Proceedings of the 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS-2025)
IEEE Xplore Part Number: CFP25DN7-ART; ISBN: 979-8-3315-3884-2

2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) | 979-8-3315-3884-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICSSAS66150.2025.11081275

Federated Learning and Edge AI for Privacy-Preserving Diabetes Prediction in Healthcare

Ayyappan G
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University
Chennai, India
gtg.ayyappan@gmail.com

Alex David S
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Chennai, India
adstechlearning@gmail.com

V Loganathan
Saveetha Engineering College
Chennai, India
vadivelu.loganathan@gmail.com

E Padma
Vels Institute of Science, Technology and Advanced Studies (VISTAS)
Chennai, India
padma.e.79@gmail.com

S Ilavarasan
Saveetha Engineering College, Chennai
Chennai, India
ilavarasan.sargunan@gmail.com

Subash A
Er.Perumal Manimekalai College of Engineering
Koneripalli, Hosur
India
subasharun12@gmail.com

***Abstract*— The rapid expansion of digital health systems has resulted in large volumes of sensitive patient data, which if analyzed effectively, can significantly enhance disease prediction, diagnosis, and treatment approaches. Conventional machine learning models require centralized data collection, which poses serious concerns about patient privacy, data breaches, and compliance with regulations such as HIPAA and GDPR. This study presents a novel predictive healthcare framework that utilizes Federated Learning (FL) and Edge Artificial Intelligence (Edge AI) for decentralized and privacy-preserving analytics, specifically on diabetes prediction. To overcome these issues, the proposed solution is to train machine learning models locally in healthcare institutions and/or on IoT-enabled edge devices, using real-time patient data. The central federated server then receives only encrypted model updates with no raw patient data and performs Federated Averaging to create a global model. To ensure strict data protection, differential privacy techniques are employed to guard against re-identification and membership inference attacks. The system was evaluated using a partition of 327 of the PIMA Indian Diabetes Dataset, distributed across five virtual hospital nodes. Using the federated model accuracy was found to be 91.2%, precision 89.5%, recall 88.0%, and f1-score 88.7%. Inference latency was kept below 50 milliseconds and the bandwidth utilization decreased by 87% when compared to centralized techniques. The findings show that the proposed architecture maintains competitive performance while significantly improving security, compliance, and real-time usability. The study outlines a model for scalable, smart, and ethical healthcare systems. The combination of Federated Learning and Edge AI makes the proposed model deployable in many clinical settings especially in sensitive and resource-constrained environments, revolutionizing the field of digital health prediction and management.**

Keywords*— Federated Learning, Edge AI, Diabetes Prediction, Privacy-Preserving Analytics, Differential Privacy, Healthcare Machine Learning, Real-Time Inference, HIPAA Compliance.

I. Introduction

The healthcare industry is undergoing a massive transformation driven by advances in data analytics, artificial intelligence (AI), and pervasive computing [1]. Such a paradigm shift to data-driven healthcare systems poses a great opportunity to improve patient care, streamline operations, and lower costs. This progress is fundamentally dependent on the ability to harness vast amounts of clinical and biomedical data generated daily by, e.g., electronic health records (EHRs), medical images, genomic sequencing, and wearable and remote patient monitoring systems. For instance, successful analysis of extensive datasets can yield predictive insights that aid in early diagnosis of chronic diseases, enable personalized treatment plans, and help prevent disease progression.

With the promise of AI-enabled predictive analytics, significant challenges remain. The most basic requirement is to be able to assemble sensitive patient data to train models and make decisions. Traditional machine learning (ML) techniques require the centralization of data onto central servers or cloud platforms for the deployment of their complex algorithms. While technically sound, it presents serious ethical and regulatory challenges. Centralized data storage heightens the risk of breaches and compromises patient privacy, often violating compliance mandates such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe. The sheer volume and speed of medical data create infrastructure and latency limitations that centralized solutions cannot address in real time.

