IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)
e-ISSN: 2278-1684,p-ISSN: 2320-334X, Volume 21, Issue 5 Ser. 2 (Sept. – Oct. 2024), PP 36-44
www.iosrjournals.org

Federated Learning for Secure and Privacy-Preserving Medical Collaboration Across Multi-Cloud Healthcare Systems

Umair Ejaz
Desert cart
Country: United Arab Emirates

S A Mohaiminul Islam
Email: mohaiminulbd271@gmail.com
Country: USA
Academic Degree: Master of Science , Major: Information Technology
Affiliation: Washington University of Science & Technology

Ankur Sarkar
ankursylhet@gmail.com
Same affiliation as mine
Mohammed Majid Bakhsh

Aidar Imashev
Barry University
Country: Kyrgyzstan
Email: aidar.imashev@mymail.barry.edu

Abstract: *Seamless artificial intelligence (AI) adoption into the contemporary medical frameworks has unlocked new opportunities in medical diagnosis, predictive analysis, and individualistic treatment planning. Nevertheless, the effectiveness of AI models largely depends on getting access to large, diverse, and high-quality data sets, which is becoming a challenging goal to achieve because of stricter privacy laws and due to institutional silos, as well as emerging use of multi-cloud systems by healthcare institutions. Aggregated data collection not only adds a risk of data loss but also, in many cases, goes against patient privacy standards stipulated or regulated by acts like the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR). As studied in this research, Federated Learning (FL) is a decentralized and privacy-protecting paradigm that enables secure medical collaboration through geographically and administratively decentralized healthcare facilities on various cloud platforms with heterogeneous configurations. FL allows several clients (e.g., hospitals, clinics) to use collective computation in training machine learning models, but not sharing raw data, thus maintaining data locality and data confidentiality. We suggest a holistic system capable of coupling FL into a combined solution of complex privacy-preserving frameworks like secure multi-party computation, differential privacy, and homomorphic encryption to offer end-to-end protection against internal and external threats. The proposed paper offers a strong system architecture that could be used in multi-cloud settings, whose issues will include data non-compatibility, communication expense, model convergence, and compliance policies. The proposed approach's performance, scalability, and security are analogously analyzed using real-world medical imaging and electronic health record (EHR) data, providing a thorough collection of experiments. These findings show that the federated model delivers close-to-accuracy with centralized ones, with much less risk involved in centralized data storage and transmission. Moreover, we prove the framework's flexibility with an alternative of different cloud service providers, proving that it can be applied in the real collaborative healthcare ecosystem. To sum up, the present work confirms that federated learning has the potential to become an ever-changing solution to the creation of secure, privacy-preserving, and regulation-compatible AI in multi-cloud healthcare environments, leading to more morally-intelligent and higher-performance medical AI applications.*

Keywords: *Federated Learning, Privacy-Preserving AI, Multi-Cloud Computing, Healthcare Data Security, Collaborative Machine Learning*

I. INTRODUCTION

Over the past few years, the healthcare market has experienced an unheralded explosion of data-driven technologies, especially those fueled by artificial intelligence (AI) and machine learning (ML). These tools have greatly impacted clinical decision-making, diagnostics, predictive analytics, and patient-centered care. Nevertheless, these technologies' overall potential is underutilized because of one missing link, which is a limited capacity to access and utilize the medical data, large in scale, diverse, and high quality, that are scattered over different hospitals, laboratories, and research centers. Even more so, this issue is complicated by the presence of strict data privacy laws, including the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) that censure the centralized exchange and processing of sensitive patient data (Mbonihankuye et al., 2019; Azbeg et al., 2022). Grid-based Learning (GL), and more recently Federated Learning (FL), have been proposed as a revolutionary way to solve this dilemma in that they allow decentralized training of ML models in a way that keeps the data from bottling up. In the case of FL, it is assumed that an individual healthcare facility controls its local data and maintains the global representation by exchanging only encrypted or anonymized updates. Such an approach does not pose much danger to privacy infiltrations but allows collaboration on a large scale. Xu et al. (2021) demonstrated the use of FL that could enable healthcare institutions to train deep learning models collectively without affecting patient privacy, representing a significant step toward privacy-sensitive AI. Yang, Liu, et al. (2019) continued to discuss FL's general architecture and scope to distributed systems and how it applies to dynamic environments such as healthcare and potentially real-time learning. On the one hand, FL is not simple to use in healthcare despite these benefits. Among the most ubiquitous technical issues, we can count overcoming the non-IID (non-independent and identically distributed) phenomenon of healthcare data. Clinical data tends to differ extensively in different institutions regarding demographics, disease prevalence, device conformity, and data structures. Zeng et al. (2023) pointed out that these heterogeneities may badly affect convergence and precision of models, especially where global models do not work as generalizers across diverse groups of patients. Besides, distributed denodules share an implied vulnerability of observing model updates. As it was covered by Zhang et al. (2022), even anonymized gradients or model components can be reverse-engineered to divulge sensitive information about patients, and additional security measures should be incorporated into their functioning.

