IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

Verifiable Secret Sharing Scheme Using Deep Learning Framework In Cloud Environment

Geetha T[1], Hariharan P[2], Matheshwaran K[3], Hariharan B[4], Mohamed Hisham R[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Dhanalakshmi Srinivasan Engineering College, Perambalur, India

Abstract: *Cross-device federated learning is a machine learning approach that enables multiple devices to collaboratively train a model without sharing their data with each other. This approach is particularly useful in medical settings where data privacy and security are paramount. In this context, medical data is sensitive and protected by law. Federated learning can help to preserve the privacy of medical data while still allowing for the development of models that can be used to improve patient outcomes. One challenge with federated learning is the need to protect the model during training and inference. Model encryption is a technique that can be used to protect the model from unauthorized access. Elliptic Curve Cryptography (ECC) is a form of encryption that is well-suited for federated learning due to its ability to efficiently encrypt and decrypt data. In this project, propose a cross-device federated learning approach that utilizes medical datasets to build a predictive model. We also employ ECC to encrypt the model during training and inference. Divide the medical dataset into subsets that are distributed across multiple devices. Train the model collaboratively across all devices using federated learning techniques. Then use ECC to encrypt the trained model to protect it from unauthorized access. The proposed system also provides a more accurate prediction of disease risk while preserving patient confidentiality. The results show that the SVM-based model can achieve high accuracy in predicting disease risk, and the encrypted data can be used effectively to train the model without compromising patient privacy. Additionally, our use of ECC encryption provides an extra layer of security for the model, ensuring that it remains protected during training and inference.*

Keywords: Federated Learning, Medical Data, Eliptic Curve Cryptography, Security And Privacy

I. INTRODUCTION

Cross-domain federated learning is a variant of federated learning that involves training a machine learning model across multiple domains or data sources that have different distributions or characteristics. In traditional federated learning, all devices or data sources have the same distribution of data, but in cross-domain federated learning, each data source has its own distribution of data. The motivation behind cross-domain federated learning is that data from different sources can provide complementary information, allowing for the development of more accurate and robust machine learning models. For example, medical data from different hospitals or clinics may have different patient populations or treatment protocols, but collectively, this data can provide a more comprehensive understanding of a particular disease or condition.

The key challenge in cross-domain federated learning is the domain shift, which refers to the differences in data distributions between different domains. To address this challenge, various techniques have been proposed, such as domain adaptation and transfer learning, which aim to reduce the differences between the data distributions of different domains. Cross-domain federated learning has several potential applications in various fields, such as healthcare, finance, and social media. For example, in healthcare, cross-domain federated learning can be used to develop more accurate predictive models for diseases that affect different populations or that have different treatment protocols. In finance, cross-domain federated learning can be used to develop more accurate fraud detection models by combining data from different financial institutions. In social media, cross-domain federated learning can be used to develop more accurate sentiment analysis models by combining data from different social media platforms. Cross-domain federated

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

learning is still an emerging research area, and there are several challenges that need to be addressed before it can be widely applied in practical settings. Some of these challenges include:

- **Data heterogeneity:** In cross-domain federated learning, data from different domains may have different features, formats, or data types. This can make it challenging to develop a unified model that can accommodate all the different types of data.
- **Privacy and security:** Cross-domain federated learning involves sharing data across different domains, which raises privacy and security concerns. Ensuring the confidentiality and integrity of data across different domains is critical for preserving privacy and maintaining trust.
- **Model bias:** Training a model across multiple domains can lead to model bias, where the model is skewed towards one domain or population at the expense of others. This can lead to inaccurate predictions or unfair outcomes for certain populations.
- **Communication overhead:** In cross-domain federated learning, devices or data sources need to communicate with each other to exchange model parameters and update the model. This can result in significant communication overhead, especially if the data sources are geographically distributed or have limited bandwidth.

