MANTEY, E.A., ZHOU, C., MANI, V., ARTHUR, J.K. and IBEKE, E. 2023. Maintaining privacy for a recommender system diagnosis using blockchain and deep learning. *Human-centric computing and information science* [online], 13, article number 47. Available from: https://doi.org/10.22967/HCIS.2023.13.047

Maintaining privacy for a recommender system diagnosis using blockchain and deep learning.

MANTEY, E.A., ZHOU, C., MANI, V., ARTHUR, J.K. and IBEKE, E.

2023

Human-centric Computing and Information Sciences

October 2023 | Volume 13

HCIS | Human-centric Computing and Information Sciences

RESEARCH (Research Manuscript) Open Access

Human-centric Computing and Information Sciences (2023) 13:47
DOI: https://doi.org/10.22967/HCIS.2023.13.047
Received: January 30, 2022; Accepted: April 17, 2022; Published: October 30, 2023

Maintaining Privacy for a Recommender System Diagnosis using Blockchain and Deep Learning

Eric Appiah Mantey[1,*], Conghua Zhou[1], Vinodhini Mani[2], John Kingsley Arthur[1], and Ebuka Ibeke[3]

Abstract

The healthcare sector has been revolutionized by blockchain and artificial intelligence (AI) technologies. AI uses algorithms, recommender systems, decision-making abilities, and big data to display a patient's health records using blockchain. Healthcare professionals can make use of blockchain to display a patient's medical records with a secured medical diagnostic process. Traditionally, data owners have been hesitant to share medical and personal information due to concerns about privacy and trustworthiness. Using blockchain technology, this paper presents an innovative model for integrating healthcare data sharing into a recommender diagnostic computer system. Using the model, medical records can be secured, controlled, authenticated, and kept confidential. In this paper, researchers propose a framework for using the Ethereum blockchain and X-rays as a mechanism for access control, establishing hierarchical identities, and using pre-processing and deep learning to diagnose coronavirus disease 2019 (COVID-19). Along with solving the challenges associated with centralized access control systems, this mechanism also ensures data transparency and traceability, which will allow for efficient diagnosis and secure data sharing.

Keywords

Blockchain, AI, Recommender Systems, Patients, Health Records, Secured Data

1. Introduction

The management of electronic health record (EHR) data access and privacy is a crucial issue. The diversity and fragmentation of e-Health data makes privacy and security issues nearly impossible to manage. Recent outbreaks of the coronavirus disease 2019 (COVID-19) have paralyzed global health and kept health at the top of our priority list. COVID-19 has been associated with numerous reports of loss of taste or smell, as well as fever, coughing, fatigue, sore throat, or even death in more serious cases [1]. Reverse transcription polymerase chain reaction (RT-PCR) is a standard diagnostic method [2] that has a high sensitivity but inconsistent in specificity. Its drawbacks include large consumption of time and limited availability. Along with pathological tests, other methods of identifying an infection in a patient

*Corresponding Author: Eric Appiah Mantey (emantey@acm.org)
[1]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China
[2]Department of Computer Science and Engineering, School of Computing, Sathyabama Institute of Science and Technology, Chennai, India
[3]School of Creative and Cultural Business, Robert Gordon University, Aberdeen, UK

include radiology exams such as computed tomography (CT), and X-ray images. Various deep learning models that analyze CT and X-ray images were proposed to improve the detection accuracy for various types of infections [3–5]. However, the healthcare centers do not maintain policies that protect sensitive data [6–8]. A flexible access rights delegation scheme that is based on attribute-based access control can be improved in dynamic healthcare environments. Due to the increasing complexity of healthcare systems, a precise tracking system that tracks how data is used throughout the process of patients' care and diagnosis is needed. A secure and reliable access control system must be provenance-based, including the histories of user flows. Access to provenance data is essential to ensuring a secure and reliable access control system. For aggregating gradients in a secure and private manner, the authors's in [9] used the federated learning global model. In the shared model, there is no way to determine whether users are genuine. Thus, due to a lack of quality data, the model performs poorly because of the lack of trust between different sources. Blockchain technology provides various trust mechanisms to overcome the trust problem of decentralized networks [10]. Using a blockchain system integrated with software-defined networking, the researchers discuss energy and security issues while exploiting the potential benefits of a system that harnesses blockchain technology [11]. Internet of things and blockchain technologies has been used as a platform for sensing and storing data; they provide security countermeasures against medical data mining threats [12, 13]. The electronic health data storage system uses blockchains as on-chain and the Interplanetary File System (IPFS) as off-chain to preserve privacy and scalability [14]. Deep learning and blockchain are used to implement the trust-based storage recommendation [15]. Identifying and authenticating data from its source is proposed in a private data sharing model [16]. The blockchain-based algorithm adopted in by the author's in [17] ensures data validity when aggregating local deep learning models. However, these techniques do not consider the gradient's privacy; they increase the risk of data leakage.

1.1 Motivation

Decentralized learning avoids transferring all data to a central server, instead using a local model to train, as opposed to traditional machine learning. Using a third party to clean, reorganize, and prepare data is essential to training the model. In the process of generating the model, the data are handled in a way that can violate privacy. Traditionally, organizations may experience delays when building models with acceptable accuracy using traditional machine learning models. The training of a traditional machine learning model requires massive amounts of historical data in addition to the huge amounts of data required to train the machine learning model. In order to solve the cold start problem and enable their data to be trained locally on their servers, clients should use a secure distributed machine learning methodology. This method allows them to have their data trained on their servers and prevents privacy violations. This method must also be efficient in terms of resources.

1.2 Contribution

The symptoms of COVID-19 vary from patient to patient and the disease spreads rapidly [18]. Information shared among hospitals can assist in diagnosing the disease. Creating a global model to detect positive cases while sharing data securely is a challenge. Furthermore, the existing studies do not provide reliable data sharing and accurate training models. In order for artificial intelligence (AI) to advance, data from various sources must be collected. Since healthcare centers do not have a privacy-preserving approach, such confidential data cannot be made available [19–21]. In addition, deep learning models need to be trained collaboratively over public networks. Detection of small areas of infection within the lungs is the first goal of our work. This is useful for radiologists in detecting infections accurately. To build a better deep learning model, data sharing is imperative, as long as the privacy of data providers are protected. Sharing data facilitates the development of an automatic COVID-19 patient detection model that is based on deep learning [22].

The rest of the paper is organized as follows. In Section 2, we discuss the related literature. Section 3 presents the proposed methodology. In Section 4, we present our experiments and results. Finally, the paper concludes in Section 5.

2. Literature Review

The authors in [23] studied the main components of traditional personalized healthcare and wellness support services. They proposed a combination of the three key health dimensions—nutrition quality, physical activity and exercises, and mental health and sleep disorders—in the same self-tracking, self-monitoring and self-quantification solution. They referred to it as a next generation of intelligent, multidimensional context-aware personalized healthcare and wellness support service. In [24], user preferences were considered, as well as how to maximize the healthy, and diminish the unhealthy compounds in food with a 3D recommendation system. Using a large set of publicly accessible collection of data (Recipe1M+), the authors built a recommendation system for recipes and ingredients. The authors in [25] highlighted the connection between sustainability adaptations, and nutrition in food security through the use of sustainable food systems. They asserted that an adaptation to sustainable food systems will necessitate shifting from an agriculture-centered to a research and food systems policy framework. In [26], the authors proposed a novel multidimensional method for context-aware recommendation in commerce, which stands for context information, items, users, and their relationships. It also characterizes an evaluation process through the implementation of the suggested approach in a food recommendation system of a restaurant, while considering some factors like day and weather as contextual information. The approach is also compared with the customary 2D one. The authors in [27] reviewed nutrition recommendation systems (NRS) and their characteristics, making use of some scientific databases such as reference sources, which enables access to different publications in the area. They proposed that with the proper design, implementation, and evaluation of NRS could be effectively utilized as a tool to enhance nutrition and promote good health. In [28], the authors identified the transformation paths that are required to contend with the key causes of the recent increase in hunger and stunted progress in decreasing malnutrition. These paths could only be attainable if they met certain stipulations, such as creating opportunities for the marginalized, fostering human health and caring for the environment.