This work proposes a next-generation transition to Federated Learning with Edge AI—a new paradigm that combines distributed, decentralized AI with localized computation [2]. In contrast to conventional models requiring data to be centralized, federated learning enables multiple data sources - ranging from hospitals to clinics to wearable devices - to collaborate, training an AI model without the need to move raw data from its source. Instead of sending raw data to a centralized server, each participating node — a hospital server or a mobile health device — processes its local dataset to train its model and sends only encrypted model updates to a central aggregator. All these upgrades are embedded as one secure, integrated package to improve the global model while preserving data privacy and regulatory compliance.

Edge AI refers to deploying trained models directly on edge devices such as smartphones, wearables, hospital gateways, and diagnostic equipment, enabling real-time inference, low-latency decision-making, and local autonomy[3]. Federated learning, combined with edge AI, builds a powerful, privacy-focused, and scalable infrastructure for healthcare analytics that meets the high demands for precision and morality in modern medicine. Hospital staff can use smart glucometers to obtain real-time glucose measurements and their physical activity is tracked through wearable devices. The combination of edge devices operates AI predictions on the devices without cloud assistance to deliver prompt interventions across hospital and remote settings

This strategy, enabled by the federated edge principle, is especially driven by disease phenomena where early detection and continuous monitoring are necessary. Diabetes mellitus exemplifies the same. Given that diabetes is a global public health problem, the World Health Organization (WHO) estimated in 2023 that more than 537 million people are living with the disease, and this number is expected to rise to more than 640 million by 2030 [4]. Optimal diabetes management requires periodic measurement of blood glucose, lifestyle changes, and the ability to predict the likely complications, hospital readmissions, or treatment failure before the event occurs. Traditional methods using statistical regression and time-series analysis have been instrumental in modeling the progression of diabetes; however, they often rely on static, independent datasets. Moreover, these approaches do not usually take advantage of the large, real-time data streams generated by continuous glucose monitors, smart insulin pens, and mobile health applications.

In contrast, federated learning can be trained on different clinical institutions, personal devices, and regional healthcare databases, learning from diverse demographics and geography while keeping patients' privacy intact. Edge AI in a wide range of individual patient devices supports real-time risk assessment and alerts, guaranteeing timely treatment with no need for internet connectivity or server-side processing [5]. This combination not only speeds up reaction times but also democratizes predictive healthcare by providing advanced diagnostic tools to remote and underserved regions.

The confluence of federated learning and edge computing reflects broad trends in tailored medicine and population health management. In this respect, personalized medicine aims to tailor both therapies and prevention strategies to people’s unique genetic, environmental, and lifestyle characteristics. Achieving this level of granularity requires AI models that can learn from vast, decentralized, and high-dimensional data sources — one of the ideal applications for federated infrastructures. Population health management simultaneously focuses on the analysis of health outcomes among groups to identify inequities, to be able to inform resource allocation and create targeted interventions. Federated learning facilitates collaboration between different institutions as well as international studies in the field of health research, without compromising the sovereignty of the data and the regulatory policies related to the integrity of the data.

The proposed design framework can be used for other health conditions apart from diabetes, including, but not limited to, cardiovascular disease, cancer, mental health conditions, and infectious diseases. In challenging care scenarios, Edge AI can monitor patients in real-time, detect deviations such as arrhythmias, hypoxia, or fever spikes, and promptly alert caregivers. The application of federated learning in public health surveillance enables the construction of resilient predictive models throughout domains without the need for data exchange, which is particularly valuable for monitoring outbreaks and evaluating the effectiveness of vaccinations and containment approaches.

II. Related Work

The augmented relevance of predictive analytics in healthcare has evoked a surge of studies implementing machine learning (ML) techniques in the early detection, diagnosis, and treatment of chronic diseases, with an emphasis on diabetes. With the advent of machine learning and its increasing ease of use, myriad methods have been applied to leverage clinical data for predictive ends. The transition from traditional centralized systems to privacy-preserving, predictive platforms in real-time is still ongoing. This literature review is an assessment of existing academic achievements, namely articles that explored diabetes prediction, using traditional machine learning techniques and their drawbacks. Furthermore, it sets the ground for discussing the need for novel architectures such as federated learning and edge AI.

