Federated Learning-Based Privacy Preservation with Blockchain Assistance in IoT 5G Heterogeneous Networks

Sampathkumar Arumugam[1,*], Shishir Kumar Shandilya[2] and Nebojsa Bacanin[3]

[1]*Department of Computer Science, Dambi Dollo University, Ethiopia, and Department of Applied Cybernetics, Faculty of Science, University of Hradec Kralove, Czech*
[2]*Visiting Researcher, Liverpool Hope University, UK, and VIT Bhopal University, India*
[3]*Singidunum University, Danijelova 32, Belgrade, 11000, Serbia*
E-mail: sampathkumar.arumugam@uhk.cz; shishir.sam@gmail.com; nbacanin@singidunum.ac.rs
[*]*Corresponding Author*

Received 30 December 2021; Accepted 02 March 2022;
Publication 18 April 2022

Abstract

In the area where privacy is of greater concern, federated learning,a distributed machine learning strategy for preserving privacy,is widely employed in several privacy concern applications. In the meantime, neural architectures became familiar with deep learning approaches for automatic tuning of the architecture of deep neural networks (DNN). While searching with neural architecture and federated learning has experienced several challenges, optimized neural architecture research in federated learning is extensively on demand. DNN faces numerous issues while training such user privacy and ensuring the integrity of the aggregated results obtained from a server. To provide solutions for the above-mentioned issues, enormous federated

Journal of Web Engineering, Vol. 21_4, 1323–1346.
doi: 10.13052/jwe1540-9589.21414

learning techniques worked towards preserving privacy and were applied in different situations. Still, it is an open challenge that enables users to verify if the cloud server functions appropriately while ensuring users' privacy while training. Federated Learning Method is a new way to improve the accuracy and precision, since the previous approach failed to opt the solutions. Here, Elliptical Curve Cryptography with Blockchain-based Federated Learning (ECC-BFL)is proposed to ensure the confidentiality of users' local gradients while performing federated learning. The parameters such as classification accuracy, running time, Communication overhead, Computation overhead, and transaction speed are considered. The values obtained for these parameters are compared against three standard methods, namely Biparing Method (BM) Homomorphic Cryptosystem (HC), and Multiple Authorities with Attribute-Based Signature scheme (MA-ABS)against proposed Elliptical Curve Cryptography with Blockchain-based Federated Learning (ECC-BFL). As a result, the proposed ECC-BFL achieved 95% of classification accuracy, 65 sec of running time, 76% of communication overhead, 63% of computation overhead, and 92% of transaction speed.

Keywords: Blockchain, 5G network, federated machine, privacy preservation, registration.

1 Introduction

In the machine learning (ML) approaches used traditionally, the model's accuracy and efficiency are based on the data trained and computing power of the centralized server. With traditional ML approaches, the central server is the storage for user data used for both training and testing, where a wide range of ML models are developed in due course (Aledhari et al., 2020). Centralized ML approaches often experience several challenges such as computational time and power, and above all, security and privacy to user data that has not been considered seriously. Federated Learning (FL), introduced in (Cabaj et al., 2018), has emerged with a solution by addressing this issue. FL provides privacy for user data where data are decentralized from the central server to end devices (Lu et al., 2020). Privacy preservation offers possibilities to influence the benefits of AI achieved by efficient machine learning models across various domains (Liu et al., 2020). Further, using the iterative local model for training on end-devices, sharing computational power among interested parties was made possible rather than a centralized server. Using this concept, FL has grown to heights in ML recently due to its

promising security and privacy features (Liu et al., 2020). Besides privacy, FL permits ML to be used in smaller domains where only inadequate training data is available to construct a standalone ML model.

When IoT devices are used, a wide range of vertical services in automation, energy, transportation, city management, manufacturing, and agriculture are provided (Xu et al., 2019). Thus, devices related to IoT are in demand in various radio technologies (Wang et al., 2020). The services differ in Quality-of-Service (QoS),ranging from ultra-low latency to ultra-dense connectivity (Ning et al., 2019). The fifth-generation (5G) mobile communication systems are recommended for connecting IoT devices with the help of HetNets(heterogeneous networks) bandwidth spectrum and accomplish vertical services (Zhang et al., 2020). The introduction of 5G includes a wide range of autonomous moving platforms (AMP), like autonomous underwater vehicles (AUV), autonomous flying vehicles (AFV), autonomous land vehicles (ALV), and autonomous surface vehicles (ASV) (Liu et al., 2020). Generally, AMPs perform some specified tasks. Few AFVs are developed to deliver goods, while others monitor (Zhang et al., 2016). ASVs and AUVsare robotic kinds of vehicles usually function on the sea's surface or under the sea, respectively. ALVs performs several tasks like mining, defense, and agriculture (Sun et al., 2019). The rapid growth of 5G has enabled the communication of various AMPs. Thus, emerging services broadly extend over space, air, sea, and ground by reshaping the fields like transport, automotive, energy, agriculture, manufacturing, and city management (Servos et al., 2017). The infrastructure of 5G is heterogeneous in terms of heterogeneous radio access networks (RANs), providing more advanced solutions to satisfy the requirements of QoS and Quality-of-Experience (QoE) several services (Sampathkumar et al., 2020). The vast amount of data generated by IoT devices is the additional burden for 5G infrastructure since the requirements of QoS and QoE have to be maintained. Deep learning (DL) approaches are extensively employed for extracting valuable information to improve the QoS of the network and QoE of the user (Sultana et al., 2019). As5G infrastructure is by nature heterogeneous, several network parts, IoT devices, as well as privacy concerns, and service data are not always available (Zhao et al., 2019). The service provided by the network is better with the availability of more data when effective privacy preservation techniques are employed. 5G network has a unique feature named Network slicing, where heterogeneous resources are shared (Niknam et al., 2020). Conventional IoTs experience several privacy issues, while 5G-bassed AMP offers several benefits to several businesses providing privacy preservation (Sampathkumar et al., 2019).

The contribution of this work is as follows:

- To develop the Elliptical Curve Cryptography-based blockchain FL technology, which helps users verify the correctness of results obtained from a server with tolerable overhead.
- To ensure the confidentiality of local gradients of the users at the time of FL. Some users exit during training for a few reasons, but the privacy of these uses is preserved.

