Chapter 18
Privacy of Medical Data Using Hierarchical Deep Learning Fused Long-Short Term Memory Based Secret Key Generation

R. Mary Sheeba and R. Parameswari

18.1 Introduction

Medical organizations worldwide increasingly provide health services to patients and manage vast amounts of electronic health records (EHR). According to consulting company EMC and research firm IDC, global healthcare data has reached approximately 153 exabytes. These clinical data need larger amounts of space for management and storage [1, 2]. The heart disease EHRs offered by numerous patients are gathered to a remote cloud-enabled eHealth scheme. The doctors worry about privacy leakages while searching EHRs on semi-trusted cloud servers, since searching functions may leak EHR contents. Thus, EHRs must be protected by utilizing a cloud-enabled eHealth scheme. With these security restrictions, a crucial quantity of laws and regulations such as the Health Insurance Portability and Accountability Act (HIPAA) used to manage as well as share EHRs. Order-preserving encryption (OPE) enables effective range query on encrypted data that balances security [3–6]. The Disease Prediction Systems (DPSs) having various data mining approaches being used, have drawn attention in recent times [2, 7]. DPSs using Single-Layer Perceptron (SLP) classifiers have benefited healthcare by improving disease prediction. However, their growth depends on managing privacy issues and prediction efficacy, especially for sensitive clinical data stored with unauthorized third parties. Privacy-preserving data mining approaches are needed to protect clinical data. Additionally, prediction methods, considered hospital assets, should not be shared with third parties to prevent misuse that could harm hospital profits. Therefore, preserving the privacy of prediction methods is crucial for DPSs. The probable privacy, as well as accuracy

R. M. Sheeba (✉) · R. Parameswari
Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies, Chennai, India
e-mail: maryrsheeba@gmail.com

S. Nagini et al. (eds.), *Proceedings of Fifth International Conference on Advances in Computer Engineering and Communication Systems*, Smart Innovation, Systems and Technologies 117, https://doi.org/10.1007/978-981-96-4410-0_18

losses increase during evaluations [8] or tuning of deep learning methods. Preserving data or its implications can be done through two primary methods: one is based on cryptography, and the other is perturbation-enabled. The perturbation-based technique [9, 10] alters data with noises to provide protection. The statistics summary remains to be about similar for helping prevent concluding information regarding any particular record. With the dramatic growth of machine learning (ML) and the medical data assessment methods utilizing ML for intelligent clinical diagnosing has become feasible. Medical diagnosing systems based on several ML classification approaches have been developed. The Support vector machine (SVM) as a dominant classification technique has acquired higher classification accuracy and effectiveness in the medical field [11–13]. The physicians can acquire outcomes for assisting medical diagnosis of online clinical diagnosing by transferring the private details of patients to a classifier on a medical cloud server (MCS) [14]. More particularly, the data privacy concept is a wide area focused on secrecy, correctness, accountability, availability, and fairness considering user data [15].

The primary goal is to introduce HDLFLSTM_KeyGen to preserve the privacy of medical data. Healthcare equipment and applications generate vast amounts of data, which is transmitted between devices and over global networks, often containing sensitive private information. Therefore, privacy and security are crucial in healthcare. Initially, a cloud system model is used, and then input data is encrypted to ensure privacy. This encryption process employs various functions such as Kronecker products and secret keys, incorporating trust factors for added security. The secret key is generated using HDLFLSTM, a combination of HDLTex and DLSTM networks.

Motivation

Particularly, the semi-trusted clinical cloud field has become a key means of hospital clinical data management as well as information services. The privacy and security problems in the clinical cloud field are highly important and must be addressed with the main concern. This motivated me to present an approach for privacy preservation in medical data. This segment reviews existing papers on privacy preservation in medical data and their limitations.

18.2 Literature Survey

Zhang et al. [13] designed a support vector machine (SVM) for privacy preservation that was proven as effective as well as practical in privacy-preserving medical diagnosis in clinical clouds, but still time cost of the decision operation computation was increased. Kwabena et al. [10] introduced a Multi-Scheme Crypto Deep Neural Network (MSCryptoNet) for privacy preservation setting. This scheme does not need any interaction amongst data providers and cloud servers for the provision of privacy-preserving predictions, but it classified only one instance at individual prediction rounds. Zhang et al. [2] presented a privacy-preserving disease prediction scheme, termed as PPDP for the healthcare system. In this scheme, the computational

cost was low, even though it did not focus on devising highly effective and privacy-preserving models. Liang et al. [6] devised multi-source order-preserving encryption (MSOPE) for cloud-enabled electronic health systems. It was highly effective, but still, it did not devise privacy-preserving range query systems for the multi-attribute EHRs.

