Cognitive Computation (2024) 16:2626–2641
https://doi.org/10.1007/s12559-024-10310-3

RESEARCH

SPEI-FL: Serverless Privacy Edge Intelligence-Enabled Federated Learning in Smart Healthcare Systems

Mahmuda Akter[1] · **Nour Moustafa**[1] · **Benjamin Turnbull**[1]

Received: 23 October 2023 / Accepted: 14 May 2024 / Published online: 17 June 2024

Abstract
Smart healthcare systems promise significant benefits for fast and accurate medical decisions. However, working with personal health data presents new privacy issues and constraints that must be solved from a cybersecurity perspective. Edge intelligence-enabled federated learning is a new scheme that utilises decentralised computing that allows data analytics to be carried out at the edge of a network, enhancing data privacy. However, this scheme suffers from privacy attacks, including inference, free-riding, and man-in-the-middle attacks, especially with serverless computing for allocating resources to user needs. Edge intelligence-enabled federated learning requires client data insertion and deletion to authenticate genuine clients and a serverless computing capability to ensure the security of collaborative machine learning models. This work introduces a serverless privacy edge intelligence-based federated learning (SPEI-FL) framework to address these issues. SPEI-FL includes a federated edge aggregator and authentication method to improve the data privacy of federated learning and allow client adaptation and removal without impacting the overall learning processes. It also can classify intruders through serverless computing processes. The proposed framework was evaluated with the unstructured COVID-19 medical chest x-rays and MNIST digit datasets, and the structured BoT-IoT dataset. The performance of the framework is comparable with existing authentication methods and reported a higher accuracy than comparable methods (approximately 90% as compared with the 81% reported by peer methods). The proposed authentication method prevents the exposure of sensitive patient information during medical device authentication and would become the cornerstone of the next generation of medical security with serverless computing.

Keywords Federated learning · Privacy-preserving · Edge intelligence · Serverless computing · Smart healthcare systems

Introduction

Technology is changing how we provide effective healthcare, driving an exponential rise in the demand for real-time, secure, cost-effective, and efficient diagnostic, treatment, and palliative services. The advancement of next-generation wireless communications, artificial intelligence (AI), cloud, fog, edge intelligence, and the rise of the Internet of Things (IoT) has created smart healthcare systems. Despite the significant advantages this new data-driven health paradigm has created, it also poses additional and unique cybersecurity considerations. According to the Protenus Breach Barometer, hackers breached 15 million patient records in 2018, nearly triple the number of reported incidents in 2017. In 2020, a ransomware attack on the Florida Orthopedic Institute resulted in the breach of 640,000 patient data records [1].

According to Gartner, cloud-based artificial intelligence grew 500% between 2019 and 2023, making it one of the most popular cloud services. Containers and serverless computing will enable machine learning models to work independently, reducing costs and overhead. However, while edge-enabled healthcare devices improve patient quality of life, create new revenue streams for healthcare providers, and provide more accurate and holistic patient diagnosis, significant privacy concerns over patient data must be overcome. In combina-

✉ Mahmuda Akter
mahmuda.akter@adfa.edu.au

Nour Moustafa
nour.moustafa@unsw.edu.au

Benjamin Turnbull
benjamin.turnbull@unsw.edu.au

[1] School of Systems and Computing, University of New South Wales, Northcott Drv, Campbell 2612, ACT, Australia

tion, edge intelligence (EI) and federated learning (FL) offer potential privacy preservation methods that may be applied to patient data.

A smart healthcare system promises to provide low-cost, low latency, secure, location-aware, and energy-efficient data-driven medical systems for diagnosis, management, treatment, and health issue prevention [2]. Emerging smart healthcare systems increasingly require dynamic architectures, charged with transmitting data securely and the development of decision-making models. Such architectures will be designed using an edge computing-based federated learning paradigm. It includes IoT devices for collecting health data and communication protocols, such as IEEE 802.11/WiFi, IEEE 802.15.1/Bluetooth, and IEEE 802.15.4/WPAN, to allow seamless transmission of data to build a secure decision-making model. Privacy preservation in the medical sector focuses on data operations that include secure transmission, patient-sensitive data encryption, and authentication to be effective against attacks. In contrast, edge intelligence can be applied to classification and prediction to get maximum accuracy [2].

Edge-based healthcare Internet of Things frameworks often feature remote monitoring systems that use several smart sensor types to construct diagnostic, sensitive, and preventive healthcare systems. Patients are tracked using dynamic patient monitoring sensors if they are outside of a hospital, office, or other static location. FL is especially appealing for smart healthcare as a new distributed collaborative AI paradigm since it allows several clients (like hospitals) to work together to accomplish AI training without sharing raw patient data. IoT devices in the physical layer generate client data, can convert analog to digital as needed, and are queried or controlled over networks.

Federated learning combined with serverless computing provides a promising approach for implementing smart healthcare systems to improve patient care and treatment outcomes. In this approach, computational resources are pooled together from multiple sources and used to train machine learning models in a distributed manner without needing a centralised server. To protect smart healthcare systems from privacy attacks, an authenticated key agreement is a crucial security measure. Cyber attackers might obtain sensitive information by gathering the identities of the devices while pretending to be part of a true client of federated learning using an edge aggregator, particularly if the devices are installed in some crucial facilities like hospitals [3].

Serverless computing enables user workloads to be executed without respect for the underlying physical and virtual infrastructure. The most popular model for this is the function-as-a-service (FaaS) paradigm. FaaS enables scalable execution of programming operations through cloud-hosted state-free platforms. Users register functions with a cloud provider, and possibly any system or language prerequisites are required to perform each function. By providing the function ID and input arguments to the cloud provider, authorised users can execute that function a single time or multiple times [4]. Users do not need to provision, set up, and operate dedicated servers that would otherwise run continuously and use resources; instead, they merely pay for the computing resources consumed (typically measured in execution time). However, as with commercial providers, these services must be built up and deployed locally, generally on a Kubernetes cluster or managed centrally. As an infrastructure requirement, FaaS functions must also be designed to be relatively stateless.

Research Motivation A fully trusted central server, as used in traditional federated learning, represents a potential limitation and security threat in implementations with privacy requirements. If there is a security breach in the central server, the entire system and all data will be under privacy threat [5]. A man-in-the-middle (MitM) attack is particularly of great concern in this paradigm during the learning phase, as parameters are passed among clients and servers over the network [6]. Regarding privacy attacks, an attacker can forge a client's identity by forging the lodging of learning parameters.

The creation and integration of the non-homogeneous data used by IoT devices can limit the effectiveness of federated learning algorithms. Multiple data types being sent at different rates from various geographic locations can limit the utility and model of such algorithms. Combined with changes in the breadth and volumes of data produced, this changes capacity requirements quickly. Such use cases require an adaptable convolution neural network (CNN) model design that can process tabular data and generate gradients to continue the FL training privately. Additional complexities exist, such as new devices becoming active online after dormant periods or changing location during updates. Any paradigm, including edge aggregator, must handle such instances robustly while investigating anomalies.

The primary research motivation of this work is to resolve constraints that prevent serverless computing and FL from being used in the health industry. Based on the literature, several requirements are necessary to consider using these technologies in real-world use cases. These include an adaptable machine learning model capable of processing structured and unstructured data as required and passing gradients as model parameters. Also required is the design and implementation of an adaptable edge intelligence-based federated learning model that can provide efficient ongoing learning continuation. Another requirement for commercial deployment is privacy-preserving and secure device-to-device authentication. A final requirement is the ability to scale based on the number of patients and data produced — this requirement lends itself to the benefits of serverless computing.

Research Contribution As a solution, a federated edge aggregator method minimises the loss of privacy and the impact of a successful adversary. With this method, edge aggregators minimise the impact if a server privacy breach occurs as they partition data. As a group of edge aggregators perform federated learning for allocated clients until converged iteration and send back to the server a blended parameter, an adversary cannot trace data back to a particular client. By adding Gaussian noise, differential privacy protects data in model parameters from alteration via MiTM attack. Finally, using edge aggregators solves direct manipulation between servers and clients, as they act as intermediaries.

Integrating an edge aggregator at the intermediate level with a conventional FL architecture can secure the privacy of individual clients and that of data [7]. This three-layered architecture comprises the following layers: a top level as the Global Aggregator, an intermediate level as the edge aggregator, and an IoT level as IoT Clients avoid direct manipulation from a central server. In such a framework, the edge aggregator delivers the blended model parameters to a central aggregator after a certain number of iterations. As a result, the global aggregator cannot track sensitive information about a particular user. However, in the current framework, the client layer is relatively static and does not adequately consider efficient privacy management for dynamically moving users with various data types. This research proposes an adaptable three-layered hierarchical privacy-preserving method, SPEI-FL, to operate over a conventional IoT-based FL paradigm. It is designed specifically for use in smart healthcare ecosystems.