The authors in [29] studied the current terrain of research in recommender systems (RS) and stipulated directions in the area of various applications. They concluded that the advent of more robust algorithms for recommendation resulted in the broader applications of RSs, and advancements in technology has accelerated the use of RSs. In [30], the authors made nutritional recommendations in all classes of food for COVID-19 quarantine. The authors in [31] conducted a multidimensional analysis and assessment of poverty measurement as an income approach using methods that are executed using collective data from varied sources. They also used methods which show joint distribution and are executed making using data where each dimension information is accessible for each item of analysis. In [32], the authors proposed a multidimensional recommendation framework in the health field making use of collaborative filtering. They also proposed a new semantic similarity feature between users, considering dimensions apart from medical problems such as level of education, health literacy, and patient's psycho-emotional state; and introduced the notion of fairness in the new aggregation technique, collecting preference scores. To better promote healthy food habits in Mauritius, the authors in [33] introduced a DASH diet recommender system that suggests a healthy Mauritian diet to patients with hypertension, considering factors like age, allergies, general user food preferences, smoking level, and blood. In [34], using Automation Anywhere, a recommender system was designed to simulate a meal planner for distributing a large amount of diet recommendations to numerous users in a database expeditiously.

The authors in [35] presented a general model for the recommendation of daily meals, with the principal feature, the concurrent management of nutritional and preference-aware information. In [36], the authors proposed a deep-learning solution for health-based medical dataset that automatically determines the kind of food to be given to a patient based on the disease and other factors like weight, gender, age, and calories. This model emphasizes on executing both machine learning and deep learning algorithms like logistic regression (LR), recurrent neural network (RNN), multi-layer perceptron (MLP), gated recurrent units (GRU), and long short-term memory (LSTM). The authors in [37] proposed a state-of-the-art fuzzy set model to produce an intelligent personalized diet for users, while considering other imputed parameters. The application produces the maximal intake of calorie range depending on the profile of the user making use of fuzzy sets. In [38], the authors applied the deep belief network to the recommender system's problem. The authors in [39] designed a hybrid algorithm that remodels a pre-filtering contextual incorporation technique and inputs the novel dimension to a deep learning-based neural cooperative filtering technique, to preserve and recover the advantages of both without their restrictions, to simplify the selection process and prevent information overloading. In [40], the authors proposed a framework for tourism recommender system, gleaned from a combined recommendation approach. The proposed model acts as a trip planner that creates a comprehensive program for a specific visit period. In [41], the authors proposed a pointer-based item-to-item collaborative filtering recommendation system.

The authors in [42] incorporated a blockchain privacy system (BPS) into deep learning for a diet recommendation system for special needs patients. It enables patients get notice about recommended medications and treatments depending on their personal data without any breach of confidentiality. In [43], the authors proposed a state-of-the-art recommender system using the benefits of a secure multiparty computation supported by blockchain. It allows for a decrease in fraud. The authors in [44] aimed at creating a secure trust-based system using the benefits of a secure multiparty computation supported by blockchain by including smart contracts with the principal blockchain protocol. In order to prevent attacks on medical plaintext data, homomorphic encryption using a privacy-preserving scheme was employed [45]. The author applied intelligent distributed storage techniques for vehicular networks that keep the system's privacy [46]. The authors used cloud-based blockchain data storage techniques to boost security, privacy, and transaction speed in [47]. An effective machine learning algorithm based on fog computing was used to classify healthcare data based on risk [48]. The system achieves low latency through task offloading [49]. The blockchain facilitates the efficient collection and analysis of patient data, while protecting patients' identities and preserving the necessary social distance. Its clients maintain control of their information since blockchain technology has no central authority. While ensuring the privacy and identification of patients, they can share data that is pertinent to the COVID-19 relief effort. Data on the COVID-19 can also be collected by the government, and clients can be assured that their data is secure, and their information will not be divulged to third parties. Our paper applies smart contracts and databases to validate data reporting in order to predict COVID-19 accurately and securely without any error. There has been a surge in social platforms claiming misinformation recently, making this use case particularly important. Therefore, information and data that are publicly communicated need to be authenticated and monitored. To identify users who are spreading rumours, conspiracy theories, and fake news, it is important to pinpoint the source of an online message. By digitally signing messages before they are added to blocks, blockchain-based systems enable the identification of information sources. A combination of the blockchain technology and recommendation systems affords virtual activities more security and privacy. Therefore, this paper tends to build a better deep learning model, data sharing is imperative, as long as the privacy of data providers are protected. We intend to create a scenario where sharing data will facilitate the development of an automatic COVID-19 patient detection model that is based on deep learning.

3. Methodology

3.1 Proposed Framework

Data is typically protected by encrypting it before it is distributed to the data requester. Only authorized parties can decrypt the encrypted data. An efficient and scalable key management system is necessary for cloud computing systems on a large scale for the distribution of public and private keys among authorized parties. Large-scale cloud systems with a large number of users have proven difficult to manage because of the collision issue. The revocation of passwords may attract a cost for users since they need to re-encrypt their data and redistribute their keys. To overcome the drawbacks listed above and to take advantage of the benefits of Blockchain technology, we propose a secure EHR data sharing and access control management system based on two-layer encryption techniques. The goal of this paper is to share data while maintaining privacy. The proposed design allows users to define their own access policies for updating and revoking access control without the need for consent from data owners. Data owners can define a list of users who require access to medical records and their access permissions and attributes by using an access control repository (ACR) as part of identity-based access control. ACRs are then placed into smart contracts and distributed over the blockchain network by data owners. The distributed ACR-smart contract eliminates the single point of failure and removes the need for a centralized third-party validator. Hence, a permissioned blockchain ensures the protection of personal information and the prevention of sensitive data leaks. Then, the input image is standardized using the normalization technique. Finally, detecting COVID-19 in x-ray images using deep learning, Keras, and TensorFlow is differentiates COVID-19 positive patients from other healthy individuals. It uses a convolutional net model built on top of different optimization techniques to enhance the results.

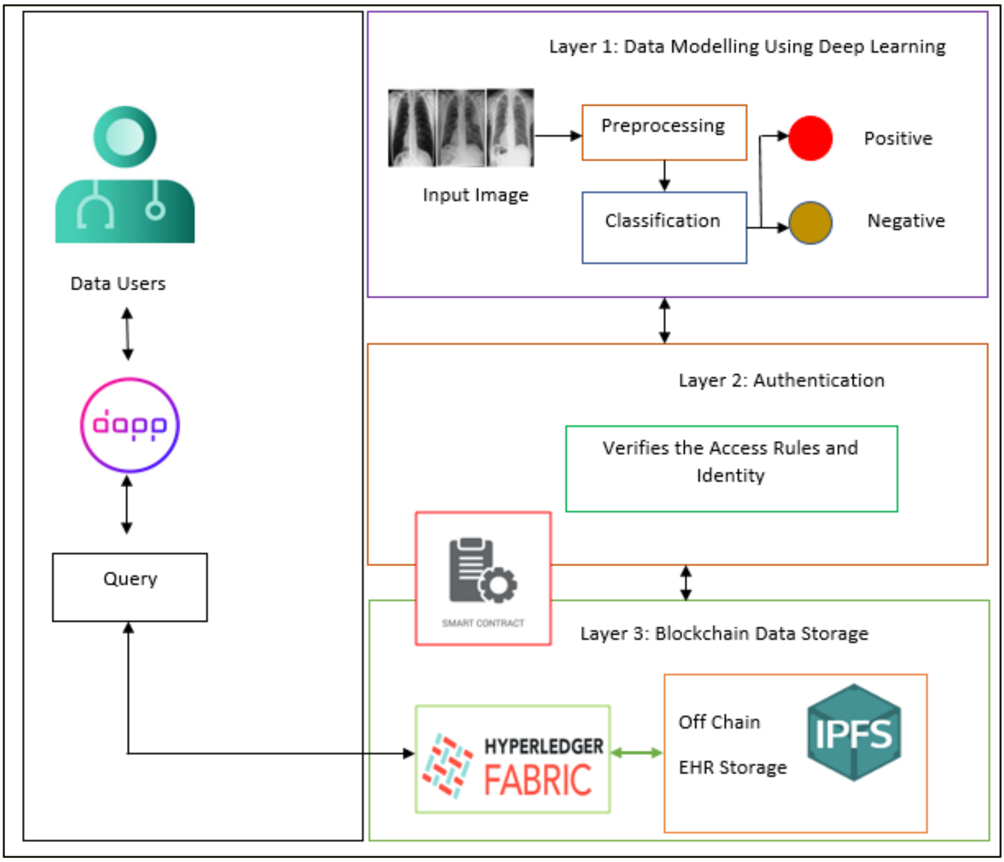

Fig. 1. Proposed framework.

