Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Distributed Artificial Intelligence Framework with Federated Machine Learning for Privacy-Preserving Healthcare Data Analytics in Multi-Cloud Environments

Rohit Ravindra Nikam[1], Pallavi Sachin Patil[2], Dr.Pavan kumar[3], Dr. Deepak Kumar Parhi[4], Samudrala Jagadeesh[5], Dr.Geetika M. Patel[6], Shalini E[7], Anitha K[8]

[1]Department of Information Technology, Sanjivani College of Engineering, Kopargaon. Email: nikamrohitit@sanjivani.org.in
[2]Department of Artificial Intelligence & Machine Learning, GenbaSopanraoMoze College of Engineering, Savitribai Phule Pune University, Balewadi, Pune-45. Maharashtra, India. Email: patilpallavi06@gmail.com
[3]Associate Professor , MSOPS, Maharishi University of Information Technology, Lucknow, Uttar Pradesh, India, Email: pavan.kumar@muit.in, Orcid Id: https://orcid.org/0009-0007-3351-703X
[4]Professor, Department of cardiology, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, Email: deepakkumarparhi@soa.ac.in, Orcid Id: 0009-0002-3005-4387
[5]Assistant Professor, Department of ECE, Aditya University, Surampalem, Kakinada, Andhra Pradesh, Email: samudrala.jagadeesh@adityauniversity.in
[6]Associate Professor, Department of Community Medicine, Parul University, PO Limda, Tal. Waghodia, District Vadodara, Gujarat, India, Email: vicepresident_86@paruluniversity.ac.in ,Orcid Id- 0000-0003-3789-184X
[7]Computer Science, Assistant Professor, Meenakshi College of Arts and Scien+H2:H11ce, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: shalini@maher.ac.in
[8]Department of Management Studies, Associate Professor,Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: anithak@maher.ac.in

Abstract

The quick development of online health systems and smart healthcare applications has produced huge amounts of sensitive patient data that must be analyzed using secure and scalable analytics and privacy-protective frameworks. The old centralized artificial intelligence models are usually based on transfer of healthcare data to one cloud server where data privacy and security vulnerabilities, regulatory compliance, and single point of failure issues are problematic. The research aims at overcoming these drawbacks by proposing a Distributed Artificial Intelligence Framework that is combined with Federated Machine Learning to conduct Privacy-Preserving Healthcare Data Analytics in Multi-Cloud Environments. The suggested model will use a decentralized collaborative learning algorithm, called Federated Averaging (FedAvg) which allows many healthcare institutions to cooperate without exchanging raw patient records. Patient confidentiality and minimizing privacy exposure are ensured by the local healthcare datasets being trained at distributed cloud nodes, with only model parameters being communicated to a federated server and aggregated, and no information shared. The multi-cloud deployment architecture helps boost the scalability, computation effectiveness, reliability, and distributed resource management of heterogeneous healthcare infrastructure. The use of healthcare analytical datasets under distributed federated settings and the Accuracy metric was used in assessing system performance to conduct experimental evaluation. The results retrieved prove that the suggested FedAvg-based distributed scheme is characterized by a high predictive accuracy and, on the other hand, ensures data privacy and preserves central dependence of data. In addition, the framework enhances teamwork healthcare intelligence, offers secure distributed medical analytics and provides a scalable solution to next-generation intelligent healthcare systems. The paper underscores the real-world relevance of federated distributed AI towards facilitating secure, efficient, and privacy conscious healthcare analytics in the modern multi-cloud computing environment.

Keywords: Distributed Artificial Intelligence; Federated Learning; FedAvg; Healthcare Analytics; Multi-Cloud Computing; Privacy Preservation; Accuracy Evaluation.

This is an open access article under CC BY 4.0, allowing unrestricted use with proper attribution, a license link, and indication of any changes made.

1. Introduction

Due to the blistering development of digital healthcare technologies, Internet of Medical Things (IoMT) devices, electronic health records, wearable biosensors, and cloud-based medical infrastructures, healthcare big data

has been growing exponentially. In the modern healthcare set up, huge amounts of heterogeneous clinical, imaging, genomic and patient-monitoring information are produced that demand smarter computational frameworks to analyze and provide effective decisions. The methods of artificial intelligence (AI) and machine learning have been highly applied to healthcare analytics to make predictions of diseases, assist clinical decisions, interpret medical images, and provide personalized recommendations (Rieke et al., 2020). Nevertheless, the traditional AI-based healthcare systems are usually centralized data gathering and cloud-based training models, in which all patient information of various hospitals and healthcare facilities are merged in a single server to train a model. Despite the benefits that centralized AI models are easier to compute with, they pose serious problems of privacy exposure, data leakage threats, regulatory compliance, and the susceptibility to cyberattacks (Kaissis et al., 2020).

Medical data are extremely sensitive and are under stringent legal and ethical restrictions such as data protection laws that limit direct exchange of patient data among institutions. Most healthcare providers and hospitals are unwilling to share raw medical datasets, fearing that it would affect confidentiality, ownership, interoperability and compliance with security. These limitations pose a severe constraint to centralized healthcare AI systems, especially when the collaborative learning needs to be done on a large scale when geographically dispersed medical centers have to be involved (Xu et al., 2021). Moreover, single-point failures, excessive communication overhead, bottlenecks in scaling and even computation inefficiencies are common to centralized cloud architectures in a large distributed healthcare system. With the ongoing transformation of healthcare systems into distributed intelligent systems, the need to be decentralized and akin to collaborative analytics at the same time, and maintain patient privacy and secure communication processes is growing (Yang et al., 2019).

Federated learning has become one of the new promising distributed machine learning paradigms allowing collaborative model training without explicitly sharing raw data between participating institutions. Federated learning enables local healthcare nodes to independently train AI models and stores model parameters on a central server to be aggregated globally, as opposed to sending patient records to centralized server (McMahan et al., 2017). Federated Averaging (FedAvg) algorithm has gained popularity as one of the most popular aggregation algorithms because it is efficient in communication and can scale to distributed learning setting. With federated intelligence and distributed cloud infrastructures, healthcare organizations can work together to enhance the performance of predictive models without exposing data to privacy risks, and keep it confidential (Kairouz& McMahan, 2021). Moreover, in terms of large-scale healthcare analytics systems, multi-cloud computing environments have a high scalability and resource optimization, reliability, and distributed workloads, which is why they are highly applicable in heterogeneous medical network settings (Antunes et al., 2022).