The contributions of this paper include the following:

- The design and implementation of a six-way handshake protocol connects clients and servers securely, expanding on a traditional four-way handshake to include an additional two-way device-to-cluster authentication. This addition allows for a more robust authentication in real-time situations where IoT data collection devices are portable.
- An adaptable CNN model is the core of the federated learning process in combination with serverless computing processes. In this, we introduce cross entropy as a loss function and stochastic gradient descent (SGD) as an optimizer.
- The proposal of a serverless privacy edge intelligence-enabled federated learning (SPEI-FL) framework that processes and evaluates both structured and unstructured data of healthcare systems.

We have evaluated the proposed framework's performance with existing techniques. The results showed stronger data privacy and scalability with durable serverless computing.

The remainder of the paper is organised as follows. The "Background and Related Work" section provides background and related work information on federated learning, IoT, serverless computing, threat models, and authentication protocols. Federated learning at the edge, its thread models, and patient data insertion and deletion during federated learning-related problem formulation are discussed in the "Federated Learning at the Edge" section. The proposed method is described in the "Proposed Framework" section. The "Experiment and Evaluation" section explains our experimental setup and our experimental findings in detail. The article is concluded in the "Conclusion" section.

Background and Related Work

Smart Healthcare Systems Using Serverless Computing

Smart healthcare refers to a health service that uses technology to provide health benefits in an automated and intelligent manner. Examples include IoT integration, wearable technologies, and wireless communication methods to connect patients with dynamic diagnostics and treatment. Smart healthcare benefits individuals, patients, professional services, and healthcare organisations by facilitating data collection, assisting with diagnosis and treatment, and allowing for robust information exchange.

Multiple forms of aggregate personal records relate to individuals' health. These include electronic health records (EHR), electronic medical records (EMR), and personal health records (PHR) [8]. A person's electronic health record (EHR) complies with nationally recognised interoperability standards and may be created, managed, and accessed by authorised medical professionals. What differentiates the EHR from the EMR is that it only contains internal information, whereas EMR data may be partially shared with employers, insurance companies, governmental organisations, and patients. A PHR is a person's electronic health record that complies with internationally accepted interoperability standards, can be accessed from various sources, and is managed, shared, and controlled by the individual to whom it is relevant. Figure 1 shows the overview of the smart healthcare system with an attack scenario.

A recent trend has been to merge data from EHRs, EMRs, and PHRs, allowing for holistic and consistent health approaches [9]. Heart et al. [9] presented the benefits and necessity of such integration, Agfa Healthcare, one of the leading healthcare IT providers in Europe, established a strategic partnership with My Personal Health Record Express Inc. as part of its entry into the market for integrated care, purchasing a 27% equity share of the smaller company.

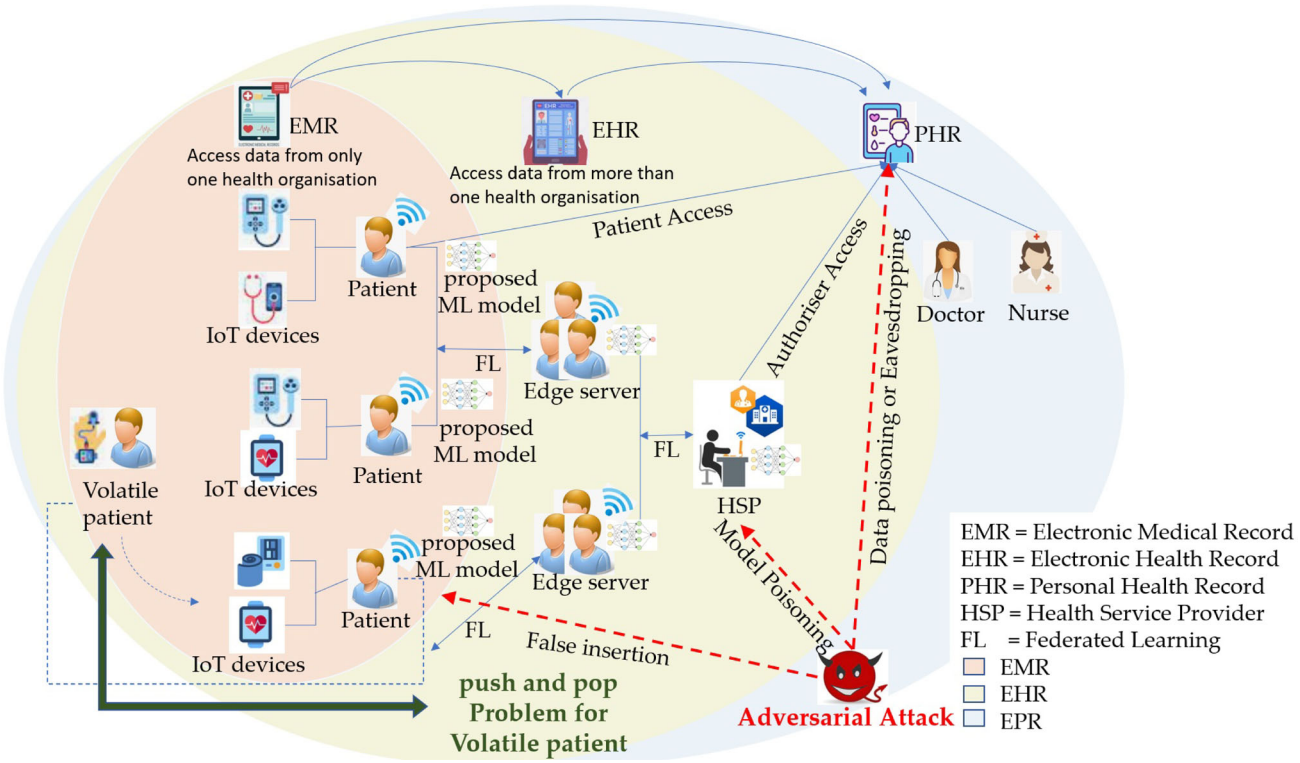

Fig. 1 Overview of SPEI-FL in smart health care system with an attack scenario

In [10], an approach is proposed to ensure smart healthcare systems offer a mechanism for patients to verify the legitimacy of the medication they are taking while ensuring that such systems are not abused. Although this approach can be made readily available via smart mobile apps for in-the-moment QR code scanning, the technique alone cannot stop the ingestion of counterfeit or unauthorised medications.

Serverless computing is an emerging cloud computing model where the cloud provider manages the infrastructure required to run and scale applications, and the users only pay for the computing resources used. It is designed to provide fast and secure development and strongly decouple development and infrastructure maintenance. Serverless computing potentially benefits the healthcare industry [11], as with any other industry with elastic data requirements. When applied to smart health solutions, serverless computing allows for rapid and secure IoT deployments. Additionally, serverless computing can make it easier to develop and integrate artificial intelligence and machine learning algorithms into smart healthcare systems, leading to more accurate diagnoses and personalised healthcare recommendations.

The Internet of Things

The Internet of Things (IoT) is the emerging lightweight bridge between the virtual and physical domains. Comprised of sensors, actuators, networking, and computing (often cloud-based), the IoT is changing how we live and interact with physical spaces around us. IoT results from several underpinning technology advances — wireless and mobile networking, cloud, edge computing, small, cheap processors, and AI/ML. The physical components, communication between components that enables data exchange and collection, and data analytics and decision-making skills that elevate a merely "connected" system to a "smart and connected" system are the three defining aspects of IoT-enabled systems [12]. From a smart health perspective, there are many existing and potential future uses for this paradigm — chronic disease management, epidemic surveillance, and control, care for the old and young, and supervision of health and fitness.

Several studies have investigated the use of clustering for adaptable IoT clients. An example of these includes the development of iterative federated clustering algorithm (IFCA) Heterogeneous clients [13]. This work's results showed that clusters improve compared to randomly scattered clients. Federated learning soft clustering (FLSC) is a combined IFCA + Soft clustering [14] technique in which clients are partitioned into overlapping clusters to achieve higher performance. The zero-knowledge clustering (ZeKoC) adaptable cluster process [15] used clustering as a filtering mechanism to mitigate the adversarial behaviour. The data-decoupling federated learning (DDFL) one-to-many matching [16] method was designed and showed

positive outcomes for privacy-preserving resource allocation. GANC was introduced in [17], a process with 149 clients, and the dynamic clustering process Pendigit. This work showed that the clustering-based handover process is 45% faster than the long short-term memory neural network (LSTM). The clustering to address data heterogeneity in the federated learning (CAFL) method introduced in [18] has been shown to lower the cost of communication while guaranteeing effective clustering outcomes. ClusterFL [19] with 184 clients depth of the performance (convergence and accuracy) demonstrates less sensitivity to the additional random noises.