An overview of the proposed framework is shown in Fig. 1. When a user requests a query at Layer 3, the data flows to that layer. An access authorization module called Layer 2 authentication receives the user request. It determines what kind of access the user should be granted based on the privileges of the user. In addition to storing and retrieving data, Layer 3 blockchain data storage is dedicated to handling different queries, including real-time queries over medical data. Following the normalization of the chest X-ray images, Layer 3 uses a deep learning model to classify them as positive or negative.

In the proposed system, there are five main constituents: EHR owner, data requester, miners, interplanetary file system, and private blockchain network.

3.2 Electronic Health Record Storage

Patients, doctors, and healthcare providers must register before they can use the system. Depending on the identity or attributes, the patient can consent to the sharing of medical information. The public and private keys of the data owner can be generated from this information. To decrypt secret information, public keys are distributed to the data users. Private keys are used to encrypt patient information. The data users are then incorporated into the ACR. A cryptographic key is used to link the encrypted data, and the meta-data of the access list will be sent to the blockchain. The encrypted data will then be sent to the off-chain IPFS database. To ensure the user is legitimate, the signature of the data owner is verified. Deep learning models are also granted permission using the same process. ACRs of deep learning models are continuously updated whenever a new user is added to the system.

3.3 The Sharing of Electronic Health Records

To initiate data sharing as a new transaction, a doctor sends an access request to a patient's image data in the system. The process of access authentication begins with checking the data owner's signature, and if found valid, the process continues, otherwise, it is rejected. Layer 2 verifies the identity or associated attributes according to the ACR, and, if the matching identity or attribute occurs, access to the EHR is given to the user. As an example, the doctor will diagnose the patient. Fig. 2 illustrates how the doctor initiates access to the deep learning model associated with the diagnosis if a second opinion is needed. Level 2 authentication confirms that the doctor has access to the model. An event is created to record each data sharing operation for ensuring data integrity. Layer 3 shows the diagnosis results of patients. It can be tracked or audited by the data owner.

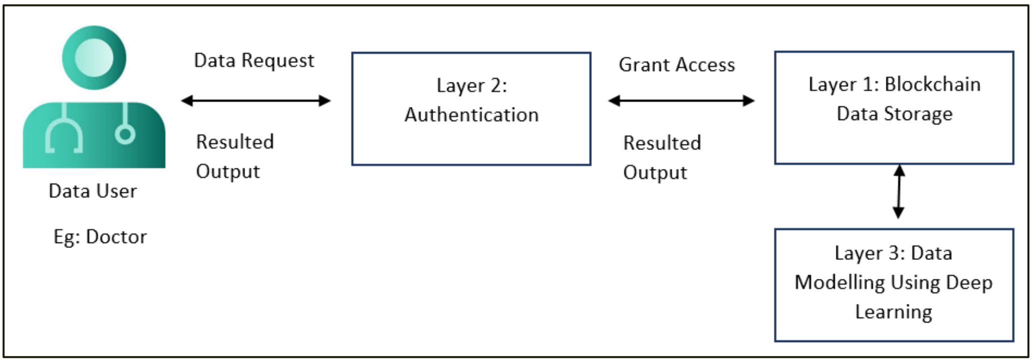

Fig. 2. Sharing of health records for diagnosis of COVID-19.

3.3.1 Encryption and decryption in permissioned blockchain

Step 1: Create public and master keys.
Step 2: Generate the user keys and set the encryption attributes.
Step 3: Encrypt the data using the key and generate the cipher text.
Step 4: The data is decrypted on the receiver's side using the private key.

3.3.2 Identity-based access generation and assignment

The attributes of a data user are defined once the user has successfully registered with the system. Detailed attribute assignments are performed by the authorization layer in response to a user's request. The authorization layer verifies the identity of each user to determine the validity of requests and identities. The service cannot be accessed by an adversary without authenticating their identity. The attributes in the attribute data store are queried following successful authorization. An attribute set is sent to the authorization layer by the attribute data store. When an associated attribute is not yet present in the data store, the trusted authority creates it. The authorization layer generates and assigns a private key based on the specified attribute.

3.3.3 Request access to the electronic health record

The authorization layer accepts access requests from users. Authorization layers verify the identity of users and their digital signatures, as well as their attributes to ensure access is authorized. Requests for service are approved or denied based on the status of the authentication layer. Data is held in blockchains using the pointer of encryption addresses and is retrieved from IPFS using an authorization layer to ensure the confidentiality of data owners. The necessary information is returned to Layer 1 based on the access level of the data requester. Using the master secret key, Layer 1 decrypts the data. Based on the access level, Layer 1 encrypts data using the public key of the data requester and returns it encrypted. Data must be decrypted using the private key of the data user before it can be accessed. By utilizing both a blockchain search and IPFS data store request, it verifies the integrity of the data requester. The data requester's public key is used to encrypt the file path in the authorization layer. An encrypted file path is sent from the authorization layer to the user. Once the user has decrypted the file, it is sent back to the authorization layer. A blockchain identifier, data access transaction, and provenance data are recorded in the authentication layer. Data with provenance identifiers are stored in IPFS data stores. The generated identities are added to the mapping data store.

3.3.4 Blockchain based data storage by data owner

To define the attributes for the data, the data owner must first register successfully. A pair of asymmetric keys (public and private) will be used to encrypt and decrypt data in the authorization layer. Public key encryption is used to encrypt medical image files and medical data. Data owners provide authorized layers to cipher text and image file paths. A pointer to the stored data is generated by the authorization layer and stored in the IPFS data store. Blockchain nodes store index data by sending the generated pointers to the authorization layer.

3.3.5 User revocation based on identity and automatic access expiration

With identity-based encryption and attribute-based encryption, integrated data can be encrypted based on pre-set attributes and access levels to enable the creation of flexible fine-grained access policies. It is important to define a mechanism which gives the patient the option of revoking access if needed. This framework introduces a temporal access condition that reduces the computational cost and complexity associated with credential revocation. Attribute keys are updated, and the new ACR is automatically distributed. Data owners can revoke access to their data immediately after they have assigned an expiration time to their access rights. A smart contract maps data from IPFS data stores to the corresponding hashes on blockchains and returns the data-to-data owners. Keeping track of the transaction and knowing exactly who uses their data and how, can be done by data owners by asking a query about their medical data usage. An authorization layer validates the user access level before retrieving data from the IPFS data store using mapping information, if there is a legitimate request as presented in Fig. 3. The trust authority entities are responsible for issuing, revoking, and updating keys in a permissioned blockchain. Each revoked user's private key must be updated by the trusted authority with a new hash code. Role-based mechanisms [45], attribute-based mechanisms [46], and capability-based mechanisms [47] have been proposed in the past to address revocation. It is problematic to involve trusted authorities and data owners in the previous approach.