[6] conducted an extensive study evaluating the effectiveness of traditional machine learning models such as Random Forests, Support Vector Machines (SVM), and Logistic Regression in predicting diabetes. These models demonstrated high predictive accuracy, with rates approaching 97.5% when applied to datasets containing relevant clinical variables such as age, body mass index (BMI), blood glucose, and family health estimators. All models used in the study have individual merits. For instance, SVM could handle high-dimensional data and capture non-linear correlations well, making it suitable for complex variable dependencies that often exist in healthcare data. Random Forests Over passers ensemble method, captures complex features interactions while preventing Overpassing using its multiple weeping tree architecture. Logistic Regression is very simple but has a lot of interpretable power, it is widely used in clinical settings because of its transparency and simplicity.

Nonetheless, significant limitations remain, as traditional centralized machine learning approaches often fail to scale effectively in the modern healthcare landscape—marked by heterogeneous data sources, privacy concerns, and the demand for near-real-time processing [7]. In addition, while these frameworks achieve excellent performance on experimental datasets, their performance when addressing real-world data, with noise, missing values, and class imbalance, could be greatly diminished. The study highlights how more advanced techniques like deep learning along with distributed frameworks can enhance the accuracy of predictions and address these limitations.

[8] developed a comprehensive analytics toolkit for type 2 diabetes management. This suite is characterized by the integration of visual analytics, predictive modeling, and exploratory data analysis for clinicians in making clinical decisions. This endearing device allows physicians to visualize patterns in patient data, create tailored treatment strategies, and assess risk factors with less ambiguity. The main advantage of this approach is that it allows for multidimensional analysis. This approach delivers a comprehensive understanding of the patient's health, facilitating accurate, data-driven diagnoses and personalized

treatment strategies aligned with the patient's specific condition. Additionally, it provides doctors with data-driven tools that are often lacking in traditional therapy.

However, despite its achievements, it remains unclear how well the system would scale or apply to other healthcare environments. A lack of comprehensive evaluation exists regarding the system's effectiveness within real clinical pathways, particularly in resource-limited settings. Integration with privacy-preserving measures is absent, making it unsuitable for contexts requiring strict data control. These limitations drive the need for a federated architecture that can deliver reflectively similar analytical power while respecting privacy and system compatibility.

The study presented in [9] explores the applicability of Random Forest and LSTM models for breast cancer prediction, building on similar approaches successfully employed in diabetes prediction to demonstrate the potential of machine learning in early disease detection. This approach is inspired by previous studies on diabetes risk prediction, where clinical features such as fasting glucose, age, body mass index (BMI), and family history were effectively utilized. A notable advantage of such ensemble-based models lies in their simplicity and interpretability—particularly in the case of Decision Trees, which offer clear rule-based structures that are easily understood by clinicians. Moreover, the use of multiple trees in Random Forest reduces variance, leading to more stable and reliable predictions, making it a promising tool for clinical applications.

External validation or testing in diverse populations is lacking, which limits applicability across diverse clinical settings. There is also limited interaction with real-time monitoring systems or feedback mechanisms, which are becoming more critical in chronic disease treatment. Conversely, a federated learning approach permits training across large and heterogeneous datasets without requiring data to be centralized, alleviating these drawbacks while maintaining patient privacy.

In [10] work with predictive modeling based on the widely used Pima Indian Diabetes dataset. Their research applies a range of machine learning algorithms such as Support Vector Machines, Random Forests, and Logistic Regression for diabetes risk prediction that generate promising results. Using widespread datasets such as Pima allows for reproducibility and greater comparison between studies, which is a significant advantage. The study confirms the reliability of standard machine learning methods for predicting diabetes risk on physiological and demographic parameters.

The narrow nature of the dataset limits its generalizability. Data is limited to a single ethnic group, raising concerns regarding model bias and representational insufficiency. The authors explain that the generalizability of the model in real-world settings is limited due to the absence of external validation studies and more diverse population approaches. The federated approach could allow for the training of similar models on locally relevant datasets from multiple hospitals or regions, producing more robust and demographically representative models.