Device-to-Device Authentication

In modern device-to-device (D2D) communication, authentication and encryption are routed through a central server. Although this seems paradoxical, using a central server provides for connection brokering when devices are on different networks [20]. On-device machine learning is made possible via federated learning, but it faces difficulties with security, resilience, and resource optimisation (both computational and communication-related).

These issues must be solved for IoT networks to use federated learning effectively. One can create federated learning protocols based on D2D communication to address concerns with resource limits. Once the D2D linkages are established, the devices are subscribed to the base stations, and while the devices are mobile, the intermediate devices communicate via the Base Station [21, 22]. When using cellular networks, improved D2D connections are available to lessen congestion. Device-to-device, a big data platform created by Wang et al. [23], accurately offered content to users, successfully discharged intelligently for operators, and effectively promoted the use of wireless networks amongst users to achieve D2D communication.

Federated Learning at the Edge

Federated learning focuses on the challenge of mapping input data I_i to output labels O_i in traditional machine learning. $(p, p+1)$ are the pair size of the input–output (I_i, O_i). Federated learning aims to optimise using the following objective function for the loss function $F_i(L)$, which measures how well a model predicts its sample using model L. The minimum $f(L)$, where

$$f(L) = \frac{1}{p}\sum_{i=1}^{p} f\,(X_i Y_i L) \tag{1}$$

And

$$f(L) = \frac{1}{p} F_i L \tag{2}$$

The objective function must be modified because the training data are dispersed over numerous remote clients, much like in federated learning. When client C receives a partition from the entire dataset D, it is called $|Dc|$. Each of the C clients taking part in the federated learning holds a data sample

$$pC = |Dc| \tag{3}$$

$$f(L) = \sum_{i=1}^{k} \frac{pk}{p} Pk(L) \tag{4}$$

$$Pk(L) = \frac{1}{Pk}\sum_{i=1}^{k} F_i L \tag{5}$$

The model for each client is updated using local data and trained for local epochs. P represents the weight of the dataset on the $k-th$ client. L, a client model, therefore, acquires alterations and transmits them to the server. The server calculates the weighted average of client updates for the subsequent training cycle based on the sample size. Here, fed averaging will have finished one round globally federated. We use FedAvg because it is commonly used for aggregation in federated learning (FL) and requires less communication overhead than other complex FedProx or SCAFFOLD algorithms. It is robust and can work reasonably well across different settings and datasets without extensive tuning. So, in our model, the requirements and constraints of the FL application, FedAvg is more compatible than others. Local epoch training is used to lower communication costs on the client's side because it would be expensive to send each client's gradient update for each round of training.

Threat Model for Federated Edge Aggregators

Edge computing offers intelligent, real-time healthcare solutions that satisfy latency and energy consumption requirements when combined with 5G and advanced smart IoT sensors [24]. Its merits include privacy, autonomy, computation, fast alerts and decisions, and resilience. However, as another significant device on a network, Edge adds additional infrastructure that is susceptible to cyber attacks potential compromise.

Edge intelligent-based federated learning relies on a threat model, which describes the assumptions about the capabilities and behaviours of the many parties involved in the process, to secure the privacy and security of the data. Recent studies [25, 26] have shown that only the transmission gradient is susceptible to different attack levels. The federated edge aggregator model shows how to avoid client-level privacy attacks by establishing an SGD of differential privacy on clients' data.

Organisations utilising IoT devices to gather or use potentially personally identifiable data or data related to health must abide by laws and regulations that define how such data should be handled. Suppose private organisations that provide IoT devices or services have access to IoT data. In such a case, there is an increased risk of the potential for information being exposed for non-public interest reasons, such as profiling, targeted advertising, leverage, or as the basis for future spear-phishing campaigns. To adequately protect users, there is a need to identify possible privacy attacks and approaches to mitigation. Although federated learning is designed for conditional training for privacy preservation, it can still reveal sensitive information [27].

An international standard's backend may become overloaded or modified by malicious players. Because only the server, in an intermediate understanding, can violate the participants' privacy, these attacks constitute a serious threat to federated learning [28]. During federated learning, insiders, such as curious servers or curious clients, and outsiders, such as adversary attacks, are possible [28]. Data transmission from the global aggregator to the system's users could be the target of internal attacks such as passive network interception attacks and Sybil attacks.

A hierarchical three-fold federated edge aggregator (FEA) method introduces an iteration-based middle association of edge intelligence, including differential privacy, before sending learning parameters [6]. This FEA method ensures data privacy by leveraging the Gaussian noise-adding mechanism before passing model parameters under certain noise perturbation levels. However, in privacy attacks during the training phase, an attacker may negotiate the validity of the training data collection during the training stage using information-infecting attacks or modify the integrity of the training technique with data poisoning.

Insertion and Deletion During Federated Learning

A FL system becomes robust when it processes stable and dynamic clients. This would require the dynamic addition and removal of IoT devices, and ensuring clients can move between geographic locations and networks. This requirement contradicts current federated learning processes, in which clients continuously run their ML model by feeding their generated data to collaborate and keep updating their learning. However, the main challenge is how inserting or deleting a new client might impact overall collaboration. These requirements raise an additional limitation in the federated edge aggregation model; if a client moves regions whilst the source edge aggregator is iterating a group of clients' data, how can the system ensure that processing is transitioned to the appropriate destination edge aggregators?

In federated learning, the user may enter before or after the training start, which is noted as 'Push', and the other side user might move away far or end his assigned task and stop being part of the training; in this case, it is noted as 'Pop'. An efficient and privacy-preserving way for users to use Push and Pop is essential for real-time applications. The proposed method investigates the true clients to allow performing the learning, and if it detects an intruder, the system deletes it immediately. The proposed method efficiently handles clients' insertion, deletion, and moving for seamless federated learning without compromising privacy.

Indicative Authentication in Federated Edge Aggregators

Although any Internet-connected device can pose general security and privacy risks to users, the nature of IoT is different [29]. There are several reasons for this; many ecosystems present difficulties for device discovery and inventory on consumer home networks, the number and variety of devices increase vulnerability surface areas, small systems may not be easily updated, and the limited nature of computation may limit security. The security of such systems impacts both the consumer's and other users' Internet access services that use shared network links, as an adversary can pivot from a compromised IoT device to adjacent devices.

In FL, authentication is the process that verifies the identity of the participating edge devices and edge aggregator servers and ensures that the communication is secure and private. Authentication in federated learning involves using a cryptographic protocol and key exchange mechanism to establish a secure communication channel between the edge devices and the edge aggregator server. For example, the secure sockets layer (SSL) or transport layer security (TLS) protocols can encrypt the communication channel and ensure user data is protected from eavesdropping or tampering.

In addition, FL would use authentication mechanisms such as digital certificates, which are cryptographic protocols that verify the identity of the devices and allow them to communicate securely. Digital certificates can also enhance the system's security by preventing unauthorised edge devices from participating in the federated learning process. Authentication in federated learning is essential for ensuring the security and privacy of user data in the collaborative learning

environment, and it involves the use of various cryptographic techniques and protocols to establish secure communication channels between the edge devices and the edge aggregator.

Proposed Framework

The procedure of the proposed framework, SPEI-FL, is illustrated in Fig. 2. In SPEI-FL, the global aggregator initialises the edge aggregators, hyperparameters, model and noise parameters, dataset, and distribution method, including the authenticated token. Consequently, edge aggregators initialise their corresponding IoT clients, passing the model's parameters and assigned data, including authenticated tokens, and initialise cluster members based on geographical location using k-means clustering. In each cluster, IoT clients load the training data with the proposed adaptable convolutional neural network model and labels to train the model to calculate the training Loss by employing an optimization technique to load the model's parameters, add Gaussian noising using differential privacy before aggregation ensures client privacy and protects against adversarial attacks to each and update the model.

