www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 2672-2679

e-ISSN : 2583-1062
Impact Factor : 7.001

PRIVACY PRESERVING FREDERATED MACHINE LEARNING IN HEALTH CARE

Sigilipalli Laxmi Prasanna[1]

[1]GMRIT Rajam.

ABSTRACT

Federated machine learning that ensures privacy is a critical tool for protecting sensitive information about health while promoting cooperative model building amongst different health care organizations. Traditional central FL may not be advantageous for private health information because it increases points of failure, bottlenecks of communication, and even incidences of data leakage. Recent developments that mitigate these constraints are decentralized FL frameworks that offer better privacy and security. Two approaches that enhance communication efficiency without compromising data integrity are ring-based structure and Ring-All reduce-based data sharing. The privacy-preserving FL architecture further leverages FAIR—findable, accessible, interoperable, and reusable—health data for safe model training without explicitly sharing data among collaborators. These approaches have been found to be efficient in predicting the risk of readmission. The related advances include federated edge and privacy-enhancing technology federated edge intelligence frameworks for medical images, which provide additional data safety guarantees and mitigate privacy concerns of IoHT-based healthcare systems. Several possibilities of safe data management and incentive schemes are provided by blockchain and NFT technologies for federated learning frameworks. These advances are well driving the construction of safe, effective, and private machine learning models with strict privacy and regulatory requirements.

Keywords: Federated learning, Privacy-preserving, Health care, Decentralized frameworks, FAIR health data, ring-allreduce, Privacy-enhancing technologies, NFT.

1. INTRODUCTION

Revolutionizing healthcare, acceleration in diagnosis and treatment advances, and wide-ranging changes in patient outcomes through high-tech machine learning. Among all these innovation trends, one new paradigm of federated learning has come into a picture that trains models at different institutions without transferring the raw data. Such innovations bring out an increasingly more serious need for patient data privacy [1][2]. Federated learning is, therefore, a very relevant field in healthcare, as such applications touch sensitive data, typically EHRs, with which most predictive analytics and clinically-related decisions are related. Essentially, Federated Learning enhances security of data by collaborative training globally distributed models on decentralized data sets and takes advantage of institutional data diversity to be more effective for model fitting [3][4]. Several techniques support Federated Learning's privacy-preserving capabilities, including differential privacy, homomorphic encryption, and k-anonymity, which ensure that the sensitive information, during processing and analysis, cannot be identified [5][6][8]. These methods therefore allow effective applications of machine learning while keeping the patient confidentiality. More critically, newly emerging privacy-enhancing technologies, like Searchable Encryption, Multi-Party Computation and Functional Encryption are under considerations to enable complex applications such as Deep Radiomics to ensure end-to-end maximally reliable utilization of AI in healthcare [9][12][13]. Federated learning also encompasses significant challenges. These issues will be client attacks; aggregation methods; and issues related to scalability, acting instead as countermeasures toward possible adversarial threats during training. This is countered by hybrid privacy-preserving technologies, optimization of the client selection processes, and improving the security of federated systems to achieve scalability and reliability [10][14][15]. This is the key in balancing data privacy protection with accuracy and utility within artificial intelligence-driven healthcare solutions, which would make Federated Learning the hub of intelligent and secure healthcare innovation [7][11].

2. PREVIOUS WORKS

Li, J., Meng, Y., Ma, L., Du, S., Zhu, H., Pei, Q., & Shen, X.a et al. [1] proposed system is designed to support the early detection of Alzheimer's disease with preservation of user privacy. The system relies on IoT devices and security mechanisms to protect user data such that data will remain on user devices. The system applies a privacy-preserving detection framework, differential privacy, and a Laplace mechanism to protect user data. The experimental results show high accuracy in AD detection with strong privacy protections. The paper argues that privacy is important within smart healthcare systems.

Abbaoud et al., M. Almuqrin, M. A., & Khan, M. F., et al., [2] for privacy and balancing clinical utility with data privacy. Using the MIMIC-III database and Syn0thea synthetic dataset for training improves the robustness of model training with patient privacy maintained. The results show significant improvements over existing models in

computational efficiency, precision, and recall. The authors state that such developments will promote federated learning across health care to better patients and researchers.

