International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

LIGHT WEIGHT PRIVACY-PRESERVING MEDICAL DIAGNOSIS IN EDGE COMPUTING

K.INDHUMATHI[1], SWAROOPA[2], T.YAMINI[3], J.ABHILASHA[4], S.SOWMYA[5], J.NANDINI[6]

[1]Assistant Professor, Dept of CSE, Princeton Institute of Engineering and Technology for Women, Hyderabad, TS, India.

[2,3,4,5,6] UG Students, Dept of CSE, Princeton Institute of Engineering and Technology for Women, Hyderabad, TS, India.

ABSTRACT:

With the development of machine learning, it is popular that mobile users can submit individual symptoms at any time anywhere for medical diagnosis. Edge computing is frequently adopted to reduce transmission latency for real-time diagnosis service. However, the data-driven machine learning, which requires to build a diagnosis model over vast amounts of medical data, inevitably leaks the privacy of medical data. It is necessary to provide privacy preservation. To solve above challenging issues, in this project, we design a lightweight privacy-preserving medical diagnosis mechanism on edge, called LPME. Our LPME redesigns the extreme gradient boosting (XG Boost) model based on the edge-cloud model, which adopts encrypted model parameters instead of local data to remove amounts of cipher text computation to plaintext computation, thus realizing lightweight privacy preservation on resource-limited edge. In addition, LPME provides secure diagnosis on edge with privacy preservation for private and timely diagnosis. Our security analysis and experimental evaluation indicates the security, effectiveness and efficiency of LPME.

Keywords:LPME, XG, Edge computing, Data.

1. INTRODUCTION:

Machine learning is taking an ever-increasing role in medical diagnosis, and has become prevalent for mobile users to submit symptoms at any time and then get diagnosis results. Compared with the shortage of experts and high cost in manual diagnosis, machine learning-based diagnosis has the great advantages in improving the quality of healthcare service and avoiding expensive diagnosis expenses. Thus, the construction of machine learning based medical diagnosis has attracted much attentions from both academic and industrial fields. With the emergence of telemedicine applications, more and more demands have blossomed in healthcare, clinical decision, and mobile telemedicine. However, the blossom has also been accompanied by various problems, i.e., the limitation of training data, vulnerabilities, and privacy concerns.In medical practice, it is a crucial issue that the collection of enough medical data is time-consuming and expensive. A single medical origination usually stores a limited number of medical data, which is hard to support the construction of data-driven machine learning. To train an accurate diagnosis model, it is necessary to share the training data distributed among various medical institutions. With the advances of extensive storage space and unlimited computing capacity in cloud computing, machine learning over outsourced medical data has been extensively studied with the adoption of cloud.

However, with the ever-increasing interactions between mobile users and the cloud, it incurs undesirable transmission latency and untimely request response. A delayed diagnosis response directly influences patients' life and health as well as medical safety, especially for patients with a diagnosis for acute disease (e.g., acute heart disease, pneumonia). To address this dilemma, edge computing, as a new computing paradigm, has been proposed to decrease latency and provide efficient computation service by using edge nodes which are close to mobile users. In the last few years, machine learning schemes based on edge computing have an extensive development, which is significant to improve the diagnosis efficiency with edge computing. Fig. 1 plots a typical edge network with several edge nodes (i.e., medical organizations) that owns restricted storage ability and limited computing power. To concentrate on the vulnerability in medical diagnosis, it is important to adopt a high-performance model on edge for real-time and reliable medical diagnosis.

Extreme gradient boosting (XG Boost) as the most state-of-the-art machine learning model enjoys the excellent prediction performance in the distributed setting, which demonstrates the outstanding ability in Kaggle competitions. Besides, with the tree-based structure, XG Boost has the advances of explain ability and ease of understanding. Therefore, there are a large number of schemes applied the XG Boost model for medical diagnoses , but they ignore the important issue of data privacy during the training phase. Actually, patients diagnosed with private diseases usually bear some psychological

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

barrier when the diagnosis results are leakage to others. It is considered as a cause to worsen the condition. Thus, it is necessary to provide privacy reservation for them. Besides, the medical data contain a large amount of sensitive information, with there lease of privacy policies (i.e., GDPR and HIPPA),more and more data are forbidden to transform in the form of plain text. Therefore, it is urgent to protect privacy of medical diagnosis in the edge computing environment.