The edge aggregator accumulates the model parameters of the corresponding set of IoT clients, broadcasts the updated model to the connected IoT clients, and calculates the accuracy of the test data and corresponding labels. This process repeats until the model is stabilised. Currently, the global aggregator aggregates the edge aggregators' model parameters, broadcasts the updated model to the connected edge aggregators, and calculates the overall testing accuracy from the test data and corresponding labels. If a new client wishes to join, they will send a join request requesting geographical coordinates with a unique ID. If the ID value matches the dataset, then send a positive acknowledgement (ACK) and accept it as a trusted client, it will be considered an intruder. In the case of acknowledgement, the trusted client will be allocated into a cluster according to minimal Euclidian distance with each cluster representative.

Proposed Device to Cluster Authentication Method

To achieve the low latency and secure requirements for smart medical systems, we propose an adaptable edge intelligence-based privacy preservation method in federated learning. Our previous work [6] achieved a 90% accuracy and a higher degree of privacy preservation when compared with existing state-of-the-art methods with medical data. To extend this work, this paper proposes several interlinked additions; a secure and flexible device insertion and deletion process that allows the machine learning model to continue updating, an intrusion detection system to isolate suspicious users, and a device to CR (cluster representative) 6-way authentication protocol. The proposed method is explicitly designed for device-to-device (D2D) communications to preserve privacy while maintaining low latency. The collection of medical devices in a geographical region is called a single cluster, and each cluster has a cluster head (CH). Through this CH, medical devices request the service provider's identity from the server for mutual authentication. The proposed Device to Cluster Head authentication updates token-based authentication, a key exchange between the device and edge aggregators cluster representative, instead of using a global aggregator. Firstly, we consider IoT devices deployed in a geographical area with distinct 2D coordinates. Then, we determine the distance between each dataset point and each initialised centroid using the K-Means technique. Points are allocated to the centroid with the shortest distance based on the values

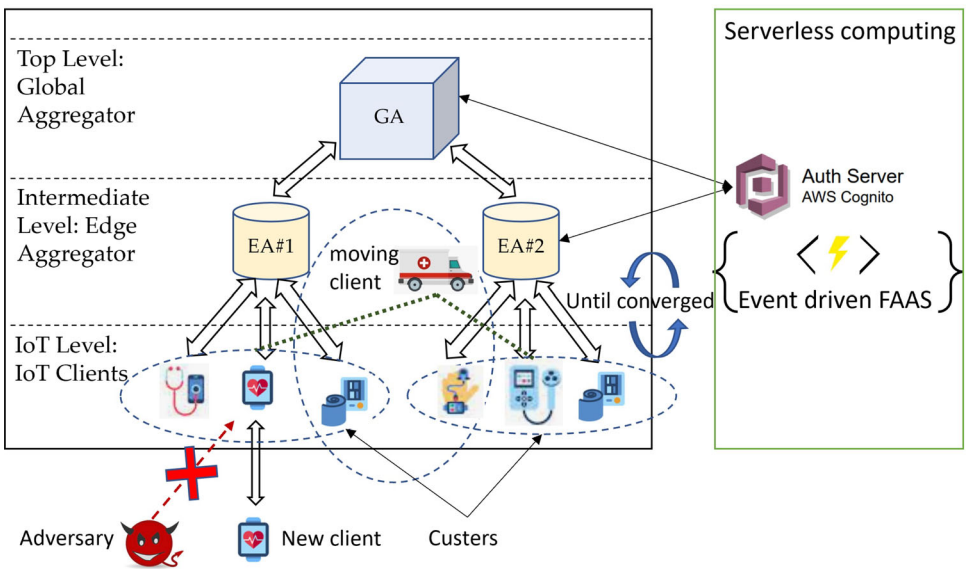

Fig. 2 Serverless privacy based edge intelligent enabled federated learning framework in smart healthcare systems

discovered. In Algorithm 1, it is mentioned that when a new client is identified for insertion, edge aggregators initialise their corresponding IoT clients, passing the model's parameters and assigned data, including Authenticated token and initialising cluster members based on geographical location using k-means clustering. The proposed research considers each device's geographic location (e.g. 2D coordinate) as the input to the K-means clustering algorithm. This means that rather than considering the individual device type, the clustering algorithm groups diverse devices into the same clusters based on their geographic closeness.

The process for this is as follows. At initialisation, clients transmit their client ID, x coordinate and y coordinate, and associated service provider ID. For new client requests, the edge aggregator checks the unique ID for verification. If the verification process fails, the edge aggregator rejects the request. For efficient edge aggregation, we have grouped user devices by k-mean clustering using Euclidean distance (Eq. (1)) based on geographical location. Consider two devices are P and Q in d dimensional area whereas, P= [p1, p2,.., pd] and Q= [q1,q2,..., qd], Euclidean distance in between P and Q as follows:

$$\|P - Q\| = \sqrt{\sum_{i=1}^{d} (pi - qi)^2} \tag{6}$$

At the beginning of the true client identification process, a request is made by a new client. A new IoT client requests to join the network's nearest devices. The cluster representative response requests the x coordinate, y coordinate, and the unique service provider ID the patient had to register for. The cluster representative will request the edge aggregator to validate which other cluster already has a database of eligible service providers that originated from the global aggregator at initialisation. If authentication is successful, the trusted client will be assigned to a cluster by estimating the minimum Euclidean distance between the cluster representative and the new client. The edge aggregator will update the client number. If authentication fails, the intruder is not responded to.

The proposed 6-way device-to-cluster authentication is one of the contributions where an individual client has to pass its geographic location along with ID to the network. The six-way handshake protocol, which connects clients and servers securely expanding on a traditional four-way handshake to include an additional two-way device-to-cluster authentication. This addition allows for a more robust authentication in real-time situations where IoT data collection devices are portable. The network then verifies whether the client is authenticated or an intruder to keep the complete network from eavesdroppers. This authentication and further clustering may cause an additional delay in adding a device to the network; however, the author prioritises network safety and stability over the initial delay. This proposed six-way handshake protocol is shown in Fig. 3. Even if a client is removed, the edge aggregator updates the members during iteration.

Algorithm 1 IoT client authentication through six-way Handshaking

Input: Dataset D, contains a set of IoT client information $IoT_k \leftarrow \{X_1, Y_1, ID_1, \ldots, X_n, Y_n, ID_n\}$ Incoming IoT Client $IoT_k \leftarrow \{X_1, Y_1, ID_1\}$
Output: Set of Clusters $\text{Cls}_{i,m} \leftarrow \{IoT_{i,1}, IoT_{i,2}, \ldots, IoT_{i,m}\}$, Incoming IoTClient IoT T_k
Status: Trusted/Intruder
Initialise: 6-way handshaking; Initial Distance, $\delta = 100000$, Intruder $\leftarrow 1$
Applying K-means Clustering
Clusters: $\text{Cls}_{k,m} \leftarrow \{\{IoT_{1,1}, IoT_{1,2}, \ldots, IoT_{1,m}\}, \ldots, \{IoT_{n,1}, IoT_{n,2}, \ldots, IoT_{n,m}\}\}$
Centroids: $\text{Cls}_{i,m} \leftarrow \{\text{IoT}_{i,1}, IoT_{i,2}, \ldots, IoT_{i,m}\}$ // Applying round () to get exact IoT client coordinates as Centroids coordinates.
for $j = 1, 2, \ldots\ldots, k$ **do** // For each IoT client in D
if $\text{IoT}_{l_{ID}} = IoTj_{ID}$ **then** //Authentication check: Incoming is Trusted Intruder $\leftarrow 0$
for $i = 1, 2, \ldots\ldots, m$ **do** // For each centroid, calculate the Euclidean distance between the incoming IoT client and the centroid
$d\,(IoT_l, Cnt_i) = \sqrt{\left(IoT_{l_X} - Cnt_{i_X}\right)^2 + \left(IoT_{l_y} - Cnt_{i_y}\right)^2}$
if $d\delta$
then $\delta \leftarrow d$// Minimum distance
index $\leftarrow i$// Store selected centroid index
end if
end for
IoT l_{lebel} = Cnt $index_{\text{lebel}}$ // Assign cluster label to IoT$_l$ **Break**
end if
end for
if Intruder = 0 **then** Incoming IoT Client IoT_l
status: **Trusted**
Update client dataset D by inserting IoT_l else Incoming IoT Client IoT_l
status: **Intruder**
end if
return D

Adaptable CNN Model for Federated Learning

The section discusses an adaptable convolutional neural network model for federated learning to classify intruders in a serverless computing manner. Handling big data is challenging for traditional machine learning. Deep neural network processes, such as transfer learning, recurrent neural network (RNN), and convolutional neural network (CNN), are advanced solutions to overcome these issues. Specifically, these neural network methods perform best for feature extraction from image datasets [30]. For clinical numerical data multilayer perceptron (MLP), random forest algorithms can be used for machine learning [31]. However, federated learning requires efficient learning parameters that are

