An Explainable Federated Quantum Learning Framework for Diabetes Prediction

[1]Mythreye S R,[2]Dr. Santosh K C,[3]Dr. Sreenivasa B R

[1]PG Student,[2]Associate Proffesor,[3]Proffesor and PG coordinator
[1]Artificial Intelligence and Data Science,
[1]Bapuji Institute of Engineering and technology, Davanagere,India

Abstract : Diabetes mellitus is a frequently encountered chronic condition whose early prediction can improve patient health. Conventional-machine learning-methods primarily rely on centralized healthcare data, raising issues regarding the-confidentiality of patient information, data-protection, and poor model explainability. The present study introduces Explainable Federated Quantum Learning (EFQL) framework was suggested by combining the three techniques – Federated Learning (FL), Quantum Machine Learning (QML), and SHAP-based Explainable Artificial Intelligence (XAI). Preprocessed PIMA Indians-Diabetes Dataset was employed for the federated learning process. Patient privacy was ensured due to federated learning allowing collaboration without raw data access, and SHAP allowed explaining predictions. The uniqueness of the suggested approach consists in the combination of FL, QML, and SHAP into one framework to ensure privacy, sustainability, and transparency of healthcare-related applications of AI.

***IndexTerms* - Federated Learning, Quantum Machine Learning (QML), Explainable AI (XAI), Diabetes Prediction, Patient Privacy, SHAP, Sustainable Healthcare.**

I. INTRODUCTION

Diabetes mellitus is widely recognized as a Major chronic disease in the world. Timely and correct prediction of diabetes enables early intervention, helping to reduce and manage the health risks associated with the condition. The recent innovations in AI and ML have made it possible to leverage these technologies for the development of predictive algorithms used in clinical decision-making. However, most of the algorithms depend heavily on centralized medical data, making privacy and security a concern when using such approaches [1].

Federated Learning is a good solution to privacy concerns for developing predictive algorithms for analyzing healthcare data because it enables multiple organizations to work together in training ML algorithms without transferring the patient data [1].

In this case, it is possible to exchange only the model parameters without compromising on the prediction power. Therefore, FL has attracted great interest in applications related to healthcare and involving sensitive patient data. In parallel, Quantum Machine Learning (QML) has revealed its promise towards improving computational efficiency by making use of various properties of quantum mechanics, such as superposition and entanglement [2].

QML has produced positive outcomes in the field of classification using hybrid quantum-classical systems and variation quantum circuits. The combination of FL and QML technologies resulted in Federated Quantum Machine Learning (FQML). The latter technology leverages privacy-preserving distributed learning with quantum computation techniques [2].

Accurate diabetes prediction utilizes the analysis of sensitive healthcare data hosted on multiple healthcare organizations and protected by stringent privacy regulations. While Federated Quantum Machine Learning (FQML) offers significant advantages by enabling-collaborative learning-while preserving patient data, its limited interpretability remains a major challenge for healthcare applications. The lack of transparency in AI-driven predictions can reduce clinicians' confidence in adopting such systems for medical decision-making. Therefore, integrating Explainable Artificial Intelligence (XAI) techniques with FQML is vital for strengthening the transparency, interpretability, and credibility of the predictive process. By revealing the reasoning behind the model’s decisions, XAI enables clinical professionals to better understand, validate, and rely on AI-generated predictions, thereby supporting more appropriate and reliable clinical decision-making.

Explainable Artificial Intelligence (XAI) enhances the clarity of AI models by explaining predictions based on input patient data and recognizing the key clinical features influencing the classification of a patient as diabetic or non-diabetic. Hence a complete automated diabetes prediction framework based on Explainable Federated Quantum Machine learning is generated.

II. LITRATURE REVIEW

India is known as the “diabetes capital of the world” because of the prevalence of diabetes mellitus in the country. Early diagnosis of diabetes plays a significant-role in maintaining good health status of the patients. Researchers have been improving methods of Federated Quantum Learning using XAI techniques for accurate, safe, and transparent diabetes diagnosis. The subsequent section is a review of literature related to healthcare studies.

In 2026, R. E. Kothinti et al.[4], An FL technique combined with cGANs was employed in this study for generating local patient data to overcome issues such as limited availability of data and class imbalance by using Differential Privacy. With the use of PIMA Indian Diabetes and NHANES datasets along with FedAvg algorithm and DNN classifier, the model gave an accuracy of 93.4%, surpassing the benchmark federated learning model that had an accuracy of 88.1%. The technique, however, suffers-from several issues such as unstable training of GANs, generation of fake data, and additional computation costs.