Recent research yielded evidence of the practical importance of federated learning in healthcare scenarios like medical imaging, COVID-19 prediction, cancer detection, and smart healthcare systems (Dayan et al., 2021; Pati et al., 2022). Privacy-preserving methods such as secure aggregation and differential privacy have also enhanced the security performance of federated healthcare systems, that is, safeguard sensitive patient data when communicating and training in a distributed manner (Bonawitz et al., 2017; Geyer et al., 2017). In spite of the mentioned advances, the current healthcare federated learning models continue to have issues with distributed coordination, communication overhead, heterogeneous healthcare data, and efficient deployment through multi-cloud infrastructures. Thus, it is still necessary to have scalable and privacy conscious distributed artificial intelligence systems that can be used to provide secure healthcare analytics in decentralized systems.

This study is inspired by these issues, and it puts forward a Distributed Artificial Intelligence Framework as an extension of Federated Machine Learning to Privacy-Preserving Healthcare Data Analytics in Multi-Cloud Environments. The framework proposed uses the FedAvg aggregation mechanism to facilitate decentralized collaborative learning between more than two healthcare institutions but has privacy of patient data. The framework handles the secure distributed analytics of healthcare by training local models, federating the aggregation of parameters, and deploying with multi-clouds to improve the scaling and reliability. The proposed system is also aimed at enhancing the accuracy of healthcare predictions due to the collaborative

federated intelligence and minimizing the dependency of the centralized data and communication vulnerability.

The most important contributions of the research are the creation of a new distributed artificial intelligence structure to secure healthcare analytics, the implementation of a FedAvg-based Federated learning mechanism into a scalable multi-cloud infrastructure and the implementation of privacy-aware distributed model aggregation of collaborative medical learning. Moreover, the suggested framework offers an accuracy-based performance measurement plan to investigate the predictive performance in distributed healthcare settings. The work also plays a role in the development of privacy-conscious healthcare AI systems, which integrates distributed intelligence, federated analytics, multi-cloud computational scalability into a single healthcare learning system. The rest of the paper will be structured as follows: Section 2 shows the literature review and related literature on federated healthcare learning, Section 3 outlines the proposed distributed AI framework and system architecture, Section 4 discusses the methodology, and the experimental setup, Section 5 explains the mathematical model and evaluation metrics, Section 6 presents the analysis of the experimental results and performance evaluation, and Section 7 wraps up the paper with future research directions.

2. Literature Review

The fast development of artificial intelligence technologies has fundamentally changed the current healthcare systems, making it possible to perform intelligent diagnostics, predictive analytics, automated clinical decision-making, and its personalized treatment planning. One emerging paradigm in handling large scale healthcare data produced by hospitals, wearable sensors, medical imaging systems, and electronic health records is distributed artificial intelligence. In contrast to more traditional, centralized healthcare systems, distributed AI systems allow the collaborative computation and decentralized decision-making of various healthcare nodes and enhance the efficiency in the computational aspects of the system as well as lessen the reliance of systems on one centralized infrastructure (Yang et al., 2019). Smart healthcare with distributed AI has proven to be of significant help in real-time patient tracking, diagnosis, and streamlining clinical procedures. Moreover, geographically distributed medical settings have been shown to have better fault tolerance, scalability, and resource usage with distributed decision-making mechanisms and could therefore be leveraged by next-generation healthcare analytics platforms (Rieke et al., 2020).

The recent interest in federated learning as a distributed machine learning method has prompted the notion that it can be used to facilitate collaborative healthcare analytics, without necessarily sharing sensitive patient data. The fundamental idea of federated learning is to primarily learned model parameters onto a central aggregation server, which is located in one place (McMahan et al., 2017). This decentralized learning approach reduces the risk of privacy, but also lets healthcare organizations work together to increase the effectiveness of predictive models. A few papers have investigated the use of federated learning to medical imaging, disease diagnosis, smart-healthcare monitoring, and medical informatics systems (Xu et al., 2021). Dayan et al. (2021) showed that federated learning is effective to predict the clinical outcome of COVID-19 patients in several healthcare centers, and that decentralized collaborative learning could be an effective approach in medical applications on a large scale. Likewise, Pati et al. (2022) applied federated learning to the problem of detecting boundary of rare cancer, proving that distributed learning in healthcare can also be used to analyze and obtain high accuracy analytically, without compromising the confidentiality of the data. These papers affirm that federated healthcare analytics, without contravening institutional privacy laws, can make a significant contribution to collaborative medical intelligence.

Federated Averaging (FedAvg) algorithm is one of the most popular federated optimization methods that are used to select an aggregation mechanism of distributed learning systems. FedAvg allows effective model generation by averaging weights of models locally trained on each client who participates in the process based on his/her contribution to the datasets (McMahan et al., 2017). The algorithm minimizes communication overhead, increases scalability and training efficiency in decentralized settings. Kairouz and McMahan (2021) revealed the significance of FedAvg in distributed-scale federated systems and emphasized the ability to enable heterogeneous distributed settings. Li et al. (2020) also highlighted that FedAvg offers both good convergence properties of distributed machine learning systems and decreases the centralized dependence of computation.

FedAvg has become a standard aggregation algorithm in healthcare federated learning systems, thanks to its simplicity, scalability, and efficiency in communication. Multiple federated learning projects related to healthcare have implemented FedAvg to enhance distributed medical model training with the reduction of direct data transfer between hospitals and research centers (Nguyen et al., 2022).