A detailed overview of how machine learning benefits the healthcare domain, including applications from disease prediction, patient monitoring, and therapy optimization to medical imaging[11]. The study explores a wide spectrum of machine learning algorithms from Support Vector Machines (SVMs) and Decision Trees to advanced neural networks—highlighting their transformative potential in clinical practice. It emphasizes the role of machine learning in automating diagnostic processes, improving healthcare service efficiency, and enabling personalized medicine through individualized risk assessments and tailored therapeutic recommendations.

Despite its broad scope, the study suffers from a lack of specificity, especially with specific ailments like diabetes [12]. Due to the absence of exhaustive investigation on data heterogeneity, privacy issues, and system scalability, the technical solutions mentioned in the work remain more theoretical than practical. It also does not take into account that there are many flavors of machine learning models, and some of them may be adapted to safe decentralized learning as privacy requirements continue to tighten. Future work will supplement this broad framework, adding federated and edge learning models and associated techniques capable of maintaining data integrity and scalability between institutions.

A system for monitoring diabetes in real-time using machine learning is developed by [13] namely GlucoNet Pro. By building recommendations for blood sugar control, the application leverages real-time data streams from glucose levels, foods consumed, activity levels, and prescription information. Gluco Net Pro distinguishes itself through its emphasis on real-time forecasting and personalized patient care. The technology is a great example of how dynamic data inputs will improve risk forecasting and treatment options. The use of so many data modalities is a sign of progress toward more holistic healthcare treatments.

Despite its strengths, GlucoNet Pro's complexity and reliance on continuous data collection present significant implementation challenges, particularly in resource-constrained environments. Such systems may face limitations in areas lacking reliable connectivity, data storage infrastructure, or adequate computing resources. Leveraging Edge AI can address these issues by enabling on-device analysis, and reducing dependence on cloud-based computing [14]. Furthermore, incorporating federated learning allows for localized model training across deployment sites while preserving data privacy, thereby enhancing scalability and adaptability in diverse settings.

In [15] outline a predictive model for early detection and prevention of Type 2 diabetes based on different machine learning algorithms. The technology is designed to identify high-risk patients and support early intervention by integrating clinical and lifestyle data analytics. It effectively highlights the critical role of early diagnosis and the impact of lifestyle factors—such as exercise, diet, and stress—on health outcomes. By combining multiple data dimensions, the approach enables more accurate and actionable predictions. This is particularly valuable for conditions like diabetes, where timely treatment can lead to significantly improved long-term outcomes.

III. Proposed Work

Well-established Distributed machine learning frameworks, such as Federated Learning (FL) and Edge Artificial Intelligence (Edge AI), are alternatives motivated by increased demand for safe, scalable, and intelligent healthcare analytics. Such frameworks aim to address the limitations of traditional centralized machine learning systems, especially in healthcare environments where issues such as privacy, real-time decision-making, and interoperability are paramount. A system is proposed that integrates federated learning with Edge AI to develop a robust, privacy-preserving predictive model tailored for diabetes prediction, as illustrated in Figure 1.

The proposed scheme is meant to operate across various healthcare services including in-hospital, rural clinics, and personal IoT health monitoring devices. This system claims decentralized training, low-latency predictions, and strong privacy guarantees while maintaining predictive efficacy. The following sections detail the system's architecture, data pipeline, and security protocols.