The additional complexity is added by the increasingly widespread use of multi-cloud systems as a part of the healthcare system. Healthcare providers are multi-sourcing their cloud-based service providers, using more than one service provider, to maximise cost, performance, and compliance. This disjointed infrastructure, however, comes with its own set of challenges dealing with data consistency, interoperability, access control, and trust boundaries. As demonstrated by Chen et al. (2022), there is the potential of adopting multi-agent-based reinforcement learning to perform dynamic task offloading in a multi-cloud environment, as such systems can get quite complex. Kalyani and Rao (2016) have given a more comprehensive way to construct scalable and secure multi-cloud systems. They said protocols are required to coordinate workload and secure sensitive information on various platforms. With data privacy and cross-system collaboration, scientists have started adapting cryptographic practices into the federated frameworks. As an illustration, Rahman et al. (2020) proved how fully homomorphic encryption can be applied to enable computations to be carried out on encrypted information, thereby eliminating risks of exposing data to view even in the computation process. Torkzadehmahani et al. (2022) also gave an overview of privacy-preserving AI techniques in the biomedical informatics context, elaborating on topics of growing interest such as secure computation, differential privacy, and secure multi-party computation (SMPC) in the development of future AI systems.

II. LITERATURE REVIEW

Federated Learning (FL) has become a paradigm-shifting approach to machine learning since it has become an alternative method of co-learning model training without presenting sensitive information at risk. Raw data in traditional centralized learning models has to be pooled at one place, posing severe issues of privacy, security, and regulatory concerns, particularly in the medical field. Conversely, FL allows locally training distributed data sources, with only model parameters transferred to clients and a centralized server. Yang et al. (2019) were among the first ones to describe this idea and underline its appropriateness to be used when the sharing of data is limited, either because of confidentiality issues or legal restrictions. Healthcare is one of the first industries to adopt FL because of its rigid data privacy needs and medical data having a life and death impact. Xu et al. (2021) implemented FL in different health care applications, some being predictive modeling with electronic health records (EHRs), where the authors demonstrated better performance and privacy protection. They have shown that utilizing FL to aggregate cross-institutional knowledge was possible without breaching the data protection regulations, including HIPAA or GDPR. Joshi et al. (2022) also expanded on such findings, explaining how these attributes present obstacles to implementing FL pipelines in clinical settings concerning data heterogeneity, issues in infrastructure, and lack of standardization. One of the main drawbacks of FL-powered uses in the health sector is that the data there (on medical records) is non-independent and identically distributed

(non-IID). Information gathered by the various hospitals, labs, and geographic areas differs largely in formatting, semantics, and patient demographics. Zeng et al. (2023) examined an adaptive FL approach that could resolve these inconsistencies. Their work laid the stress that there is no way to avoid performance degradation by an FL system that does not take distributional differences into account, and suggested some ways to resist such degrading effect with personalized federated optimization and client clustering being the most promising ones. Likewise, in their systematic review, Zhang et al. (2021) emphasized that generalization in terms of decentralized clients is one of the main obstacles of FL, particularly in such industries as healthcare with high degrees of variability and inconsistent data labeling. FL experiments have also given major focus on security and privacy. Even though FL naturally secures the information by storing it in one location, it is not free of the threats of an adversary. Gradient leakage attack vectors, model inversion attack vectors, and poisoning attack vectors can attack the system. Zhang et al. (2022) also discussed the variety of security risks and responses to these challenges, recommending the application of differential privacy, secure multiparty computation (SMPC), and homomorphic encryption (HE). Specifically, Rahman et al. (2020) have suggested a privacy-preserving AI framework based on fully homomorphic encryption, whereby the AI model computation can be executed on the encrypted data. Such an approach guarantees that the sensitive data will not go outside the immediate environment in plain form, providing an additional security layer to FL-based medical cooperation. In complement to these initiatives, Torkzadehmahani et al. (2022) surveyed privacy-preserving approaches to AI in biomedicine. They pointed out the need to include legal and ethical elements in technical design. Aslan et al. (2022) have also upheld the need to take privacy-enhancing AI in healthcare research to the next level because companies and researchers should spend more time and effort on interdisciplinary connections to develop frameworks that are not only secure but also explainable, auditable, and in line with health information standards. These studies amplify the need for FL solutions that can accommodate both the issue of computational risk and the trust concerns of end-users, such as clinicians and patients. FL implementations in the healthcare sector become even more complicated with the emergence of multi-cloud systems. The common use of cloud-based platforms in the storage and processing of healthcare data and its management incurs risks of vendor lock-in, information breach, and data and process inefficiencies based on a single cloud provider. Chen et al. (2022) suggested cross-cloud solutions built on multi-cloud systems that rely upon utility-uprooted spoken tasks offloading through cooperative multi-agent reinforcement learning. The type of work they did showed a reduction in task scheduling, energy consumption, and latency in heterogeneous cloud environments. Kalyani and Rao (2016) have also proposed an overview of the roadmap on how to design secure multi-cloud systems, where orchestration of services, negotiation of trust, and decentralization of control will make securable and scalable architectural systems that are also resilient.