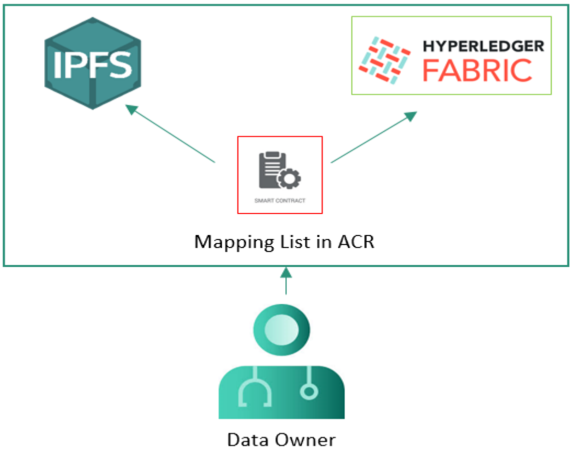

Fig. 3. User revocation based on identity.

Several studies present access expiration as a feasible solution for credential revocation [48, 49]. We propose a temporal access condition in the framework to reduce the computational cost of credential revocation and deal with the complexity of the operation. Owners of data can revoke data access immediately after expiration by assigning the access rights during an expiration or session time. As a result, the expiration time can be incorporated into the key generation phase in addition to the set of attributes. As part of the decryption process, the "time expiration" attribute is checked, and if valid, this attribute can be used to decrypt the data. The proposed "time expiration" attribute auto-updates the new ACR and does not require it to be redistributed or interact between users and trusted authorities. The revocation of consent cannot be applied to prior or future data access requests because all previous transactions have already been recorded and are immutable.

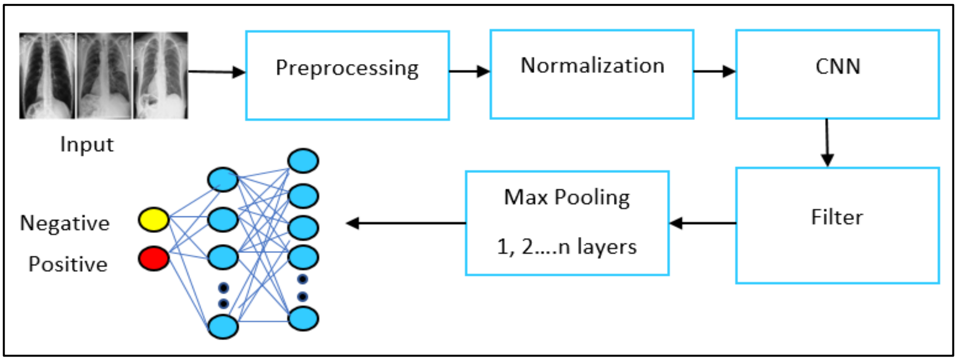

Fig. 4. Deep learning model to diagnosis COVID-19.

3.3.6 Data modelling using deep learning.

Layer 3 is used to build a model of the lung without human intervention to analyze its properties with and without infection by COVID-19 as shown in Fig. 4. Infections caused by COVID-19 are usually found in the respiratory tract, especially in the lung's lower respiratory tract. COVID-19 can cause varying degrees of respiratory issues, from mild to severe. People with other diseases, such as cancer, heart disease, and diabetes, may suffer from more severe symptoms. Scientists are learning new things about COVID-19 every day, and they are concerned about the potential effects it may have on the lungs.

The effects of these diseases are similar to those of the Middle East respiratory syndrome (MERS), acute respiratory distress syndrome (ARDS), and severe acute respiratory syndrome (SARS). An X-ray or CT scan of the chest can show signs of respiratory inflammation. It may look like the frosted glass on a shower door on a chest CT, which is called "ground-glass opacity." Convolutional neural networks (CNN) were used because the dataset included images, and its first layer can detect specific features such as edges, corners, and backgrounds, as well as more general features, such as the color of the image. To increase the dataset and to deal with differences in the real-world dataset, data augmentation methods such as rotating, zooming, and scaling were used. A CNN model with filters was used in the intermediate layer. This was followed by a clustering function and the rectified linear unit (ReLU) activation function at the end, and then a sigmoid activation unit to classify the images into two groups. In addition to training and validation losses, accuracy of the validation set, a confusion matrix, precision, recall, and F1-score were used to analyze the results. After applying dropout, batch normalization, and learning rate decay to achieve a smooth gradient descent, the CNN obtained a better result. Following this, the parameters of an already trained model which was already well optimized on a large dataset were applied by using transfer learning. As a result of freezing the parameters, the proposed model ensures that only the variables from the last classification layer are trained. In contrast, the variables from all other layers remain the same.

3.3.7 Algorithm

The framework consists of several steps. Algorithm 1 checks if the address is valid by registering it in the blockchain and verifying it only with trusted sources. A registry is created when a user is not registered by adding their Hyperledger address to a list of the repository's resources. After completing the registration, the data extracted from the off-chain database IPFS are fed into the smart contract, as shown in Algorithm 1. Then the system initializes the deep learning model to diagnose COVID-19 as shown in Algorithm 2. The pre-processing such as data augmentation is computed for a dataset and combines the augmented normal dataset and the augmented infected dataset as A. Data A is then divided into training (A_Train) and testing (A_Test) sets. Next, the CNN is computed by validating and training the set with different epochs and batch size B and then use to identify the infected and normal cases. Finally, the data us updated the in the Blockchain and IPFS database.

Algorithm 1. Validation

```
Input: Electronic Health record Infected, Recovered and Dead, IPFS Database, Hyperledger Fabric
1: Register Data owner and Data user
2: Identity and Key Generation by Hyperledger Blockchain as Hash value
3: User Login
4: Verifies Identity of User
5: Data Owner Store data in IPFS
6: Data Users request health data from database using their keys
7: Verify if the Identity Exists and Access Control
8: If Exists and Granted then
9:          Hyperledger Blockchain Allowed to transfer Health Record
10:         Initialize the Deep learning Model
11:         Go to Algorithm 2
10:         Diagnosis the Infected and Normal
11: else
12:         Register
13:         Append the Data to the Repository
14: End
14: Increment the number of Infected and Normal
15: Update in Blockchain and IPFS
16: End
```

Algorithm 2. Initialization of the model to diagnose COVID-19

Input: CNN, Dataset D, Augumented_Data AD Normal [ ], Augumented_Data AD Infected [ ], Epochs N, Batch B and Fold K

```
Divide Dataset D into Subsets S
For i=1 to S
Train and Do Validation S
For i=0 to S
   A=AD Normal Dataset Union AD Infected Dataset
   Divide A into 95% training Set and 5% test Set
   Divide the A_Train to subset Into
For i=1 to At
   A_Train=A-At_n
   A_Valid= At_n
For i=1 to N
   Batch_Training=sample (A, B)
  CNN.train (Batch_Training)
  Batch_Validate=sample (A, B)
  CNN.validate (Batch_Training)
  CNN.test (A_Test)
  Return CNN
```

4. Experiments and Results

4.1 Implementation

The proposed system is implemented in Hyperledger composer, permissioned Hyperledger Fabric, blockchain on-chain storage, and IPFS for off-chain storage. The front-end communication is carried out using DApp through Hyperledger composer. All implementation has been done through a web user interface in Python, HTML, CSS, Node JS, Kafka Ordering Service, Keras, and TensorFlow backend. A virtual machine running Ubuntu Linux 16.04 LTS with 3 GB RAM runs the blockchain on the desktop. A Windows 10 operating system with Intel Core i7-8700K 3.7 GHz processor and 16 GB of RAM was configured to implement the encryption algorithm. A smart contract is a kind of software contract that implements policies, enforces rules, and stores metadata about ownership, permissions, and data integrity. These contracts are written in Java. The permissioned Hyperledger Fabric client deploys smart contracts. Smart contracts are tested in the Hyperledger composer development environment.