The objectives of this work are as follows:

- To improve the heterogenic property among the variable data sources by improving the sensitivity among registered web data
- To enhance the classification accuracy by utilizing the FL-based architecture concerning the isolation in blockchain heterogeneous network.
- To relate the discrepancy solitude resolution on scheduled records for privacy-aware exposure.

The organization of this paper is as follows. Section 2 summarizes the related works about FL and blockchain technology for privacy improvement in 5G systems. Section 3 presents the system model developed for improving privacy with FL by adopting the Elliptical Curve Cryptography method. Section 4 details the experimental analysis by comparing the proposed method with three existing methods. Section 5 concludes the paper with future work.

2 Prior Literature

Privacy in IoT-based systems can be preserved by using the method of anonymization, and thus numerous researchers have employed this technique to preserve privacy in blockchain-based IoT applications. A few applications which commonly used this are finances, electronic health record, energy systems, and vehicular networks. Yue et al. (2016) developed an app for smart mobiles named Healthcare Data Gateway (HDG) based on blockchain with an MPC approach. The system directly computed encrypted data on the private blockchain cloud, and results were obtained where the original data was not revealed. Guo et al. (2018)introduced an Attribute-Based Signature method with Multiple Authorities (MA-ABS) for the blockchain-based healthcare system. Here, the signature confirms the identity of the patient who mentions a message rather than claiming access related to the attributes substituted using their other authorities. Moreover, the system was able to protect against collusion attacks where the secret of pseudo random function (PRF) was shared between authorities. Patel(2019) designed a cross-domain

medical image sharing system where patients could access their medical data electronically in a secured way. Moreover, a user could request access with their identification and be authorized to access without storing the original medical image in the blockchain. Yue et al. (2016) developed a simplified Indicator Centric Schema (ICS) that could easily manage any type of healthcare data by using a single simple "table," where data were uploaded only one time, but retrieval was off several times. Wu et al. (2018) adopted the bilinear pairing concept to secure the location details and identity of the patient. Additionally, revealing location information while communicating and transmitting messages among patients was prevented. Likewise, blockchain data was protected from Sybil attacks; thereby, security information was maintained. Table 1 shows the comparison of various existing methods.

Miers et al. (2013) suggested using ZeroCoin methodology, an anonymous transaction approach,to identify and anonymize the payee, transaction amount, and receiver. The objective behind this system is to eliminate data

Table 1 Comparison of existing methods

Author	Proposed Method	Merits	Demerits
Yue et al. (2018)	Healthcare Data Gateway (HGD) architecture	It does not depend on third party accessible unit, which tends to better security	Traceability of raw data is more due to extensive storage
Guo et al.(2018)	Multiple Authorities with Attribute-based signature scheme	It improves the stability of health records	The computation cost of this method is high
Patel et al. (2018)	cross-domain image sharing	The usage of double structures keys results in more security	Sometimes, this method is error-prone due to the increased number of accessible units
Yue et al.(2016)	Indicator Centric Schema (ICS)	Less computational complexity and overhead	Appendices of unwanted raw data are more
Wu et al.(2018)	Bilinear pairing Method (BM)	It is easily accessible to all types of records in heterogeneous networks	More computational storage and communication overhead
Miers et al.(2013)	ZeroCoin	cost of computing the differentiated values are less	Selection of long way trusted parties in less
Dagher et al.(2017)	data perturbation approach	The accuracy range is more with a short period	Increased amount of resources

leakage while connecting data by preserving its identity. Moreover, leakage of information was also preserved. Dagher et al. (2018) appliedthe data perturbation approach to highlight the different privacy concerns, thereby using optimal privacy measures to preserve blockchain-based healthcare systems. Besides, it was suggested that differential privacy methods could introduce noise efficiently while transmitting health-related data such that no adversary can gather any useful information from the data stored. Thus, a valuable approach to integrating differential privacy was recommended to provide security for the blockchain-based IoT healthcare systems.

The research gap exists that anonymization guarantees strong privacy to several blockchain-based IoT systems and still faces some attacks. One such is a linking attack, where data from external sources and protected anonymized data are combined to obtain critical data and limit the record details. At times analyst/receiver faces difficulties in extracting the information required from an anonymized dataset. Missinglinkability is another major concern in combined datasets. Therefore, several efforts have to be put by the researchers to make anonymization undistinguishable and effective in blockchain-based IoT systems.

3 System Model

Numerous studies have been conducted, and several operational challenges are experienced with growing consumer-related techniques. For instance, several electronic health records (EHR) systems existing today employ a centralized server model, and thus these types of deployments experience limitations towards security and privacy policies (say a bottleneck with failure and performance). Further, EHR systems have become popular, increasing the importance of data, specifically the data related to healthcare systems. Servers may secretly gather the user's personal information when they perform their routine actions. The architecture of the proposed Elliptical Curve Cryptography with Blockchain-based FL (ECC-BFL) healthcare systems is depicted in Figure 1.

3.1 Construction of Blockchain Network

Blockchain, sequential blocks, includes a list of valid, complete transaction records. Blocks are interconnected using a hash value (reference), and thus a chain is formed. The first block is the genesis block, and the one which is before the given block is the parent of the given block.

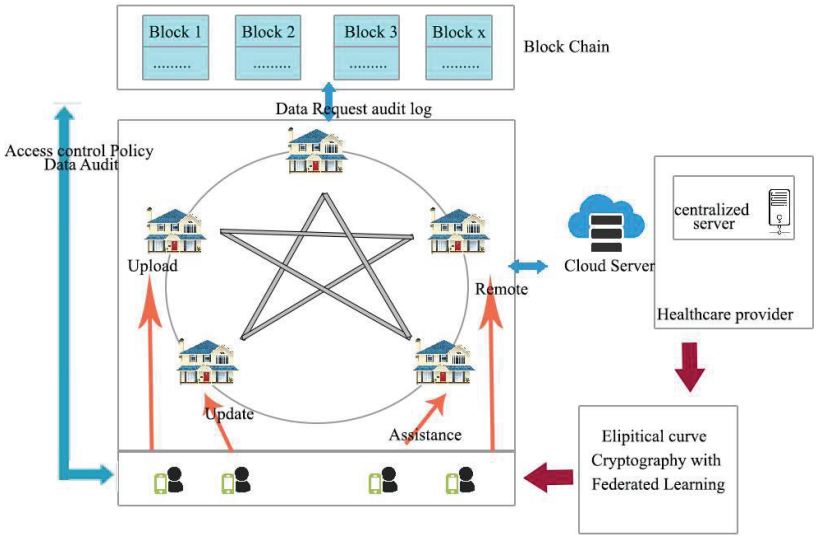

Figure 1 Architecture of ECC-BFL.