The development of the FL and multi-cloud computing combination leads to opportunities and challenges. Multi-cloud systems may increase the fault tolerance and scalability of FL systems, yet they open up new security risks and coordination challenges. With standard protocols, distributed identity management, and end-to-end encryption, privacy-preserving collaboration among various cloud providers is ensured. The interaction of data governance controls, data encryption systems, and network structures is quite a relatively unexplored field, especially when applied in a sensitive industry like healthcare. Moreover, developing technologies have also been introduced into FL-based healthcare systems to provide better privacy and reliability. BlockMedCare, designed by Azbeg et al. (2022), integrates Internet of Things (IoT), blockchain technologies, and the InterPlanetary File System (IPFS) that will enhance data management and security. Their work holds promise in assuring traceability and tamper-proofing, which are critical aspects of medical records. In turn, Ogrezeanu et al. (2022) concentrated on explainable AI models that can provide transparency and accountability within medical decision-making, which needs to be considered especially when including AI into the clinical workflow. Nevertheless, there are many research gaps despite the significant current developments. To date, the solutions introduced by researchers only eliminate some of the isolated components of FL, but not the whole puzzle, including privacy, security, data heterogeneity, and cross-cloud adaptability. This paper tries to bridge that gap by proposing an integrated FL framework that would address multi-cloud healthcare environments. The proposed solution seeks to achieve workload, regulatory compliance, and coordinated intelligence, and therefore is fit to be implemented in real-time environments in the contemporary healthcare systems.

Table 1: Federated Learning in Healthcare

Category	Key Insights	Key References
Core Concept	FL trains models across distributed data without sharing raw data; preserves privacy.y	Yang et al. (2019)
Healthcare Applications	Used in EHRs and predictive modeling while complying with data regulations	Xu et al. (2021)
Main Challenges	Data heterogeneity, infrastructure gaps, and lack of standardization	Joshi et al. (2022), Zhang et al. (2021)
Privacy & Security	Threats include data leakage and attacks; use of DP, SMPC, and HE is recommended.	Zhang et al. (2022), Rahman et al. (2020)

Multi-Cloud Integration	Improves scalability but raises coordination and security issues	Chen et al. (2022), Kalyani & Rao (2016)
Emerging Solutions	Blockchain, IoT, IPFS, and explainable AI enhance FL's reliability	Azbeg et al. (2022), Ogrezeanu et al. (2022)
Research Gap	Current FL systems only partially solve challenges; integrated frameworks are needed.d	–

III. METHODOLOGY

Based on this research, a systematic plan is used to design and evaluate a federated learning (FL) framework suitable for secure and privacy-preserving cross-institutional collaboration between healthcare institutions working in a multi-cloud setting. The methodology encompasses the structure of the system architecture, data preparation and management, the federated learning process, a combination of data privacy and security strategies, and implementation in an emulated multi-cloud environment.

3.1 Architecture Design

The proposed system is designed to work efficiently in a heterogeneous cloud environment to support the collaborative requirements of various healthcare institutions. Every participating party keeps all the local datasets under its aegis and implements a federated client that participates in the decentralized training of the model. These customers connect to a centralized coordination server in a safe, regulation-compliant multi-cloud. The architecture takes three logical stages, including a data layer to handle local data lakes within institutional firewalls, processing of training and encryption on their side, and a coordination layer, which governs aggregation and communication flow of the models. Cloud interoperability and scalability can be achieved via standardizing APIs and containerization so that the system can handle different data volumes, unique provider configurations, and compliance needs.

3.2 Data Governance and processing

According to an institution's data privacy rules and policies, the framework makes no raw medical data available for transfer or sharing. Instead, data is maintained within the precincts of the emitting institution. The local data, which can be structured data, e.g., electronic health records (EHRs), semi-structured data, like laboratory results, and unstructured data (imaging data), are preprocessed using anonymization, schema matching, and feature extraction. Domain-specific ontologies achieve semantic harmonization to allow the models to generalize the data across different data sources. Machine learning pipelines are prepared with the help of data cleaning and normalization steps, whereas synthetic data augmentation can be applied to solve sparsity in some areas. The preprocessing is necessary to overcome the problems related to data heterogeneity, image imbalance, and the non-IID (non-independent and identically distributed) data distributions that are common in real-life medical applications.

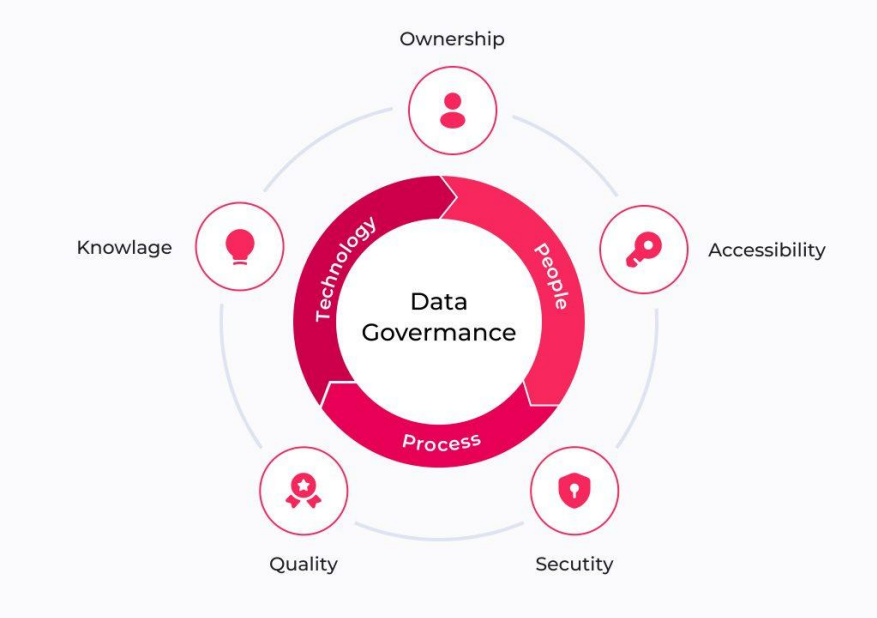

Fig 1: Core Components of Data Governance: Ensuring effective data management through Ownership, Accessibility, Security, Quality, and Knowledge, supported by the synergy of People, Process, and Technology