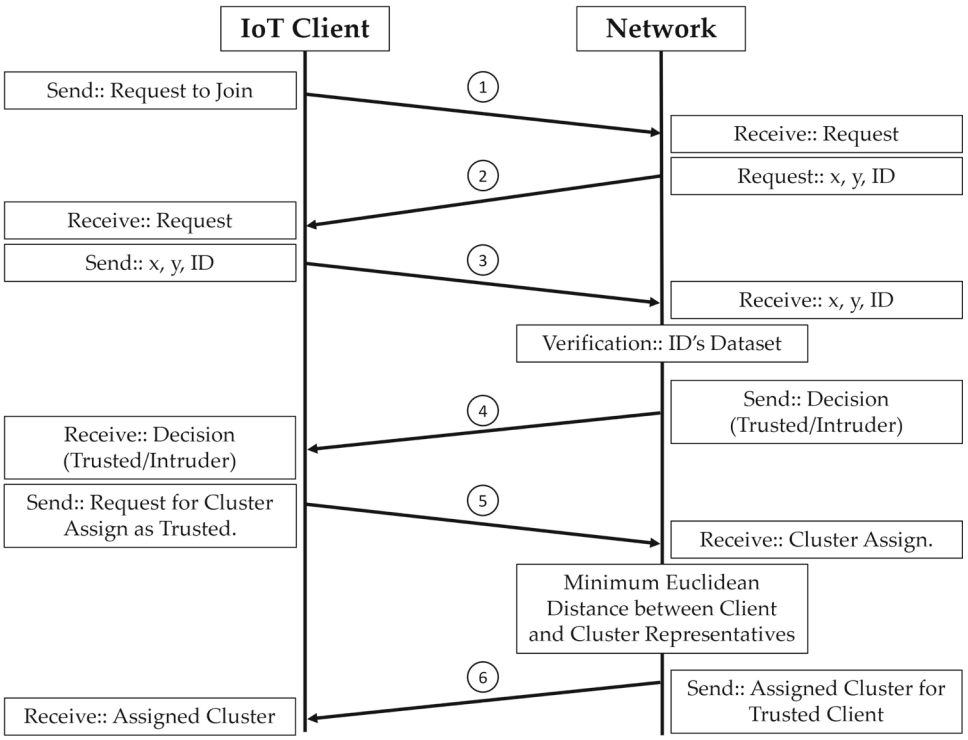

Fig. 3 6-way device-to-cluster representative authentication of SPEI-FL

easily extracted from convolutional neural network models for sending to aggregators and continuous updates on their own devices.

The proposed adaptable CNN is a 1D CNN, rather than a 2D CNN. While 2D CNNs are designed to operate exclusively on 2D data such as images and videos, 1D CNN has recently been developed to perform on tabular, signal, or times series data. The proposed adaptable CNN requires an array rather than a matrix, significantly reducing the computational cost. The relatively swallow architecture includes a few hidden layers and neurons, which makes it much easier to train and implement. The proposed adaptable CNN is well-suited for real-time and low-cost applications, especially on mobile and hand-held devices. The BoT-IoT dataset is a time series data containing spatial properties, hence the proposed adaptable CNN is designed as 1D CNN so that the kernel slides in only one dimension along the axis of time.

The proposed adaptable convolutional neural network model design presents a viable solution for structured and unstructured data in federated learning. A two-layer sequential convolutional neural network (CNN) model shown in Fig. 4 with a fully linked layer on top was used to investi-

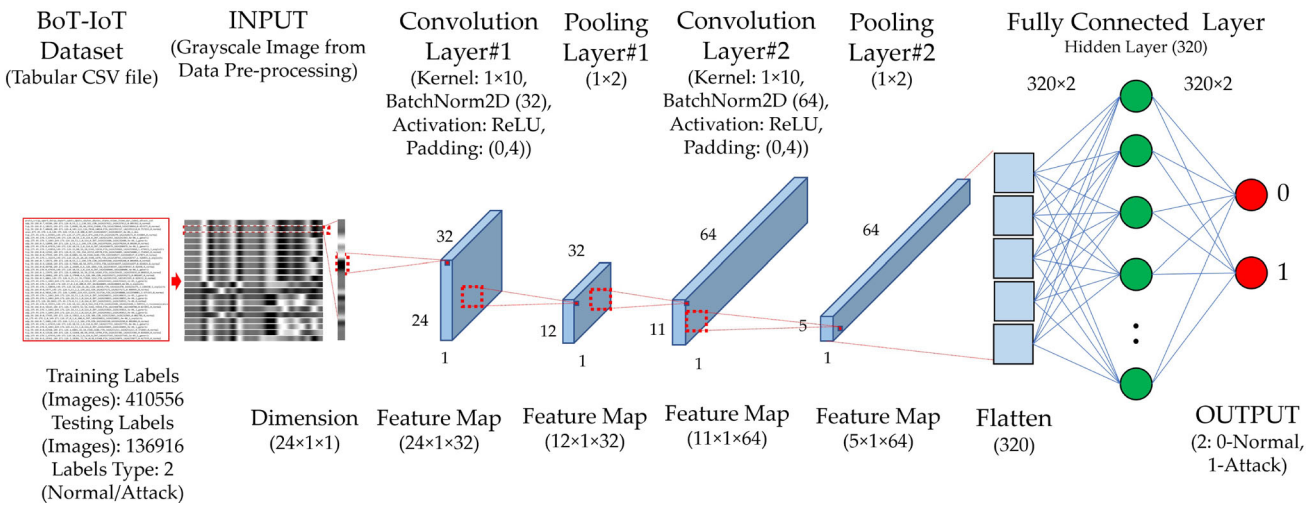

Fig. 4 Proposed adaptable convolutional neural network (CNN) Model of SPEI-FL

gate tabular data. With 1×10 Kernel, BatchNorm2D [32], Activation: ReLU, Padding 0,4, each layer has a convolution. Figure 7 depicts the CNN model's process for learning tabular data from their device.

The non-IID technique was used to divide and assign training data to each client. The non-IID-based technology ensures that each client will receive a maximum of three-digit variants data by first sorting training data according to the label before giving it to clients. Sorting is not necessary with the IID-based approach. However, after initialisation, as non-IID distribution in a cluster environment and clients move to other locations, overall accuracy will decrease because of the probability of missing data. The result shows IID still has a good accuracy in this limitation. The procedure for training the adaptable CNN model with tabular data is shown in Algorithm 2.

Algorithm 2 Training of the proposed CNN model based on tabular data

Input: Tabular Dataset(BoT-IoT) D contains a set of IoT client's ATTACK-related information at time K
Output: Image label representation of each IoT_k, Training Loss l and Testing Accuracy τ
Data Preprocessing:
For both Training and Testing data, Extract data labels
Select interested data features only
Pre-process IP addresses and ports feature
Apply type conversion of some fields to int, logarithmic
Set attack labels as 0 to Normal and 1 Attack (dos, exploits, generic)
Normalise each IoT_k and save 1D image, convert to grayscale and Tensor.
Data Distribution:
Apply Algorithm 1 to get clusters and associated IoT clients
Apply the data distribution method using algorithm 2 in reference [32] to distribute data among IoT clients.
Loss Function and Optimization:
Loss Function: CrossEntropyLoss (reduction='none')
Optimizer: Stochastic gradient descent (SGD)
Learning Rate: 0.01
CNN Model Design:
Input Image Size (24x1x1)
Convolution Layer-1 (Feature Map: 24x1x32)
Kernel Size: 1×10, Batch Normalisation: 32 channels, Activation Function: ReLU, Padding: (0, 4)
Pooling Layer-1 (Feature Map: 12x1x32)
2D Max Pool Layer with Kernel Size (1, 2)
Convolution Layer-2 (Feature Map: 11x1x64)
Kernel Size: 1×10, Batch Normalisation: 64 channels, Activation Function: ReLU, Padding: (0, 4)
Pooling Layer-2 (Feature Map: 5x1x64)
Flattening Input image to a 1D vector of size 320
Fully Connected Layer (input: 320, output: 2 (0 for Normal, 1 for Attack))
Training Loss l and Testing Accuracy τ
return l, τ

For an efficient data distribution method, this process uses Algorithm 2 from [6] to distribute training data initially with corresponding labels to each IoT client. The following three crucial criteria [33] can be used to assess rating based on every user authentication method:

- Usability: For the end-user, authentication is easy and intuitive.
- Security: Security measures how challenging it is for a hostile actor to bypass authentication.
- Deployability: For all users, regardless of platforms, devices, geography, etc., how simple is it to deploy?