4.2 Dataset

Various hospital sources are used to collect real-world data. A variety of datasets have also been collected and are available at [50]. A dictionary data structure was used to first separate COVID-19 positive case indices, then the anterior-posterior view X-ray indices were separated. After getting the indexes, the images are plotted and saved using the Python library. Kaggle's chest X-ray images (pneumonia) dataset [51] provided 25 X-ray images. COVID-19 positive X-rays provided a total of 148 images. Kaggle provided 148 additional normal X-rays. A total of 296 images were gathered, of which 16 were placed in a test folder for a final review at the end.

4.3 Results

The blockchain users will control the issuance of roles, attributes, registration of users, classification of COVID-19 positive and negative patients, and the management of those patients once the proposed infrastructure is implemented.

4.3.1 Results of identity-based access control

Based on attribute-based access control management and blockchain, the proposed scheme is evaluated to see if it is effective in ensuring security and privacy of medical data. For the provided nodes in the healthcare network, off-chain and on-chain exchanges have been tested. Patients trigger numerous transactions with healthcare provider nodes to test the network's operation and the time needed to process each transaction. The proposed scheme includes a credible and comprehensive medical history of each patient. A large number of transactions are initiated by patients, so they can observe how the network functions and how long it takes to process each transaction. Each patient's medical history is also documented within the proposed scheme. The IPFS database is updated with diagnostic information details. A healthcare provider's time to generate keys and their time to enter diagnostic reports into the database is illustrated in Fig. 5. From Fig. 5, it is apparent that encrypting data requires more computation than decrypting it. In the original data store, IPFS files are stored, but in the blockchain network, hash values are permanently stored. Blocks contain data from the internal clock as well as internal clock timestamps, diagnosis hash pointers, patient data hash pointers, and subsequent block hash values. In Figs. 5–7, the peer nodes measure how long it takes to access a transaction. According to these figures, the access time increases as the number of attributes increases.

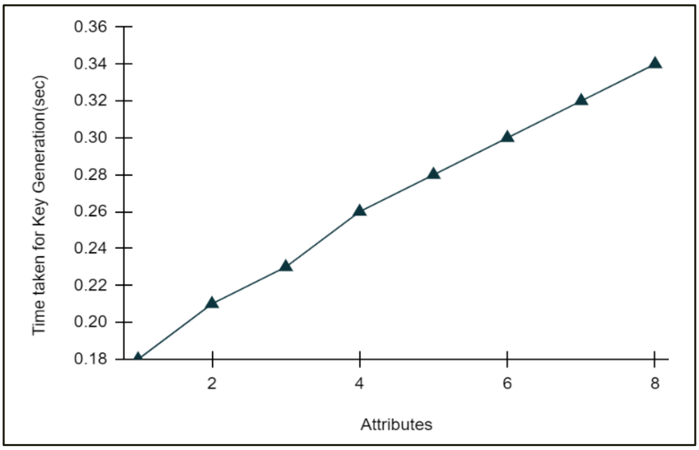

Fig. 5. Time taken to generate key.

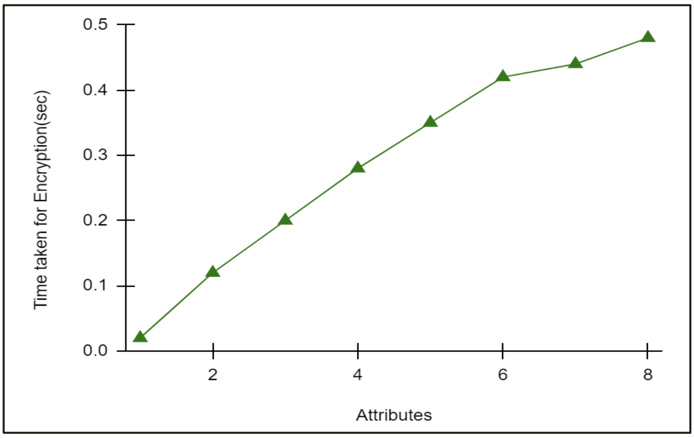

Fig. 6. Time taken for encrypting electronic health data.

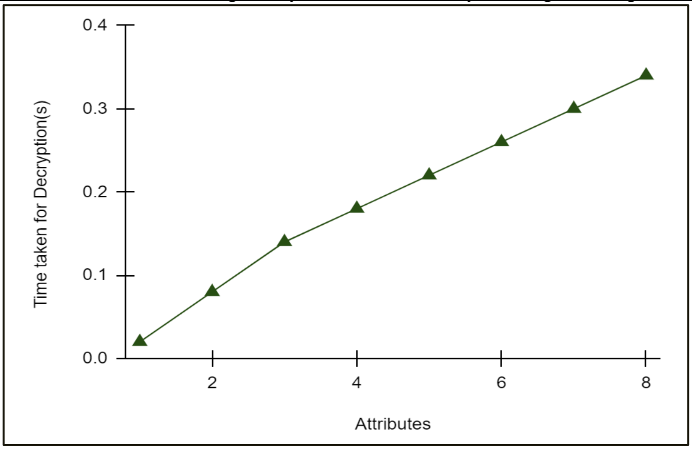

Fig. 7. Time taken for decrypting electronic health data.

4.4 Results of COVID-19 Classification

Numerous studies and research have been conducted on COVID-19 [4, 52–54], but none of these have taken advantage of sharing data for improved predictions. COVID-19 diagnoses were applied to three experiments.

4.4.1 Experiment 1

In Experiment 1, the ratio of training and validation split is 3:1, which means that 222 images are in the training set and 74 images in the testing set. Data normalization and image augmentation are the steps involved in pre-processing. A training set can be artificially augmented by applying random image transformations to existing images to increase the number of images. In order for the input parameters to have the same distribution of values, data normalization must be performed. Consequently, the learning rate can be increased. Positive numbers must appear in the input image, which means that the scale should be between [0, 1] and [0, 255]. In RGB/grayscale images, a convolutional filter was used for each color channel. In grayscale images, convolutions are applied to each color channel individually. After each convolution, a bias value is added to the result to get the result of the convolution. The same window size and stride are used for maximum pooling when working with RGB images. As with grayscale images, each color channel is subjected to maximum pooling. Maximum pooling reduces the parameter learning complexity by reducing the size of each image after each CNN layer. Deep networks are quite problematic from the viewpoint of sigmoid activation function. As a result, all values between 0 and 1 are squashed and the neuron's output and gradient disappear entirely after repeated use. The ReLU is one of the modern methods. At least on its right-hand side, the ReLU contains a derivative of 1. Activating ReLU yields nonzero gradients from some neurons, enabling the training to continue at a good pace, even if some neurons give zero gradients. After training 30 epochs, the confusion matrix for the experiment was 1 as shown in Fig. 8. The validation accuracy reached 98% and the training accuracy rate was only 81.25% based on unseen data as shown in Fig. 9. The validation loss reached 0.5923 and the training loss reached 0.543 as shown in Fig. 10.

4.4.2 Experiment 2

Experiment 2 follows the validation and pre-processing steps same as Experiment 1. The overfitting and dropout have been regularized using the Keras dropout layer. The confusion matrix for Experiment 2 after training 30 epochs is shown in Fig. 11. The training of 30 epochs yielded a validation accuracy of up to 98%. On unseen test data, however, a 96% test accuracy was achieved as shown in Fig. 12. The validation loss reached 0.23 and training loss reached 0.12 as shown in Fig. 13.