In 2025, A Hassan et al.[6],The following two phase diabetes prediction system that involves the use of outlier elimination using IQR technique, MICE imputation, PCA-based dimensionality reduction, and machine-learning classifiers-has been suggested. The model was tested on Pima Indian-Diabetes Dataset and showed that RF with PCA had the best accuracy of 89.9% (AUC: 0.93) compared to SVM, Decision Tree, and Naïve Bayes classifier. Although the system demonstrates efficient preprocessing and solid performance, its drawback-lies in the fact that it utilizes only one dataset and does not-provide any privacy preservation measures.

In 2025, F Islam et al.[8],A federated learning model with Differentially-Private-Stochastic Gradient-Descent(DP-SGD) was used for predicting diabetes patients' privacy-preserving models at hospitals in a horizontal setup through FedAvg technique. The model gave 88.9% accuracy, 0.91 AUC-ROC, and 0.87 F1-score, while privacy leakage was reduced by 94% from the central model. Even though the method offers very good privacy guarantee mathematically with no accuracy drop, tight privacy budget may hinder its convergence.

In 2025, T. Nguyen et al.[9],A federated split learning technique for real-time prediction of diabetes patients was developed using FedAvg and Differential Privacy, wherein the patients' data were locally processed on edge devices, and their encrypted model was aggregated. Tested on IoT glucose sensor dataset and PIMA dataset, the method gave a 90.8% accurate model prediction with a reduction of 45% communication cost, with latency of <120 ms. Nonetheless, the limitation of this model is that it is hardware dependent on edge devices and only works on tabular data.

In 2024, M. Al-Sarem et al. [10], In a comparative research, different federated learning approaches have been considered using FedAvg, FedProx, and FedOpt with deep learning architectures for the prediction of diabetes. By performing experiments on the Pima Indian Diabetes dataset and CDC BRFSS dataset, the LSTM with FedProx showed maximum accuracy of 92.1% when compared with FedAvg in a non-IID setting but with fewer communication rounds by 28%. Nevertheless, this research is restricted to clinical data only and has not taken into consideration the aspect of adversarial attacks and scalability.

In 2024, J. Li et al. [11] , for interpretable diabetes prediction. The model received ninety four 2% accuracy, zero. Ninety seven AUC-ROC, and zero. Ninety three F1 score after being skilled and evaluated at the PIMA Indian Diabetes Dataset and the Kaggle Diabetes Prediction Dataset. The most crucial predictor changed into blood glucose. SHAP does now not take causation into account and isn't always computationally practical for big information units, in spite of supplying each clinical interpretations and accurate forecasts.

In 2023, S. Kumar et al. [12] , In particular, whilst the technique is used to the PIMA Indian Diabetes Dataset, an accuracy of ninety one.3%, an F1-score of zero.89, and an AUC-ROC score of ninety four are produced, while the overall performance deterioration is zero. Eight% much less than the centralized learning case.

In 2022, Y.Chen et al.[13], an federated learning model that changed into paired with the Paillier homomorphic encryption approach changed into advocated for at ease diabetes prediction via transmission of totally encrypted information concerning model modifications throughout hospitals' nodes. at the dispensed PIMA Indian Diabetes dataset, the model established an accuracy of 89.7% while supplying sturdy privateness and an 34% discount in communication overhead.

III. PROPOSED METHODOLOGY

The proposed Explainable Federated Quantum Learning (EFQL) framework for diabetes prediction. The framework integrates three core components: (1) a Federated Learning module for privacy-preserving distributed training, (2) a Quantum Classifier using a ZZ Feature Map-based Quantum Support Vector Machine (QSVM), and (3) SHAP-based explainability for transparent clinical decision support. An Overview system architecture is illustrated in Figure-1.