Figure 5 presents a rating-based comparison among various authentication methods with flowing criteria.

Several key factors can be considered for being an efficient and more adaptable method. Table 1 sincerely represents how the proposed method (SPEI-FL) covers more key aspects.

Experiment and Evaluation

In this section, we have developed a serverless privacy edge intelligence-enabled federated learning framework and simulated our experiments with multiple datasets. This section also provides a performance evaluation. We evaluated the performance of our scheme in different hyperparameter settings and client variance.

We performed tests using an HP laptop with Microsoft Windows 10 Home, an Intel Core i7 processor (2.6–5.0GHz), 16GB of RAM, and an NVIDIA GeForce GTX 1650 Ti with a Max-Q design graphics processor. The development tools were PyCharm, Pysift, Pytorch, the TensorFlow library, and Python 3.6.

Datasets

To evaluate the performance of the proposed method, we considered both structured and unstructured data. Given the intended use-case in smart health care, we used the COVID-19 Chest X-ray medical dataset from Kaggle [34]. The standard unstructured MNIST data is also considered, as this is a benchmark dataset, and its inclusion allows for practical assessment using comparative methods. The MNIST dataset, with ten output classes labelled from 0 to 9, has 60,000 training and 10,000 test grayscale images with a 28 × 28 × 1 resolution. We selected the recently released COVID-19 chest X-RAY image dataset with 4 labels, 20685 training set (COVID-19: 3496, Lung Opacity: 5892, Normal: 10,072, Pneumonia: 1225), and 240 testing set, with the input image

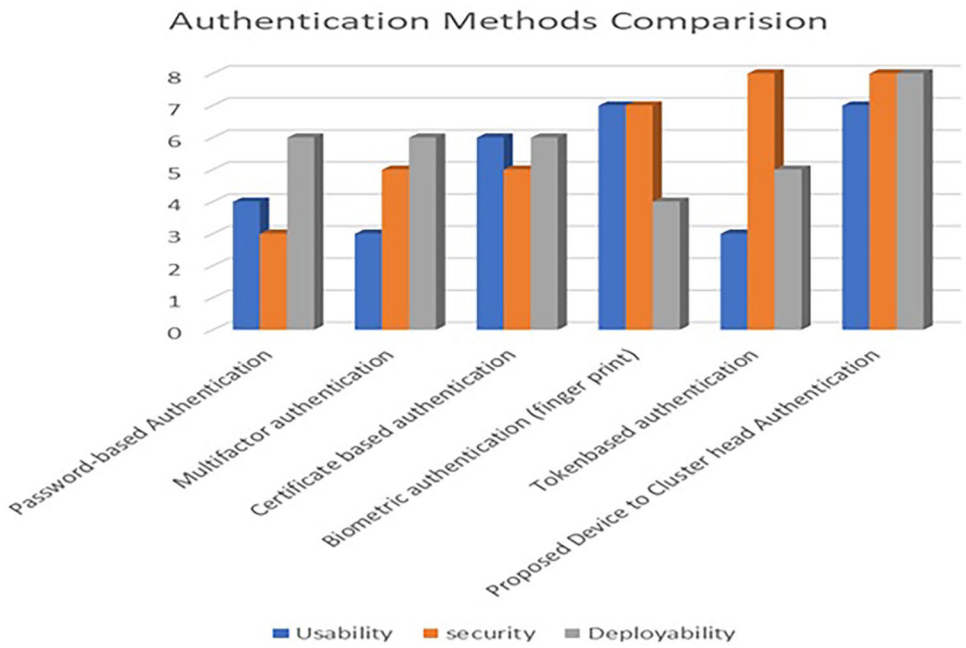

Fig. 5 Rating-based comparison of different authentication methods

dimension 299 × 299, to demonstrate the performance on medical images to support the smart healthcare system.

We resized the images to 32 × 32 during the experiment due to memory restrictions on our machine, which may have impacted the method's evaluation and overall accuracy calculations. Downsizing COVID-19 Chest in the X-RAY dataset images can lead to more efficient model training and inference, improved generalisation, enhanced interpretability, scalability, and privacy preservation, making it a valuable technique for leveraging medical imaging data in healthcare AI systems. Memory constraints are of concern in any mobile platform, so these images were scaled to 32 × 32 during the experiment. This might have affected the spatial details; however, it increased the robustness of the method's evaluations and estimations of its overall accuracy performance. We use the BoT-IoT dataset to analyse IoT-based smart healthcare systems, including attack types and structured data. In this dataset, we consider a structured CSV file with the number of training labels converted to 1D Images 410556. Labs were converted to 1D images 136916 for testing to feed the proposed machine learning model efficiently. We also considered the Normal and Attacked labels for this dataset.

Experimental Results and Discussion

We conducted several experiments using diverse datasets to examine the effect of the adaptability of federated learning on model accuracy in finding true clients and removing intruders. The proposed research's performance evaluation is based on structured and unstructured data. The publicly available datasets, such as MNIST, COVID-19 Chest X-Ray, and BoT-IoT datasets, have provided the training and testing sets separately. This confirms that datasets are already balanced and robust. Also, training datasets are sufficiently large enough to cover each output label to avoid overfitting problems. The results demonstrated here are the mean testing accuracy of 5 individual test cases. Two edge aggregators were considered to handle two clusters containing 5 and 10 clients with very close geographic locations, respectively.

Testing Accuracy Analysis

Figure 6 illustrates the proposed methods' testing accuracy regarding the BoT-IoT dataset with distribution methods IID. The figure reported 86% testing accuracy at the 6th com-

Table 1 Presence of key factors in proposed SPEI-FL and baseline FEA method

Key Factors	FEA	SAEI-FL
IoT cients authentication (6-way handshaking)	No	Yes
IoT client location (geographic: coordinate-based)	No	Yes
IoT client mobility (clustering-based: PUSH and POP)	No	Yes
Federated edge aggregator (model learning by IoT Client; model aggregation and broadcasting by EA)	Yes	Yes
User data privacy (differential privacy)	Yes	Yes

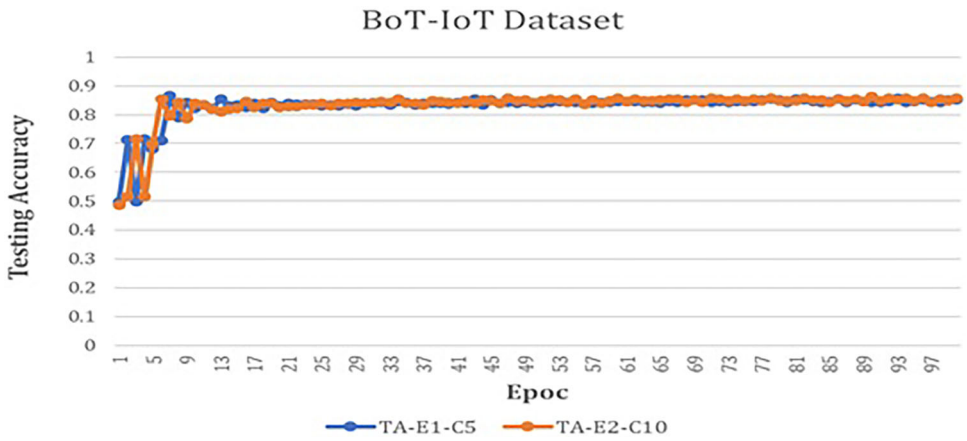

Fig. 6 Testing accuracy with proposed ML model for BoT-IoT dataset of SPEI-FL

munication round; correspondingly, the proposed method is efficiently adaptable for tabular datasets and performs better than other methods (see table).

Training Loss Analysis

Figure 7 demonstrates the training loss of the proposed method regarding the mean absolute error (MAE). We consider the tabular dataset the BoT-IoT dataset. In this experiment, we considered five clients for edge aggregator one and ten for edge aggregator two. Training loss is reducing over time.

