IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 March 2023, accepted 10 April 2023, date of publication 26 April 2023, date of current version 1 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3267431

RESEARCH ARTICLE

Blockchain-Enabled Technique for Privacy-Preserved Medical Recommender System

ERIC APPIAH MANTEY[1], CONGHUA ZHOU[1], JOSEPH HENRY ANAJEMBA[2], (Member, IEEE), YASIR HAMID[2], AND JOHN KINGSLEY ARTHUR[1], (Member, IEEE)

[1]Computer Science Department, Jiangsu University, Jingkou, Zhenjiang, Jiangsu 212013, China
[2]Information Security Engineering Technology Department, Abu Dhabi Polytechnic, Abu Dhabi, United Arab Emirates

Corresponding author: Conghua Zhou (Chzhou@ujs.edu.cn)

ABSTRACT With the proliferation of privacy issues surrounding the Internet of Medical (IoMT) recommender system data, this study presents a Secure Recommendation and Training Technique (SERTT) which is contingent on a combination of both federated learning and blockchain approaches. Firstly, the study presents a new framework for recording, sorting, and transmission of IoMT data while incorporating blockchain to ensure that the IoMT data transmitted to cloud servers is not made vulnerable by the sharing of the original data. Secondly, by utilizing medical data, the study designs a Recommender Data Management Neural Architecture (REDMANA) which is based on federated learning and model searching training framework. The proposed technique guarantees that the model gradients which are trained by each node are not disclosed all through the universal training and modeling procedure. This makes the raw data inaccessible to either the IoMT data provider or the user. Considering that the model ensures that users can only obtain their necessary inquiries, neither medical data suppliers nor users can obtain access to raw data. Thus, it reduces the issues of safeguarding medical data sets to the issues of securing data processing. Using numerical analysis and experiments the proposed technique is compared with other existing techniques, the result shows that the proposed SERTT system is efficient and secures recommender data management training and modeling technique and that it performs previously designed techniques as compared.

INDEX TERMS Recommender system, Internet of Medical Things, federated learning, blockchain, privacy assurance, neural architecture.

I. INTRODUCTION

A recommender system is a subclass of an artificial intelligent-based (AI-based) system for information filtering and prediction on a list of products for different organizations [1], [2]. Generally, such kinds of systems are common big data applications. On the internet of medical things (IoMT) system, several hospitals extensively utilize the recommender system to obtain excellent recommendations based on the interest and requests of their patients [3]. The recommender system can generate its recommendations through either collaborative filtering or content-based filtering. The former is a method of obtaining the list of predictions by establishing the interrelation between users' history and other users' interests, while the latter involves exploring both the user's profile and their corresponding items. Most hospitals and companies store users' confidential data and make use of collaborative filtering to achieve optimal recommendations [4], [5], [6]. In this kind of recommendation, the profiles of different users are designed from their respective histories coupled with the user's rating [7]. Consequently, there is the possibility of having the issue of data privacy in an AI-based system. Recently, hospitals gathered and save a massive quantity of patient data for future recommendations, however, patients are concerned about the privacy of their confidential data which are stored on different platforms (such as smart healthcare and other IoT devices). Nowadays, users' data can be disclosed through different means such as social media and through intrusion attacks.

A recent report [8] showed that almost 87 million users' data were hijacked/leaked from Facebook. The major

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

cause of this breach of data privacy is weak privacy techniques. Although several previously proposed techniques [9], [10], [11], [12], [13], [14], [15], [16], [17] have attempted to tackle this issue of data privacy in smart healthcare systems. However, they have failed to completely solve this problem of divulging patients' confidential data, as there has been continuous unauthorized accessing and exploitation of patient data. Concerning the issues that emanate from medical and clinical data sharing and guaranteeing associated medical data privacy, this study proposes a recommender and data sharing framework which utilizes federated learning and secured medical guarantee systems, by this means improving the privacy threats and model inadequacies which relate to conventional recommender systems is optimally guaranteed. The core contributions of this work include.

1. Proposing and implementing a blockchain-based recommender and training system which is referred to as Secured Recommender and Training Technique (SERTT), which locally stores data but uploads data directories and structures to the chain. Through federated learning, users are trained about the data which eventually produces a model instead of giving out the actual user data, thereby enhancing the privacy and security of user information.
2. Also, the study proposes a Classified Proxy Bond Framework (CPBF) coupled with a no-knowledge confirmation-based data analysis and validation assessment approach which provides private users and medical patients with enormous data confidentiality. This technique permits hospital management systems to authenticate the user and patient data and make absolute recommendations based on their medical history without disclosing their confidential data, significantly improving recommendation performance and data analysis privacy.
3. Additionally, a recommender model inquiry approach which is termed Recommender Data Management Neural Architecture (REDMANA) supports the automatic assembling of recommender models through neural architecture inquiry methods, consequently enhancing the automation and proficiency of the recommender data management system.

A. RESEARCH ORGANIZATION

The remaining parts of the paper are structured in this manner. The core approaches of blockchain theory coupled with the techniques of federated learning are presented in Section II, while the proposed SERTT techniques which are contingent on federated learning and blockchain design are discussed, analyzed, and presented in Section III. In section IV, the security model and design approach of the SERTT and its corresponding algorithms will be elaborated. Also, a privacy guarantee structure for a data management system that is based on no-knowledge proofs is designed in this section as well, while a wide-ranging security analysis of the SERTT approach is established in section V. Simulations to show the practical demonstration and performances of the SERTT techniques and its performance is presented in section VI. Section VII provides detailed knowledge of experiments to show the comparison and analysis of the efficiency and security performance of the proposed data management system and recommender data privacy guarantee measures. Section 8 will recapitulate the major contributions of the research coupled with the outcomes of the research, along with providing an overview of the possible future research guidelines.