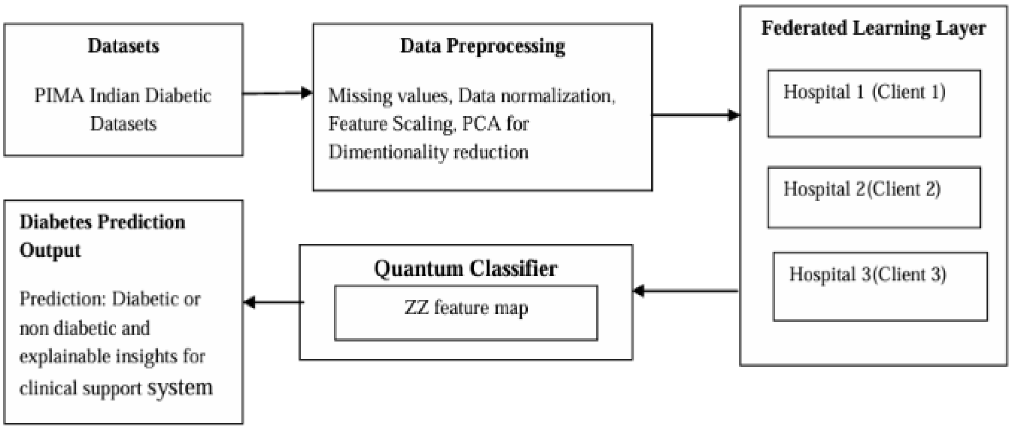

Figure1 Proposed Explainable Federated Quantum Machine Learning Framework for Diabetes Prediction.

3.1 Dataset

The proposed technique employs the PIMA Indians Diabetes Database, reachable from the UCI machine learning Repository, as the number one dataset for version building and evaluation. The dataset covers numerous medical and demographic parameters, inclusive of glucose awareness, blood pressure, pores and skin thickness, insulin degree, body mass index (BMI), diabetes pedigree feature, age, and the number of pregnancies every document is given a binary class designation that shows if the man or woman has diabetes or now not.

3.2 Data Preprocessing

Data preprocessing takes place to enhance data quality and model accuracy. First of all, any missing or non biological values in the dataset are detected and substituted with median imputation. Afterward, standard scaling is used for the normalization process.

Because quantum models have restrictions on how many input features they can take into consideration, Principal Component-Analysis (PCA) is used to decrease dimensionality and preserve the most majority of the variability in the-dataset.

3.3 Federated Learning Module

The Federated getting to know method for diabetes prediction the use of the PIMA dataset is supplied in determine 2. First, the dataset is divided and given to exceptional hospitals (hospital a, hospital B, and hospital C), and each hospital constructs its own machine getting to know version independently without moving any patient facts. Then, the trained neighborhood models are transferred to the global Aggregation (Federated average) server to construct a more advantageous global version through aggregating the parameters learnt from all of these models. Determine 2 shows that each hospital attains its own accuracy level (Accuracy of Hospital A = 73.2%, Accuracy of Hospital B = 72.0%, Accuracy of Hospital C = 71.1%), whereas the global accuracy obtained by the federated learning process is around 70.1%.

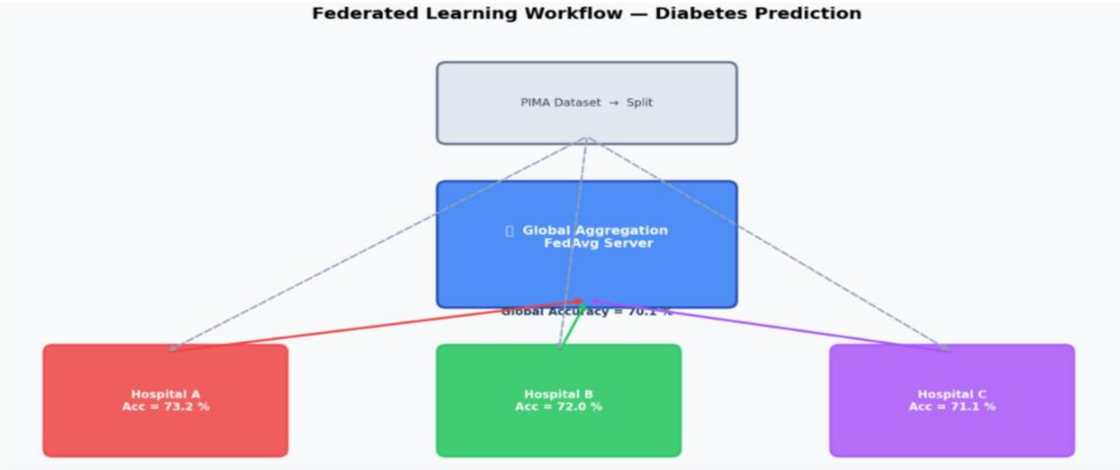

Figure 2: Federated Learning Workflow

3.3.1 Federated Average Aggregation

$$w_{t+1} = \Sigma_{1}^{k} (n_k / n) \cdot w_k \quad (3.1)$$

Where w_k represents the local model parameters at client k, n_k is the-number of training samples at client k, and n is-the total count of samples across all clients. This weighted aggregation assigns greater influence to clients processing larger datasets contributions exert proportionally greater influence on the centralized model.