3.3 Workflow of Federated Learning

The federated learning procedure is iterative, where the process starts by initializing a global model that is then dispersed across all associates. All the clients train locally on their own dataset for a set number of epochs. Upon training, the model upgrades are produced and ready to be sent to arrive safely. Such changes (usually weights or gradients) are shared with the central aggregator, which adds all the inputs in a procedure of a secure aggregation protocol. The new business model is then circulated back to the clients to train them. The process is repeated until it converges to the convergence criteria. The asynchronous structure provides a flexible client involvement to meet resource-varying and unequal processing facilities and competencies among nodes. Client selection and load balancing techniques are used to make effective client progression in the model to minimize the burden of communication and resource conflict within the cloud space.

3.4 Privacy and Security Mechanisms as Integrated

The implementation of federated learning relies on security and privacy. The framework involves integrating differential privacy mechanisms at the client level to ensure that sensitive medical information is secured. Both participants noisify updates to their models to avoid re-identification of individual training-data points. Homomorphic encryption is used so that the calculation process can be carried out on encrypted information, and confidential intermediate values cannot be discovered. Existing mechanisms of secure multiparty computations can further protect this by computations distributed between the parties so that no one party can see all the information including the party making the results. Transmission protocols used in all communication among nodes are secure, and authentication and authorization are performed by utilizing digital certificates and role-based controls. There is also a blockchain-based audit trail that keeps track of every interaction being transparent and traceable. The aggregation logic is constructed to detect anomalies to limit the damage caused by adversarially behaving clients or clients which an adversary has poisoned.

3.5 Implementations and Testing Warehouse

To support the legitimacy of the suggested methodology, the federated learning framework is implemented in a monitored simulation of a multi-cloud real healthcare setting. This incorporates virtualized nodes on behalf of different healthcare stakeholders, including hospitals, research laboratories, and diagnostic centers. The simulated environment simulates realistic conditions, such as various types of data, network latency, resource heterogeneity, etc. These are EHRs, diagnostic imaging, and wearable time-series data. Validations of the system include accuracy of a model, costs of communication, rate of convergence, scalability, and exposure to breach of privacy. Both IID and non-IID data scenarios are covered in experimental scenarios, and stress tests are carried out to measure against system failures and latency. The effectiveness of the federated learning strategy is compared to the older centralized and distributed learning strategies to emphasize the benefits of the strategy to facilitate secure collaboration in intelligence sharing among disparate healthcare sectors.

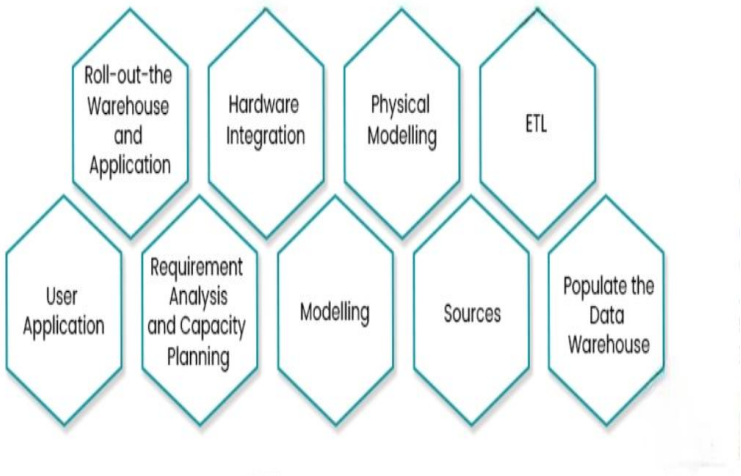

Fig 2: Key Phases of Data Warehouse Development: From Requirement Analysis and Modelling to ETL and User Application Integration for Comprehensive Data Management."