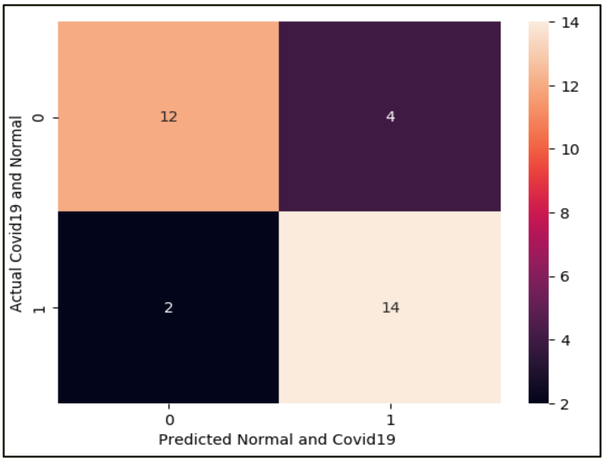

Fig. 8. Confusion matrix of prediction for Experiment 1.

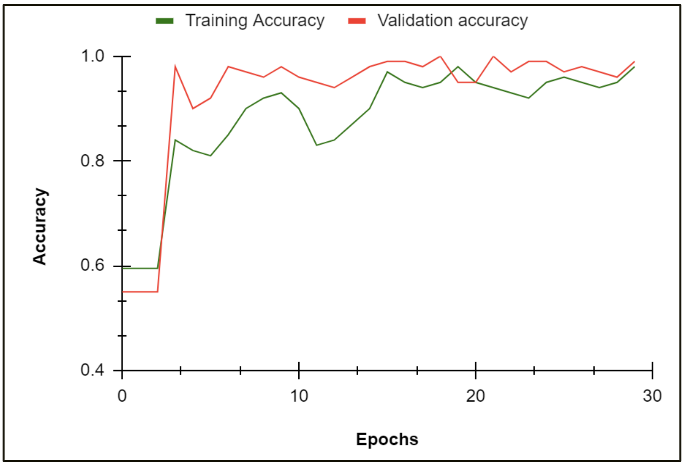

Fig. 9. Training and validation accuracy for 30 epochs (Experiment 1).

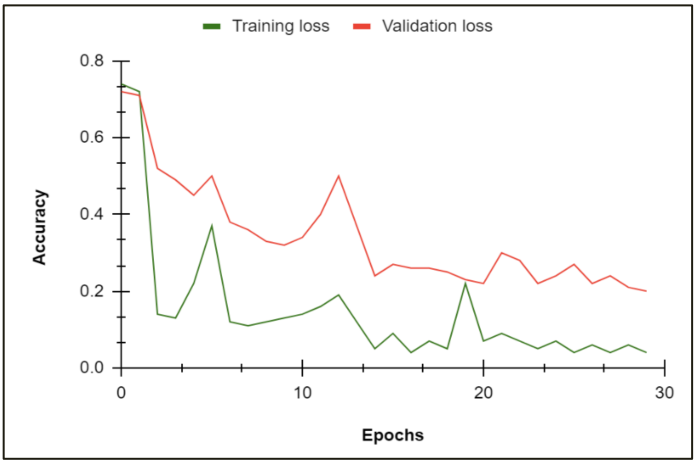

Fig. 10. Training and validation loss for 30 epochs (Experiment 1).

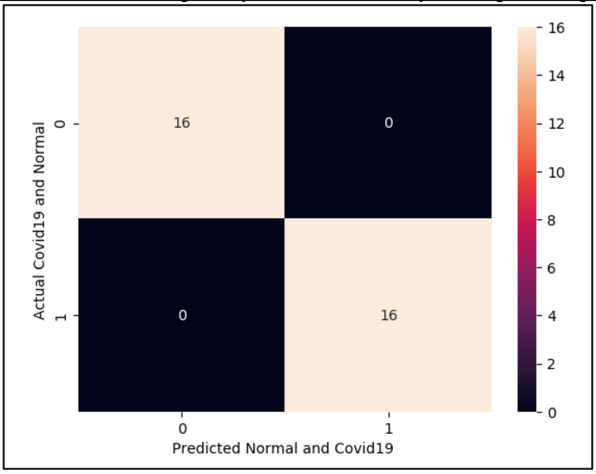

Fig. 11. Confusion matrix of prediction for Experiment 2.

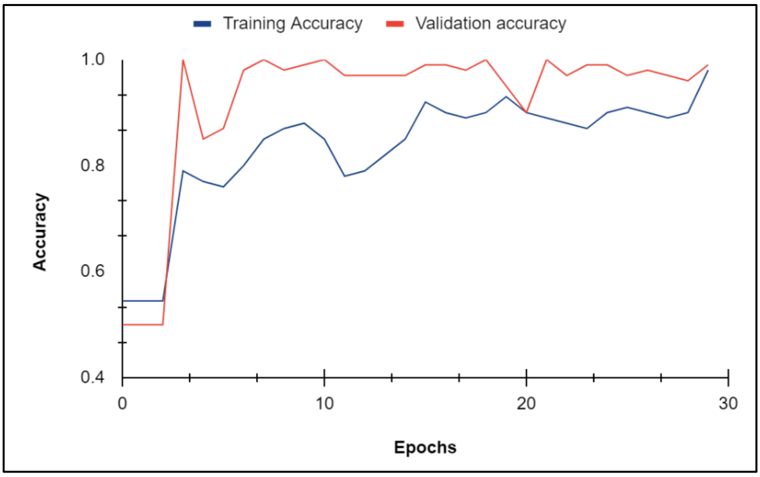

Fig. 12. Training and validation accuracy for 30 epochs (Experiment 2).

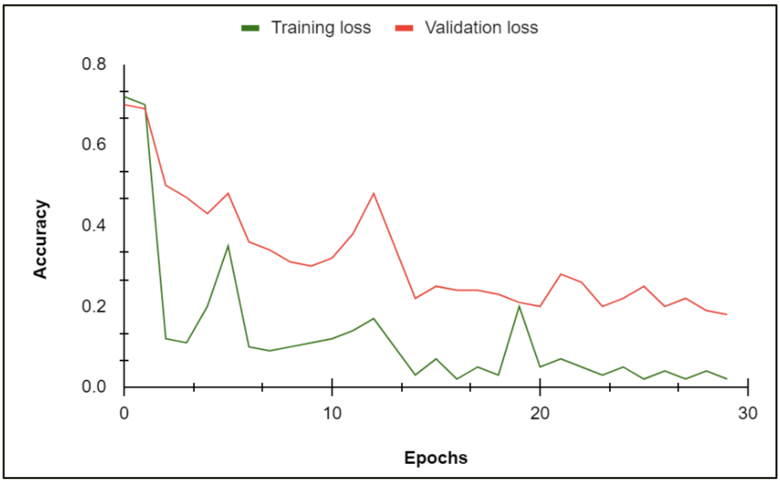

Fig. 13. Training and validation loss for 30 epochs (Experiment 2).

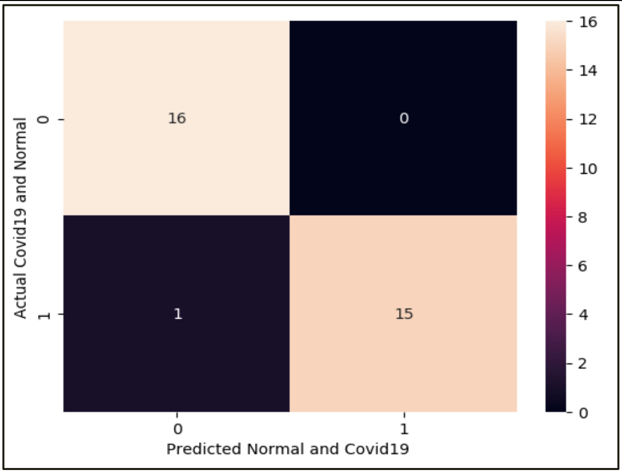

Fig. 14. Confusion matrix of prediction for Experiment 3.