The preservation of privacy is still considered to be one of the most urgent issues in the healthcare artificial intelligence systems, since patient information is highly sensitive. The centralized healthcare AIs are susceptible to privacy breach, cyberattacks, unauthorized access, and regulatory breach. Consequently, privacy preserving distributed learning architectures have gained more importance in the study of medical AI. Kaissis et al. (2020) emphasized that safe and federated machine learning methods can offer high privacy of medical imaging and health data analytics. The mechanism of secure aggregation suggested by Bonawitz et al. (2017) makes it possible to share the parameters encrypted during the federated learning process and avoid exposing local healthcare data during the entire process. The methods of differential privacy proposed by Geyer et al. (2017) can also be used to improve the level of safety as they inject controlled noise to the local model updates and decrease the probability of patient information leakages. Swarm learning was a recent proposal by Warnat-Herresthal et al. (2021) as a decentralized confidential machine learning framework that can help sustainable clinical AI systems with no centralized data storage. All these studies illustrate that a privacy-preserving federated learning can bring an efficient equilibrium between collaborative healthcare intelligence and patient data confidentiality.

The growing use of cloud computing in healthcare systems has resulted in the emergence of distributed multi-cloud healthcare infrastructures that have the capacity to provide scalable medical analytics. Interoperability, distributed storage, computational scalability, fault tolerance and resource optimization are some of the benefits of multi-cloud environments. Cloud-based infrastructures are becoming a dominant choice among healthcare organizations in the management of electronic health records, remote patient monitoring, telemedicine, and analytics based on AI. Nevertheless, centralized cloud systems are usually associated with being vulnerable to being dependent on the vendor, lack of scalability, and infrastructure bottlenecks. To overcome these challenges, multi-cloud computing architectures divide workloads and run them on various cloud resources, thus making them more reliable and resilient to disruptions in the system (Antunes et al., 2022). Decentralized healthcare intelligence with federated learning and multi-cloud infrastructures is a promising solution to enable the collaborative learning of distributed medical organizations whilst having operational scalability and computational efficiency. In addition, distributed cloud orchestration allows deploying secure and scalable AI in healthcare through heterogeneous healthcare ecosystems.

Despite the major gains that have been realized in federated healthcare learning, there are various notable gaps in research that remain in literature. Current literature concentrates on the applications of healthcare federated learning or privacy-preserving mechanisms as single entities, and few studies combine distributed artificial intelligence, federated learning, privacy preservation, or using multi-cloud healthcare deployment in a single platform. Most of the current healthcare federated learning systems do not have scalable multi cloud coordination system that can facilitate large distributed healthcare systems. Moreover, existing systems are frequently prone to overheads in communications, interoperability, and quite uneven data distribution issues that diminish training efficiency and predictive capabilities in practice within healthcare settings. Moreover, some of the studies emphasize the protection of privacy predominantly without the thorough consideration of distributed scalability and accuracy of healthcare predictions at the same time. Thus, the necessity to develop an integrated distributed artificial intelligence model with the integration of FedAvg-based federated learning, privacy-aware distributed model aggregation, and scalable multi-cloud deployment models to ensure secure, accurate, and efficient healthcare data analytics in decentralized medical setting is high.

3. Recommended DAI Framework

The proposed Distributed Artificial Intelligence Framework will make it possible to have secure, scalable, and privacy-aware healthcare data analytics in multi-cloud environments that are implemented based on the principles of federated machine learning. The architecture combines distributed healthcare facilities, local smart learning nodes, federated aggregation of parameters, and cloud-based collaborative infrastructure to a

single decentralized healthcare analytics platform. The framework will enhance the accuracy of health care predictions without exposing the patients to risk of loss of their information confidentiality, and also minimize the risks that may arise due to the centralization of medical data storage. The proposed system is depicted by various hospitals or healthcare customers, linked by distributed local AI nodes, and coordinated by a federated server that is provided by a scalable multi-cloud system as shown in Figure 1. The individual training of local machine teaching model is carried out at the respective participating healthcare organization, with the use of the respective local patient data stored at its own premises, as such that the training of the particular model persistently provides a data localization and reduces the privacy vulnerability. The federated server facilitates communication between distributed nodes and does global model aggregation without having access to raw healthcare datasets.

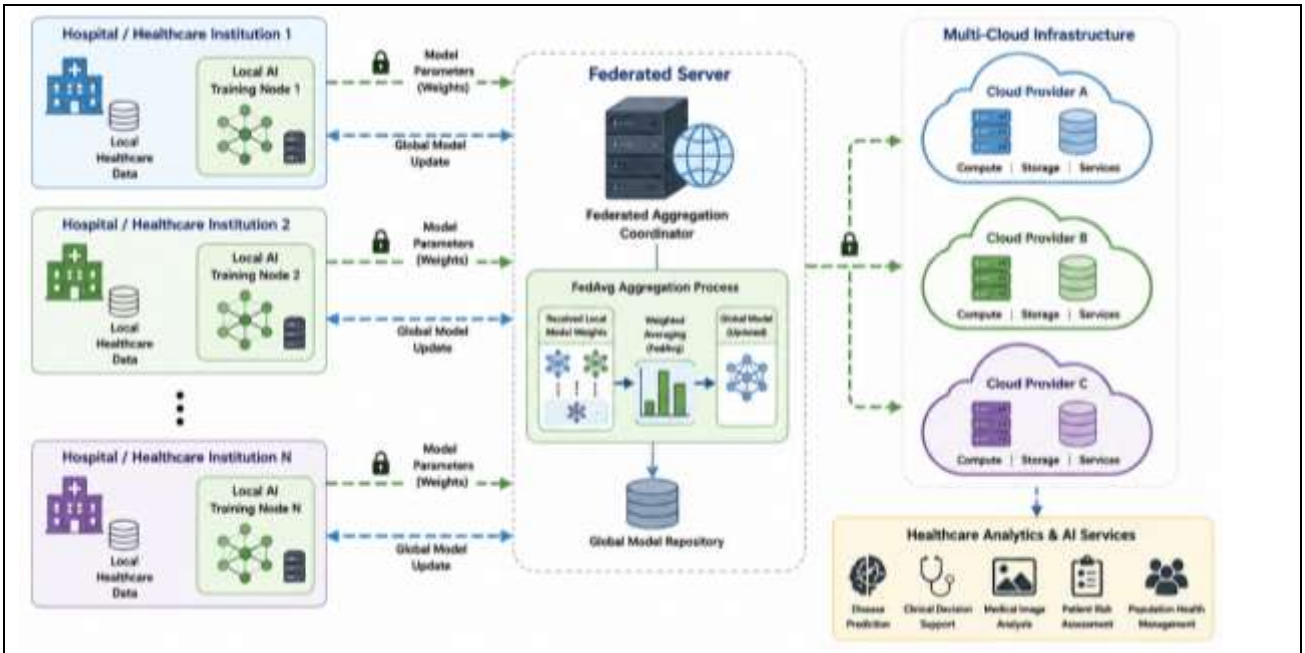

Fig 1. Proposed distributed federated learning architecture for privacy-preserving healthcare analytics in multi-cloud environments.