2. LITERATURE SURVEY

In medical practice, it is a crucial issue that the collection of enough medical data is time-consuming and expensive. A single medical origination usually stores a limited number of medical data, which is hard to support the construction of data-driven machine learning. To train an accurate diagnosis model, it is necessary to share the training data distributed among various medical institutions. With the advances of extensive storage space and unlimited computing capacity in cloud computing, machine learning over outsourced medical data has been extensively studied with the adoption

To concentrate on the vulnerability in medical diagnosis, it is important to adopt a high-performance model on edge for real-time and reliable medical diagnosis. Extreme gradient boosting (XG Boost) as the most state-of-the-art machine learning model enjoys the excellent prediction performance in the distributed setting, which demonstrates the outstanding ability in Kaggle competitions. Besides, with the tree based structure, XG Boost has the advances of explainability and ease of understanding. Therefore, there are a large number of schemes applied the XG Boost model for medical diagnoses [17], [18], [19], but they ignore the important issue of data privacy during the training phase. Actually, patients diagnosed with private diseases (e.g., HIV, Hepatitis B virus) usually bear some psychological barrier when the diagnosis results are leakage to others. It is considered as a cause to worsen the condition. Thus, it is necessary to provide privacy preservation for them. Besides, the medical data contain a large amount of sensitive information, with the release of privacy policies (i.e., GDPR [20] and HIPPA [21]), more and more data are forbidden to transform in the form of plaintext. Therefore, it is urgent to protect privacy of medical diagnosis in the edge computing environment.

Proposed System

To address the above challenges, we design a lightweight privacy-preserving XGBoost over

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

encrypted model parameters to greatly lighten computational overhead, compared with data sharing-based privacy-preserving machine learning. In this paper, we present the Lightweight Privacy preserving Medical diagnosis in Edge computing, which is termed as LPME. Specifically, our LPME mainly has the following constructions:

Lightweight XGBoost on edge: LPME system constructs a XGBoost-based diagnosis model with model parameters trained over multiple edge nodes rather than training data, which not only eliminates the drawbacks of burdensome training data storage, but also guarantees the feasibility of XG Boost. Privacy-preserving training: LPME system designs HE based secure computation with a single-cloud model, which selects optimal parameters over encrypted model parameters during the training phase. Since the secret key is randomly split into two parts, only one is stored in the single cloud. Thus, the single cloud model can not only provide strong privacy preservation for training the lightweight XG Boost, but also guarantee the reliability of the privacy preserving training on the resource-limited edges. Secure diagnosis on XG Boost at edge: LPME system provides secure diagnosis, in which a mobile user can submit his/her encrypted requests to an edge, then the edge will return the corresponding diagnosis results. During the process, HE is adopted to guarantee confidentiality of the returned diagnosis results for implementing the private and timely diagnosis.

3. METHODOLOGY

Here, we introduce the Secure Multiplication (SMUL) and Secure Comparison (SCOM) operations for secure computation. Suppose that there are two semi-honest parties (i.e., Alice and Bob) in the multiplication and comparison over encrypted data, the goals of SMUL and SCOM are that all intermediate results and final computation results cannot be disclosed to both parties. Given two encrypted numbers [[x1]] and [[x2]], Alice holds a secret share sk(1), Bob holds the other secret share sk(2) . SMUL and SCOM are defined as follows:

SMUL([[x1]], [[x2]]) → [[x1 × x2]]: Alice first generates [[x 0 1]] = [[x1]] · [[r1]], [[x 0 2]] = [[x2]] · [[r2]], where r1, r2 ∈ Z ∗ η2 are two random numbers, then uses SDecsk(1) to obtain [[x 0 1]](1) and [[x 0 2]](1). On receiving these encrypted data, Bob uses SDec and WDec with sk(2) to obtain x 0 1 and x 0 2 , and computes [[res]] = x 0 1 × x 0 2 . Then, Alice runs [[x1 × x2]] = [[res]] · [[r1 × r2]]η−1 ·[[x1]]η−r2

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

$\cdot[[x2]]\eta{-}r1$ to remove random numbers, and the multiplication result $[[x1 \times x2]]$ is returned.