II. FRAME WORK

2.1 Elliptical Curve Cryptography

Elliptical curve cryptography (ECC) is a public key encryption technique based on elliptic curve theory that can be used to create faster, smaller, and more efficient cryptographic keys. ECC generates keys through the properties of the elliptic curve equation instead of the traditional method of generation as the product of very large prime numbers. The technology can be used in conjunction with most public key encryption methods, such as RSA, and Diffie-Hellman. According to some researchers, ECC can yield a level of security with a 164-bit key that other systems require a 1,024-bit key to achieve. Because ECC helps to establish equivalent security with lower computing power and battery resource usage, it is becoming widely used for mobile applications. Elliptic curves are an algebraic structure and their use for cryptography. They feature properties which allow the setup of a problem similar to the well-known discrete logarithm problem of finite fields – also known as Galois fields (GF).ECC includes key agreement, encryption, and digital signature algorithms. The key distribution algorithm is used to share a secret key, the encryption algorithm enables confidential communication, and the digital signature algorithm is used to authenticate the signer and validate the integrity of the message

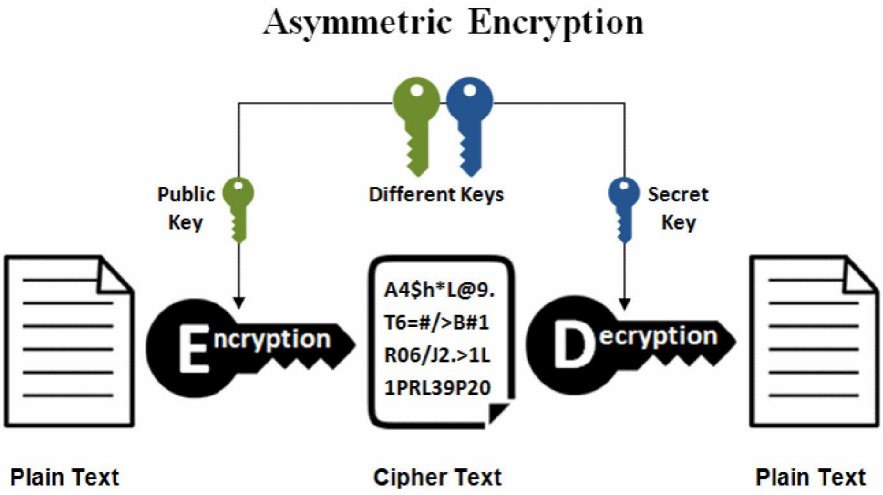

Fig.: ECC Algorithm

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

III. SYSTEM ANALYSIS

2.1 Existing System

VERSA, a verified safe aggregation protocol for cross-device federated learning, has been incorporated into the current system. VERSA allows users and the central server to both use a lightweight pseudorandom generator to prove and verify the accuracy of model aggregation while minimising the cost of verification. This eliminates the need for any trusted configuration for user-to-user verification. A centralised model is trained using a decentralised machine learning technique called federated learning (FL), which distributes the training data to each user and never shares it with anyone else. By repeating a series of rounds, a central server manages the training process. The server chooses a group of users and gives them access to the model's current parameters. These users might make it through the entire round or lose.

3.2 Disadvantages

Limited Verifiability: Some existing systems may not offer the same level of verifiability as VERSA. Without a verifiable secret sharing scheme, it may be challenging to ensure the integrity of the aggregated model and prevent malicious tampering.

Limited Scalability: Some existing systems may struggle to scale to accommodate large numbers of participating devices. As the number of participating devices increases, the computation and communication overhead may become prohibitively high.

Limited Robustness: Some existing systems may not be robust enough to handle failures or malicious attacks. For example, in some systems, a single malicious device can compromise the entire aggregation process, making the entire system vulnerable.

3.3 Proposed System

The proposed system is a framework for verifiable, secure aggregation in cross-device federated learning. It uses a combination of homomorphic encryption, secure multi-party computation, verifiable secret sharing, and the ECC algorithm to ensure the privacy, security, and integrity of the deep learning model updates and evaluation results exchanged between participating devices and the aggregation server. Each device trains its deep learning model locally using the convolutional neural network algorithm. The local model is then encrypted using the ECC algorithm before being sent to the aggregation server. The aggregation server combines the encrypted model updates from each device using a secure aggregation protocol that uses secure multi-party computation techniques. This proposed aggregated model also supports the disease prediction process. Disease prediction using symptom data and a classification disease dataset is an important task in the field of healthcare. It involves predicting the disease that a patient is likely to have based on their symptoms. SVM (Support Vector Machine) is a machine learning algorithm that can be used to build a classifier for disease prediction. SVM is a popular algorithm for classification tasks as it has high accuracy and can handle large datasets.