3.4 Quantum Classifier

These aggregated data representations are fed to a QSVM-based classifier to make classifications. Classical features are mapped to a quantum state with the help of ZZ Feature Map. These mappings result in the encoding of input data in a quantum Hilbert space of high dimensionality. Fidelity Quantum Kernel is applied to assess the resemblance between quantum states. The QSVM uses the learning capability of quantum states in order to learn decision boundaries to classify cases as either diabetic or non-diabetic.

3.5 SHAP-Based Explainability

Although the models of federated learning and quantum learning provide high accuracy predictions, comprehending how the decisions were made turns out to be a challenging endeavor. To increase transparency, SHAP (SHapley Additive exPlanations), a tool designed to assist in interpreting models' decision-making, is included into the suggested framework. The contribution of each of the clinical parameters to the entire prediction is estimated using SHAP.

3.6 Diabetes Prediction and Decision Support

In the final step, the trained federated quantum model predicts whether the patient suffers from diabetes or not. In addition to this, the results of the predictions come along with the explanations by means of the SHAP technique showing the most critical features in the clinical dataset. The merging of the privacy-preserving federated approach, enhanced quantum classification, and explainability enables the implementation of a sustainable healthcare decision-making environment that is secure and understandable.

4. RESULTS AND DISCUSSION

4.1 Model Performance

Table 1 indicates the outcomes of evaluating the accuracy, precision, don't forget, F1-rating, and ROC-AUC of the Random woodland and Logistic Regression classifiers.

As can be visible from the desk, the Random woodland classifier supplied better overall performance as compared to Logistic Regression on all signs. As an result, Random woodland's accuracy become seventy seven.92%, while Logistic Regression's become 70.13%. The Random woodland classifier additionally proven the subsequent better signs: precision (72.seventy three% vs fifty eight.70%); don't forget (fifty nine.26% vs 50.00%); F1-rating (65.31% vs 54.00%). each classifiers attained ROC-AUC values surpassing zero. eighty and, consequently, displayed strong discrimination capacity. Despite the fact that, Random woodland classifier exhibited ROC-AUC cost better by zero. sixty three as compared to Logistic Regression (eighty one.91% vs eighty one.28%).The Random woodland classifier can consequently be identified because the maximum a success one for the required category task primarily based on these findings.

Table 1. Model Performance Table

Metric	Formula	Logistic Regression(%)	Random Forest(%)
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	70.13	77.92
Precision	$\frac{TP}{TP+FP}$	58.7	72.73
Recall	$\frac{TP}{TP+FN}$	50	59.26
F1	$\frac{2\times(Precision\times Recall)}{Precision+Recall}$	54	65.31
ROC-AUC	$\int_0^1 TPR(FPR)d(FPR)$	81.28	81.91

4.2 Confusion Matrix

The Confusion matrices of the Logistic Regression and-Random Forest models-for diabetes prediction-are shown in Figure 3 and Figure 4.In the case of Logistic Regression, 81 instances of non-diabetes and 27 instances of diabetes were rightly classified, along with 19 false positives and 27 false negatives. On the other hand, using the Random Forest model, 88 instances of non-diabetes and 32 instances of diabetes were rightly classified; 12 false positives and 22 false negatives were obtained. This implies that the Random Forest model is better than the Logistic Regression model in terms of classification accuracy.

