improve prediction accuracy(82.4%). they also provides insights from a literature review on diabetes prediction and compares the performance of different algorithms.

Ali et. al [7] discussed a novel approach called the Random Forest with Best Parameters (RFWBP) algorithm, which combines the random forest algorithm with feature engineering to enhance the accuracy of diabetes diagnosis in early-stage patients.The authors shows the essentials of feature engineering in improving the functionality of machine learning models. They describe how to choose vital and relevant data, reduce data loss, and increase algorithm speed. Principal component analysis (PCA) and the random projection method (RPA) are suggested as feature extraction approaches. The authors describe how optimising parameters can lead to greater accuracy (95.83%), performance, reduced overfitting, increased efficiency, and improved interpretability.

Aslan .et.al [8] explains the process of conversion of the numerical data in the dataset into images to make it compatible with popular convolutional neural network (CNN) models such as ResNet, VGGNet, and GoogleNet.They discussed the importance of normalizing the data and the use of feature scaling, specifically the min-max normalization method, to rescale the feature values to a certain range.The accuracy achieved by the model is 77.37%.

In bibliometric analysis, the analysis of Corresponding Author's Countries provides significant insights into the worldwide landscape of research, particularly in specialised disciplines such as diabetes prediction. Understanding the geographic distribution of corresponding authors (Fig 6) in this setting provides a prism through which we may investigate international collaboration, research trends, and the contributions of various nations to increasing knowledge in diabetes prediction. This examination digs into the relevance and ramifications of this feature in diabetes prediction research bibliometric analysis.

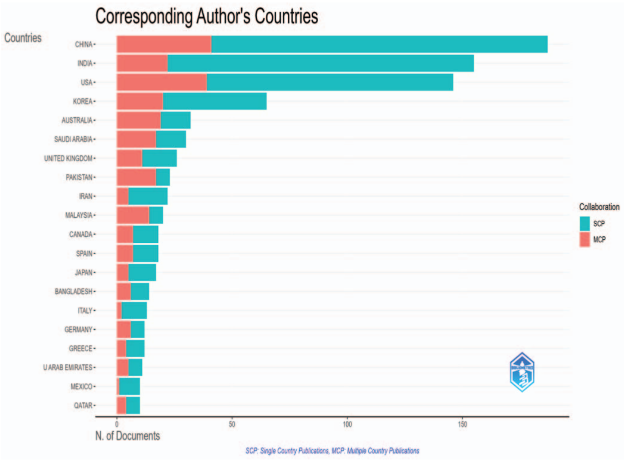

Fig. 2. Corresponding Authors Country(WoS)

Hama et .al [13] compares four different models (DTC, AdaBoost, GBC, ETC) and evaluates their performance using two datasets (PIMA and BRFSS). The main objective is to choose the best model for diabetes classification. The extra trees classifier is found to provide the highest accuracy, with a ROC of 96% for the PIMA dataset and 99% for the BRFSS dataset. The document also explains the materials and methods used, presents the results and discussion, and summarizes the conclusions. It provides details about the datasets, preprocessing techniques, evaluation metrics, and the performance of each classifier.

Alaa et.al [14] gives a summary of how machine learning algorithms are used. in diagnosing diabetes. Authors explains the concepts, techniques, and evaluation metrics involved in the prediction process.They also introduces evaluation metrics such as precision, recall, and F1-measure, which are used to assess the performance of the algorithms. The paper highlights the significance of early detection and the potential advantages of using ML in healthcare.

Gupta .et.al [17] highlights the uses of statistical metrics to analyse the success of various preprocessing procedures used to clean the dataset.Diabetes prediction models are being created using quantum machine learning (QML) and deep learning (DL).Extensive testing is used to determine the optimal number of layers necessary for diabetes prediction in both QML and DL models.To test their prediction accuracy, the constructed models are compared to cutting-edge approaches.