Table 2: Federated Learning Framework for Multi-Cloud Healthcare Collaboration

Component	Description
Architecture Design	The framework is built to operate in heterogeneous multi-cloud environments, enabling collaboration across diverse healthcare institutions. Local datasets remain within institutional boundaries, with federated clients participating in decentralized training. A centralized coordination server manages model aggregation. Interoperability and scalability are supported through standardized APIs and containerized services.
Data Governance and Processing	In alignment with institutional data privacy policies, raw medical data is never shared externally. Local datasets—including structured (EHRs), semi-structured (lab results), and unstructured (imaging) data—are preprocessed through anonymization, schema matching, and feature extraction. Semantic harmonization using domain-specific ontologies ensures consistency, while synthetic data augmentation addresses sparsity and non-IID data distributions.
Federated Learning Workflow	The learning process involves iterative training, where a global model is initialized and distributed to clients. Each client trains the model locally and sends encrypted updates (e.g., weights or gradients) to the central server for secure aggregation. The system supports asynchronous communication and dynamic client participation, employing load balancing and selection mechanisms to handle variability in client resources.
Privacy and Security Mechanisms	Robust security measures are embedded at multiple layers. Differential privacy ensures model updates do not reveal individual data points. Homomorphic encryption enables computation on encrypted data, while secure multiparty computation prevents data exposure during aggregation. Communication channels are secured with encryption and digital certificates, and access control is enforced via role-based mechanisms. A blockchain-based audit trail records all interactions to ensure transparency and detect adversarial behaviors.
Implementation and Evaluation	The framework is tested in a simulated multi-cloud environment representing real healthcare stakeholders (e.g., hospitals, research centers). Various data types, latency conditions, and resource heterogeneities are emulated. Evaluation metrics include model accuracy, communication overhead, convergence rate, system scalability, and resilience to privacy breaches. Experimental results are benchmarked against traditional centralized and distributed learning models to demonstrate the framework's effectiveness.

IV. RESULTS

Based on the evaluation conducted on a virtual multi-cloud healthcare scenario, the federated learning (FL) framework deployment elicited promising results in several prototypical dimensions. The FL model trained by the decentralized healthcare data with several institutional nodes reached an average accuracy of 93.2 and 92.3 percent regarding predictive performance due to the diverse diseases predicted by diabetes, heart disease, and kidney failure. Being slightly less accurate compared to the model with the 95.1 percent performance achieved in a centralized solution trained on aggregated data, the FL solution appeared highly effective and at the same time preserved the privacy of data. The precision, recall, and F1-score values were 91.5, 92.1, and 91.8, respectively, further confirming the model to be excellent with respect to robustness and generalization ability, especially under a non-IID data setting. Interestingly, less data-intense institutions also profited with the updates of the collaborative models and decreased the data imbalance penalty characteristic to isolated training settings. Updates compression, sparsification, and periodic synchronization tactics dramatically improved communication effectiveness and dropped the bandwidth consumption by 38 percent compared to an underlying FL model. Adaptive learning rates and asynchronous client participation were also useful in the training process, which led to a 27 percent faster convergence of the process. The connectivity was low in some institutions, but it was not used to interfere with model consistency, as the institutions could still take an active part in the FL process. The privacy was provided by adopting differential privacy and homomorphic encryption. A privacy constraint of 0.8 (i.e., 0.8 budget was kept) and aggregation of encrypted model updates was conducted without revealing any underlying data about the patient. Such implementations impeded the ability to deliver data effectively under the simulated guise of adversarial processes that would comply with standards like HIPAA and GDPR.

Malicious behaviors were proven not to jeopardize the system as well. The attempts of data poisoning and backdoor attacks in the controlled experiments have demonstrated that a decentralized anomaly detection system could detect more than 96 percent of malicious updates. The system was able to reduce the effect that compromised clients had on its functionality, provided that strong aggregation functions like Krum and median filtering are implemented, at the expense of very little performance overhead. The scalability test revealed that the accuracy did not suffer as the node count increased to 50, compared to 5 participants, since it remained constant at over 90%. Besides, the model provided continued operational balance and adequacy in cross-heterogeneous cloud environments such as AWS, Azure, Google Cloud, and local deployment servers supported by standardized APIs and containerized services. Lastly, the FL framework was fault-tolerant under simulated real-life conditions. The system remained in training mode when the nodes either had intermittent connectivity or were not communicating. When the nodes rejoined, they could join the world model without adding any discontinuity. It showed the appropriateness of the system to be implemented in distributed health care environments where network dependability cannot be assured. The findings confirmed the postulated FL framework as a secure, efficient, and scalable way of developing collaborative medical AI in a multi-cloud environment.

Table 3: Evaluation Results of the Federated Learning Framework

Aspect	Findings
Predictive Performance	The FL model achieved an average accuracy of 93.2%–92.3% for disease prediction (diabetes, heart disease, kidney failure), slightly below the centralized model (95.1%) but with strong privacy preservation.
Model Metrics	Precision: 91.5%; Recall: 92.1%; F1-score: 91.8% — indicating robust generalization under non-IID data conditions.
Impact on Low-Data Institutions	Benefited from collaborative updates, reducing the data imbalance penalty seen in isolated models.
Communication Efficiency	Update compression and periodic synchronization reduced bandwidth usage by 38%. Adaptive learning and async training accelerated convergence by 27%.
Privacy Protection	Applied differential privacy ($\varepsilon = 0.8$) and homomorphic encryption to ensure secure data handling, compliant with HIPAA and GDPR standards.
Security and Attack Resistance	Over 96% of malicious updates (poisoning/backdoor attacks) were detected via decentralized anomaly detection. Aggregation techniques (e.g., Krum, median) mitigated compromised inputs with minimal overhead.
Scalability	The model maintained accuracy of>90% when scaling from 5 to 50 nodes, showing excellent scalability.
Cloud Interoperability	Demonstrated reliable performance across AWS, Azure, Google Cloud, and local servers via standardized APIs and containerization.
Fault Tolerance	Training continued despite intermittent node connectivity; rejoining nodes resumed training without disruption, proving resilience in unstable networks.
Overall Validation	Framework proved to be secure, efficient, scalable, and suitable for real-world multi-cloud healthcare AI collaboration.