SCOM($[[x1]], [[x2]]$) $\rightarrow$ res: Alice first calculates $[[x\ 0\ 1\ ]] = [[x1]]2 \cdot [[1]]$, $[[x\ 0\ 2\ ]] = [[x\ 0\ 2\ ]]2 \cdot [[1]]$, and runs $[[res]] \leftarrow ([[x\ 0\ 1\ ]] \cdot [[x\ 0\ 2\ ]]\eta{-}1\ )\ r1 \cdot [[r2]]$, where $r1, r2 \leftarrow Z\eta$ (r2 r1) are two random numbers. Then, $[[res]](2) \leftarrow$ SDecsk(1) ($[[res]]$) is obtained.

After involving the SDec and WDec algorithms, Bob obtains res via computing the bit length of res as Eq. 3, and returns the comparison result. res = $x1 < x2$, $|res| > |\eta|/2$; $x1 \geq x2$, otherwise. (2)

Algorithm 2

Algorithm 2: Globally Optimal Split

Input: Encrypted gain parameters $\{[[\alpha\uparrow n]], [[\alpha\downarrow n]]\}Nn{=}1$,encrypted locally optimal split $\{[[s{*}n]], [[f{*}n]]\}Nn{=}1$.

Output: Globally optimal split f∗and s∗.

1 $[[score\uparrow]] \leftarrow [[0]], [[score\downarrow]] \leftarrow [[1]]$;

2 $f{*} \leftarrow [[0]], [[s{*}]] \leftarrow [[0]]$;

3 for $0 < n \leq N$ do

4 /* Compare Enc Index */

$[[score\uparrow \times \alpha\downarrow n]] \leftarrow$ SMUL($[[score\uparrow], [[a\downarrow n]]$);

5 $[[score\downarrow \times \alpha\uparrow n]] \leftarrow$ SMUL($[[score\downarrow]], [[\alpha\uparrow\ n]]$);

6 SCOM($[[score\uparrow \times \alpha\downarrow n]] \cdot [[score\downarrow \times \alpha\uparrow n]]\eta{-}1, [[0]]$);

7 if $A - B < 0$ then

8 $score\uparrow \leftarrow \alpha\uparrow n$, $score\downarrow \leftarrow \alpha\downarrow n$;

9 $[[f{*}] \leftarrow [[f{*}n]], [[s{*}]] \leftarrow [[s{*}n]]$;

10 return $[[f{*}]], [[s{*}]]$.

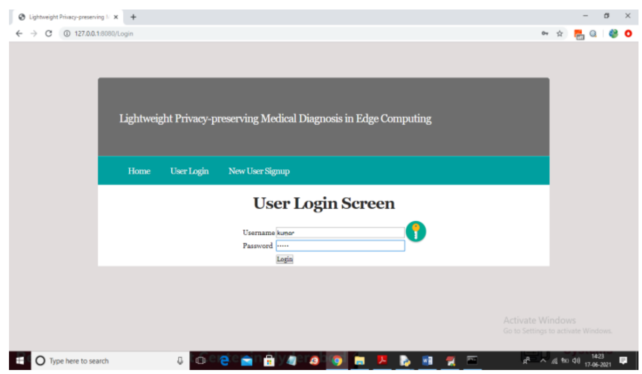

Fig 3.1 User Login Page.

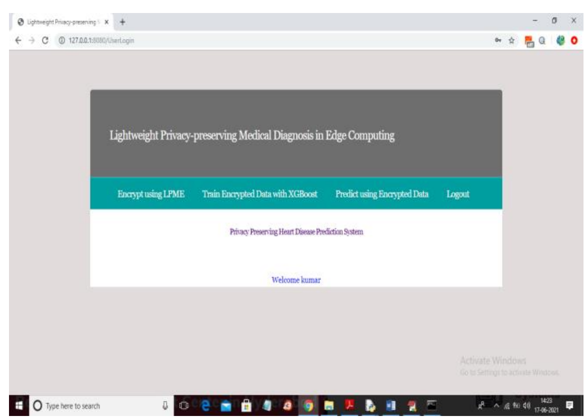

Fig 3.2 Login Success Page

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

In above screen you can click on 'Encrypt using LPME' link to encrypt dataset with LPME technique.