- Block version specifies the validation rules;
- Previous block hash provides the hash value of that block
- Timestamp describes the time taken to create the current block
- The nonce is a random field with 4-bytes which is adjusted while mining for calculating each hash, thereby providing a solution to the PoW puzzle
- Body root hash specifies the hash value of Merkle tree root generated by the transactions in the block body;
- The target hash is the threshold hash value of a new valid block that identifies the difficulties in the PoW puzzle

After sending the transaction, it is forwarded to every neighbor node via the P2P network. After receiving the transaction, the other nodes use the sender's public key to authenticate the transaction received based on the predefined block validation rules. If valid, then the transaction is broadcast to the other nodes until the transaction reaches every node in the network, and the authentication of the transaction is also verified. If not valid, the transaction is discarded. In the blockchain network, the valid transactions alone are stored in the new block. When the records are stored in an encrypted format, privacy is attained. Agencies/Hospitals use the identity of the patient for reading these health records. The patient uses an algorithm to generate

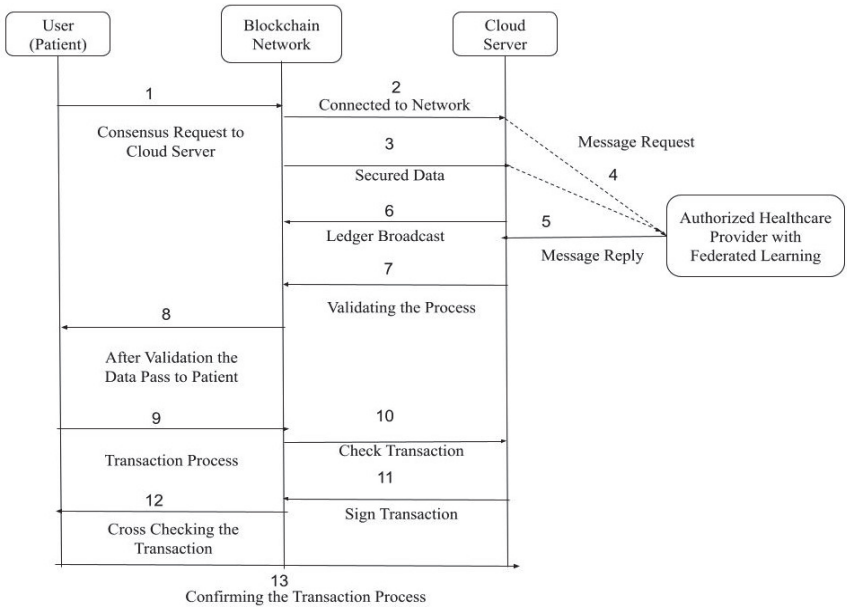

Figure 2 Execution flow diagram for ECC-BFL method.

the key, which is then shared securely to the third party while sharing the record. In healthcare applications, generating patient identity is challenging as patients are registered several times with various accounts either at the same hospital or at different hospitals, which is the cause for fragmentation. Figure 2 shows the execution flow diagram for the ECC-BFL method. The details maintained in the record are Patient Id, Doctor Id, Disease Id, Date, File Id, Record, Record Hash, and Parent record hash).

3.2 Elliptical Curve Cryptography with FL

- The global model parameters are initialized on the server and then transferred to every client connected to the network.
- These clients then learn the transferred model on its data for various training epochs. When this process is completed, updated model parameters or gradients are forwarded to the server. Gradients are the difference between the downloaded and updated models. It is to be noted that the scale of training data may differ, and computational resources may be unbalanced for clients. Thus, the server fails in receiving the information uploaded by clients.

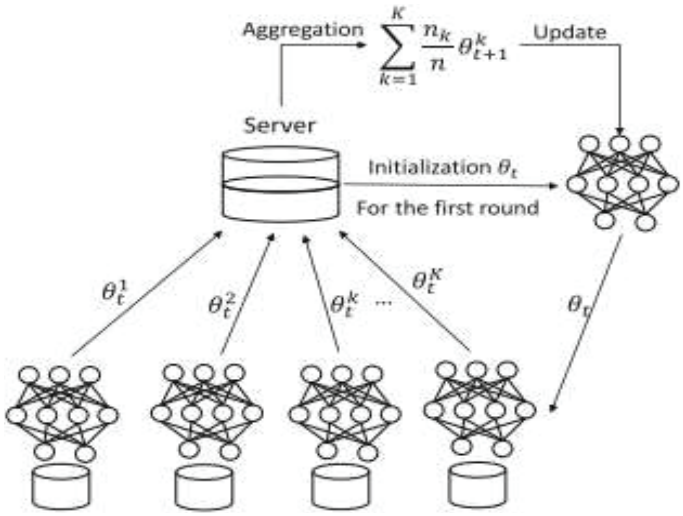

Figure 3 Federated neural network for data aggregation.

Table 2 Notations for FNN construction

Notations	Description
Θ	Global model
Nk	Data size
K	Number of clients
L_f	Loss function
D_n	Private local dataset
$i \in [1, L]$	Output layer
w_{i-1}	Weight matrix
b_{i-1}	Biases
$n \subseteq D_n$	Stochastic gradient descent

- The received uploads are aggregated by the server, which can be either in synchronous or asynchronous mode, with which the global model is updated.
- Until converging, the above two processes are repeated.

In Figure 3 federated neural network for data aggregation was illustrated. The notations of the above figure are as follows: θ, *nk*, *K*, and *t*, representing global model parameters, the data size of client *k*, total clients, and communication round in FL, respectively. Global model parameters are just randomly initialized at the start of the communication round and then use the updated model parameters.