Sharma .et.al [23] stated that how AI and ML are used to identify diabetes and treatment.The author discussed numerous data analysis and modelling methods and techniques such as OPTICS, DBSCAN, BIRCH, NB, KNN, SVM, and others. It also discusses the obstacles and potential prospects in this subject. The paper gives an overview of several research findings and approaches for detecting diabetes using machine learning.The paper emphasises the relevance of data-driven techniques in healthcare, as well as the promise of artificial intelligence in diabetes control.

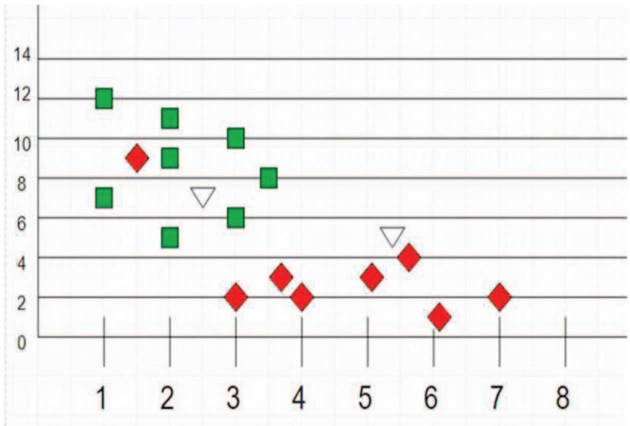

Fig. 3. kNN Example [19]

Hasan .et.al [24] presented detailed experiments and findings on preprocessing approaches, ML models, and classifier assembly. The paper highlights the dataset's intricacy in discriminating between positive and negative diabetes, the existence of outliers, and the increase in correlation following outlier rejection and filling missing values.The findings suggest that the proposed framework outperforms previous techniques and gives improved diabetes prediction performance. The author discussed future efforts and access to the diabetes prediction source code.It gives useful insights and approaches for predicting diabetes accurately.

Choi .et.al [25] evaluated the effectiveness of the suggested model.The research population is consisting of 8,454 participants without a history of diabetes who were studied over a 5-year period. The data utilised in the study were gathered from Korea University Guro Hospital's EMR database.They chose the most relevant characteristics for the prediction model using the information gain attribute assessment approach.The findings revealed that machine learning methods, namely the k-nearest neighbour algorithm, outperformed traditional prediction models in predicting the presence of T2DM.The limitations of this paper are the inclusion of participants with cardiovascular risk factors and the use of ICD codes for disease diagnosis .

Contributing countries in diabetes prediction research in WOS Database depicts that China definitely leads with 1048 publications, followed by the United States of America with 828. The Figure 5 shows the top five nations with diabetes prediction publications in the WoS database. China, without a question, has the most. In the year 2023, India had 557 publications, followed by Korea with 382 and Australia and the United Kingdom with 177 each.

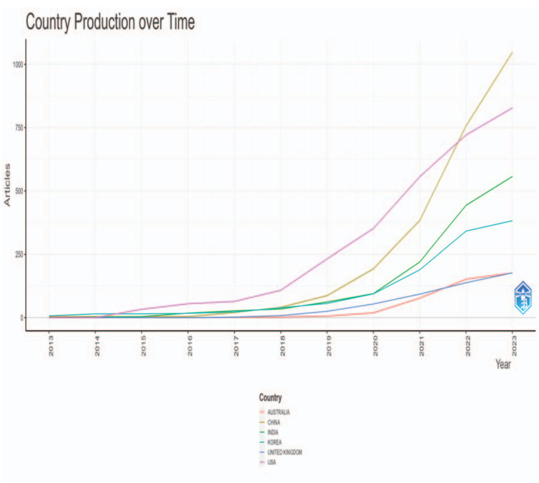

Fig. 4. Top 5 countries Production (WoS)

A. *Comparison Analysis*

The Table III. presents the performance outcomes of different algorithms It illustrates the objective of diabetes prediction through the evaluation of diverse techniques applied by various authors. The table displays the performance results of several algorithms on factors such as dataset used, feature extraction methods, classifiers employed, and the respective results achieved in different years. The variations in performances highlight the significance of selecting appropriate techniques to predict diabetes accurately, emphasizing the impact of factors like dataset quality, feature engineering, and choice of classifiers on the overall predictive efficacy.