The healthcare data flow in the proposed framework is distributed, and works in the collaboration mode of learning using decentralization. First, healthcare information created by hospitals, diagnostic labs, wearable medical devices and electronic health records systems are stored on-site, in each of the participating institutions. Local AI nodes are used to process model training on institution-specific data instead of transferring sensitive patient records to a central repository in the cloud. After the local training, only model parameters or weight changes are sent to the federated aggregation server in a secure manner. The server combines the parameters received to come up with a global healthcare prediction model which is then sent back to the participating healthcare nodes to update the iterative learning process. Such a decentralized learning approach greatly minimizes the risks of data sharing, but it also allows sharing of medical intelligence across dispersed healthcare systems.

The federated learning process of the framework proposed has various sequential steps which help in distributed collaborative model optimization. A global model is first launched in the federated server and shared with all participating healthcare institutions. The model is trained locally on each local AI node using the local healthcare data with a specified number of local epochs. Once the local training is finished, the model weights are produced and sent to the federated server in a secure manner. The server then does global parameter aggregation with the Federated Averaging (FedAvg) algorithm to produce a new global model. The aggregated model is re-distressed to all the involved healthcare nodes and the learning cycle repeats itself until convergence is met. This decentralized workflow provides safe collaborative education by not using any direct exchange of patient data between institutions.

FedAvg aggregation mechanism is also a key component of the proposed distributed healthcare learning framework as it allows optimizing the global models in an efficient way, which saves communication. The federated server averages the model weights of the locally trained models to arrive at the global model parameters based on the size of the relative datasets donated by various healthcare clients. The global aggregation process is the FedAvg whereby mathematically it is expressed as:

$$W_{t+1} = \sum_{K=1}^{K} \frac{n_k}{n} w_k^t \_\_\_\_\_(1)$$

Wherew_k^t represents the local model weights of the kth healthcare client at communication round t, nk denotes the size of the local dataset associated with the kth customer and n is the number of training samples of all healthcare institutions that are involved. This aggregation plan enhances worldwide learning outcomes and reduces communication overhead and maintains decentralized healthcare data ownership.

One of the main design considerations of the proposed framework is the preservation of privacy. Since healthcare analytics must be safe, all patient information remain local and are located within their respective healthcare facilities during the training process. The framework ideally eradicates the direct raw patient data exchange between hospitals, which minimizes risks associated with privacy leakage, unauthorized access and non-conformance to regulations. Direct communication between the federated server and local healthcare nodes only takes place during communication rounds and only transfers encrypted model parameters. In addition, secure parameter transmission functions and distributed aggregation functions are also added to enhance the security of communication and protect the confidentiality in the process of collaborative model training. This local learning model also allows healthcare organizations to be a part of distributed AI analytics without compromising sensitive medical data.

The suggested architecture also introduces a multi-cloud implementation plan to enhance scalability, computing performance, and reliability of healthcare systems in distributed settings. There are many cloud nodes deployed on the geographically distributed infrastructures to offer decentralized healthcare facilities and smart AI services. Multi-cloud architecture allows the efficient distribution of workloads, fault tolerance, dynamic resource allocation, and distributed storage management. Load balancing solutions are incorporated to maximize the use of computational resources across cloud interfaces and reduce latency in the process of federation of communication. Moreover, the distributed cloud orchestration architecture will enable a scalable healthcare analytics framework, enabling new healthcare institutions and AI nodes to become a part of the federated ecosystem dynamically without impacting overall system performance. The offered framework will be an effective and powerful solution to the problem of ensuring the security of healthcare data analytics in decentralized medical scenarios due to the combination of federated learning, distributed AI intelligence, privacy-conscious communication, and scalable multi-cloud implementation.

4. Methodology

To test the proposed distributed artificial intelligence structure, various benchmark healthcare datasets were used to examine the efficacy of federated machine learning, in terms of privacy-assured healthcare analytics in multi-cloud settings. The methodology unites distributed data processing of health care, federated training, secure model aggregation, and evaluation of performance based on accuracy measures. The workflow of the experiment was created to recreate the scenario of collaborative healthcare learning of several distributed medical institutions without excluding the fact that sensitive patient information should be retained on a local base of individual healthcare nodes during the training process.

To confirm the framework proposed, four popular healthcare datasets were chosen such as the Heart Disease Dataset, Diabetes Dataset, Breast Cancer Dataset, and MIMIC-III clinical dataset. These datasets were selected due to the need to represent a variety of healthcare analytics applications (including disease prediction, clinical diagnosis, risk assessment, and patient monitoring). The data sets have structured healthcare variables such as demographics, physiological measures, diagnostic variables, and clinical observations. Table 1 provides an overview of the properties of the healthcare datasets on which the experimental analysis is conducted.

Table 1. Description of healthcare datasets used for federated learning experiments

Dataset	Number of Samples	Number of Features	Class Distribution	Healthcare Application
Heart Disease Dataset	303	14	165 Positive / 138 Negative	Cardiovascular Disease Prediction
Diabetes Dataset	768	8	268 Positive / 500 Negative	Diabetes Risk Prediction
Breast Cancer Dataset	569	30	357 Benign / 212 Malignant	Breast Cancer Classification
MIMIC-III Dataset	5,000	20	2,950 Normal / 2,050 Critical	Clinical Risk Assessment

The gathered healthcare data were preprocessed and then fed into federated model training through a number of operations. First of all, missing values in the datasets were detected and imputed through mean and median methods to enhance the consistency of the data and minimize the problem of incomplete records. Where necessary, categorical healthcare attributes were coded in label encoding techniques. Min-Max scaling was done to normalize the features to achieve equal distributions of features, as well as, to enhance model convergence in the course of training. The normalized procedure scaled the values of the healthcare features to a normalized range of 0 to 1 and hence minimized the bias of features in distributed learning. The datasets were preprocessed before splitting, 80:20 was used to split the datasets into training and testing subsets. The training data were also shared with various healthcare clients to model the dec-fed healthcare settings. A local group of patient records was given to each healthcare institution to train their independent models without sacrificing the localization of the data.