Liu et al. [16] developed a privacy-preserving reinforcement learning method called Preyer for a patient-centric dynamic treatment system. It assisted patient dynamic treatment policy-making without leakage of sensible information to unauthorized users. Ma et al. [17] designed privacy-preserving clinical decisions with cloud support (PPCD) that securely conducted disease system training and prediction for patients. The enhanced efficacy was feasible for practical utilization, but it did not consider a balance between efficiency and security. Almaiah et al. [18] introduced a Bidirectional Long Short-Term Memory and Convolutional Neural Network (BLSTM + CNN) for privacy preservation. This method was suggested for the cross-domain networks in healthcare schemes, even though it failed to utilize the particle swarm optimization (PSO) approach incorporated with federated learning. Gomathi and Karlekar [19] devised Whale based Sine Cosine Algorithm with a Support Vector Neural Network (WSCA-SVNN) for the privacy preservation of medical data classification. It obtained minimal computation time, though it failed to utilize deep learning techniques for enhancing classification performance.

18.3 HDLFLSTM_KeyGen for Medical Data Privacy Preservation

Cloud computing allows extensive medical data storage but raises concerns about unauthorized access due to resource sharing. Therefore, effective privacy preservation measures are essential. To address this, we introduce HDLFLSTM_KeyGen for securing medical data. Initially, the cloud system model processes input data through a privacy preservation stage, where encryption, the Kronecker product, and a secret key are employed. The secret key is generated using HDLFLSTM, which combines HDLTex and DLSTM. Figure 18.1 shows the HDLFLSTM_Keygen model for securing medical data.

Data Acquisition Method: Input data is acquired from a particular database specific for the privacy preservation of medical data that can be modeled as, $P = \{P_1, P_2, \ldots P_m \ldots P_d\}$. Here, the total data present in the dataset P is signified as P_d whereas mth data is symbolized by P_m. The acquired input data A is processed with a polynomial, which can be illustrated by, $D = A * Y$, Where D specifies processed parameters and Y indicates a polynomial that can be given by, $Y = 50u^9 + 20u^6 + 51u^4 + 30u^2 + 10u$. The privacy-preserved data can be formulated by, $P_{v \times w} = E(A, \kappa)$. Here, E specifies encryption and κ denotes key. However, the key is generated by employing the newly presented HDLFLSTM.

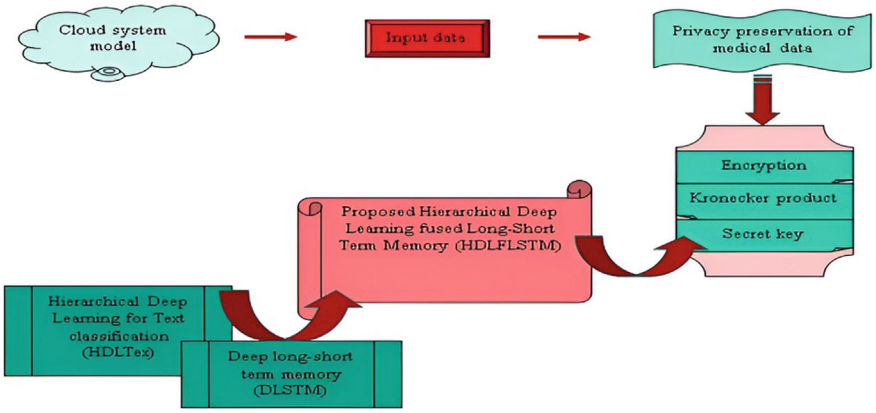

Fig. 18.1 HDLFLSTM_KeyGen model for securing medical data

$$\Im_{v\times w} = (\boldsymbol{D}_{v\times 1} \times \mathbf{P}_{v\times w}) * \boldsymbol{R}_{1\times w} \tag{18.1}$$