II. SYSTEM MODEL

As a result of the incorporation of cryptographic techniques and distributed algorithms, Bitcoin which is an offset of Blockchain technology is fundamentally a distributed system that is difficult to interfere with and upholds reliability without the need for a central control [18]. Likewise, contractual algorithms are a vital aspect of blockchain technology and have been attracting cutting-edge research ideas [19], [20], [21], [22]. Considering that the operations of blockchain technology are implemented in a vulnerable environment, there is a need for its activities to tackle the likelihood that transmitting nodes can be adversely affected, thus they require extensive fault tolerance and privacy-enabled algorithms in the workplace.

A. FEDERATED LEARNING AND RECOMMENDER SYSTEM DATA TRAINING AND MODELING

Federated learning (FL) is a kind of machine learning (ML) technique that permits ML models to acquire knowledge from diverse datasets that are traceable in different locations (such as, a central server or a local data centers) and devoid of revealing the training data [23]. By this approach, confidential data are securely stored in the local sites, minimizing the probability of private data access. This permits different enterprises to form a mutual universal model without enclosing the training data in a central site. FL process topples the challenges of data breaches by permitting recurrent learning on end-user devices while guaranteeing that end-user data does not leave end-user devices [24]. Federated schools comprise model modelling and training and model inference [20]. A fundamental context for FL and centralized training is described in Fig.1.

In the framework, conventional ML models are trained on local heterogeneous datasets. For instance, as users make use of an ML application, several errors could be spotted in the application's predictions and these mistakes are automatically modified. A local training dataset is generated in each user's connected device. Then, the parameters of the models are periodically switched over between these local data centers. Several models encrypt these parameters before sharing. However, local data samples are not exchanged. This enhances data security and cybersecurity. In the end, a shared global model is created, and the features of the global model are distributed to local data centers to incorporate the global model into their machine learning local models.

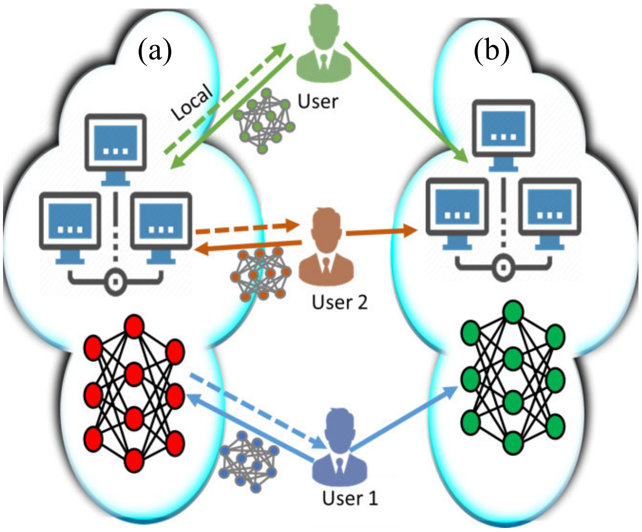

FIGURE 1. A fundamental framework for two different training systems: federated training and centralized training. (a) Federated Training. (b) Centralized Training.

III. SERTT DESIGN AND SECURITY MODEL

This section will address the SERTT framework to realize the core security features that were pointed out earlier. The four stages of the modules' exchanges include contract deployment, requirements matching, execution preparation, and contract execution.

A. CLASSIFIED SHAPLEY PROXY PROOF-OF-STAKE (CSPPoS) MECHANISM

In recommender system modeling and training, the major way of realizing the optimal success of the goals is by the constant participation of the participants in the data-sharing process. By taking part in the FL process, the data providers contribute to the general system while the trained models produce income. The main question is how to measure the merits that participants bring to the system and make it viable. To achieve this, the study employs a combination of the values of Shapley and the Proxy Proof-of-Stake (PPOs) this combination results in the formulation of a model referred to as Classified Shapley Proxy Proof-of-Stake (CSPPoS) which benefits the interests of all participants and accomplishes consensus on the blockchain. This research employed a classified proxy mechanism in the SERTT system. The primary node no longer enjoys higher energy, and it may be removed by many other nodes, if most nodes consider the primary node as malicious, then a new form of choice for the primary node is created. At the beginning of a new term, the system creates an election operation. Each node that intends to be involved in the election process first transmits a message to other nodes through local logs to demonstrate that it comprises the system's better log information and that it has all the requirements to assume the position of a leader or primary node. As soon as the mainstream of nodes come to an agreement that an operational node holds the dominant list of logs in the network, the operational node obtains a particular required number of votes and is successfully chosen and made the primary node.

The representation in Algorithm 1 illustrates the performance of CSPPoS mechanism of the proposed SERTT system. Primarily, the performance of the algorithm is based on the election of representatives, who have the right to recordkeeping, monitoring, and the capability to act as controllers. After every voting stage, the one that records the highest number of votes assumes the position of the system representatives. The core responsibilities of these representatives include block stuffing and functioning as the federated learning managers and acquiring incentives. Also, the data-sharing in the mechanism of the Shapley game is implemented as a contribution-based marginal technique. Compared with the union game profit-sharing approach, the Shapley game mechanism measures the effect of the variance in compensating participants who joined the group in different orders, thus, making available a reasonable and more precise evaluation of each member's contributions. The proxy framework is dependent on the election of representatives. These representatives must have privileged access to recordkeeping, and administration and function as managers.

As soon as the FL is complete, the manager, which also serves as the recordkeeping node, is presented with a static contribution while the data-providing node is rewarded based on the contribution. For every proxy cycle, there is only one election process in the Shapley data sharing-based CSPPoS algorithm, however, there is always a fresh election of group representatives after each proxy cycle.

IV. PERFORMANCE OF SERTT MODEL

A. THE BASIC PROCEDURES OF FEDERATED TRAINING IN THE SERTT SYSTEM

In this section, the three vital security features on which the SERTT system is designed will be extensively discussed. All these features are targeted at safeguarding data against theft from nodes. Originally, three core types of data make up the SERTT system and they are source data, intermediate data which is generated in the process of executing the smart contract, and final data analysis findings. The following description is the main security procedural phase of the system.