IV. Challenges and Limitations

After selecting the best-performing model based on evaluation findings, further optimization of its parameters is pursued through hyperparameter tuning approaches, such as random or grid searches. These methods systematically explore the parameter space to enhance the model's effectiveness. Subsequently, the chosen model undergoes evaluation using an independent testing dataset to assess its real-world performance. This critical stage validates the model's generalizability and provides insights into its ability to accurately forecast diabetes. By rigorously testing on unseen data, the model's reliability and predictive power are thoroughly examined, ensuring its applicability beyond the training dataset and enhancing confidence in its predictive capabilities. [25].Also using machine learning approaches to predict diabetes involves a number of challenges and limitations. Among the most significant issues facing this field are:

- Imbalanced Data: Diabetes datasets often suffer from class imbalance, where non-diabetic instances outnumber diabetes instances, potentially biasing model outcomes.
- Data Availability and Quality: Data quality varies across sources such as health records, surveys, and genetic data, posing challenges like incomplete information and entry errors.
- Dimensionality and Feature Selection: Selecting significant features for diabetes prediction amidst high-dimensional data is challenging and requires careful consideration to avoid overfitting.
- Generalisation and External Validation:Models trained on a single dataset may struggle to generalize, necessitating validation on external datasets to ensure real-world applicability.
- Perceivability and Explicitability: Interpretability issues arise with complex algorithms like deep neural networks, hindering understanding of diabetes prediction mechanisms.
- Ethical Considerations: Concerns about privacy, bias, and transparency in prediction models highlight the importance of responsible implementation in healthcare.
- Validation and Comparative Studies:Variability in datasets and methodologies complicates model performance comparison, necessitating rigorous validation and study designs.
- Longitudinal and Temporal Analysis:Incorporating temporal dynamics into models is crucial for understanding the progression of diabetes over time.
- External Factors and Individual Variability: Lifestyle and socioeconomic factors influence diabetes prediction, posing challenges in model development.
- Clinical Adoption and Integration: Overcoming integration hurdles like electronic health record compatibility is essential for successful adoption of machine learning models in clinical settings.

To address these challenges and constraints, interdisciplinary cooperation, rigorous data management and preparation, strong validation processes, interpretability methodologies, and ethical concerns are required. Overcoming such challenges can lead to more accurate and reliable diabetes prediction models, allowing for earlier identification and intervention for better patient outcomes.

V. Research Gap

Some potential research gap in the diabetes prediction using machine learning techniques are

- Integration of Patient-Generated Health Data : Most diabetes prediction studies now depend on clinical data and electronic health records (EHRs), it could not give a complete picture of a person's lifestyle and health. PGHD (Patient-Generated Health Data) integration has the potential to improve forecast accuracy, offer real-time monitoring, and enable more personalised therapies [16].
- Explainable AI:While machine learning models have showed promise in diabetes prediction, their lack of interpretability and explainability remains a hurdle [14]. There is a need for research focused on building explainable AI algorithms targeted particularly to diabetes prediction models.