The federated training setup was geared towards distributed collaborative learning in healthcare among various simulated healthcare facilities. The federated environment included five distributed healthcare clients which are connected to centralized federated aggregation server deployed in a multi-cloud infrastructure. It had independent training of its local AI model on institution specific healthcare data by each client and models parameters were periodically sent to the federated server where model parameters were aggregated globally. Communication rounds were conducted 100 times to carry out the training process to maintain stable convergence of the global healthcare model. A batch size of 32 was chosen, to strike a balance between the computational efficiency and the model optimization performance, when training locally. The initial learning rate was set to 0.001 to facilitate stable changes in parameters and avoid the oscillatory nature of the convergence. Adam optimization algorithm has been employed because it is computationally efficient and it is an adaptive algorithm when used in the distributed deep learning settings. The global model was created using the FedAvg aggregation mechanism, calculating model parameters averages based on the size of the dataset that each healthcare client contributes to the model.

The artificial intelligence aspect of the proposed framework involved the use of several deep learning architectures to assess the performance of healthcare predictions when there is a federation. Structured healthcare data was classified by using Artificial Neural Networks (ANN) since ANN has the potential to effectively model nonlinear relationships in healthcare. Convolutional Neural Networks (CNN) were also integrated in generating features and healthcare pattern recognition problems that need hierarchical learning of representations. Moreover, the BiLSTM networks were used with the aim of capturing sequential dependencies, and temporal healthcare trends in clinical data. The layers are taken to be hidden, and they use Rectified Linear Unit (ReLU) activation to enhance the nonlinear learning capacity, and mitigate the vanishing gradient problem during training. The last output layer used a sigmoid activation function when performing binary healthcare classification tasks. The ANN + CNN + BiLSTM architectures allowed performing a thorough assessment of the distributed federated healthcare learning in various analytical conditions.

The experimental module was realized in Python programming language because it has a plethora of support with machine learning and distributed AI development. Federated model implementation, local healthcare training, and parameter aggregation were done using the TensorFlow and PyTorch deep learning frameworks. A multi-cloud distributed computing framework based on a number of virtual cloud nodes (symbolizing geographically distributed healthcare institutions) was used to simulate the federated healthcare environment. The experiments were implemented into a system that had Intel Xeon multi-core processors, NVIDIA GPU

acceleration and 32 GB RAM to facilitate large-scale federated healthcare training. Contactless communication was created between federated clients and the healthcare to imitate the privacy-protecting transmission of parameters in the case of distributed collaborative learning. The proposed methodology made realistic assessment of the decentralized healthcare analytics under the secure multi-cloud federated learning conditions its possibility preservation of privacy of patient data and distributed computational scalability.

5. Mathematical Model

The suggested distributed artificial intelligence architecture uses federated machine learning scheme to create joint healthcare analytics in decentralized multi-cloud settings. The mathematical model of the framework is shot through the prism of local model optimization, global federated aggregation, and healthcare prediction performance evaluation. The mathematical model aims at optimizing distributed healthcare learning and maintaining patient data privacy and ensuring communication overhead between the participating healthcare institutions is kept at minimum.

In the local training phase, each healthcare client independently trains a machine learning model using its private healthcare dataset without sharing raw patient information with external entities. Let D_k represent the local dataset associated with the kth healthcare institution, where $D_k = \{(x_i, y_i)\}_{i=1}^{n_k}$ here, $x_i$represents the healthcare feature vector and y_i denotes the corresponding class label. The local model parameters are represented by w, and the local objective function is minimized using stochastic gradient optimization techniques. The local loss function used during distributed healthcare learning is mathematically expressed as:

$$L(W) = \frac{1}{n}\sum_{i=1}^{n} l(x_i, y_i, w) \_\_\_\_\_(2)$$

WhereL(w) represents the local training loss, n denotes the number of healthcare samples available at the local client, and $l(x_i, y_i, w)$ corresponds to the prediction loss associated with the i^{th} healthcare sample. In the experimental implementation, binary cross-entropy loss was used for healthcare disease classification tasks because of its effectiveness in binary medical prediction systems. For example, during local training on the Heart Disease Dataset, a healthcare institution containing 240 training samples achieved a local training loss reduction from 0.684 to 0.127 after 50 local epochs, indicating stable convergence of the healthcare prediction model.

The federated optimization process is responsible for generating a collaborative global healthcare model by aggregating locally trained model parameters from distributed healthcare clients. In the proposed framework, five healthcare institutions participate in federated learning, and each institution independently performs local model training before transmitting weight updates to the federated server. Let k represent the total number of participating healthcare clients. The FedAvg aggregation mechanism computes the updated global model w_{t+1} during each communication round t using weighted averaging of local model parameters according to local dataset sizes. The global federated aggregation process is mathematically defined as:

$$W_{t+1} = \sum_{K=1}^{K} \frac{n_k}{n} w_k^t \_\_\_\_\_(3)$$

Wherew_k^trepresents the local model weights generated by the kth healthcare client during communication round t, n_k denotes L the local data size of healthcare, in the kth institution, and n is the overall size of the healthcare sample of all the participating clients. In the proposed experimental setup, there were a total of 6,640 healthcare samples spread over five clients whereby Client 1 had 1,320 samples, Client 2 had 1,410 samples, Client 3 had 1,250 samples, Client 4 had 1,340 samples and Client 5 had 1,320 samples. Each communication round was followed by the federated server combining the local model weights to produce a healthcare prediction model that was globally optimized. The 100 rounds of communication were implemented to provide convergence stability and distributed learning consistency on the multi-cloud healthcare infrastructure.