Yazdinejad,A., Dehghantanha,A., & Srivastava, G.et al,[3] federation with non-independent and identical distributions in Federated learning. It has several users together with one central host that has an auditor checking the accounts. The process is carried out in phases of client update, authentication and audit accompanied by a threat model and the performance evaluation. Blockchain technology as applied in future work.

Sinaci, A. A., Gencturk, M., Alvarez-Romero, C., Erturkmen, G. B. L., Martinez-Garcia, A., Escalona-Cuaresma, M. J., & Parra-Calderon, C. L. et al., [4] The architecture is the federated ML Agents with a federated ML manager implemented in a secure network of the healthcare organizations involved; thus, data privacy is protected by not transferring sensitive data while enhancing model quality. This was tested through a real-world application of five European healthcare organizations. The potential that could be unlocked with using FAIR health data in federated learning is one way that will open doors for further improvement in the domain of healthcare analytics and data-driven decision-making.

Akter, M., Moustafa, N., Lynar, T., & Razzak, I. et al., [5] Proposed a novel framework: A hierarchical three-fold architecture: Differential Privacy and Edge Intelligence. A new integration based on Edge Intelligence, combining it with Federated learning, this framework does prove better protection of the private information than in the regular case of Federated Learning that takes care of two considerations, both privacy preservation as well as model precision.

This approach combined the framework of Sai, S., Hassija, V., Chamola, V., & Guizani, M. et al.,[6] with NFT-based blockchain technology to solve the fragmented issue of medical data in health systems. The system allows safe data sharing and ownership control while preserving patients' privacy. NFTs manage data ownership, and a blockchain-based incentive mechanism encourages users to contribute their data for intelligent diagnosis. The proposed model includes workflow, NFT marketplace, incentive mechanism, and aggregation of local models. This model works better than traditional centralized models and popular FL algorithms in terms of security and privacy. In the future, one could extend this framework for other types of healthcare data.

Rahman, A., Hossain, M. S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M.,. & Band, S. S. et al [7] transformed healthcare through the use of connected devices to collect real-time data. Federated Learning (FL) provides a solution by enabling decentralized AI models that can be trained locally, thereby preserving patient data privacy. The authors propose a comprehensive framework for implementing FL in the health sector, including decentralised AI model training, data privacy, blockchain, and XAI. Challenges include data heterogeneity, communication overhead, security, and compliance with regulations.

Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. et al.,[8] revolutionized healthcare by connecting wearable and medical devices for real-time monitoring, enhancing patient care. However, this has also introduced challenges related to data privacy. Generally speaking, traditional AI frameworks consume centralized data processing; due to this, privacy risk amplifies. Federated learning then, as a solution provides with the possibility of data being trained without sharing the unprocessed data, ensuring and supporting user privacy. Among modern complex FL architectures, are especially addressed a few: Digital Twin(DT), Generative adversarial networks(GAN), and Deep reinforcement learning(learned). FL is used in many smart healthcare scenarios, such as EHR, Medical Image Analysis, and COVID-19 detection. Some of the future research directions are optimizing communication in FL, universal standards for FL, and improving aggregation of models.

Ahamed, S. K., Nishant, N., Selvaraj, A., Gandhewar, N., Srithar, A., & Baseer, K. K. et al., [9] that addresses the problems of health research, particularly data sharing. It suggests the use of PPML techniques, such as Federated Learning, to have safe data sharing without violating the patient's privacy. The study highlights a research gap on real-world implementation and suggests collaboration for future work and understanding the comorbidity indices distribution.

Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. et al.,[10] Uncertainty discusses the integration of Federated Learning (FL) in medical imaging, focusing on privacy preservation and uncertainty estimation. It discusses techniques like differential privacy and homomorphic encryption, and introduces a knowledge transfer method called PrivateKT. The paper also addresses challenges like communication overhead, data heterogeneity, and security threats.

Sedghighadikolaei, K., & Yavuz, A. A. et al., [11], Importance of Privacy and Security in Deep Radiomics Deep Radiomics medical images are analysed using techniques for Deep Learning. It presents various PETs and integration in the pipeline of Deep Radiomics. Challenges such as scalability, accuracy, and security risks are highlighted while proposing hybrid PET compositions that improve security and performance.