We present our results through box plots to statistically analyse our observations of testing accuracy and training loss in a different setting of the client's insertion and deletion of SPEI-FL. In Fig. 8, Ev represents Evaluation, EA1 represents the 1st Edge Aggregator, which contains 5 IoT devices, and EA2 represents the 2nd Edge Aggregator, which includes 10 IoT devices. Here, we can see that for the MNIST dataset, testing accuracy demonstrated higher mean and median accuracy values than the COVID-19 dataset for 4 random observations. Figure 8a shows that median Testing Accuracy for all the settings observed for the MNIST dataset was reported around 85–90% whereas 30–50% was observed for the COVID-19 dataset in Fig. 8b. Regarding training error, Fig. 8c shows around 0.6 median training error was observed for different settings for the MNIST dataset, whereas around 0.8 was observed for the COVID-19 dataset in Fig. 8d. For all the settings in any dataset, reported median testing accuracy and median training error were almost similar, proving the proposed method's robustness. However, in the COVID-19 chest x-ray dataset, the result is a bit low because of the large image size, so we needed to resize the images.

Figure 9 shows the visualisation of the proposed method's outcome message. By entering x-y coordination and a valid unique service provider ID, we can see that the system authenticates the client and assigns it a suitable cluster for continuing federated learning.

Comparative Accuracy

The table shows the comparative accuracy for different methods. All methods have been tested with the same dataset MNIST. Here, we can see that the proposed method shows more accuracy, around 90% than others. Table 2 presents

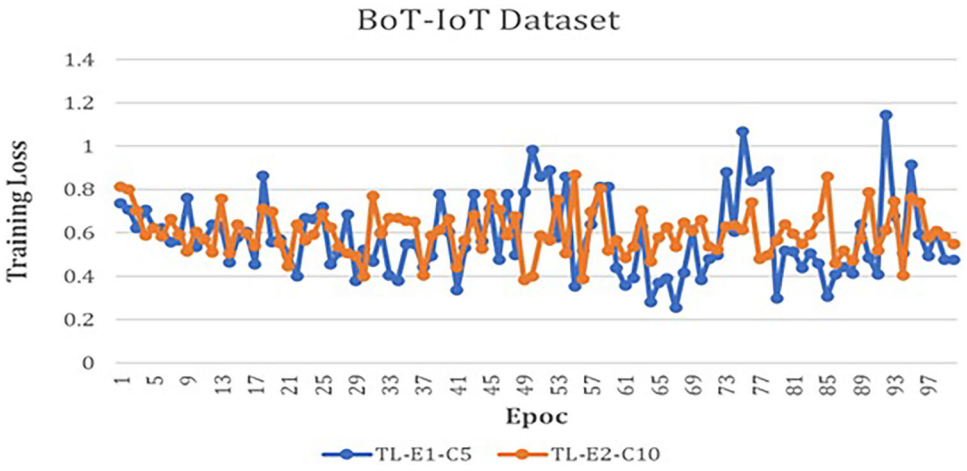

Fig. 7 Training loss with proposed ML model for BoT-IoT dataset of SPEI-FL

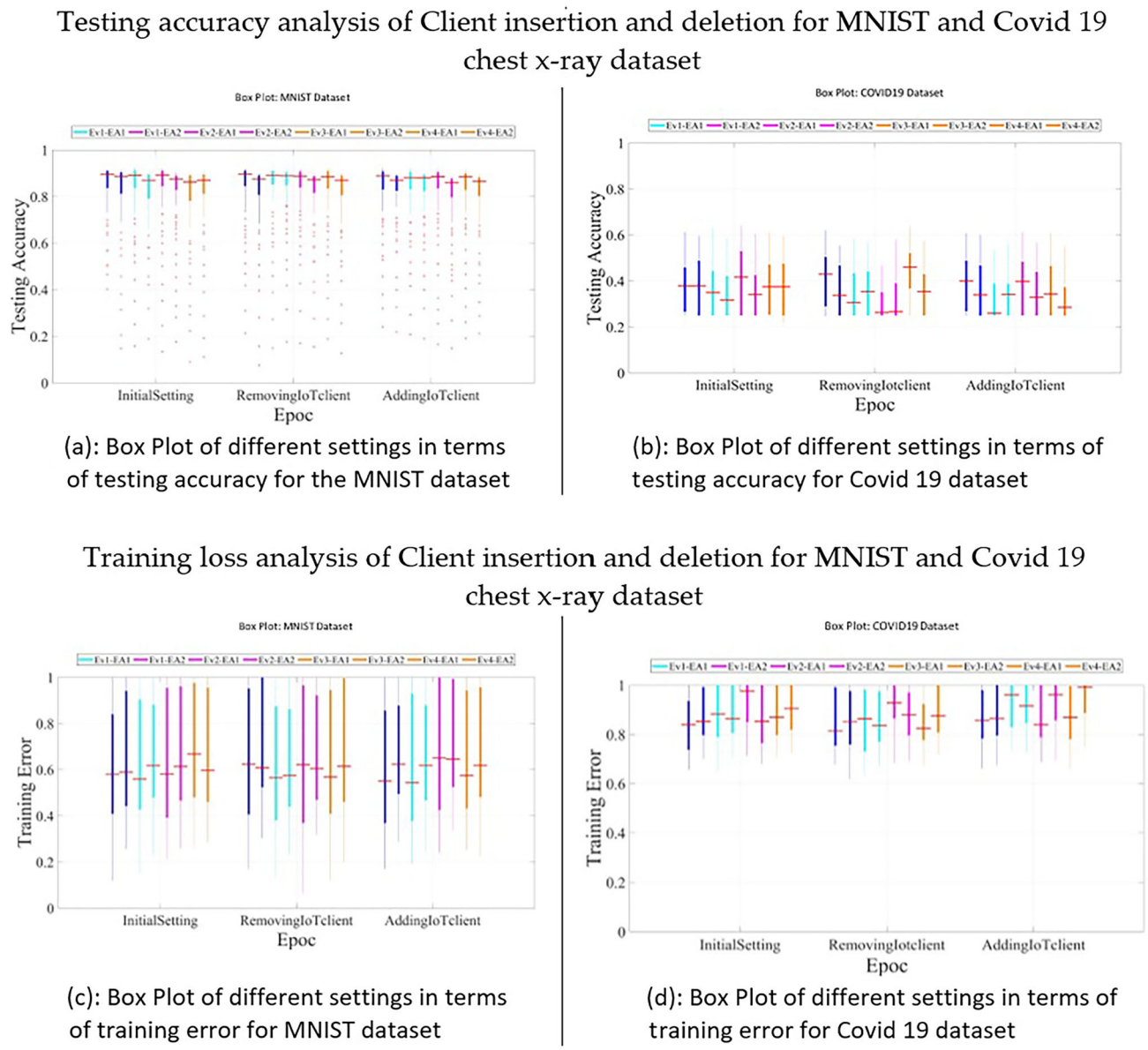

Fig. 8 Box plot of testing accuracy and training loss in a different setting of clients insertion and deletion of SPEI-FL

Fig. 9 Illustration of the options regulation and visualisation of client's insertion and intruder detection massage in SPEI-FL

Table 2 Comparative accuracy of different methods

Method	Dataset	Accuracy
NbAFL (Private) [34]	MNIST	81%
FEA (Base method) [7]	MNIST	90%
SPEI-FL (Proposed method)	MNIST	90%

Table 3 Comparative performance at the 100^{th} iteration with MNIST dataset

Method	Processing time (train)	Processing time (test)	Training loss	Testing accuracy	Precision	Recall	F1Score
NbAFL (Private)	5.4329	5.4329	1.1393	0.9231	0.8213	0.8071	0.8141
FEA (Base method)	5.4942	5.4942	1.0792	0.2861	0.9104	0.9093	0.9098
AEI-FL (Proposed method)	5.3287	5.3287	1.0681	0.3832	0.9007	0.9004	0.9006

the comparative accuracy of different methods that are conducted with the same dataset MNIST and the result shows that the proposed method has a promising outcome with 90% accuracy than others.

Table 3 lists the comparative performance of different methods, including precision, recall, and F1 Score. The result shows that the proposed method performs better than others and takes less training and testing processing time.

Discussion of Results

Using a well-known large dataset, the proposed framework produces promising outcomes. The results show client insertion requests can be made to aggregators with geographical coordination for suitable clusters. After the proposed cluster representative to device authentication, the true client added its assigned cluster according to the location, to continue federated learning. The system can successfully insert and remove true clients and detect intruders. We considered a unique service provider ID as a token for authentication in the proposed six-way handshaking protocol. In the base method, the three-layered architecture shows an overall testing accuracy of around 90%. In contrast, the proposed SPEI-FL is designed to handle dynamic clients with structured and unstructured data without compromising overall accuracy. We convert each tuple into a 1D image for preprocessed structured data to fit the CNN model for efficient federated learning.