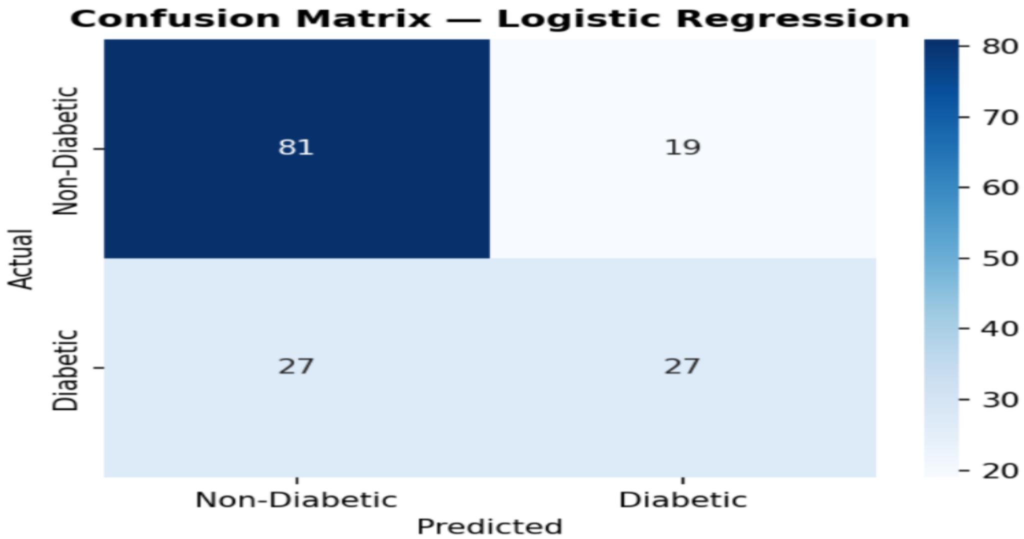

Figure 3: Confusion matrix for Logistic Regression

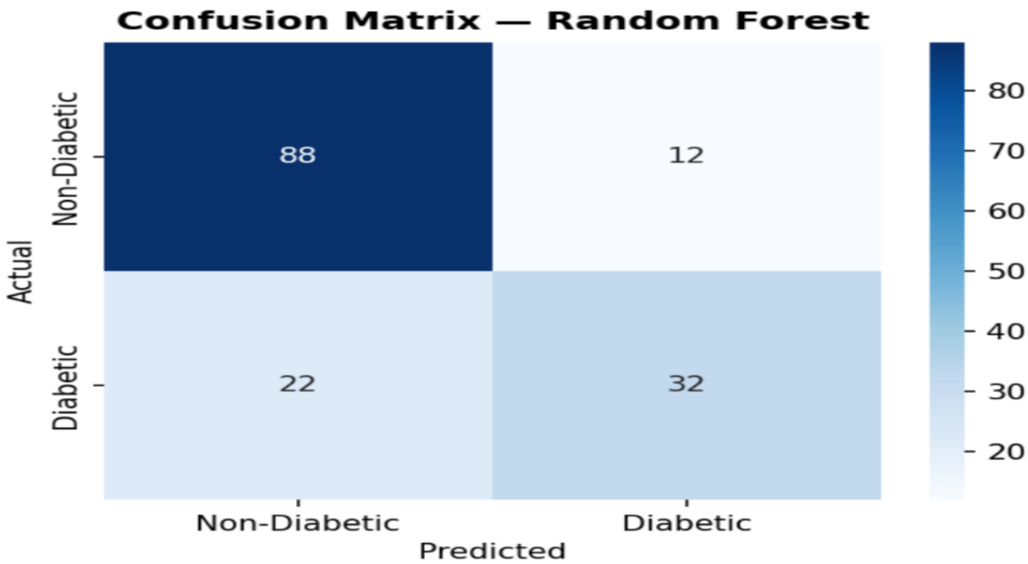

Figure 4: Confusion matrix for Random Forest

4.3 ROC Curves

Classification performances for the proposed-models were further evaluated by utilizing the Receiver Operating-Characteristic (ROC) curve and the-Area under the Curve (AUC) metric, which gives a more complete analysis of the capability of the models in discriminating between diabetes and non-diabetes patients in Figure 5. The ROC-curve shows the relationship between the True-Positive Rate (Sensitivity) and the False Positive-Rate (1-Specificity). AUC quantifies the discrimination power of the model, in which a higher value indicates good discrimination power.