TABLE II
A COMPARISON OF SEVERAL DIABETES PREDICTION ALGORITHMS

Ref	Author	Year	Dataset	ML Classifier	Features	Accuracy
[1]	B. Ahamed et al.	2022	PIMA ,DMS	LGBM,RF,GBM	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	92.5%,98.9%
[2]	S. S. Bhat et al.	2022	Clinical Data	RF, KNN, MLP, SVC, GB, DT, LR	403 Instances and 11 Attributes	98%
[3]	Rashi Rastogi et al.	2022	Diabetes	LR	Pregnancy, Sugar level, BMI, Skin thickness	82.46%
[5]	A. Abdalrada et al.	2022	DISCRI	LR, EVIMP, CART	Common risk factors for DM and CVD co occurrence	94.09%
[7]	Md Shahin Ali et al.	2023	PIMA	FT RFWBP	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	95.83%, 90.68%
[8]	Muhammet Fatih Aslan et al.	2023	PIMA	SVM	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	77.37%
[9]	T. Abegaz et al.	2023	AOU	RF	Demographic, Biomarkers and Hematological Indices	80%
[10]	Antonio Agliata et al.	2023	NHANES, MIMIC-III, MIMIC-IV	Binary Classifiers	Blood Glucose, HDL ,BP, Gender, and Weight	86%
[11]	Chun-Yang Chou et al.	2023	Clinical Data	LR,NN,DJ,DT	Pregnancies, PGL, BP, ST, Insulin,BMI, DPF, and Age	95.3%
[12]	B. Kurt et al.	2023	Clinincal Dataset	SVM, RF, RNN LSTM	Pregnancies, Age ,etc.	98%
[13]	Mariwan Ahmed et al.	2023	PIMA, BRFSS	GB, Adaboost ,DT, ET	Pregnancies, Gender, Age, BMI	89%, 96%
[14]	Fayroza Alaa Khaleel et al.	2021	PIDD	SVM, NB, DT	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	94%
[15]	Isfafuzzaman Tasin et al.	2022	PIMA, Clinical Data	XGboost	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	90%
[16]	Victor I. Chang et al.	2022	PIMA	NB, RF, and J48 DT	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	79.57% and 79.13%
[17]	Himanshu Gupta et al.	2021	PIMA	QML	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	95%
[19]	Umair Muneer Butt et al.	2021	PIMA, Clinical Data	MLP	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	86.08%, 87.26%
[20]	Quan Zou1, et al.	2018	Clinical Data, PIMA	DT, RF, NN	Age, Pulse rate, Breathe, LSP, RSP, LDP, RDP, Height, Weight, Fasting glucose, LDL, HDL	80.84%
[21]	Huaping Zhou et al.	2020	PIMA	Forward Propagation DNN	Pregnancies, Gender, Age, BMI, BP, Skin, Insulin, DPF, Age	94.02174%, 99.4112%
[24]	Md Kamrul Has et al.	2020	PIMA	LDA, QDA, NB, GPC, SVM, ANN, AB, LR, DT, RF	Pregnancies, Gender, Age, BMI, BP,Skin,Insulin, DPF,Age	95%
[25]	Byoung Geol Choi et al.	2019	Clinical Data	LR, LDA, QDA, KNN	Sex, Age, BMI, Dyslipidemia, Chronic renal failure, Myocardial infarction, and Cardiovascular medication	-
[26]	Faranak Kazerouni et al.	2020	Clinical Data	K-NN, SVM, ANN, LR	Blood Test	95%
[27]	Pei et al.	2019	Physical Examination Reports	J48 Decision Tree, Adaboostm1,Sequential Minimal,Optimization, Bayes Net, Naïve Bayes	Age, gender, body mass,index (BMI), hypertension,history of cardiovascular disease or stroke, family history of diabetes,physical activity, work stress, and salty food,preference	95% by J48

- Long-term Risk Assessment and Prognostication: Current diabetes prediction research frequently focuses on identifying those who are most vulnerable to acquiring diabetes in the near future. Future research might look towards the creation of machine learning models that predict the risk of long-term diabetes, measuring disease progression, and identifying individuals who may advance from prediabetes to diabetes over time [13].