Generally, a neural network is defined as a function $f(x, \omega) = y'$, where *x* and y' represent the inputs of the user and its respective outputs respectively through function f with ω as its parameter. With no generality loss, let every data record be (*x*,*y*), an observation pair, and $D = \{\Box x_i, y_i \Box, \forall i = 1, 2, \ldots, T\}$ represents the complete training set. The loss function (L_f) for the training set is described as

$$L_f(D, \omega) = L_f(x_i, y_i, \omega), \tag{1}$$

where $L_f(x, y, \omega) = l(y, f(x, \omega))$ for a particular L_f which is sets as $l(y, y') = \|y, y'\|^2$, where $\| \cdot \|^2$ is the l^2 norm of a vector. The neural network is trained with the objective of finding the optimum parameters ω, thereby minimizing the loss function. Particularly, every parameter is estimated iteratively as:

$$\omega^{j+1} \leftarrow \omega^j - \lambda \nabla L_f(D^j, \omega^j), \tag{2}$$

where ω^j, *D*, and λ represent the parameters after iteration *j*, an arbitrary subset of *D*, and learning rate parameters, respectively. In the FL technique used here, every user $n \in N$ has a private local dataset D_n, which is trained using a specific neural network where $D = N_j$ is an arbitrary subset selected by the server at iteration *j*, and then every user *n* of N_j selects subset D_j at random $(n \subseteq D_n)$ for executing stochastic gradient descent. Hence, the parameter update can be given by

$$\omega^{j+1} \leftarrow \omega^j = \lambda * \rho \; j \; n, \tag{3}$$

where $\rho \; j \; n = |D_j n| \nabla L_f(D^j n, \omega^j)$ is calculated by every user,shared with the cloud server. $\nabla L_f(D^j n, \omega^j)$ is the iterative process, and $D_j n$ is the weight parameter. The global parameters ω^{j+1} are returned by the cloud server to every patient. $x \in F^{n0 \times b \; p}$ is taken as the input with dimension n_0 by the network, here *b* indicates the size of the batch. Layer $i \in [1, L]$ produces n_i output neurons which are stated using a weight matrix $w_{i-1} \in F_p^{ni \times ni-1}$, and biases $b_{i-1} F^{ni}$.

$$\text{The output of Layer } i \in [1, L], \; y_i \in F_p^{ni \times ni-1} \text{ is :} \tag{4}$$

$$Y\mathrm{i} = \alpha_{\mathrm{quad}}(w_{i-1}, y_{i-1+} b_{i-1})^T \forall i \in [1, L-1] \tag{5}$$

$$yL = \sigma_{out}(w_{L-1 \cdot yL-1} + b_{L-1} T) \tag{6}$$

where $\alpha_{quad}(\cdot)$ is the quadratic activation function, $\sigma_{out}(\cdot)$ is the activation function of the output layer, and $1 \in F_b^p$ is the vector of all ones. Typically, softmax activation is used in output layer.

Table 3 Notations for ECC

Notations	Description
ME	Medical Expert
MSj	Medical Sensor
mxandmy	Master Key
UID	User Identity
SMi	Smart Card
CR	Card Reader
$r1$ and $r2$	Random Number
TSi	Current Timestamp
Usr	User
Usk	User Session Key
$\rightarrow$	Insecure Channel
$\oplus$	Bitwise XOR operation

3.3 User Registration Phase in ECC

Table 3 shows the ECC notations.

Step 1: MS initiates BTi to extract $\langle Ri, Pi\rangle$ from GEN(BTi) $\rightarrow \langle Ri, Pi\rangle$ and then the *Pi* values are stored in memory. Next, MS sends $\langle$UID, Adi = H(*Ri*)$\rangle$ to WGAc over the communication channel, which is secured.

Step 2: Once the registration request $\langle$UID, Adi$\rangle$ is received from MS, WGAc computes the user authentication parameters that are as follows:

$$\mathrm{C1} = \mathrm{H}(\mathrm{UID} \parallel \mathrm{mx} \parallel \mathrm{my}),$$

$$\mathrm{MSI} = \mathrm{H}(\mathrm{CI})\square\mathrm{Adi},$$

$$\mathrm{NI} = \mathrm{mx}\square\mathrm{CI}\square\mathrm{my},$$

$$\mathrm{VRI} = \mathrm{H}(\mathrm{UID} \parallel \mathrm{Adi}).$$

Step 3: WGAc stacks the user authentication parameters, namely MSI, NI, VRI, and H($\cdot$) into smartcard SMI.

Step 4: Finally, MS stores Pi into the smartcard.MS then initiates BIO *to extract *Ri* from REP(BIOi *, *Pi*) $\rightarrow \langle Ri\rangle$. Then, SMI uses a fuzzy extractor to compute Adi * and VRI *. At last, VRI * is compared with VRI as $Ri^* =$ REP(BIOi* , *Pi*); Adi * = H(Ri^*); VRI * = H(UID $\parallel$ Adi *); then Verifies,

whether VRI = VRI* or not

$$Y_i = r1 \times P,$$

$$\mathrm{H(CI)} = \mathrm{MSI}\square\mathrm{Adi}*, \mathrm{ADi} = \mathrm{UID}\square\mathrm{H}(r2),$$

$$\mathrm{MS1} = r2\square\mathrm{H(CI)},$$

$$\mathrm{MS2} = \mathrm{H(ADi} \parallel \mathrm{H(CI)} \parallel Y_i \parallel r2 \parallel \mathrm{TSi}),$$

$$\mathrm{MS3} = \mathrm{Ni}\square(r1 \times xP).$$

WGAc tries to compute the parameters such as *Kg*, *Cg*, and *Wg* to validate whether the communication between Usr and MSj is authentic or not. The expressions are defined as,

$$\mathrm{Kg} = \mathrm{H(H(SDj} \parallel \mathrm{my)} \parallel \mathrm{TSg}),$$

$$\mathrm{Cg} = \mathrm{ECkg(ADi} \parallel r2 \parallel Y_i),$$

$$\mathrm{Wg} = \mathrm{H(H(SDj} \parallel \mathrm{my)} \parallel \mathrm{ADi} \parallel \mathrm{Cg} \parallel \mathrm{TSg}).$$