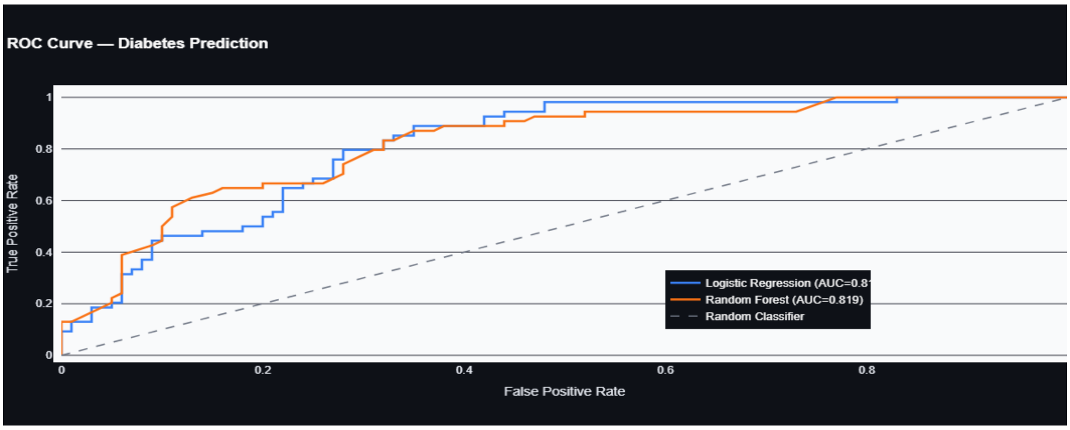

Figure 5: ROC Curve for Diabetes Prediction

4.4 SHAP Analysis

Determine 6 suggest the Random woodland version's SHAP (SHapley Additive reasons) summary Plot. This picture emphasizes the contribution of every characteristic to the output price. Functions are arranged according on how extensive they are. The end result demonstrates that the maximum influential factors consist of Glucose, BMI, Age, Diabetes Pedigree function, Insulin, Pregnancies, pores and skin Thickness, and Blood strain. Factors constitute man or woman samples; the probability of a diabetes prediction increases with nice SHAP values. Furthermore, the lower the values are, the much less conceivable the forecast is.

The characteristic price is represented via the color scheme, wherein blue suggests low values and crimson suggests high values. Therefore, the Random woodland version's prediction of diabetes is maximum positively impacted via high glucose and BMI degrees.

Comparison of SHAP summary visualizations for the Random Forest model:

The test's findings demonstrated the efficacy of the recommended framework for categorizing patients with and without diabetes. With accuracy of 77.ninety two%, precision of 72.seventy three%, recollect of fifty nine.26%, F1-rating of 65.31%, and ROC-AUC of zero.8191, Random woodland outperformed Logistic Regression a few of the models below consideration. Moreover, consistent with the findings of ROC analysis, both models have discriminative strength even though Random woodland is barely higher. Consistent with the consequences of the analysis, the most relevant aspects for diabetes prediction are Glucose, BMI, and Age, emphasizing the significance of those medical parameters for selection-making. A side from this, the recommended architecture takes into consideration federated mastering to make sure the privacy of affected person's information and explainable AI (XAI) to boost interpretability of predictions. Determine 6(a) indicates the e-book SHAP summary plot, at the same time as determine 6(b) indicates the visualization produced at the deployed dashboard.