1) INITIALIZATION

By initializing some of the system parameters, firstly, the coordinators will make use of the installation process to select the parameters for the public recommendation model of the blockchain. The first generated block which is referred to as the genesis block will comprise $\{H, V, v, r, e, f, \mathbb{N}, \zeta\}$ parameters. The second phase will be contract deployment. At this stage, the coordinators can utilize the Enroll technique to create a long-term account such as, $(pl_y = R_y, cl_y = G_y)$. This account is dependent on the system features as obtainable in the genesis block. At this point, the

Algorithm 1 Classified Shapley Proxy Proof-of-Stake (CSPPoS)

```
Input: mutual recommender data for nodes participating in FL
Declare: and as previous and current blocks, respectively
while in the proxy cycle do
    Each participant casts their vote depending on their respective contribution.
    Categorize the vote results to realize the list referred to as 'sorted_vote_list'.
    Choose the $X$ highest-voted candidates from the sorted_vote_list.
    Acquire $Y$ candidates from the sorted_vote_list → Candidates.
    Shamble candidates → randomly disarrange candidates.
    for select the stuffing node do
        $\#P_B * (C_B) \rightarrow$ to obtain the slot
        To get representative index slot → mod $Y$
        if present node is index, then.
            if the present node rank is manager, then.
                Add the administrator's contribution level of FL to the record.
            end if
            Stored all candidate's contributions, authenticate that it makes use of each.
            party's public key, then approve the transmission with the public key.
            Add candidate $n$ to the group alongside other candidates in altered orders.
            and estimate their average marginal returns and estimate.
                $y(n) = \frac{1}{x!}\sum C\left[d\left(X_n^C \cup n\right) - d\left(X_n^C\right)\right].$
            create block ($sign_{\mathrm{cl}}$ (authenticated_transmission))
        else
            skip
        end if
    end for.
end while
```

coordinator utilizes this account to transmit the designed smart contract in the form of a transaction to the blockchain. The transmitted smart contract will be inputted to the blockchain mechanism the moment it is successfully approved by the proof procedure. In the third stage, both private and public keys are generated and shared. First, the public/private keys are generated as.

$$\{(\delta, \rho), (X_x^{PL}, X_x^{CL}), (P_x^{PL}, P_x^{CL}), L = (L_1, L_2)\} \qquad (1)$$

for each user x, ($x \in \mathrm{L}$, $|\mathrm{L}| = \mathrm{X}$), considering (δ, ρ) and $L = (L_1, L_2)\}$ as the private keys utilized in homomorphic hash and pseudorandom functions, respectively. Thus, the system utilizes $(X_x^{PL}, X_x^{CL}), (P_x^{PL}, P_x^{CL})$ to exploit the local gradient (x_a) of the user x. Secondly, through a secure channel, the user x broadcasts the public key (X_x^{PL}, X_x^{CL}) to the cloud server. Thirdly, broadcasted data from at least i users (which is denoted as $\mathrm{L}_1 \subseteq \mathrm{L}$) is received by the server slide. This point i represents the threshold the Shamir's $i - \text{out} - \text{of} - \mathrm{X}$ protocol utilized by the system. Else, stop the operation and restart again. Transmit $\{y, X_y^{PL}, P_y^{PL}, \tau = \text{sum}\}_{y \in \mathrm{L}_1}$ to respective users considering $\tau = \text{sum}$ the statistical tag to be estimated. For better understanding, a detailed description of all mathematical symbols is presented in Table 1.

TABLE 1. Description of mathematical symbols.

S/N	Symbols	Descriptions
1	(x_a)	Local gradient
2	$\{H, V, v, r, e, f, \square, \zeta\}$	Initially generated blocks
3	$L = (L_1, L_2)\}$	Private key for pseudorandom functions
4	(δ, ρ)	Private key for homomorphic hash
5	(X_x^{PL}, X_x^{CL})	Public key
6	i	threshold the Shamir's protocol
7	$\tau = \text{sum}$	Estimated statistical tag
8	cl_y	Secret key utilized to track record
9	$\frac{\mathrm{L}_2}{\mathrm{L}_3}$	Server-data broadcasting users
10	$\mathrm{L}_4 \subseteq \mathrm{L}_2$	User acquired data
11	(m, n)	Directed Edge between different tensor
12	Φ_n	Computational cost of every element
13	$-A(\Phi_{m,n})$	Number of floating-point activity

2) REGISTRATION FOR THE AGGREGATED RECOMMENDER MODEL

At this stage, the system needs to register two different entities (which are the Administrator and User). To register an Administrator, the system starts with fetching the basic settings from the blockchain (for example, the basic system setting for a transmitter could be Alice), then, it launches the long-term account ($pl_y = R_y, cl_y = G_y$) which was already created at the initialization stage by invoking Enroll command. Similarly, the registration process for the user is the same as that of an administrator. In the proposed method, the ownership of the long-term account is mutual to the administrators, and it is only used for tracking and not for issuance of transactions or proxy. Consequently, the registered task is only a one-time operation, and the generated long-term account pl_y can only be utilized once.

3) CHAIN OPERATION

This stage is strictly managed by the administrators. They perform operations at this stage to chain the awaiting operations against the blockchain. Their operation begins with confirming the authenticity of the received data (tx, c). For instance, to check the data of Alice (txa, ca), the administrators must first construe txa, ca $ap.l_a = (R'_a, R''_a)$, a. Then, the $\mathrm{cl_y}$ secret key is utilized to track the long-term directory of $apl_a = (R'_a, R''_a)$. Therefore, the Trace command is employed to fetch pla. Furthermore, the *isLegal*function is executed to confirm and ascertain if pla is forbidden or approved. If forbidden, then this awaiting operation tx will be prohibited; else, the administrators will advance to the next phase. Also, the search function of the model executes a search for trained models by first querying the Bloom filter to check if it exists on the cache server of the model, and then proceeds to the cache server to retrieve the model.