A collision-free one-way hash function is considered to specify the significance of an arbitrary value $r2$ and control and monitor secret session-keys mx and my of $WGAc$. Let a collision-free one-way hashing function be described as: $\{0,1\}* \rightarrow \{0,1\}n$. A binary string a $\in \{0,1\}$ *is taken as input which produces a length of $H(a) \in \{0,1\}n$. The requirements are satisfied as follows. Let $b \in B$, which cannot determine the computation of $a \in A$ such that $b = (a)$. Let $a \in A$ which cannot determine the computation of $a' \neq a \in A$ such that $(a') = H(a)$. The disadvantage is not being able to find the computation of a string pair $(a', a) \in A' \times A$ with $a' \neq a \in A$ such that $(a') = H(a)$. The message transmission is as follows,

Msg1: Usri $\rightarrow$ DC: {UIDi , X}H(UIDj $\parallel$ my)

Msg2: Usri $\rightarrow$ DC: {UIDi , X, SIDj , Y}H(SIDj $\parallel$ my)

Msg3: DC $\rightarrow$ Usri: $\langle$UIDj , SIDj , X, Y, Usri Y $\leftrightarrow$ WGAc$\rangle$ H(UIDj $\parallel$ my)

Msg4: DC $\rightarrow$ WGAc: $\langle$UIDj , SIDj , X, Y, Usri X $\leftrightarrow$ WGAc$\rangle$ H(SIDj $\parallel$ my)

Msg5: WGAc $\rightarrow$ Usri: $\langle$UIDj , SIDj , X, Y, Usri USK $\leftrightarrow$ WGAc$\rangle$ USK

Msg6: Usri $\rightarrow$ WGAc: $\langle$SIDj , UIDj , X, Y, Usri USK $\leftrightarrow$ WGAc$\rangle$ US

In this federal learning, the view of a party is described as its internal state (inputs and randomness) and every message received from other parties. It is noteworthy that when this party exits from executing at any point, a party stops receiving the messages immediately.

3.4 Algorithm- ECC-BFL

The algorithm takes the inputs asthe set of secret keys between two users with the identification of secret and random numbers. They getto share with n number of users, say 't.' Then t number of users receive the upcoming message and cross-check with the broadcast list. After cross-checking, the function key assignment phase comes in to activate the process of a blockchain transaction. Under this traction, the process called ECC takes place. This ECC makes to store in a database and check for the server's computational process.

Generate the public/secret keys
$(\delta, \rho), (NPKn, NSKn), (PPKn, PSK)$
user$n$, (n 2 U, $jUj = N$)
Select a random number βn
Generate the shares of βn
Receive messages from at least t users
$HF(xn) = (An, Bn) = (gHF\_;\_(xn), hHF\_;\_(xn))$.
Broadcast the list $U3$ to each user 2 $U2$.
Check whether $U3$ _ $U2$ and $jU3j$ _ t.
Functional key assignment ()
{
If the patient confirm transaction over blockchain then
Generate a key kusing ECC
A ← get values from ECC
a sends request with appendend value to b
b computes from a to b
b sends reply with c
The server check for c
If
$c = ab$
return to server
else
start transaction
}

---

4 Experimental Analysis

The experimental result is carried out, and the parameters used for analysis are classification accuracy, running time, Communication overhead, Computation overhead, and transaction speed. These values obtained for these parameters are compared against three standard methods, namely Biparing

Method (BM) [Wu et al., 2018], Homomorphic Cryptosystem (HC) [Dagher et al., 2017], and Multiple Authorities with Attribute-Based Signature scheme (MA-ABS) [Guo et al., 2018] with proposed Elliptical Curve Cryptography with Blockchain-based FL(ECC-BFL). The medical reports of patients are the datasets that we are considering in this work. A record of 100 MB was uploaded at a time and has been successful which determined the scalability of the system. Considering the machine configuration, the system also verified that the average time taken by multiple users for the uploading and retrieval of the record was less than 60 seconds. The below parameter discussion shows that how the proposed method has outperformed the state of art methods.

- Classification accuracy

The total users taking part in training and for every user, the local gradient size is considered. Generally, the accuracy (A) produced by the model is the ratio of the Total Gradients (TG) to the Total Users (TU) participating in training.

$$A = \frac{TG}{TU} \tag{7}$$

- Running time is the time taken by the application server to respond to the user request. Some factors that affect this time are the number of users, network bandwidth, type and number of requests made by the user, and the thinking time.

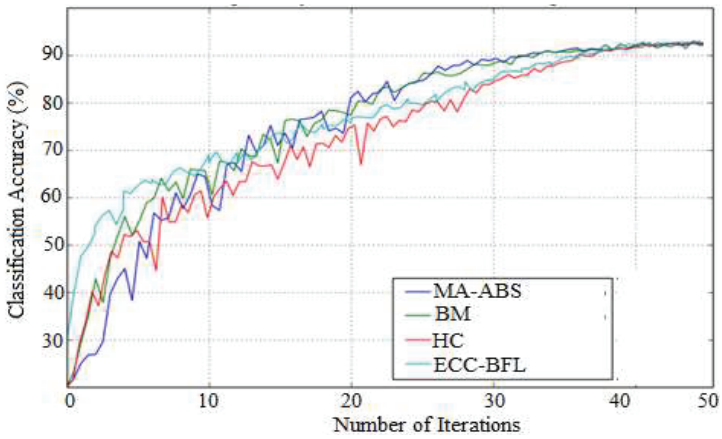

Figure 4 Analysis of classification accuracy.

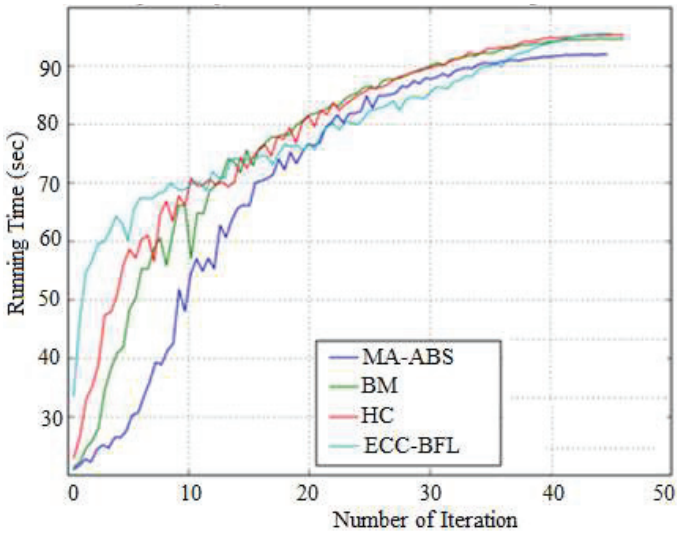

Figure 5 Analysis of running time.