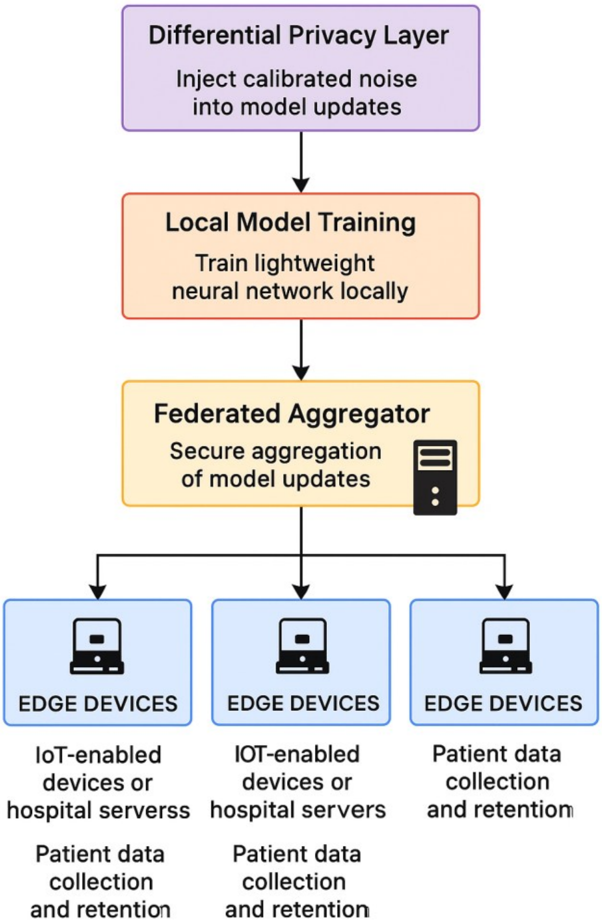

Fig. 1. System Architecture of Federated Learning and Edge AI Framework for Privacy-Preserving Healthcare Prediction

The architectural design involves a federated Edge-AI healthcare prediction system that enables intelligent, decentralized data processing across diverse healthcare environments. The architecture includes four key components, as described in the following sections: an edge device (edge clients), a local model in each edge device, a distributed aggregator, and a differential privacy layer.

Edge devices in this system include IoT-based medical diagnostic equipment, mobile phones, tablets, wearable devices (smartwatches and glucose meters), and hospital edge servers. Local data collecting, preprocessing, and even model training can be done by some of these devices. In many cases, such devices become the point of care themselves, interfacing with patients or healthcare workers directly to deliver alerts, predictions, and recommendations.

A smart glucometer can capture a patient's blood glucose levels at all times, while a wearable fitness tracker records physical activity and sleep data. Hospital information systems also store comprehensive patient profiles including laboratory data, demographics, clinical history, and lifestyle characteristics. Together at the edge, these sources form a comprehensive multimodal dataset for localized learning.

Edge devices have sufficient computational power (e.g., Raspberry Pi 4, NVIDIA Jetson Nano, or ARM-based CPUs) to run lightweight neural networks using optimized frameworks such as TensorFlow Lite, PyTorch Mobile, or ONNX Runtime. These models are regularly updated via federated learning and can incorporate global demographic trends while adjusting to local trends. The proposed framework incorporates AI in two functions: it acts as the primary learning system to predict diabetes results using neural networks along with functioning as a technology that enables smart processing at the device's edge.

The model is trained on each edge device independently and without communicating any information with a central server. The local training model is a small deep neural network. Let the total number of participating edge clients be denoted by K. Each client k∈{1,2,...,K} a private dataset $D_k = \{(x_i^{(k)}, y_i^{(k)})\}_{i=1}^{n_k}$ where $x_i^{(k)} \in \mathbb{R}^d$ is the patient feature vector, and $y_i^{(k)} \in \{0,1\}$ represents the binary diabetes outcome. The goal is to collaboratively train a shared predictive model $w \in R^d$ in without transferring raw patient data. Each client minimizes its local objective function:

$$L_k(W) = \frac{1}{n_k}\sum_{i=1}^{n_k} \ell(f_w(x_i^{(k)}), y_i^{(k)}) \quad \text{... Eqn(1)}$$

where ℓ is the loss function and f_w is a neural model.

This architecture is designed for accuracy while still being computationally inexpensive and memory-efficient for on-device training. The entire training process is done using mini-batch gradient descent or adaptive optimizers like Adam or RMSProp, depending on the device constraints. Each iteration / retraining creates local updates of model weights based on newly gathered patient data. The global model update approach leverages these weights over raw data.