4) FEDERATED TRAINING GRADIENTS UPDATE

In this stage, the system checks if $\mathrm{L}_3 \geqslant i$ and $\mathrm{L}_3 \subseteq \mathrm{L}_2$. Assuming the expression is negative, stop operation and restart. Then, decrypt each;

$$\frac{\mathrm{P}_{x,y}, y \in \mathrm{L}_2}{\{x\}} \tag{2}$$

as,

$$x||y||X^{CL}_{x,y}||\beta_{x,y} \leftarrow \mathrm{AE.dec}\left(\mathrm{LA.agree}\left(P^{CL}_x, P^{PL}_y\right), \mathrm{P}_{x,y}\right). \tag{3}$$

Proceed and transmit $\left\{\left(X^{CL}_{x,y}\right)|y \in \frac{\mathrm{L}_2}{\mathrm{L}_3}\right\}$ and $\left\{\left(\beta_{x,y}\right)| y \in \mathrm{L}_3\right\}$ to the cloud server. At this point, $\frac{\mathrm{L}_2}{\mathrm{L}_3}$ denotes the users who have broadcasted data to the server but withdrew prior to uploading data to the cloud server. Then, acquire data from at least i users which denotes $\mathrm{L}_4 \subseteq \mathrm{L}_2$. Else, stop operation and restart. Estimate $\beta_x \leftarrow \mathrm{C.recon}\left(\beta_{x,y}\right)_{y \in \mathrm{L}_4}, i$ and the Proof of accumulated gradients $\{W, Z, K, R, \Omega)$ as follows.

$$W = \prod\nolimits_{x=1}^{x=|\mathrm{L}_3|} W_x; \tag{4}$$

$$Z = \sum\nolimits_{x=1}^{x=|\mathrm{L}_3|} Z_x; \tag{5}$$

$$K = \prod\nolimits_{x=1}^{x=|\mathrm{L}_3|} K_x; \tag{6}$$

$$R = \sum\nolimits_{x=1}^{x=|\mathrm{L}_3|} R_x; \tag{7}$$

$$\Omega = \sum\nolimits_{x=1}^{x=|\mathrm{L}_3|} \Omega_x; \tag{8}$$

Then, send result to each of the user $\in \mathrm{L}_4$ as;

$$B_{\text{result}} = \{\delta = \sum\nolimits_{x \in \mathrm{L}_3} a_x, W, Z, K, R, \Omega\} \tag{9}$$

5) VERIFICATION OF MODEL

In the verification process, the following generated parameters are check;

$$PA_{L_1}(x) = (\gamma_x, v_x) \tag{10}$$

$$PA_{L_2}(\tau) = (\gamma, v) \tag{11}$$

The system then calculates;

$$\mu = \sum\nolimits_{x \in \mathrm{L}_3} (\gamma_x \gamma + v_x v) \tag{12}$$

and

$$\varphi = a(h, k)^{\mu} \tag{13}$$

Therefore, the system verifies;

$$\begin{aligned} &(W, Z) \overset{?}{=} \left(W', Z'\right), \\ &a\left(W, k\right) \overset{?}{=} a\left(h, Z\right); a\left(K, h\right) \overset{?}{=} a\left(h, R\right), \\ &\varphi \overset{?}{=} a\left(W, k\right) \cdot a(K, h)^g \end{aligned} \tag{14}$$

Assuming any of the above commands are invalid, the system rejects the result of the aggregation. If not, the obtained result is accepted, and the operation proceeds to the initial round.

B. THE RECOMMENDER DATA MANAGEMENT NEURAL ARCHITECTURE TECHNIQUE

This section will elaborate on the techniques employed by the Recommender system for data sorting and the overall data management proof analysis.

1) DATA SORTING AND MANAGEMENT APPROACH USING REDMANA

Another key component incorporated in the proposed SERTT model is a recommender data management neural architecture (REDMANA) which fundamentally minimizes extreme human intervention. For example, in the areas of data management, model design and modification the techniques provide an automated, process-based and proficient data management solution. The technique of neural architecture search is basically an automatically designed neural networks, which permits algorithms to automatically build high-performing network frameworks according to sample arrays. These designed frameworks tend to match-up with the level of human professionals in some jobs and can even discern some network frameworks which are not formerly discovered by humans, and which can successfully minimize the implementation and usage costs of neural networks.

Due to the number of time and resources wasted while applying neural architecture search directly to data sorting and management, REDMANA is proposed in this study in order to minimize unnecessary time and resources consumed on creating data sorting and management model. In a given structural search scenario, REDMANA targets at identifying the optimally performing training hyper-parameters. This study implements the frequently used cell-based architecture search environment. Typically, a network comprises of a particular number of cells which can either be referred to as norm or reduction cells. This architecture is made of a network where the outputs of the previous two cells are initialized in each cell. A cell in this architecture is considered as a $Y-$ node coordinated acyclic graph which is completely connected to (DAG)$\{X_1, X_2, \ldots, X_Y\}$. Each node X_n accepts the dependent nodes as input and generates an output through a sum operation as follows.

$$X_m = \sum\nolimits_{n<m} o^{(n,m)}\left(X_n\right) \tag{15}$$

Each node signifies a different tensor, and each directed edge (m, n) between X_n and X_m illustrates an operation $o^{(m,n)}$ which is realized from the equivalent operation search space $\mathcal{O}(m, n)$.