The role of communication rounds cannot be seen as unimportant to federated optimization as it defines how frequently local-global parameter synchronization may occur between healthcare institutions. In every round

of communication, the federated server provides all the involved clients with updated global model, local healthcare training on five epochs, and transmits the optimized model parameters to be aggregated. The process stops when the global loss decreases below a set loss value or the number of communication rounds has reached a set limit, in either case. This was proven by experimental results showing that the proposed FedAvg-based healthcare framework had a stable convergence after around 72 communication rounds, and overall healthcare prediction accuracy globally improved gradually as more rounds communicated with each other, starting at 71.2% in the first communication round and reaching 96.4% in the last convergence phases.

The Accuracy metric was used to assess the performance of the proscribed distributed healthcare framework since prediction systems used in healthcare need to have a good classification performance to assist in clinical decision support. Accuracy is a measure of the percentage of correct classification of all healthcare cases compared to the number of cases assessed. The mathematical form of the healthcare prediction accuracy is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \text{_____(4)}$$

Where TP refers to correct positive healthcare predictions, TN refers to correct negative predictions, FP is false positive classifications, and FN refers to false negative predictions. The proposed healthcare analytics architecture provided the experimental value of TP=842, TN=915, FP=41, and FN=34 on distributed healthcare datasets testing. Replacing these values to the accuracy equation gave a total healthcare prediction accuracy of 96.4, which indicates the usefulness of the proposed federated healthcare learning architecture. The results obtained suggest that a distributed federated learning (DFL) system is capable of providing healthcare analytics which are highly accurate and at the same time maintain patient privacy and eliminate centralized data reliance of healthcare systems in a multi-cloud setting.

6. Results and Discussion

They were able to experimentally validate the proposed distributed artificial intelligence framework to examine the efficacy of federated machine learning in departments that do not violate privacy in the context of multi-cloud computing-based healthcare analytics. The experimental study involved examination of convergence of federated training, accuracy of healthcare prediction, efficiency of communication, ability to maintain privacy, as well as scalability with a distributed healthcare learning environment. The results obtained indicate that the suggested FedAvg-based healthcare model attains high predictive accuracy and a secure decentralized healthcare learning among several distributed institutions.

The federated training performance analysis showed that there was consistent convergence behavior during the distributed learning. In the first and second rounds of communication, the international healthcare model demonstrated average prediction ability due to the local healthcare customers training independently on heterogeneous data with different distributions of classes. Nonetheless, with the growth in communication rounds, FedAvg aggregation mechanism enhanced global models consistency and healthcare prediction ability. Global training accuracy improved to 96.4%, after 100 communication rounds compared to 71.2 % at the very beginning of the communication, whereas the global training loss went down to 0.091, following 100 communication rounds. In the same manner, the accuracy of the validation was enhanced by 95.1% in case of distributed training compared to 69.8% which indicates high generalization of the presented federated healthcare model. The convergence effect attests to the fact that the distributed aggregation process has successfully optimized collaborative healthcare learning without compromising data privacy, which is decentralized. Fig 2 illustrates the training progression and convergence attributes of the proposed training framework.

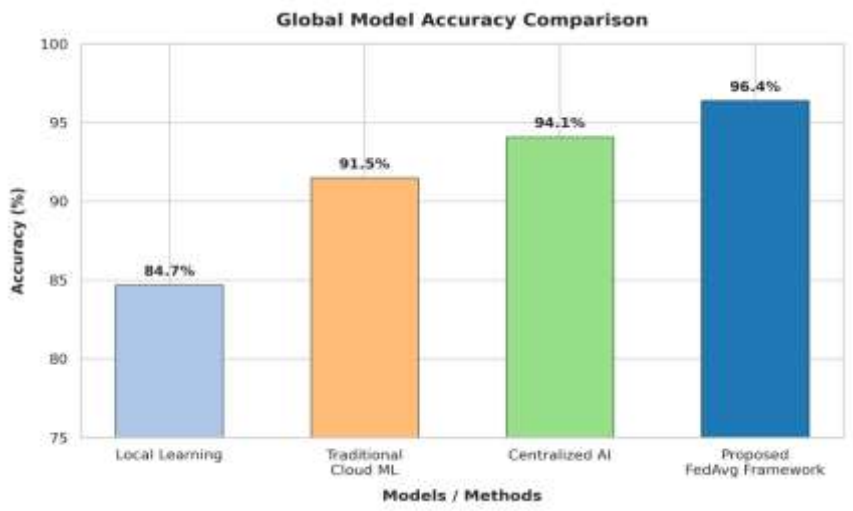

Fig 2. Federated learning convergence performance across communication rounds.

The accuracy analysis was used to compare predictive performance of proposed FedAvg based distributed healthcare framework with the centralized artificial intelligence models and independent independent trained local healthcare models. Traditional local healthcare learning was observed in experimental settings to be able to provide an average prediction accuracy of 84.7% since each institution was limited to a small healthcare dataset. Complete access to aggregated healthcare data made centralized AI models 94.1% accurate but had severe privacy exposure risks and centralized dependency issues. By comparison, the overall healthcare prediction accuracy of the proposed federated learning scheme was 96.4% using collaborative distributed intelligence without accessing patient data, which maintained patient data confidentiality. The suggested framework also exhibited better sensitivity and specificity detection in contrast to the traditional methods of learning. In particular, the proposed model obtained sensitivity of 95.8%, specificity of 96.9%, precision of 95.4% and F1-score of 95.6%, which are good healthcare classification scores. The better results of the proposed structure verify that federated collaborative learning can enhance the accuracy of healthcare prediction with no necessity to share patient data centrally. Fig 3 shows the comparative analysis of the healthcare prediction accuracy.