3.4 Advantages

Improved Privacy: The use of ECC encryption ensures that data remains encrypted throughout the entire process, which helps to protect the privacy of users and their sensitive data.

Enhanced Security: The use of secure multi-party computation techniques and verifiable secret sharing schemes ensures that model updates and evaluation results are combined without revealing any private data, which helps to prevent attacks by malicious actors.

Efficient Collaboration: The use of federated learning allows devices to collaborate on machine learning tasks without needing to share their data with a centralised server, which helps to reduce communication and computation costs.

Increased Accuracy: The ability to aggregate model updates and evaluation results from multiple devices can help improve the accuracy of the overall model by incorporating diverse perspectives and data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

IV. SYSTEM ARCHITECTURE

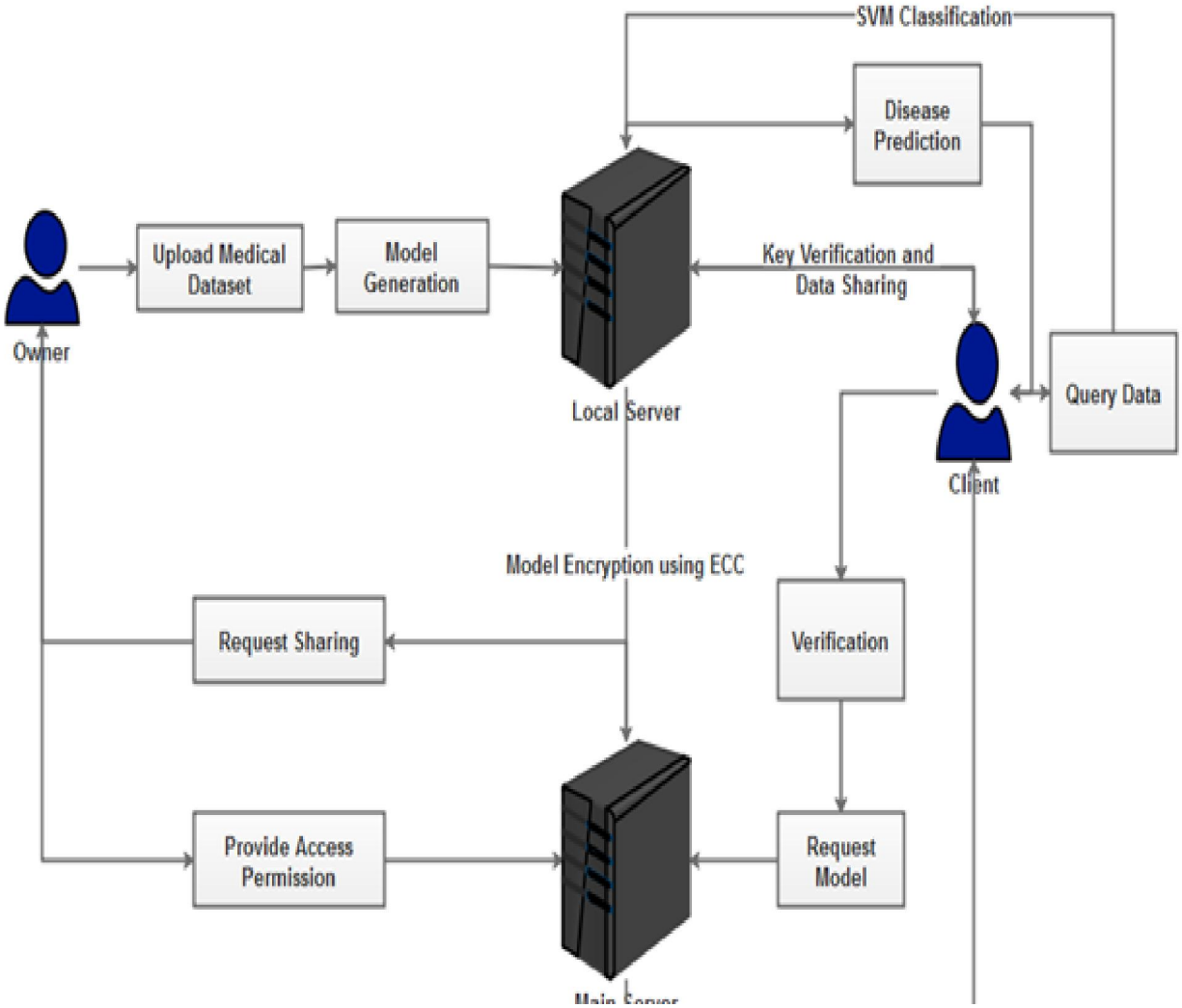

Figure 4 System Architecture

Prepare the medical dataset by collecting and preprocessing the data from different sources. This includes cleaning the data, removing any outliers or inconsistencies, and normalising the data to ensure that it has a consistent format and scale. The next step is to design the machine learning model that will be trained using the medical dataset. The model should be optimised to