Also, the proposed system provides a minimum data essential trimming approach. This begins by sampling the hyper-parameters which consumes minimal time. The effectiveness level of the selected hyper-parameters is calculated by means of random forest (RF) using the sampled cases. Those that are least effective are then pruned by setting the value with the least consumption time. This pruning phase is completed as soon as the search space records only a single hyper-parameter. Thus, a classification sharing associated with the computational cost for every element in Φ_n is introduced and presented as;

$$b\left(\xi_{m,n}\right) = \frac{\exp\left\{-A\left(\Phi_{m,n}\right)\right\}}{\sum_n \exp\left\{-A\left(\Phi_{m,n}\right)\right\}} \tag{16}$$

considering the quantity of floating point operations as function $-A\left(\Phi_{m,n}\right)$. The operations produces a set of J with diverse subsets $j^{\mathrm{ref}}, j^{\mathrm{pos}}$ and j^{neg} following continuously iterating the initial operations Lfor 30 times which is implemented as a training category for the RF. The RF tree is consisting of a group of J which also contains auxiliary samples from J. The reduction in node adulteration, which is measured by the quantity of samples made it to the node, is utilized to calculate the significance of the selected parameter for each Y node in the regression tree. Therefore, the parameter significance for each Y node is expressed and presented as.

$$\begin{aligned} N_y = {} & \left|P_y\right| K\left(P_y\right) - \left|P_{\{pos,y\}}\right| K\left(P_{\{pos,y\}}\right) \\ & - \left|P_{\{neg,y\}}\right| K\left(P_{\{neg,y\}}\right) \end{aligned} \tag{17}$$

$$K\left(P_y\right) = \frac{\sum_{j_{ref},n\in P_y}\left(j_{ref}, n - \overline{j_{ref}, P_y}\right)^2}{\left|P_y\right|} \tag{18}$$

After the significance approximation procedure, the hyper-parameter which has the lowest N_y is pruned according to the setting. This process of trimming considerably enhances search proficiency. By modifying the less significant hyper-parameter to a value that requires a smaller quantity of resources, the model may dedicate other system resources to more important parameters. The REMANA data management model design is presented in Algorithm 2. With reference to the process in Algorithm 2, subsequent to a total number of l epochs, the system will ultimately select the best performing M data management models from a larger number of participant models via hyper-parameter trimming and a competent search approach. In the succeeding investigation of this research, the system sets $M = 1$ to enable evaluation with other data management models, which implies that the research selects the optimal data management achieved from each search for comparative evaluation.

2) DATA SORTING AND MANAGEMENT PROOF ANALYSIS

Generally, different regulators and institutions employs data sorting and management proof with no-knowledge to authenticate the validity of each user's data access. Both the data and model that are used in this study are those achieved in the previous part by means of federated average method. The overall performance of the data sorting and management proof is described in Algorithm 3. The study also employed a Weight of Proof (WOP) mechanism in coding the original autonomous variables into the algorithm. These variables must be discretized or grouped, while the WOP value for group N can be estimated after discretization. This process is detailed in Algorithm 2.

V. SECURITY ANALYSIS

In line with the proposed privacy protection system and its related blockchain-mechanism, the proposed SERTT system can fulfill the following security conditions;

The proposed system incorporates a privacy mechanism which guarantees optimal data security with respect to blockchain DSM application unlike previous approaches that only utilized blockchain for data transmission. In the system, the entire user data is not stored directly on the chain, rather only the data index is saved on the chain, and the entire model is transmitted instead of data. Additionally, the unknown secret key which relates to the transmitter's secret address provides an SPK proof, which is exploited in the SERTT mechanism to chain communications.

Another important feature of the proposed system is the finding and sorting of recommender data and communication. This feature enables the administrator in the SERTT system to make use of the credential confirmation to validate the authenticity of each participant in the communication. This validation process maps a long-term address to the participants' actual identity. As soon as a malicious operation is detected, the administrator circulates the unknown address to recover the long address via trace and find operation by means of the smart contract's secret code, and then obtains the corresponding credentials from the smart contract.

Also, the technique allows for devolution of trust entities to minimize the threat that comes with data leakage. The SERTT system utilizes and corporation of federated learning and blockchain to substitute the conventional integrated data center. Similarly, the system makes use of the CSPPoS proxy mechanism which is incorporated in the FL process to appoint a provisional trusted manager for cooperation. This process can also minimize the threat of data leakage which is

Algorithm 2 REMANA Data Management Modelling in SERTT Framework

```
Input: $B_A = \phi$; $B_{2A} = \phi$; $B_{3A} = \phi$
Set: $\eta(n)$ and $\lambda(n)$ as initial data escalation factors
Declare: $B_l$ as DSM network trained for $l$ periods
    Train $w$, $z$ DSM models
    Declare: DSM training functions from period $w$ for the sum of $z - w$ periods    $k = \phi$; $R = \phi$
    Declare: $k$ to represent the network history group and $Q$ as the hyper-parameters group
    while $|Q_A| <$ lowest $N_{\text{init}}$ do
        Trim and filter redundant hyper-parameters
        Remove the threesome loss parameters of the DSM valuation focus $Q_i$
        Add $Q_i$ to $Q$
        DSM = Random architecture()
        DSM.accuray = Train (DSM, 0, $A$)
        Add DSM model to $B_A$ and $k$
        end while
        for $n = 1 \to X$ do
          for $n = 1 \to X_0$ do
        Randomly display DSM model from $B_A$; $B_{2A}$; $B_{3A}$
        offspring model = Random mutate of DSM model
        offspring.accuray = Train (offspring model, 0, $A$)
        Add offspring model to $B_A$ and $k$
      end for
    for DSM model = top 1 to $X_2$ DSM model in $B_A$ do
        DSM model.accuracy = Train(DSM model, $2A$, $3A$)
          Transfer DSM model $B_{2A} \to B_{3A}$
        end for
        Remove dead DSM model from $B_A$; $B_{2A}$; $B_{3A}$
    end for
    return optimal $M$ DSM models in $k$, $M = 1$ in the system model
```