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. et al., [12], Deep Radiomics, hybrid medical imaging and Deep Learning techniques, their importance in the augmentation of diagnosis and treatment. In this paper, the authors will introduce a number of PETs and their likely usages, categorize them into four groups depending upon their functionality. The authors discuss some challenges and future directions: scalability, accuracy, and risks in security.

Ganadily, N. A., & Xia, H. J. et al., [13], the importance of privacy preservation in AI-based healthcare applications highlighting its potential to improve diagnosis, treatment, and patient outcomes. It discusses challenges like non-standardized medical records, limited dataset availability, and stringent legal requirements. The authors propose privacy-preserving techniques like federated learning, hybrid techniques, cryptographic techniques, and non-cryptographic techniques. Grama, M., Musat, M., Muñoz-González, L., Passerat-Palmbach, J., Rueckert, D., & Alansary, A. et al., [14]. the utilitarian adoption of ML in EHR for improved healthcare outcomes; Privacy-protecting Techniques-the utilization of Synthea as well as the Kaggle datasets followed by the application of some basic models of ML predicting patients charges by drawing also discussion on differential privacy as well as Pseudonymisation plus, more discussion on FL implementations and finally Federated Analytics with the implementation based upon various sub-specialised models. Tian, Y., Wang, S., Xiong, J., Bi, R., Zhou, Z., & Bhuiyan, M. Z. A. et al.,[15] Federated Learning (FL in healthcare, emphasizing the advantages such as collaborative learning, better model accuracy, and improved patient outcomes. It also deals with the challenges like data privacy concerns, adversarial clients, and data poisoning attacks. The authors give an experiment setup and results, pointing out the efficacy of robust aggregation methods and k-anonymity in identifying and rejecting malicious clients.

3. METHODOLOGY

3.1 Federated Learning Framework: A privacy-preserving federated machine learning system where health care data stays within the local servers of each organization. Only model parameters that are trained locally are shared with a centralized federated manager, holding no real data. This allays some of the privacy concerns due to the centralization of data but still uses diverse datasets for training the ML model.

3.2 FAIR Data Principles: Converted local health datasets into standardized HL7 FHIR, compliant with the FAIR data principles. It should ensure that the data was in a machine-readable format with interoperability and sharing capabilities across different healthcare organizations. The FAIRification workflow was applied to transform existing health data into FAIR-compliant datasets.

3.3 Agent-Based Architecture: Created an apparatus comprising two main entities Federated ML Manager or Centralized Orchestration component. Federated ML Agent: This is the local component that is to be installed at each of the participating healthcare organizations. The coordinator of the training process is the manager, and the agent prepares the data, trains local models, and communicates with the manager.

3.4Data Preparation: FHIR Search Queries & FHIR Path Statements: They use to fetch the data from the local HL7 FHIR repositories. For example, using FHIR search parameters, I can get the demographics, conditions, and medication for a patient.The extracted data was transformed into tabular form and then prepared for the ML task. That is, FHIR resources were translated into rows that represent records, and columns that represent features.

3.5 Privacy-Preserving Federated Learning: Used federated learning where a number of clients were permitted to train the same model cooperatively without sharing any data. It uses a local model trained on its internal dataset and sends only the model parameters, not raw data, to the Federated ML Manager at the company's end. Boosting Algorithm: This applied a new aggregation method that aggregates local models into a single global model. I aggregated the confusion matrix to determine how best to assign weights to local models on the accuracy level. Created a global model by computing a weighted sum of local models.

3.6 Logistic Regression: Classification algorithm to predict probabilities of outcomes (e.g., readmission risk).In federated learning, the hospitals train a local model and share only coefficients of the model instead of sharing raw patient data. The aggregator central server aggregates the coefficients to form a global model. Simple, efficient, and interpretable, due to which it is suitable to healthcare datasets with relatively fewer features.

3.7 Support Vector Machine (SVM): Classification algorithm that aims to find the optimal hyperplane in order to distinguish classes. Each medical institution trains a local SVM model and reports only the support vectors, protecting their privacy. The central server collects support vectors and works them out to improve a global SVM model. It does well on high-dimensional data. Hence, it is useful for complex medical data.