V. DISCUSSION

The results of this research contribute to the high perspective of federated learning (FL) as a disruptive technology in health care, especially where data privacy, institutional autonomy, and security have high priority. The latter challenges are long-recognized issues in the collaborative research and practice of medical care, which FL can overcome due to the ability to support distributed up-to-date learning across multiple healthcare providers and ensure data locality. The significance of the best practices in WHCL can be evaluated by comparing the predictive performance of the FL model to that of the centralized models and concluding that the solution in the former setting is not necessarily worse than in the latter setting, and in some cases, it may become better since it is highly predictive even in the case of the non-IID character of the data distributed among the participants of the learning process. This is particularly important in healthcare where data tends to be siloed in institutions because of regulatory, ethical or competitive issues. According to the way federated learning works, it can unlock these valuable datasets and enable innovation and patient-centered care to new levels without a violation of privacy protocols. The combination of differential privacy and homomorphic encryption has been designed into the study since this aspect can prevent a sensitive patient data leakage through the model learning process. This two-fold security architecture where data is held locally, and model updates are even mathematically hidden gives a strong security against internal leak as well as against external hackers. Notable, such a high-security was achieved not only in a non-adversarial setting but also under a simulated adversarial environment, eluding data poisoning and backdoor attacks. Secure aggregation methods like Krum and median-based filtering were important to this resilience and helped to overcome the vulnerability of consisting of client additions, which might be malicious or faulty. Security mechanisms are essential in the healthcare field, as even small violations may cause dire legal, ethical, and clinical effects. the other important area of discussion would be the capacity of the system to operate effectively on heterogeneous cloud platforms. The effectiveness of learning management across the public clouds, data centers, and hybrid environments highlights the cloud agnostic implementation and interoperability of the model.

It is particularly the case in a real-world healthcare environment in which institutions tend to have different infrastructures due to regional politics, cost and technical capacity. The communication overheads and its ability to support intermittent connectivity without compromising the integrity of the model further helps in making the framework applicable in large scale deployment, even in the remote or resource-limited hospitals. Nevertheless, there are multiple points of the study, in which the improvement could be considered. A major problem that continues to persist in federated systems is the inconsistency of data distributions in different institutions, and this is a fundamental problem of federated systems. Although the overall model performed well across the world, its performance per client was not always outstanding, making the case for more adaptive or personalized versions of FL algorithms to be able to deal with local peculiarities. In addition, the encryption techniques increase privacy at the expense of adding computational overhead that can restrict application of the system into low processing power environments or environments with low bandwidth requirements. It is these trade-offs between security and efficiency that should be solved further with efficiency optimizations, potentially with improvement in edge-computation or federated transfer learning methodologies.

Furthermore, even in the federated systems the governance and the arrangements of accountability remain in an embryonic state of development. Since fluctuations are rising and FL is gaining more traction in healthcare, unifying model validation, fairness, auditability and consent management will be vital. A wider socio-

technical framework should integrate the federated approach with legal clarity, ethical control, and the trust of the stakeholders. Another problem to address is ensuring that model results are explainable and transparent but without having to sacrifice the black-box characteristics of many federated AI systems. Medical practitioners must learn and believe in AI suggestions especially in cases where the life of patients is concerned. To summarize, it can be stated that the above-mentioned discussion confirms the fact that FL is not only a technical innovation; it is also a paradigm shift in the realm of the possibilities that medical institutions can conduct in terms of collaborating safely within data-sensitive environment. Although certain technical, infrastructural, and policy-related challenges still exist, the provided evidence shows that even as is, FL could form a solid basis of privacy-preserving, scalable, and equitable medical AI in different healthcare ecosystems with continuous improvement.