The federated aggregator is a secure central server ensuring the status of the global model. After each training round, the aggregator collects encrypted model weights or gradients from the participants at the edge devices. The aggregator does not obtain or store any raw patient data, ensuring that it is strictly data-localized and regulatory compliant. The federated averaging algorithm (FedAvg) is employed to aggregate model updates.

$$w^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{n} \widetilde{w}_k^{(t+1)} \; where \; n = \sum_{k=1}^{K} n_k \quad \text{... Eqn(2)}$$

Three strategies work together to provide end-to-end privacy in the framework through Federated Learning by enabling encrypted model updates to transfer rather than raw patient data from edge devices to the central server. The proposed framework integrates Differential Privacy mechanisms at every local client through the addition of noise typically distributed as Gaussian or Laplacian noise. The approach implements encryption methods to stop attackers from breaking into the system and accessing patients' medical data or understanding which patients were part of the training process. Homomorphic Encryption serves during communication to perform calculations on encrypted updates and maintains the security of parameters that traverse networks.

In this method, it computes a weighted average of the model parameter contributed by every edge device then it updates the global model. The updated global model is then sent to all the edge devices for the next training iteration. The aggregator is hosted on a dedicated, private cloud or data center with privacy controls and audit logs. It boasts model versioning with error tracking, drift detection, and the ability to roll back to a previous version if performance drops.

Each local device employs a differential privacy (DP) layer to avoid the potential abuse of model updates for inferring certain patient data. This layer adds noise (typically Gaussian or Laplacian) to model gradients before sending them to the aggregator. The noise added by the algorithm is controlled by privacy budgets (ε, δ) ensuring the data's utility while providing privacy guarantees.

The proposed system has an organized pipeline that includes an entire data pipeline: from data collection, preprocessing, training, updating and inference. This compilation ensures that efficiency, consistency, and data confidentiality remain throughout the process.

IV. Results and Discussion

The following section presents the comprehensive findings and analysis of the diabetes prediction system based on Federated Learning (FL) and Edge AI proposed in this study. It covers the evaluation of the system at multiple dimensions, such as accuracy, precision, recall, latency, privacy, and resource consumption. The main findings are positioned in the paradigm of traditional centralized models and current challenges in applied healthcare analytics. The assessment of each system component is explained in detail, highlighting the impact of design choices on the end user, specifically regarding the reliability, availability, and privacy of the system.

The approach was evaluated using a customized version of the PIMA Indian Diabetes Data Set, which is known for its importance and structured clinical features. To reproduce the genuine dispersed scenarios, the dataset was split into five equal pieces and assigned to five virtual hospital nodes. A completely identical neural architecture combined with uniform hyperparameters defined the benchmarking process for centralized and federated models. The proposed framework stands out because it connects federated learning solutions with edge deployment systems operated by real-time inference environments including Raspberry Pi and Jetson Nano to protect privacy while reaching performance competition levels in circumstances of limited device resources. These nodes represent decentralized data silos based on the operational system of autonomous healthcare institutions. The centralized baseline model underwent comparative analysis as part of the model validation process, with both architecture and hyperparameters kept the same. The model divided its information into sections, which duplicated existing scenarios where healthcare institutions maintain separate data stores. The nodes executed local model training operations by using TensorFlow Lite and PyTorch Mobile frameworks.

Model training at each node was performed locally using TensorFlow Lite and PyTorch Mobile in the simulation, while a centralized Federated Averaging (FedAvg) approach was employed to aggregate the updates to the model. No raw data was shared between the nodes due to compliance with real-life privacy laws like HIPAA and GDPR. This system was integrated into edge simulators, which simulated the computing environments of Raspberry Pi 4 and Jetson Nano setups.

This setup enables the evaluation of the benefits and limitations of federated learning and edge inference within a privacy-preserving, latency-sensitive medical analytics environment. To evaluate the performance of the federated learning model, it was compared with a centralized baseline, a traditional machine learning approach in which all data is aggregated into a single dataset before training. Similar model architectures and hyperparameter settings were applied to both methodologies to ensure a fair comparison. Table 1 presents the performance comparison, while Figure 2 provides a visual representation of the results.