3.8 Decision Trees: Tree-based approach that partitions data based on feature values to produce a prediction, say, disease diagnosis In the federated scenario, hospitals report only the tree structures and not patient records The central server combines those trees either to better the prediction or use ensemble methods. Highly interpretable. This is critical in clinical decision-making

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 2672-2679
e-ISSN : 2583-1062
Impact Factor : 7.001

3.9 Random Forest: An ensemble learning approach that generates multiple decision trees and combines their output. Organizations train a local version of Random Forest and share only the trees while keeping the patients' data local The central server aggregates those trees into a robust global model This algorithm is noise-resistant for noisy health care data and handles data imbalance very well.

3.10 Model Performance Metrics:

The application will evaluate each classifier for the following four key performance metrics:

Accuracy: Ratio of correctly classified samples of all samples.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

Precision: the percentage of correct positive predictions out of the actual number of predicted positives, describes how the model can suppress false positives.

$$Precision = \frac{TP}{TP + FP}$$

Recall: the percentage of positive predictions of the total to be actual positive, illustrating how good a model is at locating the positive samples.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: it describes the harmonic mean of both precision and recall. Even if class imbalance occurs, the F1 score is balanced.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

All of these metrics can be available when the model outputs the results on the testing dataset. The user then compares all these metrics in these multi-classifiers in order to find the one which best fits his/her data set.