$$R = entropy(A_{1\times n}) \tag{18.2}$$

$$B_{v\times v} = R \times R^{Z} \tag{18.3}$$

where, $\times$ implies matrix multiplication and $B_{v\times v}$ indicates parameter. The bilateral matrix can be modeled by, $T_{v\times w} = B_{v\times v} \times P_{v\times w}$. The Kronecker matrix is specified as, $K_{(v\times v)\times w} = T_{v\times w} \otimes R_{v\times 1}$. Here, $\otimes$ indicates the Kronecker product. The privacy-preserved data can be formulated as $E_w = (\Im_{v\times w} + A_{v\times w}) * \left(\left(K_{w\times(v\times v)} \right)^{Z} + I_{w\times(v\times v)} \right)$ where I signifies identity matrix. Then, this encrypted data is XORed with a key that can be given by, $\varepsilon = E_w \oplus \kappa$. The retrieval of privacy-preserved data can be illustrated by

$$E_w = \frac{W}{\kappa + I} - \Im \tag{18.4}$$

The key retrieval can be represented by, $\kappa = E_w \oplus \varepsilon$. Here, E_w symbolizes privacy-preserved data whereas κ implies key.

Key Generation Utilizing HDLFLSTM: The secret key generation is crucial and time-consuming, involving the creation of secure, non-derivable keys. Our approach uses HDLFLSTM, which combines HDLTex and DLSTM, for this task. Figure 18.2 shows the structure of the HDLFLSTM Network.

First, the HDLTex model processes the input data to produce an output. This output and a processed parameter are fed into the HDLFLSTM layer for fusion. The resulting output is then passed to the DLSTM model, generating the final key.

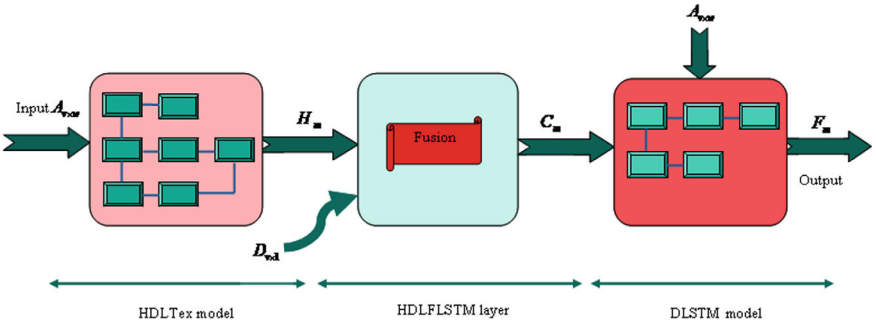

Fig. 18.2 Structure of HDLFLSTM network

HDLTex Model: In comparison to classical non-neural-based systems, the hierarchical neural-based system HDLTex shows superior performance. The bidirectional LSTM is initially utilized for extracting features of data.

$$\overrightarrow{q_\tau} = \overrightarrow{LSTM}\left(\omega_\tau, \overrightarrow{q_{\tau-1}}\right) \tag{18.5}$$

The data depiction for level h is attained by multiplying a multi-head attention matrix with features, which is formulated by $A_k = G_{n3} M_h \overline{Q_h}$.

The two-layered multi-layer perceptron (MLP) is utilized for classifying classes at the level h.

$$a_h = RELU\left(G_A[A_h, a_h - 1]\right) \tag{18.6}$$

$$H_m = softmax(G_h a_h) \tag{18.7}$$

The above equation specifies an output attained from HDLTex.

HDLFLSTM Layer

The fusion is done in the HDLFLSTM layer, wherein two models namely HDLTex and DLSTM are fused. Here, K_m is given as an input to perform fusion, such that,

$$K_m = \{H_m, D_{v\times 1}\} \tag{18.8}$$

The output can be given by $g = \sum_m D_m \times H_m$, where g represents output from ℓth interval whereas g_1 indicates output from $(\ell - 1)$th interval. Applying fractional calculus (FC),

$$\chi(\ell + 1) = \hbar \cdot \chi(\ell) + \frac{1}{2}\hbar\chi(\ell - 1) \tag{18.9}$$

The above equation becomes $C_m = \hbar \cdot g + \frac{1}{2}\hbar H_m$ where $g = \sum_m D_m \times \mathrm{H}_m$ and C_m denotes an output from the HDLFLSTM layer.