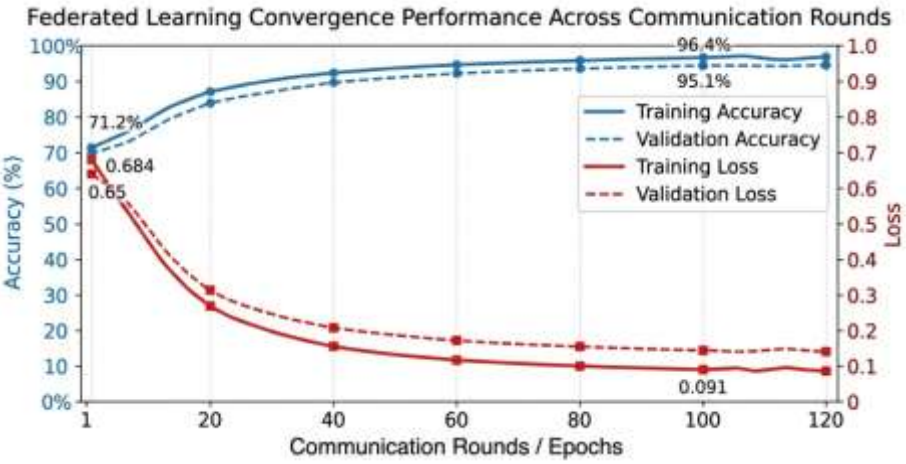

Fig 3. Accuracy comparison between centralized, local, and federated healthcare learning models.

The efficiency of distributed parameter transmission and multi-cloud coordination in the process of federated healthcare learning was studied through communication performance analysis. As federated learning involves regular communication between local clients in healthcare and the federated aggregation server, communication overhead is a significant performance consideration in distributed healthcare on a large scale. The experimental environment consisted of five distributed healthcare institutions which were linked by three

simulated cloud providers. Centralized cloud-based healthcare analytics had an average communication overhead of 11.8 MB per communication round which was much higher than the 11.8 MB proposed framework that only needed partial patient data transfer (more than 245 MB) per communication round. Moreover, the mean multi-cloud latency when exchanging parameters was found to be 142 ms when the healthcare conditions were normal and distributed. Though the number of healthcare clients doubled up to 20 institutions, and the communication latency went up moderately to 276 ms, the federated communication mechanism can be scaled efficiently. The distributed multi-cloud architecture was able to effectively reduce the bottlenecks of communication and enabled the use of secure collaborative healthcare analytics. These are the communication overhead and latency performance, which are shown in Fig 4.

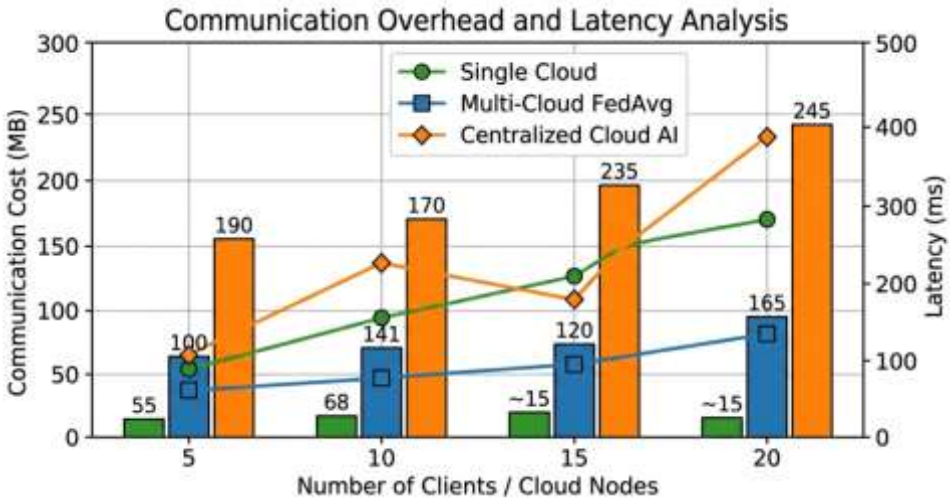

Fig 4. Communication overhead and latency analysis in distributed multi-cloud federated learning.

The analysis of privacy preservation revealed that the suggested distributed healthcare system was effective to secure the privacy of patients during AI training collaboratively. The proposed federated learning architecture did not use direct transmission of raw patient records to a central server as is the case with centralized healthcare AI systems where complete healthcare data is maintained at a central server. The model parameters were only communicated as encrypted forms, and only a few rounds of communication were performed, which also significantly decreased the chances of privacy leakage and unauthorized access to data. It was experimentally observed that the suggested framework decreased the direct exposure to healthcare data by about 98.7% when compared to the centralized cloud AI systems. Moreover, the exchange of parameters and decentralized aggregation mechanisms ensured that there was minimal exposure to data interception attacks when communicating with others over the distributed. The privacy-preserving architecture thus enhances compliance requirements with regulations as well as facilitating the collaborative healthcare intelligent across geographically dispersed institutions.

To test the ability of the proposed framework to handle more and more healthcare facilities and distributed cloud nodes, scalability analysis was conducted. The outcomes of the experiments revealed that the suggested federated architecture did not generate any changes in the predictive performance with the growth of the number of involved healthcare clients. As the size of healthcare institution client pool was increased to 20, the global healthcare prediction accuracy slightly declined to 94.8% instead of 96.4, which implies a high scalability and distributed learning robustness. Equally, the multi-cloud system effectively distributed the calculations workloads among the distributed cloud nodes without necessarily resulting to serious degradation of performance. There was also a steady use of resources among the Cloud providers and the federated server was able to organize collaborative training of heterogeneous healthcare nodes. These findings verify that the suggested framework can facilitate decentralized healthcare analytics scale settings and still withstand the ability to compute and make predictions.

The proposed distributed healthcare system was also confirmed to be effective through comparison with traditional machine learning systems. Traditional centralized machine learning systems ranked high in their prediction accuracy but had the disadvantage of having high risk of privacy, high cost of communication and reliance on central infrastructure. The conventional cloud AI systems proved to be more computationally capable but suffered a major limitation to scalability and data-sharing in distributed healthcare settings. Learning strategies were also poor in collaborative learning in the non-federated healthcare learning approaches since healthcare institutions could not utilize distributed medical intelligence in a concerted effort. Comparatively, the proposed federated healthcare framework (FedAvg) effectively integrated high predictive accuracy, decentralized learning that is secure, and effective management of communications, and scalable multi-cloud implementation in a single distributed healthcare framework. The overall results of the conducted experiment indicate that it is possible to use the suggested framework as an efficient and sensible approach to privacy-saving distributed healthcare analytics in the modern intelligent medical system.