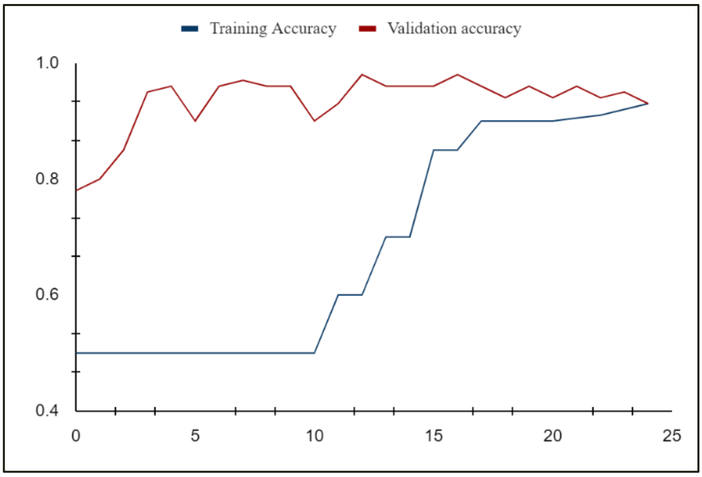

Fig. 15. Training and validation accuracy for 25 epochs (Experiment 3).

Fig. 16. Training and validation loss for 25 epochs (Experiment 3).

4.4.3 Experiment 3

Experiment 3 also follows the validation and pre-processing steps same as Experiment 1. To start fast and decay the learning rate exponentially, learning rate decay is the optimal solution. Batch normalization is concerned with the distribution of neuron output based on the normalized activation function of neurons. The confusion matrix for the Experiment 3 after training 25 epochs is shown in the Fig. 14. The training of 25 epochs yielded a validation accuracy of up to 94%. On unseen test data, however, a 96% test accuracy was achieved as shown in Fig. 15. The validation loss reached 0.03 and training loss reached 0.01 as shown in Fig. 16.

5. Conclusion

In this paper, we propose a framework for improving the recognition of X-ray images with the use of up-to-date data and preserving privacy while sharing the data among all users. An extensive experiment was performed on the training and testing sets using various deep learning models. The experimental results show that the proposed obtained a good performance, attaining accuracy of 90%–95% and compared to state of the art method in the literature reviewed. By sharing ownership of the data to train a global and more accurate model, the proposed model can be used to detect COVID-19 patients as shown in the confusion matrix. However, the limited dataset was a major obstacle to overcome. In the future, modifying the general structure of the CNN model, as well as increasing the number of training dataset could lead to better image analysis.

Author's Contributions

Conceptualization, EAM. Investigation and methodology, EAM, CZ. Project administration, EAM, JKA, MV, EI. Resources, EAM, JKA, MV, EI. Supervision, CZ. Writing of the original draft, EAM. Writing of the review and editing, EAM, CZ. Software EA, JKA, MV, EI. Validation, EAM, CZ, JKA. Formal analysis, EAM, JKA, MV, EI. Data curation, EAM, CZ, JKA, MV, EI. Visualization, EAM, JKA, MV, EI.

Funding

This paper was supported by the key research and development plan (social development) projects of Jiangsu province (No. BE2016630 and BE2017628) and the scientific research project of Wuxi health and family planning commission (No. Z201603).

Competing Interests

The authors declare that they have no competing interests.

References

[1] World Health Organization, "Advice for the public: coronavirus disease (COVID-19)," 2023 [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.

[2] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on COVID-19: an update: radiology scientific expert panel," *Radiology*, vol. 296, no. 2, pp. E113-E114, 2020. https://doi.org/10.1148/radiol.2020200527

[3] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, and W. Zhang, "Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2584-2594, 2020. https://doi.org/10.1109/TMI.2020.2996256

[4] X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, et al., "Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2595-2605, 2020. https://doi.org/10.1109/TMI.2020.2995508

[5] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, et al., “Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2606-2614, 2020. https://doi.org/10.1109/TMI.2020.2992546
[6] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin, “VerifyNet: secure and verifiable federated learning,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 911-926, 2019. https://doi.org/10.1109/TIFS.2019.2929409
[7] M. A. Azad, J. Arshad, S. M. A. Akmal, F. Riaz, S. Abdullah, M. Imran, and F. Ahmad, “A first look at privacy analysis of COVID-19 contact-tracing mobile applications,” *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15796-15806, 2021. https://doi.org/10.1109/JIOT.2020.3024180
[8] T. C. Chiu, Y. Y. Shih, A. C. Pang, C. S. Wang, W. Weng, and C. T. Chou, “SemisuVpervised distributed learning with non-IID data for AIoT service platform,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9266-9277, 2020. https://doi.org/10.1109/JIOT.2020.2995162
[9] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, “Blockchain empowered asynchronous federated learning for secure data sharing in Internet of vehicles,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4298-4311, 2020. https://doi.org/10.1109/TVT.2020.2973651
[10] H. Sun, S. Li, F. R. Yu, Q. Qi, J. Wang, ad J. Liao, “Toward communication-efficient federated learning in the Internet of Things with edge computing,” *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11053-11067, 2020. https://doi.org/10.1109/JIOT.2020.2994596
[11] K. Yang, T. Jiang, Y. Shi, and Z. Ding, “Federated learning via over-the-air computation,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2022-2035, 2020. https://doi.org/10.1109/TWC.2019.2961673
[12] Z. Zhou, S. Yang, L. Pu, and S. Yu, “CEFL: online admission control, data scheduling, and accuracy tuning for cost-efficient federated learning across edge nodes,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9341-9356, 2020. https://doi.org/10.1109/JIOT.2020.2984332
[13] D. Ngabo, W. Dong, E. Ibeke, C. Iwendi, and E. Masabo, “Tackling pandemics in smart cities using machine learning architecture,” *Mathematical Biosciences and Engineering*, vol. 18, no. 6, pp. 8444-8461, 2021. https://doi.org/10.3934/mbe.2021418
[14] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, K. (2017, October). “Practical secure aggregation for privacy-preserving machine learning,” in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, Dallas, TX, 2017, pp. 1175-1191. https://doi.org/10.1145/3133956.3133982
[15] Y. Liu, J. Q. James, J. Kang, D. Niyato, and S. Zhang, “Privacy-preserving traffic flow prediction: a federated learning approach,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7751-7763, 2020. https://doi.org/10.1109/JIOT.2020.2991401
[16] Y. Qu, L. Gao, T. H. Luan, Y. Xiang, S. Yu, B. Li, and G. Zheng, “Decentralized privacy using blockchain-enabled federated learning in fog computing,” *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5171-5183, 2020. https://doi.org/10.1109/JIOT.2020.2977383
[17] S. A. Latif, F. B. Xian Wen, C. Iwendi, L. F. Wang, S. M. Mohsin, Z. Han, and S. S. Band, “AI-empowered, blockchain and SDN integrated security architecture for IoT network of cyber physical systems,” *Computer Communications*, vol. 181, pp. 274-283, 2022. https://doi.org/10.1016/j.comcom.2021.09.029
[18] C. Iwendi, S. Mohan, E. Ibeke, A. Ahmadian, and T. Ciano, “Covid-19 fake news sentiment analysis,” *Computers and Electrical Engineering*, vol. 101, article no. 107967, 2022. https://doi.org/10.1016/j.compeleceng.2022.107967
[19] 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, article no. 2110, 2021. https://doi.org/10.3390/electronics10172110
[20] V. Mani, P. Manickam, Y. Alotaibi, S. Alghamdi, and O. I. Khalaf, “Hyperledger healthchain: patient-centric IPFS-based storage of health records,” *Electronics*, vol. 10, no. 23, article no. 3003, 2021. https://doi.org/10.3390/electronics10233003
[21] V. Mani, C. Kavitha, S. S. Band, A. Mosavi, P. Hollins, and S. Palanisamy, “A recommendation system based on AI for storing block data in the electronic health repository,” *Frontiers in Public Health*, vol. 9, article no. 831404, 2022. https://doi.org/10.3389/fpubh.2021.831404

[22] R. L. Kumar, F. Khan, S. Din, S. S. Band, A. Mosavi, and E. Ibeke, "Recurrent neural network and reinforcement learning model for COVID-19 prediction," *Frontiers in Public Health*, vol. 9, article no. 744100, 2021. https://doi.org/10.3389/fpubh.2021.744100

[23] H. Sellak and M. Grobler, "Towards intelligent, multidimensional context-aware personalised healthcare and wellness support services," *Studies in Health Technology and Informatics*, vol. 266, pp. 168-173, 2019. https://doi.org/10.3233/SHTI190790

[24] L. Rita, "Building a food recommendation system," 2020 [Online]. Available: https://towardsdatascience.com/building-a-food-recommendation-system-90788f78691a.