18.4 Results and Discussion

HDLFLSTM_KeyGen for privacy preservation of medical data obtained better outcomes that are interpreted in this segment along with experimentation setup and performance measures for assessment. It was implemented using the Python tool.

Description of Dataset

Blood Bank Directory-India dataset comprises state and city-wise blood bank lists and other necessary information such as area pin code, address, contact details, e-mail address, and so on. The files are in .csv format.

Performance Metrics, Comparative Techniques and Evaluation

The evaluation focused on two metrics: privacy and utility, to assess the effectiveness of HDLFLSTM_KeyGen for privacy preservation of medical data. SVM [13], MSCryptoNet [10], PPDP [2], and MSOPE [6] are the comparison methods to evaluate HDLFLSTM_KeyGen to reveal its effectiveness. The assessment compared HDLFLSTM_KeyGen with existing techniques using different data sizes.

Setup-1 Assessment: Fig. 18.3 illustrates the performance of HDLFLSTM_KeyGen with a data size of 100 and 200 kb, focusing on privacy and utility metrics. At iteration, HDLFLSTM_KeyGen achieved a privacy score of 0.894, outperforming SVM (0.811), MSCryptoNet (0.816), PPDP (0.845), and MSOPE (0.882). The HDLFLSTM_KeyGen achieved a utility score of 0.833, compared to SVM (0.793), MSCryptoNet (0.798), PPDP (0.809), and MSOPE (0.812).

Setup-2 Assessment: At iteration, HDLFLSTM_KeyGen achieved a privacy score of 0.956, outperforming SVM (0.826), MSCryptoNet (0.838), PPDP (0.862), MSOPE

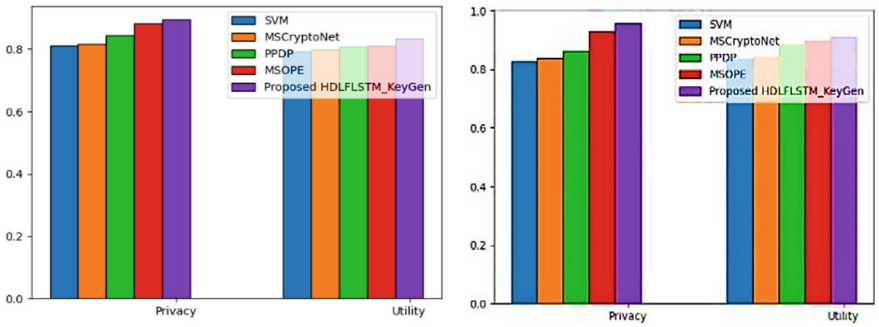

Fig. 18.3 Performance of HDLFLSTM with a data size of 100 and 200 kb

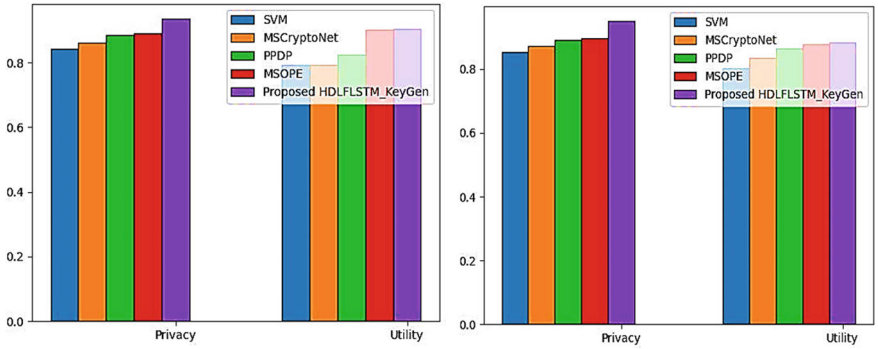

Fig. 18.4 Evaluation metrics of HDLFLSTM with a data size of 300 and 400 kb

(0.928) and attained a utility score of 0.909, compared to SVM (0.836), MSCryptoNet (0.839), PPDP (0.884), and MSOPE (0.897).