Figures 4 and 5 illustrates the classification accuracy and running time for various iterations. When the number of iterations is increased, the system performs better, and the classification accuracy of the model is also improved. It is even more clear that when the number of gradients is increased, high accuracy is produced by the model.

$$\mathrm{S(n)}^n = \int_{t=0}^{n=1} \binom{n(t)}{Ri} x^{(n-1)} \tag{8}$$

Figure 6 illustrates the data selected at random, which are discrete points where the real distribution differs slightly from the ideal distribution. For every user, the cloud server estimates the aggregated outputs together with its respective server. At the verification request, the distribution of data uploaded, which assists in generating the aggregated outputs, must be identical to the real data. The aggregated outputs are further used while calculating the mean and variance of the data uploaded.

- Communication overhead

Total messages and data that have to be exchanged are estimated in communication overhead. Data involved in the proposed model occupies 2 bytes and 5 bytes for identifier and time stamp, respectively, where q, elliptic

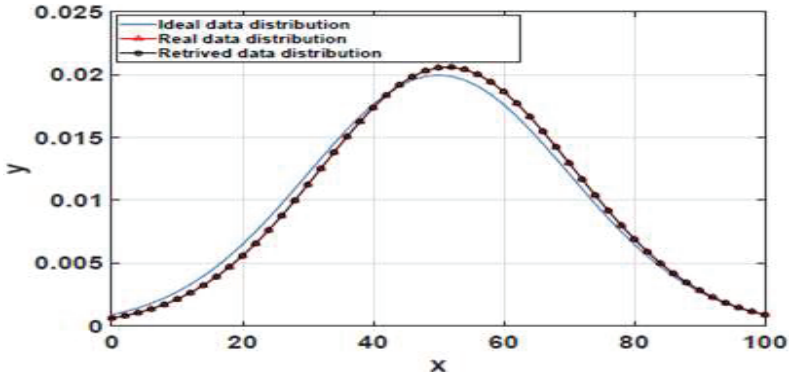

Figure 6 Analysis of data distribution.

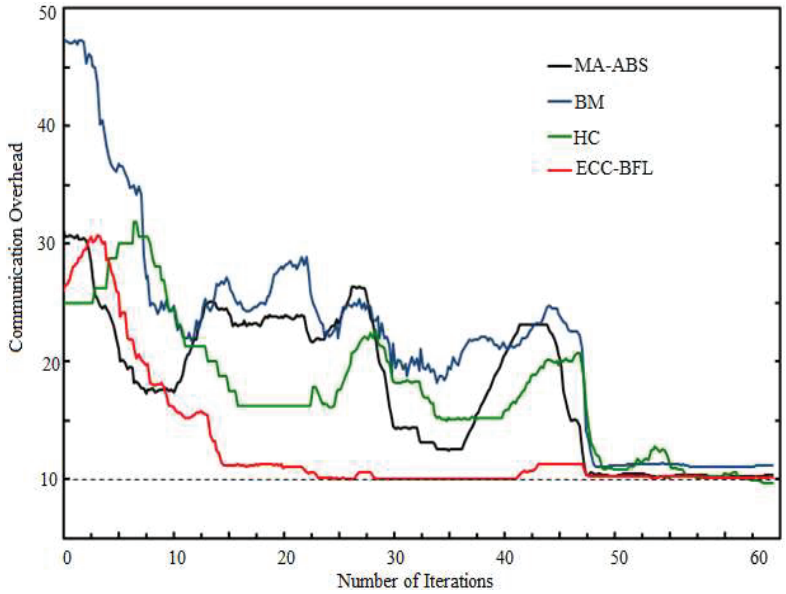

Figure 7 Analysis of communication overhead.

curve point, and signature occupy 20 bytes each.

$$S^n = \frac{S(t)}{T} \tag{9}$$

Figure 7 compares the communication overhead between existing MA-ABS,BM,HC, and proposed ECC-BFL methods where the *X*-axis indicates

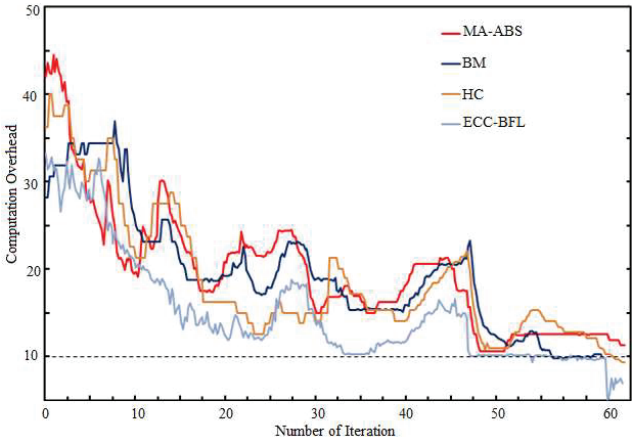

Figure 8 Analysis of computation overhead.

the number of iterations and the Y-axis the values of communication overhead. When compared, the proposed method achieves less communication overhead

- Computation overhead

The time taken by the 5G node for executing the functions involved in the authentication procedure is termed computation overhead. This reduces the congestion as well as the security risks in the core network.

$$Y = \frac{\text{node failures } (\%)}{\text{number of delivered packets}(\%)} \tag{10}$$

Figure 8 compares the computation overhead between existing MA-ABS,BM,HC, and proposed ECC-BFL, where the *X*-axis indicates the number of iterations and the *Y*-axis the values of computation overhead. When compared, the proposed method achieves less computation overhead

Figure 9 compares the transaction speed between existing MA-ABS,BM,HC, and proposed ECC-BFL, where the *X*-axis indicates the number of iterations and the *Y*-axis the transaction speed. When compared, the transaction speed for the proposed method is higher. Table 4 indicates the comparison of existing BM, HC, MA-ABS, and proposed method ECC-BFL

Here, the proposed algorithm, namely ECC-BFL which to ensure the confidentiality of users' local gradients while performing federated learning. Using ECC-BFL method the Communication overhead, Computation