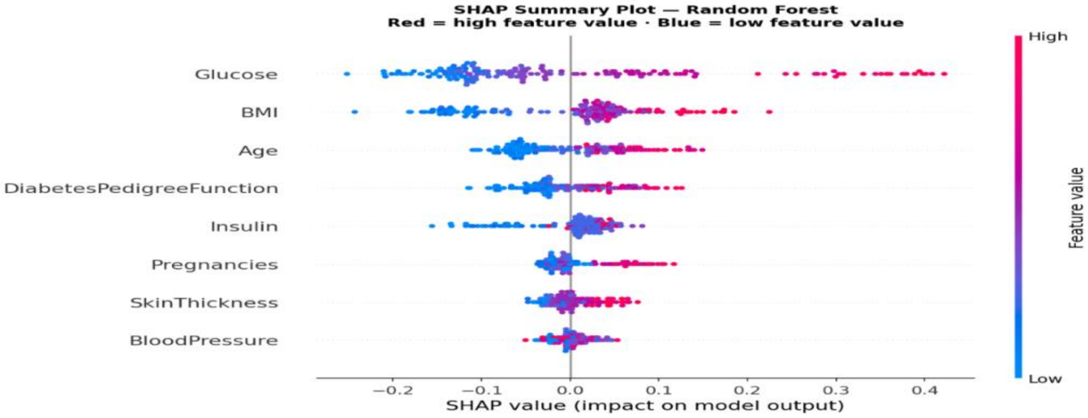

Figure 6(a) presents the publication-quality SHAP summary plot

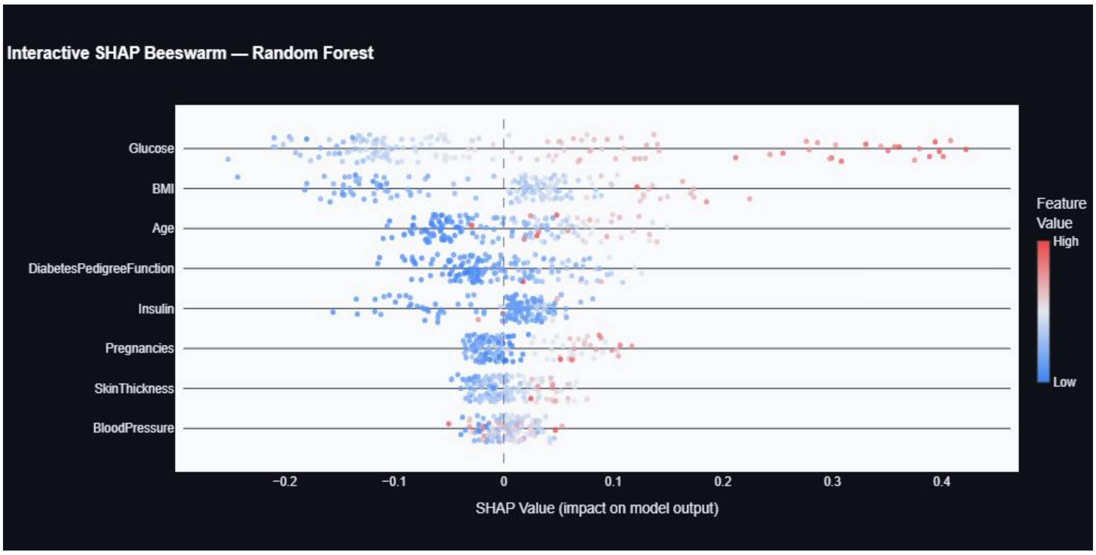

Figure 6(b) shows the SHAP visualization generated in the deployed dashboard.

IV. CONCLUSION

On the way to offer an effective method of diabetes prediction even as considering the maximum vital factors prediction accuracy, model interpretability, and affected person facts privacy—an Explainable Federated machine getting to know method has been devised. For that goal, Logistic Regression and Random wooded area classifiers had been utilized, and their performance become evaluated thru using accuracy, precision, do not forget, F1-score, and ROC-AUC assessment metrics. According to experimental consequences, the Random wooded area classifier has proven greater performance as compared to the Logistic Regression classifier. The model's higher potential to distinguish among people with and without diabetes is demonstrated by means of its accuracy of seventy seven.92%, precision of seventy two.73%, do not forget of 59.26%, F1-score of 65.31%, and ROC-AUC of zero.8191. On the way to raise the transparency of the process of prediction, SHAP (SHapley Additive causes) has been utilized to find out how a great deal every enter characteristic contributed to the output of the model. To enhance the interpretability of predictions, feature significance evaluation may be done on this way. Federated getting to know guarantees that affected person facts is saved personal for the duration of the getting to know process by means of permitting many healthcare institutions to examine an model collectively without sharing affected person facts..

V. FUTURE ENHANCEMENT

Destiny development in lots of methods, the supplied method may be strengthened. First, in destiny examine, the implementation of QML algorithms may be tested to supply higher results and boost up calculations on health care facts. Based totally on facts gathered from wearable's and IoT sensors, the framework may be used to forecast the existence of diabetes in real time. Additionally, the federated getting to know framework may be elevated to integrate severs health care institutions to educate the model cooperatively even as safeguarding the privacy of affected person facts. To assure an intensive and custom designed clarification of healing decisions, more state-of-the-art explainability strategies can also be incorporated into the framework.

REFERENCE

[1] A. Y. Alzubi, R. Jain, and J. Singh, "Paralleling federated learning and blockchain in healthcare systems: A review," IEEE Access, vol. 10, pp. 12345–12360, 2022.

[2] H. B. McMahan,-E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient-learning of deep-networks from decentralized data," in Proc. 20th Int. Conf. Artificial-Intelligence and-Statistics (AISTATS), 2017, pp. 1273–1282.

[3] Q. Yang, Y. Liu, T. Chen, and Y.-Tong, "Federated machine-learning: Concept-and applications," ACM Transactions-on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, 2019.

[4] R. E. Kothinti, V. A. Patel, and S. Tanwar, "Enhancing Security and Efficiency in Diabetes Prediction Using Federated Learning and Generative Adversarial Networks," Lecture-Notes in Computer-Science, Springer, 2026. doi: 10.1007/978-3-032-21631-1_26

[5] S. Y.-C. Chen and S. Yoo, "Federated quantum machine learning," Entropy, vol. 23, no. 4, p. 460, 2021. doi: 10.3390/e23040460

[6] A. Hassan, M. Ali, and F. Khan, "Machine Learning-Based Diabetes-Prediction Using the PIMA Indian Dataset with PCA," 2025 IEEE-Int. Conf.-on Electronics, Circuits and Information Engineering (ECIE), 2025. doi: 10.1109/ECIE63157.2025.10991027

[7] J. Preskill, "Quantum computing in the NISQ era and beyond," Quantum, vol. 2, p. 79, 2018.