VI. Conclusion and Future Scope

Millions of individuals worldwide are impacted by diabetes, a health issue. and strains healthcare systems across the world. Population growth, co-occurring health conditions, and fiscal constraints highlight the significance of improved preventative, early diagnosis, and management techniques. The use of ML and technology breakthroughs has the potential to revolutionise diabetes treatment and improve outcomes for patients.It is important to fill present research gaps, particularly in the areas of interpretability, integration of Patient-Generated Health Data (PGHD), and long-term risk assessment. Bridging these gaps can aid in the creation of highly accurate and trustworthy diabetes prediction models, allowing for targeted treatment interventions and, eventually, improving patient well-being. By interpreting complicated model outputs, utilising real-time patient-generated data, and assessing long-term risk factors, the future of diabetes treatment can be defined by personalised care, proactive care, and predictive care.

References

[1] Ahamed, B. S., Arya, M. S., and Nancy, A. O. (2022). Diabetes mellitus disease prediction using machine learning classifiers with Oversampling and feature augmentation. Advances in Human-Computer Interaction, 2022, 1-14. https://doi.org/10.1155/2022/9220560

[2] Bhat, S. S., Selvam, V., Ansari, G. A., Ansari, M. D., and Rahman, M. H. (2022). Prevalence and early prediction of diabetes using machine learning in north Kashmir: A case study of district Bandipora. Computational Intelligence and Neuroscience, 2022, 1-12. https://doi.org/10.1155/2022/2789760

[3] Rastogi, R., and Bansal, M. (2023). Diabetes prediction model using data mining techniques. Measurement: Sensors, 25, 100605. https://doi.org/10.1016/j.measen.2022.100605.

[4] Ismail, L., Materwala, H., Tayefi, M., Ngo, P., and Karduck, A. P. (2021). Correction to: Type 2 diabetes with artificial intelligence machine learning: Methods and evaluation. Archives of Computational Methods in Engineering, 28(7), 5039-5039. https://doi.org/10.1007/s11831-021-09654-y

[5] Abdalrada, A. S., Abawajy, J., Al-Quraishi, T., and Islam, S. M. (2022). Machine learning models for prediction of Co-occurrence of diabetes and cardiovascular diseases: A retrospective cohort study. Journal of Diabetes and Metabolic Disorders, 21(1), 251-261. https://doi.org/10.1007/s40200-021-00968-z.

[6] Ahmed, N., Ahammed, R., Islam, M. M., Uddin, M. A., Akhter, A., Talukder, M. A., and Paul, B. K. (2021). Machine learning based diabetes prediction and development of smart web application. International Journal of Cognitive Computing in Engineering, 2, 229-241. https://doi.org/10.1016/j.ijcce.2021.12.001.

[7] Ali, M. S., Islam, M. K., Das, A. A., Duranta, D. U., Haque, M. F., and Rahman, M. H. (2023). A novel approach for best parameters selection and feature engineering to analyze and detect diabetes: Machine learning insights. BioMed Research International, 2023, 1-15. https://doi.org/10.1155/2023/8583210

[8] Aslan, M. F., and Sabanci, K. (2023). A novel proposal for deep learning-based diabetes prediction: Converting clinical data to image data. Diagnostics, 13(4), 796. https://doi.org/10.3390/diagnostics13040796.

[9] Abegaz, T. M., Ahmed, M., Sherbeny, F., Diaby, V., Chi, H., and Ali, A. A. (2023). Application of machine learning algorithms to predict uncontrolled diabetes using the all of us research program data. Healthcare, 11(8), 1138. https://doi.org/10.3390/healthcare11081138

[10] Agliata, A., Giordano, D., Bardozzo, F., Bottiglieri, S., Facchiano, A., and Tagliaferri, R. (2023). Machine learning as a support for the diagnosis of type 2 diabetes. International Journal of Molecular Sciences, 24(7), 6775. https://doi.org/10.3390/ijms24076775

[11] Chou, C., Hsu, D., and Chou, C. (2023). Predicting the onset of diabetes with machine learning methods. Journal of Personalized Medicine, 13(3), 406. https://doi.org/10.3390/jpm13030406