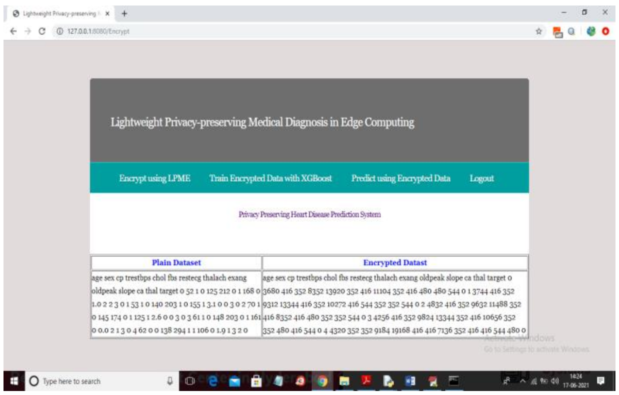

Fig 3.3 Encrypted data

In above screen first column showing original dataset and second column showing encrypted format of that original plain data and now dataset is encrypted and now click on 'Train Encrypted Data with XG Boost' link to train dataset and to build XGBOOST secure disease prediction model

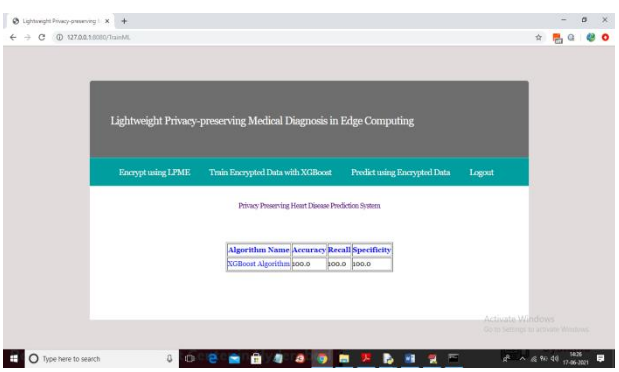

Fig 3.4 Data Prediction

In above screen XGBOOST training completed and we got accuracy of the model on test data is 100% and in below screen you can training and testing of XGBOOST

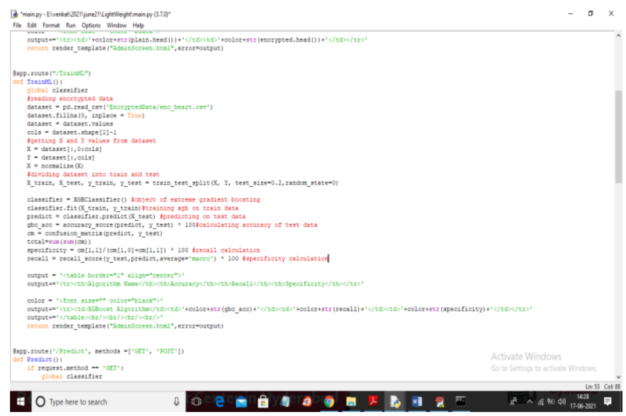

Fig 3.5XGBoost training on encrypted data

In above screen, the code is used to create XGBOOST training on encrypted data and now go back to previous application and click on 'Predict Using Encrypted Data' link to predict disease from new test data and below is the test data screen

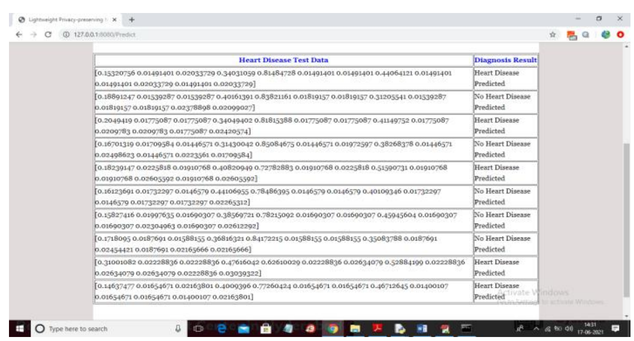

Fig 3.6 Result Analysis

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

In above screen in first column you can see then encrypted test data and in second column you can see prediction result as 'No Heart Disease Detected' or 'Heart Disease Detected'