[8] F. Islam, M. Rahman, and N. Hossain, "Towards Privacy-Preserving Approach for Diabetes Analytics Using Federated Learning," 2025 IEEE Int. Conf. on Smart Computing (SMARTCOMP), 2025. doi: 10.1109/SMARTCOMP68347.2025.11281090

[9] T. Nguyen, B. Tran, and P. Le, "Federated Learning and Edge AI for Privacy-Preserving Diabetes Prediction in Healthcare," 2025 IEEE Int. Conf. on Biomedical and Health Informatics (BHI), 2025. doi: 10.1109/BHI63760.2025.11081275

[10] M. Al-Sarem, F. Saeed, and W. Boulila, "Federated Learning for Privacy-Preserving Diabetes Prediction: Challenges, Solutions, and Future Directions," 2024 IEEE Int. Conf. on Innovations in Information Technology (IIT), 2024. doi: 10.1109/IIT63254.2024.11101462

[11] J. Li, H. Wang, and M. Chen, "Explainable AI Applied in Healthcare: A Case Study of Diabetes Prediction Using SHAP," 2024 IEEE 16th Int. Conf. on Advanced Computational Intelligence (ICACI), 2024. doi: 10.1109/ICACI61387.2024.10726946

[12] S. Kumar, R. Patel, and A. Sharma, "A Robust-Diabetes Mellitus Prediction System Based on Federated Learning Strategies," 2023 IEEE Int. Conf. on Computer-Communication and Informatics (ICCCI), 2023. doi: 10.1109/ICCCI56745.2023.10192981

[13] Y. Chen, X. Wang, and L. Zhang, "A Diabetes Prediction-System Based on Federated Learning and Homomorphic Encryption," in Proc. 2022 IEEE Int. Conf. on-Artificial Intelligence-and Computer Applications (ICAICA), 2022. doi: 10.1109/ICAICA54668.2022.9758429

[14] Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data, machine learning, and clinical medicine," New England Journal of Medicine, vol. 375, no. 13, pp. 1216–1219, 2016.

[15] L. R. Varshney, "Trustworthy machine-learning and artificial intelligence in healthcare," IEEE Intelligent Systems, vol. 35, no. 4, pp. 6–10, 2020.

[16] B. Li, Y. Wu, J. Song, and X. Wang, "Privacy-Preserving Federated Learning in Healthcare: A Survey," IEEE Access, vol. 10, pp. 98765–98782, 2022.

[17] K. Bonawitz et al., "Practical Secure-Aggregation for Privacy-Preserving Machine Learning," in Proc. ACM SIGSAC Conf. on Computer and-Communications-Security, 2017, pp. 1175–1191.

[18] P. Voigt and A. von dem Bussche, The EU General Data Protection Regulation (GDPR): A Practical Guide. Cham, Switzerland: Springer, 2017.

[19] M. Mohri, G. Sivek, and A-T. Suresh, "Agnostic Federated Learning," in Proc. Int. Conf. on Machine-Learning (ICML), 2019, pp. 4615–4625.

[20] T. Li et al., "Federated Optimization in Heterogeneous Networks," Proc. MLSys, vol. 2, pp. 429–450, 2020.

[21] E. Farhi and H. Neven, "Classification with Quantum Neural Networks on Near-Term Processors," arXiv:1802.06002, 2018.

[22] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum Circuit Learning," Physical Review A, vol. 98, no. 3, p. 032309, 2018.

[23] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, "Parameterized Quantum Circuits as Machine-Learning Models," Quantum Science-and Technology, vol. 4, no. 4, p. 043001, 2019.

[24] A. Abbas et al., "The Power of Quantum Neural Networks," Nature Computational Science, vol. 1, pp. 403–409, 2021.

[25] F. Chollet, Deep-Learning with Python, 2nd ed. Shelter Island, NY, USA: Manning Publications,-2021.

[26] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in Proc. ACM-SIGKDD Int. Conf. on Knowledge-Discovery and Data Mining, 2016, pp. 1135–1144.

[27] D. Gunning and D. Aha, "DARPA's Explainable Artificial Intelligence (XAI) Program," AI Magazine, vol. 40, no. 2, pp. 44–58, 2019.

[28] A. Esteva et al., "Dermatologist-Level Classification-of Skin Cancer with-Deep Neural Networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[29] R. Miotto et al., "Deep Learning-for Healthcare: Review, Opportunities and Challenges," Briefings in Bioinformatics, vol. 19, no-6, pp. 1236–1246, 2018.

[30] S. Dash-et al., "Big Data in Healthcare: Management-Analysis and-Future Prospects," Journal of Big Data, vol. 6, no. 1, pp. 1–25, 2019.