achieve high accuracy while minimising the risk of overfitting. To ensure the privacy and security of the medical data, the model parameters are encrypted using the ECC (Elliptic Curve Cryptography) algorithm. ECC is a public-key cryptography algorithm that provides strong encryption while minimising the computational overhead. Finally, the trained model is deployed for use in the healthcare industry. The model can be used for disease diagnosis, personalised treatment plans, clinical trials, and healthcare analytics, among other applications.

V. FLOW DIAGRAM

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

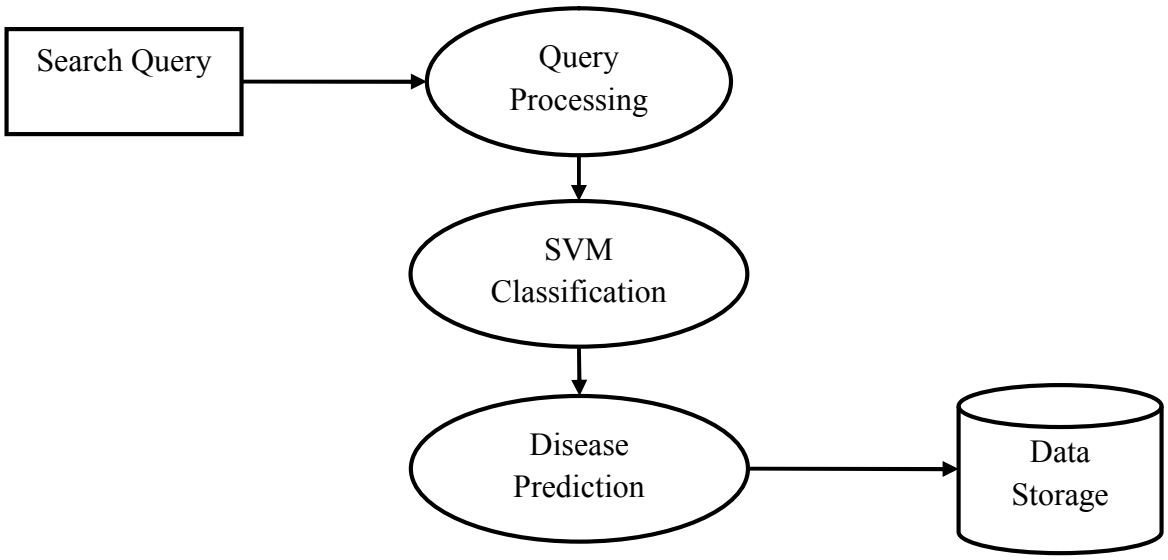

Fig.: Flow Diagram

VI. SYSTEM IMPLEMENTATION

6.1 Modules

- Model Initialization
- Encrypted Model Storage
- Client Process
- Model Request
- Model Distribution
- Disease Prediction

6.2 Module Description

6.2.1 Model Initialization

Model initialization is a critical step in federated learning for cloud data sharing. It involves creating an initial machine learning model on a central server, which will be used as the starting point for the collaborative training process. There are several ways to initialize the model in federated learning. One common approach is to use a pre-trained model that has already been trained on a large, representative dataset. The pre-trained model can be fine-tuned on the cloud data using the federated learning approach.