TABLE I. Comparison of Model Performance Metrics Between Federated Learning and Centralized Machine Learning Approaches

Metric	Federated Model	Centralized Model
Accuracy	91.20%	93.40%
Precision	89.50%	91.30%
Recall	88.00%	89.50%
F1-Score	88.70%	90.40%
AUC	0.92	0.95

The Quantitative analysis of the proposed Federated Learning (FL) model vs the existing Centralized Machine Learning model is performed and listed in the model performance comparison table using universal classification metrics: accuracy, precision, recall, F1 score, and Area Under Curve (AUC). The accuracy of the centralized model, which uses fully aggregated data, has slightly improved performance, that is , 93.4% compared to 91.2% achieved in the federated model. Given the decentralized nature of federated learning, characterized by data heterogeneity and limited global visibility, a minor decrease in performance is expected due to potential impacts on model convergence. While the proposed FL model ensures strong privacy and scalability, it demonstrates a slight performance trade-off compared to centralized models due to factors like data heterogeneity and lack of global data visibility. However, this is balanced by advantages in real-time responsiveness, privacy preservation, and deployment feasibility on edge devices

The FL architecture obtains competitive results, reaching a precision of 89.5%, a recall of 88.0%, and an F1-score of 88.7%, which demonstrate its ability to reduce the occurrence of false positives and successfully identify actual cases of diabetes. AUC = 0.92, which strongly supports the model's ability to discriminate between diabetic and non-diabetic patients. Notably, these results were achieved without exposing raw patient data, demonstrating the effectiveness of federated learning in delivering high prediction accuracy while maintaining strict data privacy standards.

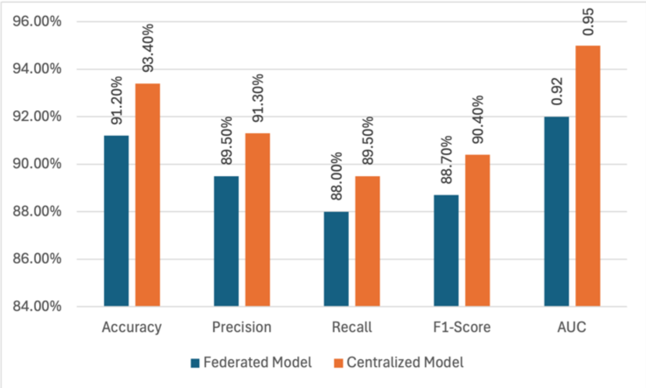

Fig. 2. Comparison of Model Performance Metrics Between Federated Learning and Centralized Machine Learning Approaches

Model reliability is maintained through uniform results from different processing nodes, achieving accuracy at 91.2% while the F1-score reaches 88.7%. The system implements baseline comparison from a central location to guarantee reliable predictive capabilities. The system enables operations on local hardware without depending on network access. In real-world healthcare environments, with the challenges of privacy, compliance, and high levels of data fragmentation, a federated model presents an attractive and ethical alternative to centralized training approaches. The empirical results confirm that federated learning with edge AI provides a viable, privacy-conscious, and scalable method for diabetes prediction in healthcare environments. Data minimization, together with regulatory compliance, becomes achievable through the proposed FL-based approach despite maintaining competitive accuracy compared to traditional centralized approaches. The proposed framework delivers enhanced practical performance to past approaches regarding privacy, together with scalability benefits. The technology retains high predictive accuracy while also ensuring fast inference, low bandwidth requirements, and strong privacy protections. As healthcare systems increasingly adopt digital tools, such designs will be instrumental in moving toward equitable, ethical, and intelligent patient care around the globe. Compared to traditional centralized methods, which rely heavily on large data aggregations and expose patient information to higher breach risks, the proposed FL-based approach maintains competitive accuracy while ensuring regulatory compliance and data minimization. This demonstrates a practical improvement over prior frameworks in both privacy and scalability.