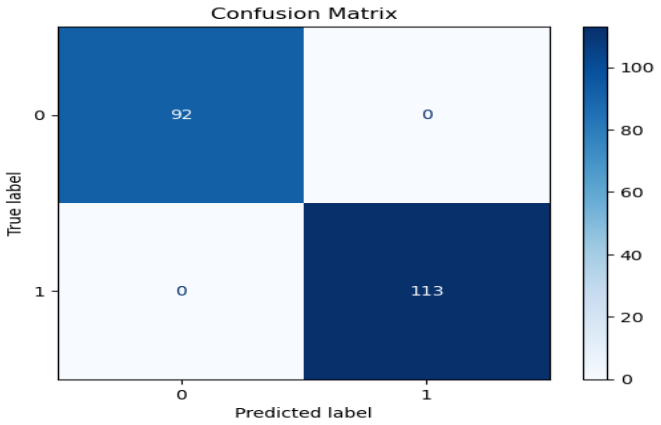


Fig.1 Model performance of "*heart*" dataset

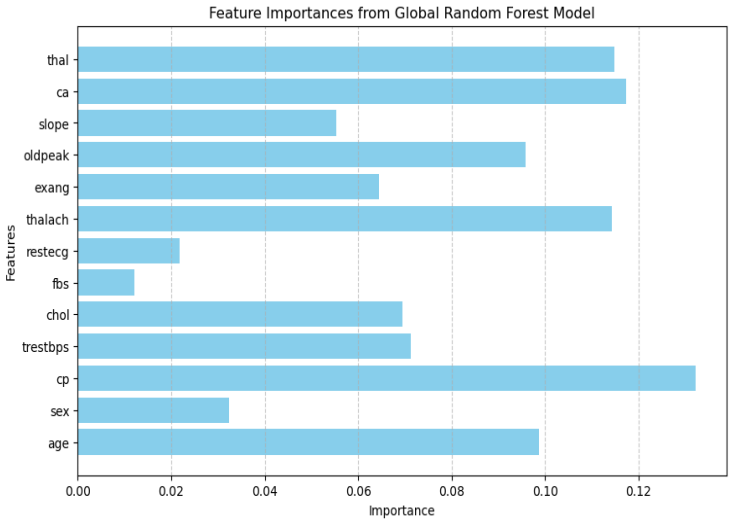


Fig .2 Data of "*Heart"* Dataset

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 2672-2679

e-ISSN : 2583-1062
Impact Factor : 7.001

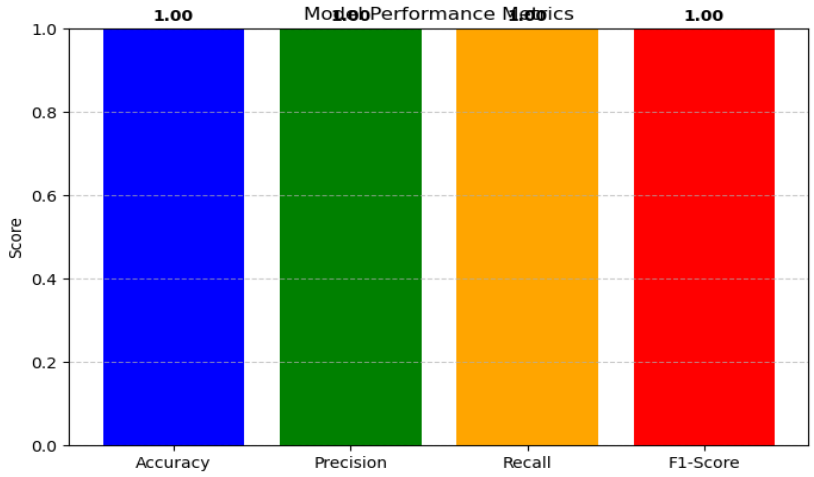


Fig 4 Model performance metrices

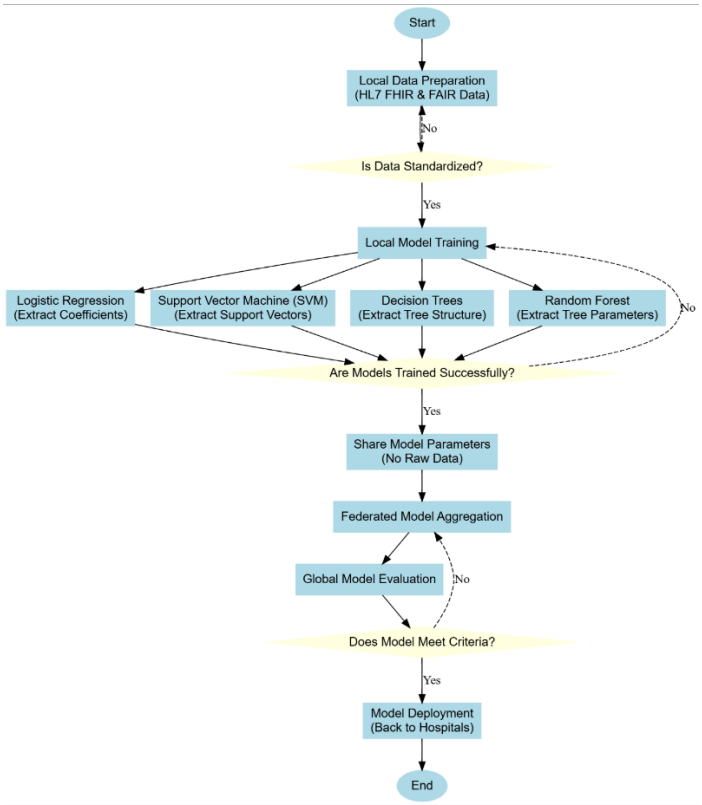


Fig.3 Flowchart.

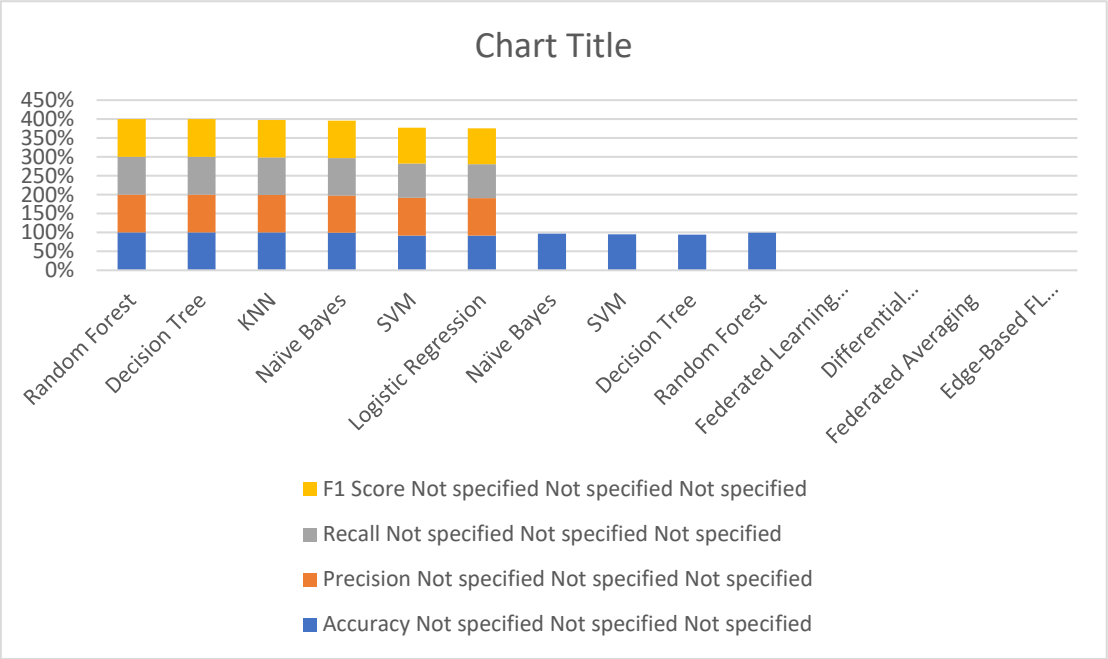
4. COMPARSION AND RESULTS:

Ref No	Objectives	Limitations	Advantages	Gaps
[1]	It is to design a privacy-preserving system called ADD ETECTOR that enables the early detection of Alzheimer's disease.	It is to develop a privacy-preserving system called ADD ETECTOR that facilitates the early detection of Alzheimer's disease	ADD ETECTOR effectively maintains user privacy by keeping raw data on local devices and using federated learning to train models.	The paper centers on federated learning for privacy in smart healthcare but doesn't investiga te scalability in large datasets.
[2]	Develop and evaluate federate d learning models to effectively	A significant limitation is the trade-off between privacy and model	The primary advantage of the proposed models is their ability to maintain high privacy standards while	The study does not account for the real-world challenges of

	preserve privacy while analyzing healthcare data to improve patient care and medical research without compromising the sec urity of the data.	performance, where implementing strong privacy measures can potentially affect the accuracy and efficiency of the models.	delivering significant improvements in computational efficiency and predictive accuracy compared to existing models.	deployment into diverse healthcare environments while advancing the mechanisms of privacy preservation.
[3]	The main objective of the AP2FL framework is to enhance federated learning by addressing privacy preservation and the challenges of non-independent and identically distributed data	the complexity of integrating and managing multiple components, including the auditor and the Trusted Execution Environments	The main advantage of the AP2FL framework is its comprehensive approach to privacy and data integrity, incorporating auditing mechanisms and secure processing environments to ensure that sensitive healthcare data is protected.	The framework is audita ble for privacy, but i t doesn't consider the computational overhead in devices with resource constr aints.
[4]	Develop a federated machine learning architecture that ensures data privacy while allowing collaborative model training using FAIR .	The framework currently supports only binary classification algorithms and does not specifically address time-series data, which is common in healthcare.	Ensures high data privacy and compliance with regulations by avoiding the centralization of sensitive health data, while achieving effective collaborative model training and interoperability using FAIR principles.	Real-world application is demonstrated, and l ong-term data drift or evolving data privacy standards remain unraveled in the paper.
[5]	Develop a framework integrating edge intelligence with federated learning to enhance privacy protection in smart healthcare systems while maintaining high model performance.	The framework's evaluation needs further expansion to address various privacy attacks and improve the flexibility of client participation.	e Federated Edge Aggregator (FEA) framework provides robust privacy protection by keeping sensitive data local and integrating differential privacy with edge intelligence, while also improving model accuracy compared to traditional federated learning methods.	Real-world application is demonstrated, and l ong-term data drift or evolving data privacy standards remain unraveled in the paper.

Ref.No	Model	Accuracy	Precision	Recall	F1 Score
[1]	Federated Learning Framework	Not specified	Not specified	Not specified	Not specified
	LSTM	-	-	-	-
	RNN	-	-	-	-
[2]	Random Forest	100%	100%	100%	100%
	Decision Tree	100%	100%	100%	100%
	SVM	92%	100%	90%	95%
	Logistic Regression	92%	99%	90%	94%
[3]	Logistic Regresion	97%	-	-	-
	SVM	95%	-	-	-

	Decision Tree	94.7%	-	-	-
	Random Forest	99%	Not specified	Not specified	Not specified
[4]	Federated Learning Model	Not specified	Not specified	Not specified	Not specified
	Differential privacy based approach	Not specified	Not specified	Not specified	Not specified
[5]	Federated averaging	Not specified	Not specified	Not specified	Not specified
	Edge-Based FL Framework	Not specified	Not specified	Not specified	Not specified