Algorithm 3 Data Sorting and Management Proof Technique

```
Input: DSM records which are shared using federated training model
  1. Estimate the user's probability of default
       $= \frac{d}{1-d}, d = \frac{1}{1+a^{-z}}.$
  2. Compute the weight of proof $WOP_n = \text{In}\left(\frac{m_n/m_i}{x_n/x_i}\right)$
  3. Compute the DSM groups by the data
       record record $= M - N \log(\text{odd } c)$
  4. group$(c)$ = $(c \geqslant record_o) + (c \geqslant record_1) + \ldots + (c \geqslant c_{y-1})$
  5. Transform all individual gate circuit to an
       R1CS constraint. Create these constraints.
       using Lagrange interpolation, and then utilized the
       no-knowledge evidence technique to construct
       proofs.
  6. Utilize Lagrange interpolation algorithm to
       convert the R1CS constraints to Polynomial
       function
  7. return data sorting and management proof outcomes
```

associated with the centralized method. With the newly designed system, it is difficulty for data on the chain to be interfered with because these data are signed with the secret code which is generated by the data vendor, thus making it difficult for an attacker to either "poison" or obtain access to the model by modifying the directory, likewise, the data owner cannot disagree about the data considering that it is sealed with its own secret key.

VI. PERFORMANCE ANALYSIS

With the help of experiments and simulations the performance of the proposed SERTT system is evaluated and the result of its design is demonstrated in this section.

A. DATASET AND EXPERIMENTAL BACKGROUND

This study utilized two datasets as provided from certified pharmaceutical web sites known as Druglib.com and Drug.com. The datasets contain patient recommendations and reviews of individual medications pertaining to health conditions such as, blood pressure, acne, pain, anxiety, and the rest of other diseases. The content of these dataset includes several patients' recommendations and reviews, thus, the contents were filtered and sorted by means of python soup library and automatic web crawler. After the crawling outcomes, the total dataset has 231,206 drug recommendations, out of 628 total number of targeted drugs. This dataset was implemented for the training and testing of the system models. The experimental environment is set-up using a Python (3.7) programming language, while the Sklearn library is utilized to implement the data partitioning and classification model. The PySyft is employed in the experiment as a platform for federated learning communication while, the CPU model is Intel(R) Core(TM) i7-6700K 4.00GHz.

B. ANALYSIS AND COMPARISON OF SYSTEM PERFORMANCE

In Table 2, the percentage of the values of contribution which are generated from the federated learning of the contributors is presented. The first three rows shows the no-attacked case of the local dataset, while the last row indicates a case where even though there are four data providers, but the dataset of the last data provider Z is attacked.

From the analysis presented in the Table above, the contribution values can easily be seen, and they are negatively computed each time the attacked nodes are concerned. Generally, the proposed SERTT training framework guarantees proficient and accurate sharing and training of data management model without triggering substantial rise in training time, thus, the system contains superior stability mechanism. The study also measured the performance of the SERTT

TABLE 2. Contribution portion comparison for varying federated learning.

Types of Federated Learning	Contribution (%)			
	W	X	Y	Z
Two-party	42.66	49.50	-	-
Three-party	29.54	31.32	33.44	-
Four-party	22.65	25.32	20.20	23.12
Sum	62.45	60.50	60.21	-85.82

mechanism for both two-party, three-party and four party data providers with respect to their respective federated learning. Each of these parties possess a percentage of medical record data, data users search via the blockchain to the two parties that possess the expected data for federated learning modeling. Furthermore, the performance of the recommender data management proof and DSM proof algorithm were analyzed using six 64-bit Ubuntu18 servers which are powered with 8GB of RAM each and 8-core CPUs. In a setting with a total of 30 nodes, each machine propelled four docker containers which represents autonomous nodes of blockchain. For a period of two hours, the test chain of these 30 autonomous nodes was permitted to run its operation. During this operation, each node transmitted a no-knowledge evidence transaction every 15 seconds. While a node is transmitting transaction, the rest of the nodes evaluate and arrange the outcomes by means of proxy procedure.

The description in Fig.2 shows that the trained model is basically the same as if the models were federated provided that each party reliably supplies their respective local data for federated learning. The key variance is in the time of training. If there is an increase in the number of contributors transmission overhead and also an intensity in the computation of the federation average, and more contributors in the federation learning, there is an increase in the computational time.

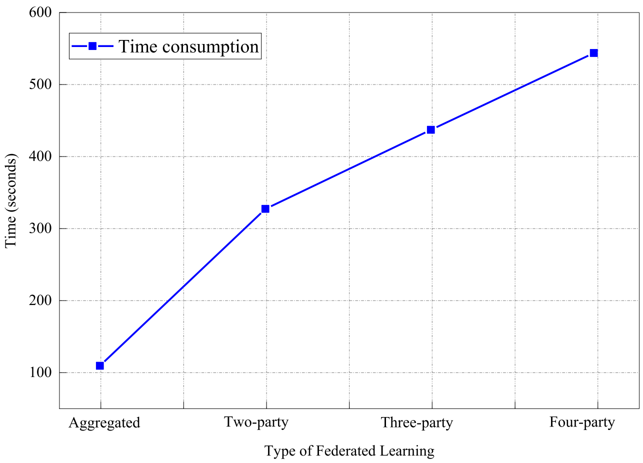

FIGURE 2. Comparison of consumption time for different types of federated learning.

The experimental results which illustrate the performance comparison result of the proposed CSPPoS are shown in Fig. 3 and 4. In Fig. 3, the number of concurrency is set at 800, while the data throughput is measured against varying nodes qualities. The experiment shows that in a cluster of different nodes, the overall throughput of the CSPPoS algorithm drops while the latency rises with every rise in the number of nodes. The metric utilized in testing the throughput of the algorithm is comparatively higher when $\alpha = 0.4$. On the other hand, the latency of the proposed algorithm is lower at $\alpha = 0.6$ and $\alpha = 0.8$. Similarly, the illustration in Fig.4 indicates that when the number of preliminary devious nodes in the cluster rises, the computational time necessary for the alteration of a primary node to a devious node reduces.