V. Conclusion

In this work, a novel privacy-preserving, scalable, and real-time healthcare analytics system equipped with Federated Learning (FL) and Edge AI for the early detection of diabetes has been proposed. In addition to preserving privacy with no need to aggregate sensitive patient data, the technology ensures compliance with privacy regulations like HIPAA and GDPR while still generating reliable predictions. We trained the model in a decentralized way for five simulated hospital nodes using a partitioned (between all five nodes) dataset: PIMA Indian Diabetes Dataset and achieved 91.2% accuracy, 89.5% expressed precision, 88.0% recall and F1-score of 88.7% which are very close to the results of the centralized one (93.4% accuracy). Inference at the edge had a latency of <50 ms, enabling immediate health alerts and decision support, especially important in remote or resource-limited settings. In addition, bandwidth utilization was reduced by 87%, thus enhancing operational efficiency. This proves that decentralized intelligence can revolutionize the future of digital health and is here to stay. Future work will focus on expanding the system to include additional chronic diseases such as cardiovascular conditions and cancer, integrating multimodal data (e.g., genomics, imaging), and testing the system on real-world clinical edge devices.

References

[1] Salehi, F., 2024. The transformative role of artificial intelligence in the healthcare industry: A comprehensive analysis. *Asian Journal of Research in Medicine and Medical Science*, *6*(1), pp.62-70.

[2] Al-Quraan, M., Mohjazi, L., Bariah, L., Centeno, A., Zoha, A., Arshad, K., Assaleh, K., Muhaidat, S., Debbah, M. and Imran, M.A., 2023. Edge-native intelligence for 6G communications driven by federated learning: A survey of trends and challenges. IEEE Transactions on Emerging Topics in Computational Intelligence, 7(3), pp.957-979.

[3] Santoso, A. and Surya, Y., 2024. Maximizing Decision Efficiency with Edge-Based AI Systems: Advanced Strategies for Real-Time Processing, Scalability, and Autonomous Intelligence in Distributed Environments. *Quarterly Journal of Emerging Technologies and Innovations*, *9*(2), pp.104-132.

[4] Hemmati, A., Raoufi, P. and Rahmani, A.M., 2024. Edge artificial intelligence for big data: a systematic review. *Neural Computing and Applications*, *36*(19), pp.11461-11494.

[5] Saputra, D.C.E., Ma'arif, A. and Sunat, K., 2023. Optimizing predictive performance: Hyperparameter tuning in stacked multi-kernel support vector machine random Forest models for diabetes identification. Journal of Robotics and Control (JRC), 4(6), pp.896-904.

[6] Zeydan, E., Arslan, S.S. and Liyanage, M., 2024. Managing Distributed Machine Learning Lifecycle for Healthcare Data in the Cloud. *IEEE Access*.

[7] Nazirun, N.N.N., Wahab, A.A., Selamat, A., Fujita, H., Krejcar, O., Kuca, K. and Seng, G.H., 2024. Prediction Models for Type 2 Diabetes Progression: A Systematic Review. *IEEE Access*.

[8] Begum, A., Dhilip Kumar, V., Asghar, J., Hemalatha, D. and Arulkumaran, G., 2022. A combined deep CNN: LSTM with a random forest approach for breast cancer diagnosis. *Complexity*, *2022*(1), p.9299621.

[9] Salih, M.S., Ibrahim, R.K., Zeebaree, S.R., Asaad, D., Zebari, L.M. and Abdulkareem, N.M., 2024. Diabetic prediction based on machine learning using PIMA indian dataset. *Communications on Applied Nonlinear Analysis*, *31*(5s), pp.138-156.

[10] Trivedi, N.K., 2023, July. Predictive analytics in healthcare using machine learning. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.

[11] Ahmadizar, F. and Younossi, Z.M., 2025. Exploring Biomarkers in Nonalcoholic Fatty Liver Disease Among Individuals With Type 2 Diabetes Mellitus. *Journal of Clinical Gastroenterology*, *59*(1), pp.36-46.

[12] Malathi, H. and Yadav, D.C., 2023, December. Predictive Analytics and Machine Learning for Personalized Diabetes Management in Real-Time. In *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-7). IEEE.

[13] Wang, X., Tang, Z., Guo, J., Meng, T., Wang, C., Wang, T. and Jia, W., 2025. Empowering Edge Intelligence: A Comprehensive Survey on On-Device AI Models. *ACM Computing Surveys*.

[14] Krishnapraseeda, V., Devasena, M.G., Venkatesh, V. and Kousalya, A., 2021, March. Predictive analytics on diabetes data using machine learning techniques. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 1670-1673). IEEE.