CONCLUSION

This project has proposed a lightweight privacy-preserving XG Boost framework on edge, which could not only provide lightweight XG Boost over edge nodes with strong privacy preservations, but also achieve privacy-preserving and real-time medical diagnosis on edge. The proposed LPME system with secure computation could securely construct XG Boost model with lightweight overhead, and efficiently provide medical diagnosis without privacy leakage. Experimental results over real-world datasets verified the efficiency and security of the LPME system on edge computing.

FUTURE SCOPE

As we normally use automatic rapid test for detecting the malaria a person can know the status of malaria using strips in rapid test as we do test of malaria we can also use a method to send the details of a patient as he/she was infected or not to the patient mobile and let the patient know the details of him/her through what's app or message. We can also send the appointment to meet the doctor.

Firstly, we create a excel sheet which has patient details and send the details to the person directly without any delay. There will be no waiting of patient. The patient receives the message of the test directly with a message and the patient also receives appointment to meet doctor at a specific time and date. It becomes to the patient and doctor to meet and discuss.

For this, we use visual studio management to create a message and send the message to the patient. Visual studio plays a major role to create a message and send it to the person. The details of a person which we need to send is already placed in excel sheet. So, it becomes easy to send to person and allot appointment to the patient.

REFERANCES

[1] X. Wang, J. Ma, Y. Miao, X. Liu, and R. Yang, "Privacy-preserving diverse keyword search and online pre-diagnosis in cloud computing," IEEE Transactions on Services Computing, 2019.

[2] Y. Zhang, C. Xu, H. Li, K. Yang, J. Zhou, and X. Lin, "Health Deep: An efficient and secure duplication scheme for cloud-assistede health systems," IEEE Trans. Industrial Informatics, vol. 14, no. 9,pp. 4101–4112, 2018.

[3] B. Fu, P. Liu, J. Lin, L. Deng, K. Hu, and H. Zheng, "Predicting invasive disease-free survival

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

for early-stage breast cancer patients using follow-up clinical data," IEEE Transactions on Biomedical Engineering, 2018.

[4] A. Galletta, L. Carnevale, A. Bramanti, and M. Fazio, "An innovative methodology for big data visualization for telemedicine, "IEEE Transactions on Industrial Informatics, vol. 15, no. 1, pp. 490497, 2018.

[5] I. Kononenko, "Machine learning for medical diagnosis: history state of the art and perspective," Artificial Intelligence in medicine, vol. 23, no. 1, pp. 89–109, 2001.

[6] C. P. Friedman, A. K. Wong, and D. Blumenthal, "Achieving anation wide learning health system," Science Translational Medicine, vol. 2, no. 57, pp. 57cm29–57cm29, 2010.

[7] Y. Miao, X. Liu, K.-K. R. Choo, R. H. Deng, H. Wu, and H. Li, "Fairand dynamic data sharing framework in cloud-assisted internet of everything," IEEE Internet of Things Journal, 2019.

[8] Y. Miao, Q. Tong, K.-K. R. Choo, X. Liu, R. H. Deng, and H. Li,"Secure online/offline data sharing framework for cloud-assisted industrial internet of things," IEEE Internet of Things Journal, 2019.

[9] Y. Miao, J. Weng, X. Liu, K.-K. R. Choo, Z. Liu, and H. Li, "Enabling verifiable multiple keywords search over encrypted cloud data, "Information Sciences, vol. 465, pp. 21–37, 2018.

[10] T. Ouyang, R. Li, X. Chen, Z. Zhou, and X. Tang, "Adaptive user-managed service placement for mobile edge computing: A non line learning approach," in Proc. IEEE Conference on Computer Communications (INFOCOM'19). IEEE, 2019, pp. 1468–1476.

[11] P. Dai, K. Liu, X. Wu, H. Xing, Z. Yu, and V. C. Lee, "A learning algorithm for real-time service in vehicular networks with mobile edge computing," in Proc. IEEE International Conference on Communications (ICC'19). IEEE, 2019, pp. 1–6.