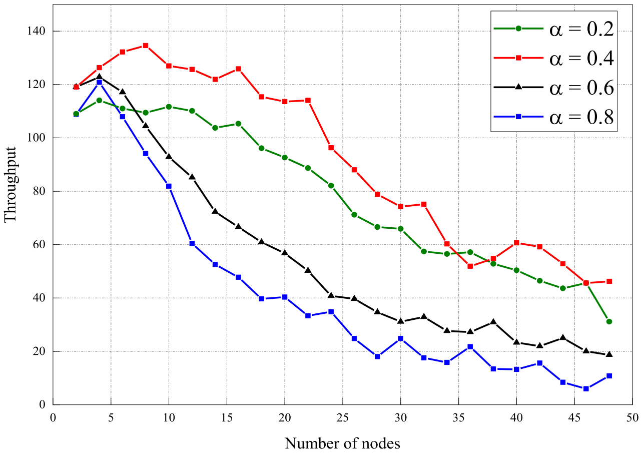

FIGURE 3. Performance comparison of throughput CSPPoS algorithm using different values of α.

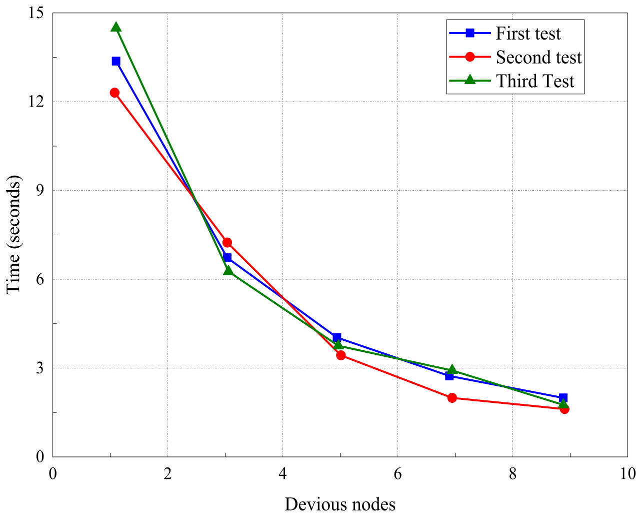

FIGURE 4. The effect of devious nodes on the computational time consumption in CSPPosS.

C. COMPARISON WITH OTHER METHODS

The experiment in Fig.5 compares the performance of the proposed SERTT technique with other recently proposed methods such as, KiRKi [1], ANCILE [2], TCUGA [3] and Private-Rec [4]. As observed in the Figure, the overall time grows correspondingly in proportion to the quantity of the

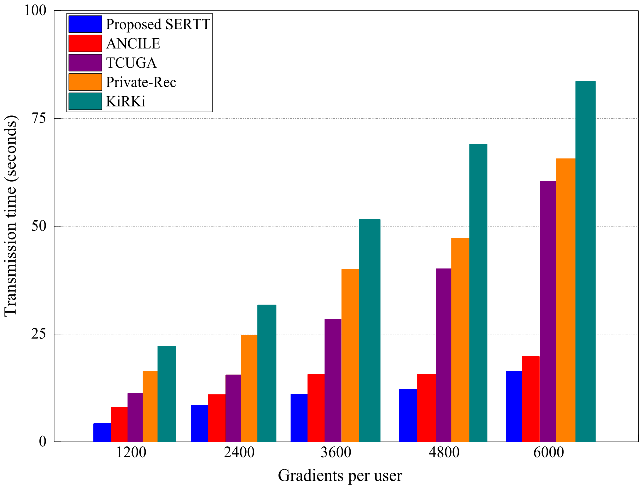

FIGURE 5. Comparison of gradients per user effect on transmission time.

utilized datasets. As soon as there is a modification in SERTT technique, it takes approximately 10 seconds to attain a proxy operation.

In conclusion, the experiment indicates that in comparison to other existing methods, the proposed SERTT training mechanism outperforms them with respect to transmission and computational time. Likewise, with the same metric for gradients per user and dropout metrics, the compared existing methods show greater amount of training success for the proposed SERTT.

VII. CONCLUSION

With the recent rapid evolution of artificial intelligence technology, secure medical data recommender training and modelling techniques are more and more important. Preserving user data from unauthorized access has occurred to be a pressing concern. Therefore, this chapter of the thesis was focused on providing a new solution for handling the issue of analyzing user medical data for training while sustaining user privacy. This section proposed a SERTT technique, which is an incorporation of a federated learning and blockchain for data training and modelling. For proficient data model designing, the study proposed a REMANA approach which is based on neural structural search, and which provides an efficient data sorting management (DSM) model design. For a secured training of federated learning, the study utilized the blockchain's classified proxy mechanism CSPPoS and the no-knowledge proof system. Results of the experiment show that SERTT technique is secure with optimal performance. Nonetheless, there are still some latitudes for improving performance and security in the proposed technique. In order to achieve this in further research, the study will further optimize the efficiency of the distributed learning system of federation learning and will also reduce federated learning's training time using a proficient gradient distribution algorithm. Additionally, the study will further require optimizing the proficiency of data management model search and design to further improve the accuracy performance of the system's DSM model. Moreover, the no-knowledge proof process in this study was deployed in an ethereum smart contract, which is a less efficient verification approach at the core layer, thus, further attention can be given to integrating no-knowledge proof algorithms in the blockchain source code layer, to optimize the proficiency and performance of the proposed SERTT system.

REFERENCES

[1] S. B. Patel, P. Bhattacharya, S. Tanwar, and N. Kumar, "KiRTi: A blockchain-based credit recommender system for financial institutions," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1044–1054, Apr. 2021, doi: 10.1109/TNSE.2020.3005678.

[2] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustain. Cities Soc.*, vol. 39, pp. 283–297, May 2018.

[3] C. Lin, D. He, X. Huang, M. K. Khan, and K.-K.-R. Choo, "A new transitively closed undirected graph authentication scheme for blockchain-based identity management systems," *IEEE Access*, vol. 6, pp. 28203–28212, 2018.