[12] Kurt, B., Gürlek, B., Keskin, S., Özdemir, S., Karadeniz, Ö., Kırkbir, İ. B., Kurt, T., Ünsal, S., Kart, C., Baki, N., and Turhan, K. (2023). Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. Medical and Biological Engineering and Computing, 61(7), 1649-1660. https://doi.org/10.1007/s11517-023-02800-7

[13] Hama Saeed, M. A. (2023). Diabetes type 2 classification using machine learning algorithms with up-sampling technique. Journal of Electrical Systems and Information Technology, 10(1). https://doi.org/10.1186/s43067-023-00074-5.

[14] Alaa Khaleel, F., and Al-Bakry, A. M. (2023). Diagnosis of diabetes using machine learning algorithms. Materials Today: Proceedings, 80, 3200-3203. https://doi.org/10.1016/j.matpr.2021.07.196.

[15] Tasin, I., Nabil, T. U., Islam, S., and Khan, R. (2022). Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters, 10(1-2), 1-10. https://doi.org/10.1049/htl2.12039

[16] hang, V., Bailey, J., Xu, Q. A., and Sun, Z. (2022). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, 35(22), 16157-16173. https://doi.org/10.1007/s00521-022-07049-z

[17] Gupta, H., Varshney, H., Sharma, T. K., Pachauri, N., and Verma, O. P. (2021). Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. Complex and Intelligent Systems, 8(4), 3073-3087. https://doi.org/10.1007/s40747-021-00398-7

[18] Tuppad, A., and Patil, S. D. (2022). Machine learning for diabetes clinical decision support: A review. Advances in Computational Intelligence, 2(2). https://doi.org/10.1007/s43674-022-00034-y

[19] Tigga, N. P., and Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science, 167, 706-716

[20] Nomura, A., Noguchi, M., Kometani, M., Furukawa, K., and Yoneda, T. (2021). Artificial intelligence in current diabetes management and prediction. Current Diabetes Reports, 21(12). https://doi.org/10.1007/s11892-021-01423-2

[21] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., and Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. Frontiers in Genetics, 9. https://doi.org/10.3389/fgene.2018.00515

[22] Zhou, H., Myrzashova, R., and Zheng, R. (2020). Diabetes prediction model based on an enhanced deep neural network. EURASIP Journal on Wireless Communications and Networking, 2020(1). https://doi.org/10.1186/s13638-020-01765-7

[23] Sharma, T., and Shah, M. (2021). A comprehensive review of machine learning techniques on diabetes detection. Visual Computing for Industry, Biomedicine, and Art, 4(1). https://doi.org/10.1186/s42492-021-00097-7

[24] Hasan, M. K., Alam, M. A., Das, D., Hossain, E., and Hasan, M. (2020). Diabetes prediction using Ensembling of different machine learning classifiers. IEEE Access, 8, 76516-76531. https://doi.org/10.1109/access.2020.2989857

[25] Choi, B. G., Rha, S., Kim, S. W., Kang, J. H., Park, J. Y., and Noh, Y. (2019). Machine learning for the prediction of new-onset diabetes mellitus during 5-Year follow-up in non-diabetic patients with cardiovascular risks. Yonsei Medical Journal, 60(2), 191. https://doi.org/10.3349/ymj.2019.60.2.191

[26] Kazerouni, F., Bayani, A., Asadi, F., Saeidi, L., Parvizi, N., and Mansoori, Z. (2020). Type2 diabetes mellitus prediction using data mining algorithms based on the long-noncoding RNAs expression: A comparison of four data mining approaches. BMC Bioinformatics, 21(1). https://doi.org/10.1186/s12859-020-03719-8.

[27] Pei, et.al., Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. (2005). New England Journal of Medicine, 353(25), 2643-2653. https://doi.org/10.1056/nejmoa052187