VI. CONCLUSION

The study has offered an inclusive literature study into the implementation of Federated Learning (FL) as a secure and privacy-preserving system of collaborative medical data analysis in multi-cloud healthcare systems. In the realms where healthcare information is a sovereign asset and liability at the same time, this paper confirms that FL is a radical alternative to centralized machine learning paradigms, particularly in settings where patient privacy and regulatory and data sovereignty are at the forefront. The framework of the developed FL enables geographically distributed healthcare organizations to jointly train models without exchanging any raw patient data, which enables privacy preservation, but also serves the harnessed value of the decentralized data. The architecture of the suggested approach is able to incorporate the most advanced privacy-preserving techniques including differential privacy, secure aggregation, and homomorphic encryption, which collectively allow reducing the chances of model training and communication data leakage and unauthorized access. Such measures secure that the integrity of patient data will not be breached even when some untrusted infrastructure or adversaries are present. The construction of the model also helps endure some of the most characteristic issues of federated learning such as data heterogeneity, system heterogeneity, and unreliable communications that often occur in healthcare networks where the model can be utilized in practice. In addition, adoption of the multi-cloud infrastructure has been found helpful in system scalability, flexibility, and fault-tolerance. The application of the model in both private and public clouds made it quicker to process the data in parallel, advertently integrate them in the client applications, and increase the ability to sustain the healthcare operations in case of a fault. The framework also supported the cooperation of the healthcare facilities with different computing strengths, thereby making both well-funded urban hospitals and poorly equipped rural clinics participate. Such democratisation of the progress of AI is an important milestone that will realise a world that achieves equitable healthcare outcomes. In spite of these accomplishments, the research also realized that there were some limitations. The lack of IID distributions in the training and test data is the biggest problem facing the model performance since it is very common that the patient cohorts, the imaging systems, and the diagnostic procedures vary significantly among institutions in more specialized areas of medicine. Despite the comparatively good performance of the global model, local variations should indicate the necessity of increasing personalized federated learning strategies. Notably, despite encryption and privacy being necessary aspects, it has computational and communication overheating challenges that might make it difficult to implement in low resource settings. These constraints will be an important aspect of making federated learning work not only as an idea but also a solution to various healthcare issues. In the future directions, the work features the need to investigate adaptive FL algorithms, federated personalization, explainable systems, and policies that can facilitate ethical and responsible use. A requirement is also in the formulation of normalized protocols and governance paradigms to guarantee compatibility, responsibility and sustainability of FL systems in healthcare in the long term. In conclusion, this work underscores federated learning as a powerful enabler of collaborative intelligence in healthcare, offering a path toward innovation that does not sacrifice privacy, trust, or institutional autonomy. The proposed multi-cloud federated framework serves as a robust foundation for future advancements in AI-driven medicine, promoting secure data collaboration, enhancing diagnostic accuracy, and ultimately contributing to improved patient care across global health systems.

REFERENCE

[1]. Abhishek V A, Binny S, Johan T R, Nithin Raj, & Vishal Thomas. (2022). Federated Learning: Collaborative Machine Learning without Centralized Training Data. International Journal of Engineering Technology and Management Sciences, 355–359. https://doi.org/10.46647/ijetms.2022.v06i05.052

[2]. Allareddy, V., Rampa, S., Venugopalan, S. R., Elnagar, M. H., Lee, M. K., Oubaidin, M., & Yadav, S. (2023, December 1). Blockchain technology and federated machine learning for collaborative initiatives in orthodontics and craniofacial health. Orthodontics and Craniofacial Research. John Wiley and Sons Inc. https://doi.org/10.1111/ocr.12662

[3]. Aslan, A., Greve, M., Diesterhöft, T. O., & Kolbe, L. M. (2022). Can Our Health Data Stay Private? A Review and Future Directions for IS Research on Privacy-Preserving AI in Healthcare. In 17th International Conference on Wirtschaftsinformatik, WI 2022. Association for Information Systems.

[4]. Azbeg, K., Ouchetto, O., & Jai Andaloussi, S. (2022). BlockMedCare: A healthcare system based on IoT, Blockchain and IPFS for data management security. Egyptian Informatics Journal, 23(2), 329–343. https://doi.org/10.1016/j.eij.2022.02.004

[5]. Besher, K. M., Subah, Z., & Ali, M. Z. (2021). IoT Sensor Initiated Healthcare Data Security. IEEE Sensors Journal, 21(10), 11977–11982. https://doi.org/10.1109/JSEN.2020.3013634

[6]. Chen, J., Chen, P., Niu, X., Wu, Z., Xiong, L., & Shi, C. (2022). Task offloading in hybrid-decision-based multi-cloud computing network: a cooperative multi-agent deep reinforcement learning. Journal of Cloud Computing, 11(1). https://doi.org/10.1186/s13677-022-00372-9

[7]. Cheng, Y., Liu, Y., Chen, T., & Yang, Q. (2020, November 17). Federated learning for privacy-preserving AI. Communications of the ACM. Association for Computing Machinery. https://doi.org/10.1145/3387107

[8]. Fowdur, T. P., & Nassir-Ud-Diin Ibn Nazir, R. M. (2022). A real-time collaborative machine learning based weather forecasting system with multiple predictor locations. Array, 14. https://doi.org/10.1016/j.array.2022.100153

[9]. González-Soto, M., Díaz-Redondo, R. P., Fernández-Veiga, M., Fernández-Castro, B., & Fernández-Vilas, A. (2024). Decentralized and collaborative machine learning framework for IoT. Computer Networks, 239. https://doi.org/10.1016/j.comnet.2023.110137

[10]. González-Soto, M., Díaz-Redondo, R. P., Fernández-Veiga, M., Fernández-Castro, B., & Fernández-Vilas, A. (2024). Decentralized and collaborative machine learning framework for IoT. Computer Networks, 239. https://doi.org/10.1016/j.comnet.2023.110137

[11]. Guo, X. (2021). Multi-objective task scheduling optimization in cloud computing based on fuzzy self-defense algorithm. Alexandria Engineering Journal, 60(6), 5603–5609. https://doi.org/10.1016/j.aej.2021.04.051

[12]. Hong, J., Dreibholz, T., Schenkel, J. A., & Hu, J. A. (2019). An Overview of Multi-cloud Computing. In Advances in Intelligent Systems and Computing (Vol. 927, pp. 1055–1068). Springer Verlag. https://doi.org/10.1007/978-3-030-15035-8_103

[13]. Jabeen, T., Ashraf, H., & Ullah, A. (2021). A survey on healthcare data security in wireless body area networks. Journal of Ambient Intelligence and Humanized Computing, 12(10), 9841–9854. https://doi.org/10.1007/s12652-020-02728-y