[25] H. El Bilali, C. Callenius, C. Strassner, and L. Probst, "Food and nutrition security and sustainability transitions in food systems," *Food and Energy Security*, vol. 8, no. 2, article no. e00154, 2019. https://doi.org/10.1002/fes3.154

[26] M. Hosseini-Pozveh, M. Nematbakhsh, and N. Movahhedinia, "A multidimensional approach for context-aware recommendation in mobile commerce," 2009 [Online]. Available: https://arxiv.org/abs/0908.0982.

[27] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: algorithms, challenges, metrics, and business opportunities," *Applied Sciences*, vol. 10, no. 21, article no. 7748, 2020. https://doi.org/10.3390/app10217748

[28] World Health Organization. *The State of Food Security and Nutrition in the World 2020: Transforming Food Systems for Affordable Healthy Diets*. Rome, Italy: Food & Agriculture Organization of the United Nations, 2020.

[29] S. Abhari, R. Safdari, L. Azadbakht, K. B. Lankarani, S. R. N. Kalhori, B. Honarvar, et al., "A systematic review of nutrition recommendation systems: with focus on technical aspects," *Journal of Biomedical Physics & Engineering*, vol. 9, no. 6, article no. 591, 2019. https://doi.org/10.31661/jbpe.v0i0.1248

[30] G. Muscogiuri, L. Barrea, S. Savastano, and A. Colao, "Nutritional recommendations for CoVID-19 quarantine," *European Journal of Clinical Nutrition*, vol. 74, no. 6, pp. 850-851, 2020. https://doi.org/10.1038/s41430-020-0635-2

[31] S. Alkire, J. M. Roche, P. Ballon, J. Foster, M. E. Santos, and S. Seth, *Multidimensional Poverty Measurement and Analysis*. Oxford, UK: Oxford University Press, 2015.

[32] M. Stratigi, H. Kondylakis, and K. Stefanidis, "Multidimensional group recommendations in the health domain," *Algorithms*, vol. 13, no. 3, article no. 54, 2020. https://doi.org/10.3390/a13030054

[33] R. Sookrah, J. D. Dhowtal, and S. D. Nagowah, "A DASH diet recommendation system for hypertensive patients using machine learning," in *Proceedings of 2019 7th International Conference on Information and Communication Technology (ICoICT)*, Kuala Lumpur, Malaysia, 2019, pp. 1-6. https://doi.org/10.1109/ICoICT.2019.8835323

[34] P. Vasireddy, "An autonomous diet recommendation bot using intelligent automation," in *Proceedings of 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2020, pp. 449-454. https://doi.org/10.1109/ICICCS48265.2020.9121120

[35] R. Y. Toledo, A. A. Alzahrani, and L. Martinez, "A food recommender system considering nutritional information and user preferences," *IEEE Access*, vol. 7, pp. 96695-96711, 2019. https://doi.org/10.1109/ACCESS.2019.2929413

[36] 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. https://doi.org/10.1109/ACCESS.2020.2968537

[37] N. Mulla, S. Kurhade, M. Naik, and N. Bakereywala, "An intelligent application for healthcare recommendation using fuzzy logic," in *Proceedings of 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2019, pp. 466-472. https://doi.org/10.1109/ICECA.2019.8821959

[38] B. Akay, O. Kaynar, and F. Demirkoparan, "Deep learning based recommender systems," in *Proceedings of 2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, Turkey, 2017, pp. 645-648. https://doi.org/10.1109/UBMK.2017.8093489

[39] K. Al Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: a conceptual framework," *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 47-55, 2021. https://doi.org/10.26599/BDMA.2020.9020015

[40] 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 in Public Health*, vol. 9, article no. 737269, 2021. https://doi.org/10.3389/fpubh.2021.737269

[41] C. Iwendi, E. Ibeke, H. Eggoni, S. Velagala, and G. Srivastava, "Pointer-based item-to-item collaborative filtering recommendation system using a machine learning model," *International Journal of Information Technology & Decision Making*, vol. 21, no. 1, pp. 463-484, 2022. https://doi.org/10.1142/S0219622021500619

[42] R. Frey, D. Worner, and A. Ilic, "Collaborative filtering on the blockchain: a secure recommender system for e-commerce," in *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS)*, San Diego, CA, 2016, pp. 1-5.

[43] T. Y. Yeh and R. Kashef, "Trust-based collaborative filtering recommendation systems on the blockchain," *Advances in Internet of Things*, vol. 10, no. 4, pp. 37-56, 2020. https://doi.org/10.4236/ait.2020.104004

[44] M. Arora, A. B. Chopra, and V. S. Dixit, "An approach to secure collaborative recommender system using artificial intelligence, deep learning, and blockchain," in *Intelligent Communication, Control and Devices*. Singapore: Springer, 2020, pp. 483-495. https://doi.org/10.1007/978-981-13-8618-3_51

[45] M. M. Salim, I. Kim, U. Doniyor, C. Lee, and J. H. Park, "Homomorphic encryption based privacy-preservation for IOMT," *Applied Sciences*, vol. 11, no. 18, article no. 8757, 2021. https://doi.org/10.3390/app11188757

[46] S. K. Singh, J. H. Park, P. K. Sharma, and Y. Pan, "BIIoVT: blockchain-based secure storage architecture for intelligent Internet of vehicular things," *IEEE Consumer Electronics Magazine*, vol. 11, no. 6, pp. 75-82, 2022. https://doi.org/10.1109/MCE.2021.3089992

[47] A. Kishor, C. Chakraborty, and W. Jeberson, "Reinforcement learning for medical information processing over heterogeneous networks," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 23983-24004, 2021. https://doi.org/10.1007/s11042-021-10840-0

[48] A. Kishor and C. Chakarbarty, "Task offloading in fog computing for using smart ant colony optimization," *Wireless Personal Communications*, vol. 127, pp. 1683-1704, 2022. https://doi.org/10.1007/s11277-021-08714-7

[49] J. Cha, S. K. Singh, T. W. Kim, and J. H. Park, "Blockchain-empowered cloud architecture based on secret sharing for smart city," *Journal of Information Security and Applications*, vol. 57, article no. 102686, 2021. https://doi.org/10.1016/j.jisa.2020.102686

[50] J. P. Cohen, "Covid-19 image data collection," 2020 [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset.

[51] P. Mooney, "Chest X-ray images (pneumonia)," 2018 [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

[52] A. Kuzdeuov, D. Baimukashev, A. Karabay, B. Ibragimov, A. Mirzakhmetov, M. Nurpeiissov, M. Lewis, and H. A. Varol, "A network-based stochastic epidemic simulator: controlling covid-19 with region-specific policies," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2743-2754, 2020. https://doi.org/10.1109/JBHI.2020.3005160

[53] L. Sun, Z. Mo, F. Yan, L. Xia, F. Shan, Z. Ding, et al., "Adaptive feature selection guided deep forest for covid-19 classification with chest CT," 2020 [Online]. Available: https://arxiv.org/abs/2005.03264.

[54] N. Deepa, Q. V. Pham, D. C. Nguyen, S. Bhattacharya, B. Prabadevi, T. R. Gadekallu, P. K. R. Maddikunta, F. Fang, and P. N. Pathirana, "A survey on blockchain for big data: approaches, opportunities, and future directions," *Future Generation Computer Systems*, vol. 131, pp. 209-226, 2022. https://doi.org/10.1016/j.future.2022.01.017