While evaluating the impact on overall accuracy, we consider various settings for clients to add and remove in the edge aggregator. After a converged iteration, the cluster representative will reset. As a result, cluster representative to-device authentication can avoid being the target of privacy attacks. It is expected that smart healthcare systems will hold multiple data types, including images, signal data, and numerical data. The proposed on-device machine learning design can process image and numerical data for federated learning. To minimise the execution cost while achieving the deadline for such applications, considerations such as the influence of network latency, bandwidth restrictions, and data placement must be made [34]. In this regard, the proposed serverless privacy can also bring viable solutions.

We consider two edge aggregators with five and ten clients for experimental purposes. However, vast numbers of clients are expected in real-time deployments. In the evaluation, we consider cluster-based intrusion detection. Patients can produce signals such as ECG data in a smart healthcare system. In our proposed method, the CNN model is not designed to learn signal data. In future work, we will consider overcoming these limitations and testing for scale.

Conclusion

This paper has proposed a versatile solution for a smart healthcare system by leveraging adaptable edge intelligence in a federated learning model over a serverless computing paradigm. It shows seamless learning continuation of federated learning over moving patients. It demonstrates a serverless privacy edge intelligence-based federated learning framework to protect overall systems privacy and allocate proper resources for each client at the edge of a network. To avoid complicated infrastructure management load on a central server, lost computation, and incurred costs during authentication demands from inactive clients, serverless computing brings a viable solution. Experiments have shown that this strategy offers noteworthy learning accuracy and time complexity outcomes.

Additionally, it defends against Global Aggregator manipulation by enhancing privacy protection. As part of the experiment, popular datasets with both IID and non-IID distributions were initially distributed among clients, who were trained and tested, and the training loss and privacy costs were calculated. However, non-IID distribution among existing client sets before model learning during client insertion and deletion in a cluster is challenged because missing

data will affect overall accuracy. The proposed framework showed a significantly better influence on learning accuracy with less computing power, has effective data load control, and provides better privacy by evaluating privacy costs. It keeps outstanding performance and provides a more adaptable privacy-preserving paradigm at the same privacy cost as the standard way. Fine-grained microservices-based edge federated learning will be investigated to continue this work.

Author Contribution M.A conducted all experimentation. N.M. and M.A. conceived the initial concept. All authors refined initial concepts and designed experiments. All authors conducted data analysis and validated the outcomes. M.A. led data analysis. M.A. wrote the main manuscript. All authors reviewed the manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability All data in this study are available upon request by contact with the corresponding author.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Stankevičiūtė G. Identity verification in the healthcare industry. https://www.idenfy.com/blog/identity-verification-healthcare/.
2. Hartmann M, Hashmi US, Imran A. Edge computing in smart health care systems: review, challenges, and research directions. Trans Emerg Telecommun Technol. 2022;33(3):e3710.
3. Seok B, Sicato JCS, Erzhena T, Xuan C, Pan Y, Park JH. Secure D2D communication for 5G IoT network based on lightweight cryptography. Appl Sci. 2019;10(1):217.
4. Kotsehub N, Baughman M, Chard R, Hudson N, Patros P, Rana O, et al. Flox: federated learning with faas at the edge. In: 2022 IEEE 18th International Conference on e-Science (e-Science). 2022. pp. 11–20.
5. Wang Z, Pang X, Chen Y, Shao H, Wang Q, Wu L, et al. Privacy-preserving crowd-sourced statistical data publishing with an untrusted server. IEEE Trans Mob Comput. 2018;18(6):1356–67.
6. Akter M, Moustafa N, Lynar T, Razzak I. Edge intelligence: federated learning-based privacy protection framework for smart healthcare systems. IEEE J Biomed Health Inform. 2022;26(12):5805–16.
7. Akter M, Moustafa N, Lynar T. Edge intelligence-based privacy protection framework for IoT-based smart healthcare systems. In: IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2022. pp. 1–8.
8. Calcium. What are personal health records? https://calciumhealth.com/what-are-personal-health-records/.
9. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy. Health Policy Technol. 2017;6(1):20–5.
10. Pandey P, Litoriya R. Securing e-health networks from counterfeit medicine penetration using blockchain. Wirel Pers Commun. 2021;117:7–25.
11. Arkhipov A. Reasons to use serverless architecture in healthcare. https://www.techmagic.co/blog/serverless-in-healthcare/.
12. Kontar R, Shi N, Yue X, Chung S, Byon E, Chowdhury M, et al. The internet of federated things (IoFT). IEEE Access. 2021;9:156071–113.
13. Ghosh A, Chung J, Yin D, Ramchandran K. An efficient framework for clustered federated learning. Adv Neural Inf Process Syst. 2020;33:19586–97.
14. Li C, Li G, Varshney PK. Federated learning with soft clustering. IEEE Internet Things J. 2021;9(10):7773–82.
15. Chen Z, Tian P, Liao W, Yu W. Zero knowledge clustering based adversarial mitigation in heterogeneous federated learning. IEEE Trans Netw Sci Eng. 2020;8(2):1070–83.
16. Khan LU, Han Z, Niyato D, Hong CS. Socially-aware-clustering-enabled federated learning for edge networks. IEEE Trans Netw Serv Manage. 2021;18(3):2641–58.
17. Kim Y, Hakim EA, Haraldson J, Eriksson H, da Silva JMB, Fischione C. Dynamic clustering in federated learning. In: ICC 2021 - IEEE International Conference on Communications. 2021. pp. 1–6.
18. Luo Y, Liu X, Xiu J. Energy-efficient clustering to address data heterogeneity in federated learning. In: ICC 2021 - IEEE International Conference on Communications. 2021. pp. 1–6.
19. Ouyang X, Xie Z, Zhou J, Xing G, Huang J. ClusterFL: a clustering-based federated learning system for human activity recognition. ACM Trans Sens Netw. 2022;19(1):1–32.
20. Schlegel R, Kumar S, Rosnes E, i Amat AG. CodedPaddedFL and CodedSecAgg: straggler mitigation and secure aggregation in federated learning. IEEE Trans Commun. 2023;71(4):2013–27.
21. Khan LU, Saad W, Han Z, Hossain E, Hong CS. Federated learning for internet of things: recent advances, taxonomy, and open challenges. IEEE Commun Surv Tutor. 2021;23(3):1759–99.
22. Balasubramanian V, Aloqaily M, Reisslein M, Scaglione A. Intelligent resource management at the edge for ubiquitous IoT: an SDN-based federated learning approach. IEEE Netw. 2021;35(5):114–21.
23. Wang X, Zhang Y, Leung VC, Guizani N, Jiang T. D2D big data: content deliveries over wireless device-to-device sharing in large-scale mobile networks. IEEE Wirel Commun. 2018;25(1):32–8.
24. Amin SU, Hossain MS. Edge intelligence and Internet of Things in healthcare: a survey. IEEE Access. 2020;9:45–59.
25. Yin L, Feng J, Xun H, Sun Z, Cheng X. A privacy-preserving federated learning for multiparty data sharing in social IoTs. IEEE Trans Netw Sci Eng. 2021;8(3):2706–18.
26. Ganju K, Wang Q, Yang W, Gunter CA, Borisov N. Property inference attacks on fully connected neural networks using permutation invariant representations. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018. pp. 619–33.
27. Chen D, Xie LJ, Kim B, Wang L, Hong CS, Wang L, et al. Federated learning based mobile edge computing for augmented reality applications. In: 2020 International Conference on Computing, Networking and Communications (ICNC). 2020. pp. 767–73.

28. Alam T, Gupta R. Federated learning and its role in the privacy preservation of IoT devices. Future Internet. 2022;14:246.
29. Tawalbeh LA, Muheidat F, Tawalbeh M, Quwaider M. IoT privacy and security: challenges and solutions. Appl Sci. 2020;10:4102.
30. Ahsan MM, Alam TE, Trafalis T, Quwaider M. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. Symmetry. 2020;12:1526.
31. Aslan MF, Sabanci K, Durdu A. A CNN-based novel solution for determining the survival status of heart failure patients with clinical record data: numeric to image. Biomed Signal Process Control. 2021;68:102716.
32. Toosi AN, Sinnott RO, Buyya R. Resource provisioning for data-intensive applications with deadline constraints on hybrid clouds using Aneka. Futur Gener Comput Syst. 2018;79:765–75.
33. Inamdar S. Comparison of user authentication methods on three parameters. https://www.cyberark.com/resources/blog/comparison-of-user-authentication-methods-on-three-parameters.
34. Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: algorithms and performance analysis. IEEE Trans Inf Forensics Secur. 2020;15:3454–69.