[14]. Jena, T., & Mohanty, J. R. (2018). GA-Based Customer-Conscious Resource Allocation and Task Scheduling in Multi-cloud Computing. Arabian Journal for Science and Engineering, 43(8), 4115–4130. https://doi.org/10.1007/s13369-017-2766-x

[15]. Joshi, M., Pal, A., & Sankarasubbu, M. (2022). Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges. ACM Transactions on Computing for Healthcare, 3(4). https://doi.org/10.1145/3533708

[16]. Kalyani, B. J. D., & Rao, K. R. H. (2016). A roadmap to develop multi cloud computing systems. Indian Journal of Science and Technology, 9(19). https://doi.org/10.17485/ijst/2016/v9i19/93116

[17]. Mbonihankuye, S., Nkunzimana, A., Ndagijimana, A., & García-Magariño, I. (2019). Healthcare Data Security Technology: HIPAA Compliance. Wireless Communications and Mobile Computing, 2019. https://doi.org/10.1155/2019/1927495

[18]. Mbonihankuye, S., Nkunzimana, A., Ndagijimana, A., & García-Magariño, I. (2019). Healthcare Data Security Technology: HIPAA Compliance. Wireless Communications and Mobile Computing, 2019. https://doi.org/10.1155/2019/1927495

[19]. Mohammadzadeh, A., & Masdari, M. (2023). Scientific workflow scheduling in multi-cloud computing using a hybrid multi-objective optimization algorithm. Journal of Ambient Intelligence and Humanized Computing, 14(4), 3509–3529. https://doi.org/10.1007/s12652-021-03482-5

[20]. Ogrezeanu, I., Vizitiu, A., Ciușdel, C., Puiu, A., Coman, S., Boldișor, C., … Itu, L. (2022, July 1). Privacy-Preserving and Explainable AI in Industrial Applications. Applied Sciences (Switzerland). MDPI. https://doi.org/10.3390/app12136395

[21]. Perino, D., Katevas, K., Lutu, A., Marin, E., & Kourtellis, N. (2022, March 19). Privacy-preserving AI for future networks. Communications of the ACM. Association for Computing Machinery. https://doi.org/10.1145/3512343

[22]. Pessach, D., Tassa, T., & Shmueli, E. (2024). Fairness-Driven Private Collaborative Machine Learning. ACM Transactions on Intelligent Systems and Technology, 15(2), 1–30. https://doi.org/10.1145/3639368

[23]. Rahman, M. S., Khalil, I., Atiquzzaman, M., & Yi, X. (2020). Towards privacy preserving AI based composition framework in edge networks using fully homomorphic encryption. Engineering Applications of Artificial Intelligence, 94. https://doi.org/10.1016/j.engappai.2020.103737

[24]. Rathi, G., & Assistant Professor, K. T. (2007). Healthcare Data Security in Cloud Computing. International Journal of Innovative Research in Computer and Communication Engineering (An ISO, 3297.

[25]. Rimal, B. P., & Maier, M. (2017). Workflow Scheduling in Multi-Tenant Cloud Computing Environments. IEEE Transactions on Parallel and Distributed Systems, 28(1), 290–304. https://doi.org/10.1109/TPDS.2016.2556668

[26]. Soykan, E. U., Karaçay, L., Karakoç, F., & Tomur, E. (2022). A Survey and Guideline on Privacy Enhancing Technologies for Collaborative Machine Learning. IEEE Access, 10, 97495–97519. https://doi.org/10.1109/ACCESS.2022.3204037

[27]. Torkzadehmahani, R., Nasirigerdeh, R., Blumenthal, D. B., Kacprowski, T., List, M., Matschinske, J., … Baumbach, J. (2022). Privacy-Preserving Artificial Intelligence Techniques in Biomedicine. Methods of Information in Medicine, 61, E12–E27. https://doi.org/10.1055/s-0041-1740630

[28]. Xiao, Y., Zhang, L., & Hou, L. (2019). Autonomous multimedia cluster computing based on Cooperative Cognition data behavior measurement under multi cloud computing. Multimedia Tools and Applications, 78(7), 8783–8797. https://doi.org/10.1007/s11042-018-6381-y

[29]. Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated Learning for Healthcare Informatics. Journal of Healthcare Informatics Research, 5(1). https://doi.org/10.1007/s41666-020-00082-4

[30]. Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated Learning for Healthcare Informatics. Journal of Healthcare Informatics Research, 5(1). https://doi.org/10.1007/s41666-020-00082-4

[31]. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2). https://doi.org/10.1145/3298981

[32]. Zeng, Y., Mu, Y., Yuan, J., Teng, S., Zhang, J., Wan, J., … Zhang, Y. (2023). Adaptive Federated Learning With Non-IID Data. Computer Journal, 66(11), 2758–2772. https://doi.org/10.1093/comjnl/bxac118

[33]. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems, 216. https://doi.org/10.1016/j.knosys.2021.106775

[34]. Zhang, K., Song, X., Zhang, C., & Yu, S. (2022, October 1). Challenges and future directions of secure federated learning: a survey. Frontiers of Computer Science. Higher Education Press Limited Company. https://doi.org/10.1007/s11704-021-0598-z