5. CONCLUSION

Federated machine learning has emerged as one of the most prominent transformative approaches in health care, in which multiple organizations collectively train models on data and safeguard patient privacy. FL decentralized the data and used FAIR-compliant HL7 FHIR datasets to address the inherent privacy issues in traditional centralized systems. Security and robustness are enhanced by Ring AllReduce, blockchain integration, and various other privacy enhancing technologies such as Differential Privacy and Homomorphic Encryption used in these frameworks. Among the several algorithms tested, Random Forest turned out to be the most effective one for healthcare data due to its tolerance towards imbalances and noise. Evaluated with accuracy, precision, recall, and F1-score metrics for the FL framework, it reflected on the development of models with security and reliability without affecting the privacy of data. These innovations create avenues through which compliant AI solutions may evolve within healthcare systems: robust systems that are scalable and balance data protection with collaborative analytics.

6. REFERENCES

[1] Li, J., Meng, Y., Ma, L., Du, S., Zhu, H., Pei, Q., & Shen, X. (2021). A federated learning based privacy-preserving smart healthcare system. *IEEE Transactions on Industrial Informatics*, *18*(3).

[2] Abaoud, M., Almuqrin, M. A., & Khan, M. F. (2023). Advancing federated learning through novel mechanism for privacy preservation in healthcare applications. *IEEE Access*, *11*, 83562-83579.

[3] Yazdinejad, A., Dehghantanha, A., & Srivastava, G. (2023). AP2FL: Auditable privacy-preserving federated learning framework for electronics in healthcare. *IEEE Transactions on Consumer Electronics*.

[4] Sinaci, A. A., Gencturk, M., Alvarez-Romero, C., Erturkmen, G. B. L., Martinez-Garcia, A., Escalona-Cuaresma, M. J., & Parra-Calderon, C. L. (2024). Privacy-preserving federated machine learning on FAIR health data: A real-world application. *Computational and Structural Biotechnology Journal*, *24*, 136-145.

[5] Akter, M., Moustafa, N., Lynar, T., & Razzak, I. (2022). Edge intelligence: Federated learning-based privacy protection framework for smart healthcare systems. *IEEE Journal of Biomedical and Health Informatics*, *26*(12), 5805-5816.

[6] Sai, S., Hassija, V., Chamola, V., & Guizani, M. (2023). Federated learning and NFT-based privacy-preserving medical data sharing scheme for intelligent diagnosis in smart healthcare. *IEEE Internet of Things Journal*.

[7] Rahman, A., Hossain, M. S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M., ... & Band, S. S. (2023). Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Cluster computing*, *26*(4), 2271-2311.

[8] Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. *IEEE journal of biomedical and health informatics*, *27*(2), 778-789.

[9] Ahamed, S. K., Nishant, N., Selvaraj, A., Gandhewar, N., Srithar, A., & Baseer, K. K. (2023). Investigating privacy-preserving machine learning for healthcare data sharing through federated learning. *The Scientific Temper*, *14*(04), 1308-1315.

[10] Koutsoubis, N., Yilmaz, Y., Ramachandran, R. P., Schabath, M., & Rasool, G. (2024). Privacy Preserving Federated Learning in Medical Imaging with Uncertainty Estimation. *arXiv preprint arXiv:2406.12815*

[11] Sedghighadikolaei, K., & Yavuz, A. A. (2024). Privacy-Preserving and Trustworthy Deep Learning for Medical Imaging. *arXiv preprint arXiv:2407.00538*.

[12] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, 106848.

[13] Ganadily, N. A., & Xia, H. J. (2024). Privacy Preserving Machine Learning for Electronic Health Records using Federated Learning and Differential Privacy. *arXiv preprint arXiv:2406.15962*

[14] Grama, M., Musat, M., Muñoz-González, L., Passerat-Palmbach, J., Rueckert, D., & Alansary, A. (2020). Robust aggregation for adaptive privacy preserving federated learning in healthcare. *arXiv preprint arXiv:2009.08294*.

[15] Tian, Y., Wang, S., Xiong, J., Bi, R., Zhou, Z., & Bhuiyan, M. Z. A. (2023). Robust and privacy-preserving decentralized deep federated learning training: Focusing on digital healthcare applications. *IEEE/ACM Transactions on computational biology and bioinformatics*.