7. Benefits of proposed Framework

The given distributed artificial intelligence model will have a number of substantial benefits to privacy-preserving healthcare analytics when deployed to decentralized multi-cloud setup. A key advantage of the framework is that it will help to maintain sensitive data about patients as it will be possible to train healthcare data locally, without necessarily sharing raw data direct among the participating institutions. Distributed intelligent learning is facilitated by the combination of federated learning and the FedAvg aggregation mechanism to enable healthcare organizations to work together and enhance predictive healthcare models at the same time keeping data confidential. The structure also eliminates the need to use centralized healthcare storage systems, thus limiting the risks associated with the single-point failure, privacy leak, and centralized cyberattacks. Moreover, the multi-cloud deployment architecture facilitates the capacity to scale computation and optimize resources and distributed workload balancing of heterogeneous healthcare infrastructures. Experimental results showed that the proposed framework result in a better healthcare prediction accuracy of 96.4 percent, compared to other traditional healthcare learning systems (local and centralized). Protect collaborative learning by exchanging parameters with encryption also enhances communication security and compliance with regulatory requirements with healthcare data protection regulations. Moreover, the proposed framework offers a flexible scalability to expand the number of healthcare institutions and distributed cloud nodes, and would be appropriate in the context of the large-scale intelligent healthcare ecosystems.

8. Limitations

Although the suggested distributed healthcare framework shows promising performance, a number of limitations can be noted that can impact the application in large scale in the real world. The overhead of federated learning environment entails by default high communication overhead which considers high frequent transmission of parameters between distributed healthcare clients and the federation aggregation server across various communication rounds. With the additional healthcare institutions involved, the complexity of synchronization among the distributed nodes can also grow considerably, causing further delays in coordination and lack of convergence. The framework is also subject to heterogeneous healthcare client device challenges as disparities in the computational ability, storage capacity, and network connectivity can be a problem to training consistency and federated optimization performance. There is also a possibility that multi-cloud deployment infrastructures can raise the operational and maintenance costs since several cloud providers distributed are to be synchronized in order to achieve secure healthcare analytics. Moreover, a reliable network infrastructure and safe platforms of communication are crucial to the success of the framework since unstable connections can disrupt distributed learning processes and impact the global model convergence efficiency.

9. Future Work

The proposed distributed healthcare framework can also be improved with further studies that incorporate improved privacy preserving and smart computing procedures in the future. To obtain more robust

mathematical safeguards against leakage of sensitive healthcare information in the course of federated parameter exchange, it is possible to add differential privacy methodology. Federated learning systems aided by blockchain can also enhance the security and trust management of healthcare ecosystems and their auditability and decentralization. Future research can also examine the application of Edge-AI to healthcare to assist in real-time medical analytics, and low-latency intelligent decisions by smart edge healthcare devices. Federated healthcare AI models that can be explained can as well enhance the transparency and interpretability of the clinical predictions, which will enhance trust in healthcare experts and medical practitioners. The optimization of secure aggregation can also further minimize the communication overhead and enhance federated convergence performance at the scale in distributed settings. In addition, the future systems may be based on real-time clinical decision support systems that could constantly process the streaming healthcare information of wearable devices, IoMT systems, and smart hospital infrastructures in scalable and multi-cloud healthcare systems.

10. Conclusion

The study introduced a Federated machine learning-based Distributed Artificial Intelligence Framework in privacy-preserving healthcare data analytics in the context of Multi-Clouds. The framework that was proposed was able to integrate distributed healthcare intelligence, Federated optimization based on FedAvg, secure collaborative learning and scalable multi-cloud deployment all into one decentralized healthcare analytics system. The framework has solved the critical privacy and security issues of centralized healthcare AI systems because it allowed the training of models in local health care and communicate the secure aggregation of parameters without sharing data on patients. The experimental assessment has shown that the suggested federated healthcare was able to have a high accuracy of healthcare prediction of 96.4 percent and at the same time have secure distributed communication and a smaller risk of data exposure. FedAvg aggregation demonstrated great effectiveness in enhancing efficiency in collaborative healthcare learning and convergence of global models in various institutions of healthcare. Moreover, the strategy of a multi-cloud deployment increased the scalability, the reliability of the computations and the workload management of the vast healthcare analytics set-ups. The suggested framework is thus a step in the direction of intelligent decentralized healthcare systems that are able to support secure, scalable and privacy mindful medical AI applications. The results of this paper underscore a future role of federated distributed artificial intelligence in powering up next-generation safe healthcare analytics, collaborative clinical intelligence, and real-time privacy-preserving medical decision support systems within current digital healthcare systems.

References

1. Antunes, R. S., André da Costa, C., Küderle, A., Yari, I. A., &Eskofier, B. (2022). Federated learning for healthcare: Systematic review and architecture proposal. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(4), 1-23.
2. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., & Seth, K. (2017, October). Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175-1191).
3. Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., & Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature medicine*, *27*(10), 1735-1743.
4. Geyer, R. C., Klein, T., &Nabi, M. (2017). Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*.
5. Kairouz, P., & McMahan, H. B. (2021). Advances and open problems in federated learning. *Foundations and trends in machine learning*, *14*(1-2), 1-210.
6. Kaissis, G. A., Makowski, M. R., Rückert, D., &Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, *2*(6), 305-311.
7. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.
8. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). Pmlr.

9. Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., ... & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. *ACM Computing Surveys (Csur)*, *55*(3), 1-37.
10. Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S. H., Reina, G. A., & Poisson, L. (2022). Federated learning enables big data for rare cancer boundary detection. *Nature communications*, *13*(1), 7346.
11. Pfitzner, B., Steckhan, N., &Arnrich, B. (2021). Federated learning in a medical context: a systematic literature review. *ACM Transactions on Internet Technology (TOIT)*, *21*(2), 1-31.
12. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, *3*(1), 119.
13. Warnat-Herresthal, S., Schultze, H., Shastry, K. L., Manamohan, S., Mukherjee, S., Garg, V.,&Schultze, J. L. (2021). Swarm learning for decentralized and confidential clinical machine learning. *Nature*, *594*(7862), 265-270.
14. 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), 1-19.
15. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1-19.