[4] R. Bosri, M. S. Rahman, M. Z. A. Bhuiyan, and A. Al Omar, "Integrating blockchain with artificial intelligence for privacy-preserving recommender systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1009–1018, Apr. 2021.

[5] Q. Wang and M. Su, "Integrating blockchain technology into the energy sector—From theory of blockchain to research and application of energy blockchain," *Comput. Sci. Rev.*, vol. 37, Aug. 2020, Art. no. 100275.

[6] U. Bodkhe, S. Tanwar, K. Parekh, P. Khanpara, S. Tyagi, N. Kumar, and M. Alazab, "Blockchain for industry 4.0: A comprehensive review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020.

[7] D. Berdik, S. Otoum, N. Schmidt, D. Porter, and Y. Jararweh, "A survey on blockchain for information systems management and security," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102397.

[8] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A blockchained federated learning framework for cognitive computing in industry 4.0 networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2964–2973, Apr. 2021.

[9] M. Zhaofeng, W. Xiaochang, D. K. Jain, H. Khan, G. Hongmin, and W. Zhen, "A blockchain-based trusted data management scheme in edge computing," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2013–2021, Mar. 2020.

[10] L. Xu, T. Bao, and L. Zhu, "Blockchain empowered differentially private and auditable data publishing in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7659–7668, Nov. 2021.

[11] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, 3rd Quart., 2020.

[12] L. U. Khan, S. R. Pandey, N. H. Tran, W. Saad, Z. Han, M. N. H. Nguyen, and C. S. Hong, "Federated learning for edge networks: Resource optimization and incentive mechanism," *IEEE Commun. Mag.*, vol. 58, no. 10, pp. 88–93, Oct. 2020.

[13] H. Kim, J. Park, M. Bennis, and S.-L. Kim, "Blockchained on-device federated learning," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1279–1283, Jun. 2020.

[14] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–7, Sep. 2020.

[15] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir, and F. Noor, "Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model," *IEEE Access*, vol. 8, pp. 28462–28474, 2020, doi: 10.1109/ACCESS.2020.2968537.

[16] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, H. B. McMahan, T. Van Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," 2019, *arXiv:1902.01046*.

[17] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106775.

[18] C. Iwendi, J. H. Anajemba, C. Biamba, and D. Ngabo, "Security of things intrusion detection system for smart healthcare," *Electronics*, vol. 10, no. 12, p. 1375, Jun. 2021.

[19] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Gener. Comput. Syst.*, vol. 115, pp. 619–640, Feb. 2021.

[20] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for Internet of Things: Recent advances, taxonomy, and open challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1759–1799, 3rd Quart., 2021.

[21] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu, and X. Zheng, "Privacy-preserving federated learning framework based on chained secure multiparty computing," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6178–6186, Apr. 2021.

[22] D. Ngabo, D. Wang, C. Iwendi, J. H. Anajemba, L. A. Ajao, and C. Biamba, "Blockchain-based security mechanism for the medical data at fog computing architecture of Internet of Things," *Electronics*, vol. 10, no. 17, p. 2110, Aug. 2021.

[23] E. A. Mantey, C. Zhou, J. H. Anajemba, I. M. Okpalaoguchi, and O. D.-M. Chiadika, "Blockchain-secured recommender system for special need patients using deep learning," *Frontiers Public Health*, vol. 9, Sep. 2021, Art. no. 737269.

[24] E. A. Mantey, C. Zhou, V. Mani, J. K. Arthur, and E. Ibeke, "Maintaining privacy for a recommender system diagnosis using blockchain and deep learning," *Hum.-Centric Comput. Inf. Sci.*, to be published.

[25] Y. Qu, M. P. Uddin, C. Gan, Y. Xiang, L. Gao, and J. Yearwood, "Blockchain-enabled federated learning: A survey," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–35, Apr. 2023.

ERIC APPIAH MANTEY received the Bachelor of Education degree in computer network management from Koforidua Technical University, Ghana, and the M.Sc. degree in computer science and technology from the Jiangsu University of Science and Technology, China. He is currently pursuing the Ph.D. degree in computer science with Jiangsu University, China. His research interests include deep learning/machine learning, recommender systems, artificial intelligence, computer networking and management, rough set and fuzzy set theory, information technology, and data science.

CONGHUA ZHOU received the M.S. and Ph.D. degrees from Nanjing University, China, in 2001 and 2006, respectively. His research interests include big data and artificial intelligence.

JOSEPH HENRY ANAJEMBA (Member, IEEE) received the Ph.D. degree in information and communication engineering from Hohai University, China, in 2021. Prior to the Ph.D. degree, he was a Lecturer and the Dean of Students Affairs with the Department of Information Technology, Uma Ukpai College of Business and Technology (UUCBT), Nigeria. He is currently an Assistant Professor with the Department of Information Security Engineering Technology, Abu Dhabi Polytechnic. He has authored or coauthored many top scientific researches. His research interests include the Internet of Things, physical layer security, and artificial intelligence. He is an Associate Fellow of the Higher Education Academy (AFHEA), U.K. He has served as a reviewer and the guest editor for several journals.

YASIR HAMID received the Ph.D. degree in computer science and engineering from Pondicherry University, in 2019. He is currently an Assistant Professor with the Department of Information Security Engineering Technology, Abu Dhabi Polytechnic. His research interests include machine learning, deep learning, and big data analysis. He is an editorial board member of many journals and is also a member of many scientific societies in the above areas.

JOHN KINGSLEY ARTHUR (Member, IEEE) received the bachelor's degree in computer science from Valley View University, Ghana, and the master's degree in information technology from Open University, Malaysia. He is currently pursuing the Ph.D. degree with the Faculty of Computer Science and Telecommunication Engineering, Jiangsu University, China. His research interests include recommender systems, machine learning, algorithm optimization, and network security.