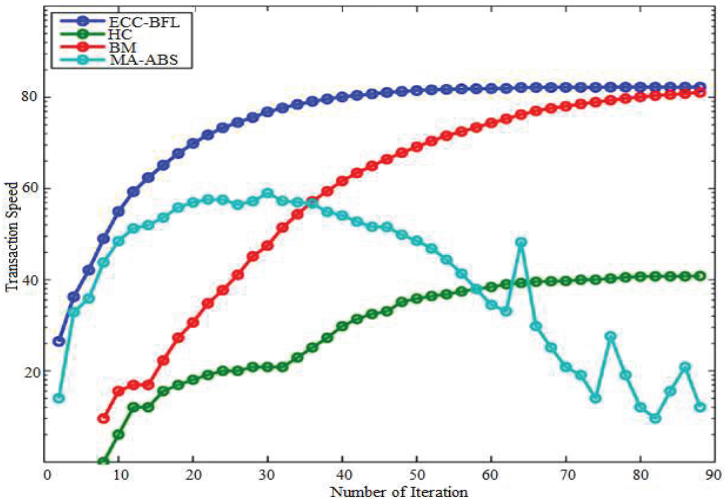

Figure 9 Analysis of transaction speed.

Table 4 Comparison of existing and proposed method

Parameters	BM	HC	MA-ABS	ECC-BFL
Classification accuracy (%)	90	93	91	95
Running time (sec)	80	72	73	65
Communication overhead (%)	85	80	81	76
Computation overhead (%)	72	68	69	63
Transaction speed (%)	85	90	88	92

overhead are reduced which then advances the transaction speed when compared with the existing systems BM ,HC and MA-ABS. Table 3 shows the Comparison of existing and proposed method in terms of parameters such as classification accuracy, running time, Communication overhead, Computation overhead, and transaction speed. While analyzing classification accuracy, existing method achieves 90%, 93% and 91% while the proposed method achieves 5% better than BM, 2% better than HC and 4% better than MA-ABS. While analyzing running time, existing method achieves 80 sec, 72 sec and 73 sec while the proposed method achieves 15 sec better than BM, 7 sec better than HC and 8 sec better than MA-ABS. While analyzing communication overhead, , existing method achieves 85%, 80% and 81% while the proposed method achieves 8% better than BM, 4% better than HC and 5%better than MA-ABS. While analyzing computation overhead, existing method achieves 72%, 68% and 69% while the proposed method achieves 9% better than BM,

5% better than HC and 6%better than MA-ABS.While analyzing transaction speed, existing method achieves 85%, 95% and 88% while the proposed method achieves 6% better than BM, 2% better than HC and 4% better than MA-ABS.

5 Conclusion

Blockchain potentially transforms conventional healthcare industries. Even there are several challenges that require a solution, FL is practically important for various real-time problems.

5.1 Theoretical Contribution

The main contribution is to develop the Elliptical Curve Cryptography-based blockchain FL technology, which helps users verify the correctness of results obtained from a server with tolerable overhead.Moreover, it is responsible for ensuring the confidentiality of local gradients of the users at the time of FL. Some users exit during training for a few reasons, but the privacy of these uses is preserved. Deep neural networks have failed to produce apt solutions; thus,Elliptical Curve Cryptography with Blockchain-based FL (ECC-BFL) is proposed, which helps in verifying the calculation results of the server for every user. Moreover, ECC-BFL supports users to drop out of the training process.

5.2 Practical Contribution

Further, experiments on real-time data practically demonstrated the performance of the ECC-BFL approach. This proposed ECC-BFLapproach achieved 95% of classification accuracy, 65 sec of running time, 76% of communication overhead, 63% of computation overhead, and 92% of transaction speed.

5.3 Limitation and Future Work

It is difficult to find and rely on such a trustworthy third membership service party that validates user identity in heterogeneous network.In the future, the reduction of communication overhead has to be focused on the entire protocol. The future concentrates on investigating various privacy-guarantee methods and apply regression-oriented FL to improve the privacy-guarantee data sharing among heterogeneous networks.

References

Aledhari, M., Razzak, R., Parizi, R. M., and Saeed, F. (2020). Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access*, *8*, 140699–140725.

Cabaj, K., Caviglione, L., Mazurczyk, W., Wendzel, S., Woodward, A., and Zander, S. (2018). The new threats of information hiding: The road ahead. *IT professional*, *20*(3), 31–39.

Dagher, G. G., Mohler, J., Milojkovic, M., and Marella, P. B. (2018). Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable cities and society*, *39*, 283–297.

Guo, R., Shi, H., Zhao, Q., and Zheng, D. (2018). Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems. *IEEE access*, *6*, 11676–11686.

Liu, Y., Huang, A., Luo, Y., Huang, H., Liu, Y., Chen, Y., ... & Yang, Q. (2020, April). Fedvision: An online visual object detection platform powered by federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 08, pp. 13172–13179).

Liu, Y., James, J. Q., Kang, J., Niyato, D., and Zhang, S. (2020). Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, *7*(8), 7751–7763.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. (2020). Blockchain empowered asynchronous federated learning for secure data sharing on internet of vehicles. *IEEE Transactions on Vehicular Technology*, *69*(4), 4298–4311.

Miers, I., Garman, C., Green, M., and Rubin, A. D. (2013, May). Zerocoin: Anonymous distributed e-cash from bitcoin. In *2013 IEEE Symposium on Security and Privacy* (pp. 397–411). IEEE.

Niknam, S., Dhillon, H. S., and Reed, J. H. (2020). Federated learning for wireless communications: Motivation, opportunities, and challenges. *IEEE Communications Magazine*, *58*(6), 46–51.

Ning, Z., Dong, P., Wang, X., Rodrigues, J. J., and Xia, F. (2019). Deep reinforcement learning for vehicular edge computing: An intelligent offloading system. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(6), 1–24.

Patel, V. (2019). A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. *Health informatics journal*, *25*(4), 1398–1411.

Sampathkumar, A., and Vivekanandan, P. (2019). Gene selection using parallel lion optimization method in microarray data for cancer classification. *Journal of Medical Imaging and Health Informatics*, *9*(6), 1294–1300.