6.2.2 Encrypted Model Storage

Encrypted model storage involves encrypting the machine learning model before it is stored on a central server or in the cloud. This ensures that the model remains private and secure, even if the server or storage system is compromised. In proposed system asymmetric encryption techniques that can be used for model storage in federated learning, including ECC (Elliptic Curve Cryptography) algorithm. ECC allows computations to be performed on encrypted data without the need to decrypt it first, while secure multi-party computation allows multiple parties to perform a computation collaboratively without revealing their inputs.

VII. CONCLUSION

Here implemented cross domain federated learning system with disease dataset model generation and disease prediction approach. Cross-domain federated learning can enable providers to collaborate across different healthcare domains, such as hospitals, clinics, and research institutions. This can lead to the development of more accurate and robust models, as the data used to train the models will be more diverse. The use of disease datasets can provide valuable insights into disease prevention, diagnosis, and treatment. By using federated learning to generate predictive models

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, May 2023

based on disease datasets, healthcare providers can improve their ability to predict disease outcomes and develop more effective treatment plans. A disease prediction approach can be highly beneficial in helping healthcare providers to identify patients who are at high risk of developing certain diseases. This can enable providers to intervene early and prevent the progression of the disease, ultimately leading to better healthcare outcomes.

VIII. FUTURE ENHANCEMENT

There are several avenues for future research in the area of cross-domain federated learning system with disease dataset model generation and disease prediction approach. Privacy-preserving techniques: While federated learning offers strong privacy guarantees, there is always room for improvement. Future research could explore new privacy-preserving techniques that can further enhance the privacy and security of healthcare data. In federated learning, the models trained on each participant's data need to be aggregated to create a global model. Future research could focus on developing more efficient and accurate model aggregation techniques that can handle larger and more complex datasets.

REFERENCES

[1]. Fu, Anmin, Xianglong Zhang, Naixue Xiong, Yansong Gao, Huaqun Wang, and Jing Zhang, (2020) "VFL: A verifiable federated learning with privacy-preserving for big data in industrial IoT." IEEE Transactions on Industrial Informatics 18, no. 5 (2020): 3316-3326.

[2]. Guo, Xiaojie, Zheli Liu, Jin Li, Jiqiang Gao, Boyu Hou, Changyu Dong, and Thar Baker, (2020) "VeriFL: Communication-efficient and fast verifiable aggregation for federated learning." IEEE Transactions on Information Forensics and Security 16 (2020): 1736-1751.

[3]. Lei Yu, Ling Liu, Calton Pu, Mehmet Emre Gursoy, Stacey Truex, (2019) "Differentially Private Model Publishing for Deep Learning"

[4]. Melis, Luca, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov,(2018) "Exploiting unintended feature leakage in collaborative learning." In 2019 IEEE symposium on security and privacy (SP), pp. 691-706. IEEE, 2019.

[5]. Peng, Zhe, Jianliang Xu, Xiaowen Chu, Shang Gao, Yuan Yao, Rong Gu, and YuzheTang,(2021) "Vfchain: Enabling verifiable and auditable federated learning via blockchain systems." IEEE Transactions on Network Science and Engineering 9, no. 1 (2021): 173-186.

[6]. Sav, Sinem, Apostolos Pyrgelis, Juan R. Troncoso-Pastoriza, David Froelicher, Jean-Philippe Bossuat, Joao Sa Sousa, and Jean-Pierre Hubaux,(2021) "Poseidon: Privacy-preserving federated neural network learning." arXiv preprint arXiv:2009.00349 (2020).

[7]. Shai Halevi,NaliniRatha, Sharath Pankanti,Karthik Nandakumar, (2020) "Towards Deep Neural Network Training on Encrypted Data"

[8]. Sheller, Micah J., Brandon Edwards, G. Anthony Reina, Jason Martin, Sarthak Pati, AikateriniKotrotsou, Mikhail Milchenko et al,(2020) "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data." Scientific reports 10, no. 1 (2020): 1-12.

[9]. Tian L,MaziarSanjabi,AhmadBeirami,Virginia Smith, (2020) "Fair resource allocation in federated learning Fair resource allocation in federated learning"

[10]. Tomer Gafni, Nir Shlezinger, Kobi Cohen, Yonina C. Eldar, and H. Vincent Poor, (2021) "Federated Learning: A Signal Processing Perspective"