Sampathkumar, A., Maheswar, R., Harshavardhanan, P., Murugan, S., Jayarajan, P., and Sivasankaran, V. (2020, July). Majority Voting based Hybrid Ensemble Classification Approach for Predicting Parking Availability in Smart City based on IoT. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–8). IEEE.

Servos, D., and Osborn, S. L. (2017). Current research and open problems in attribute-based access control. *ACM Computing Surveys (CSUR)*, *49*(4), 1–45.

Sultana, N., Chilamkurti, N., Peng, W., and Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, *12*(2), 493–501.

Sun, G., Sun, S., Sun, J., Yu, H., Du, X., and Guizani, M. (2019). Security and privacy preservation in fog-based crowd sensing on the internet of vehicles. *Journal of Network and Computer Applications*, *134*, 89–99.

Wang, X., Han, Y., Leung, V. C., Niyato, D., Yan, X., and Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *22*(2), 869–904.

Wu, H. T., and Tsai, C. W. (2018). Toward blockchains for healthcare systems: Applying the bilinear pairing technology to ensure privacy protection and accuracy in data sharing. *IEEE Consumer Electronics Magazine*, *7*(4), 65–71.

Xu, C., Lin, H., Wu, Y., Guo, X., and Lin, W. (2019). An SDNFV-based DDoS defense technology for smart cities. *IEEE Access*, *7*, 137856–137874.

Yue, X., Wang, H., Jin, D., Li, M., and Jiang, W. (2016). Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *Journal of medical systems*, *40*(10), 1–8.

Zhang, K., Long, J., Wang, X., Dai, H. N., Liang, K., and Imran, M. (2020). Lightweight searchable encryption protocol for industrial internet of things. *IEEE Transactions on Industrial Informatics*.

Zhang, L., Wu, Q., Domingo-Ferrer, J., Qin, B., and Hu, C. (2016). Distributed aggregate privacy-preserving authentication in VANETs. *IEEE Transactions on Intelligent Transportation Systems*, *18*(3), 516–526.

Zhao, J., Chen, Y., and Zhang, W. (2019). Differential privacy preservation in deep learning: Challenges, opportunities and solutions. *IEEE Access*, *7*, 48901–48911.

Biographies

Sampathkumar Arumugam has received his Bachelor's in Information Technology in 2009; Master's in Mainframe Technology in 2012 and completed Ph.D. degree in Anna University Chennai in 2019. He has 10 years of academic experience in various repudiated institution. He has published more than 10 SCI articles and more than 13 Scopus articles in peer-reviewed journals. He has been editor in the some of the book series and published several book chapters in springer and Elsevier which are Scopus indexed. He has published Indian and Australian government patents. He has been actively participating as reviewers in some of the international journals and member of CSI societies. His research interest includes Bioinformatics, Artificial Intelligence, Data Mining, Machine Learning, Data Analytic and Optimization Techniques.

Shishir Kumar Shandilya is the Division Head of Cyber Security and Digital Forensics at VIT Bhopal University. He is working as a Principal Consultant to the Govt. of India for Technology Development and Assessment in Cyber Security. He is also a Visiting Researcher at Liverpool Hope University-United Kingdom, a Cambridge University Certified Professional Teacher and Trainer, ACM Distinguished Speaker and a Senior Member of

IEEE. He is a NASSCOM Certified Master Trainer for Security Analyst SOC (SSC/Q0909: NVEQF Level 7) and an Academic Advisor to National Cyber Safety and Security Standards, New Delhi. He has received the IDA Teaching Excellence Award for distinctive use of technology in Teaching by Indian Didactics Association, Bangalore (2016) and Young Scientist Award for two consecutive years, 2005 and 2006, by Indian Science Congress and MP Council of Science and Technology. He has seven books published by Springer Nature-Singapore, IGI-USA, River-Denmark and Prentice Hall of India. His recently published book is on Advances in Cyber Security Analytics and Decision Systems by Springer. Dr. Laxman Singh obtained his B. Tech in Electronics and Communication Engineering from C.R. State (Govt.) College of Engineering, Murthal, Sonipat (Haryana) and M.Tech in Instrumentation and Control from M.D. University, Haryana, India in 2004 and 2009 respectively. He received his PhD degree from Jamia Millia Islamia (a central Govt. of India University) in 2016. Presently he is working as Associate Professor in the Department of Electronics & Communication Engineering at Noida Institute of Engineering & Technology (NIET), Greater Noida. He has total teaching experience of more than seventeen years. Dr. Laxman Singh has published about 35 research articles in the field of image processing, AI, and machine learning in various refereed international/national journals as well as in international conferences of repute. His current research interests are in the areas of Wavelet analysis, Artificial Intelligence, Image processing, and Optimization techniques.

Nebojsa Bacanin received his Ph.D. degree from Faculty of Mathematics, University of Belgrade 1 in 2015 (study program Computer Science, average grade 10,00). He started University career in Serbia 13 years ago at Graduate

School of Computer Science in Belgrade. He currently works as an associate professor and as a vice-dean at Faculty of Informatics and Computing, Singidunum University, Belgrade, Serbia.

He is involved in scientific research in the field of computer science and his specialty includes stochastic optimization algorithms, swarm intelligence, soft-computing and optimization and modeling, as well as artificial intelligence algorithms, swarm intelligence, machine learning, image processing and cloud and distributed computing. He has published more than 120 scientific papers in high quality journals and international conferences indexed in Clarivate Analytics JCR, Scopus, WoS, IEEExplore, and other scientific databases, as well as in Springer Lecture Notes in Computer Science and Procedia Computer Science book chapters. He has also published 2 books in domains of Cloud Computing and Advanced Java Spring Programming.

He is a member of numerous editorial boards, scientific and advisory committees of international conferences and journals. He is a regular reviewer for international journals with high Clarivate Analytics and WoS impact factor such as Journal of Ambient Intelligence & Humanized Computing, Soft Computing, Applied Soft Computing, Information Sciences, Journal of Cloud Computing, IEEE Transactions on Computers, IEEE Review, Swarm and Evolutionary Computation, Journal of King Saud University “C Computer and Information Sciences, SoftwareX, Neurocomputing, Operations Research Perspectives, etc. He actively participates in 1 national and 1 international projects from the domain of computer science. He has also been included in the prestigious Stanford University list with 2% best world researchers for the year 2020.