Setup-3 Assessment: Fig. 18.4 shows the evaluation of metrics with a data size of 300 and 400 kb. At iteration, HDLFLSTM_KeyGen achieved a privacy score of 0.935, outperforming SVM (0.843), MSCryptoNet (0.862), PPDP (0.886), MSOPE (0.891) and attained a utility score of 0.903 for HDLFLSTM, compared to SVM (0.791), MSCryptoNet (0.793), PPDP (0.822), and MSOPE (0.902).

Setup-4 Assessment: The privacy values acquired by SVM, MSCryptoNet, PPDP and MSOPE are 0.853, 0.873, 0.891 and 0.895 while HDLFLSTM_KeyGen attained 0.950. Utility obtained by HDLFLSTM_KeyGen is 0.882 while utility attained by SVM is 0.801, MSCryptoNet is 0.835, PPDP is 0.865 and MSOPE is 0.876.

18.5 Conclusion

Medical data sharing is crucial for smart medicine but is complicated by heterogeneous information schemes and privacy concerns. To address these challenges, we introduce HDLFLSTM_KeyGen for medical data privacy preservation. This method uses a cloud system model to encrypt data with functionalities like encryption, the Kronecker product, and a secret key generated by the HDLFLSTM, which combines HDLTex and DLSTM. Future work will explore its applicability in cloud environments with extensive concurrent clinical data access and sharing.

References

1. M. Li, S. Yu, K. Ren, and W. Lou, Securing personal health records in cloud computing: patient-centric and fine-grained data access control in multi-owner settings, in *Security and Privacy in Communication Networks: In the Proceedings of 6th International ICST Conference, SecureComm 2010, Singapore, September 7–9, 2010*, vol. 6 (Springer Berlin Heidelberg, 2010), pp. 89–106
2. C. Zhang, L. Zhu, C. Xu, R. Lu, PPDP: "An efficient and privacy-preserving disease prediction scheme in cloud-based e-Healthcare system." Futur. Gener. Comput. Syst. **79**, 16–25 (2018)
3. R. Agrawal, J. Kiernan, R. Srikant, Y. Xu, Order preserving encryption for numeric data, in *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, pp. 563–574 (2004, June)
4. A. Boldyreva, N. Chenette, Y. Lee, A.O. Neill, Order-preserving symmetric encryption, in *Proceedings of Advances in Cryptology-EUROCRYPT 2009: 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26–30, 2009*, vol. 28 (Springer Berlin Heidelberg, 2009), pp. 224–241
5. A. Boldyreva, N. Chenette, A.O. Neill, Order-preserving encryption revisited: improved security analysis and alternative solutions, in *Proceedings of Advances in Cryptology—CRYPTO 2011: 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14–18, 2011*, vol. 31 (Springer Berlin Heidelberg, 2011), pp. 578–595
6. J. Liang, Z. Qin, S. Xiao, J. Zhang, H. Yin, K. Li, Privacy-preserving range query over multi-source electronic health records in public clouds. J. Parallel Distrib. Comput. **135**, 127–139 (2020)
7. J.W. Bos, K. Lauter, M. Naehrig, Private predictive analysis on encrypted medical data. J. Biomed. Inform. **50**, 234–2432 (2014)
8. N. Phan, X. Wu, D. Dou, Preserving differential privacy in convolutional deep belief networks. Mach. Learn. **106**(9–10), 1681–1704 (2017)
9. R. Shokri, V. Shmatikov, Privacy-preserving deep learning, in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1310–1321 (2015, October)
10. O.A. Kwabena, Z. Qin, T. Zhuang, Z. Qin, Mscryptonet: multi-scheme privacy-preserving deep learning in cloud computing. IEEE Access **7**, 29344–29354 (2019)
11. N. Bouguila, Hybrid generative/discriminative approaches for proportional data modeling and classification. IEEE Trans. Knowl. Data Eng. **24**(12), 2184–2202 (2011)
12. M.A.Ł.G.O.R.Z.A.T.A. Cwiklinska-Jurkowska, Performance of the support vector machines for medical classification problems. Biocybern. Biomed. Eng. **29**(4), 63–81 (2009)
13. M. Zhang, W. Song, J. Zhang, A secure clinical diagnosis with privacy-preserving multiclass support vector machine in clouds. IEEE Syst. J. **16**(1), 67–78 (2020)
14. D. Conforti, R. Guido, Kernel based support vector machine via semidefinite programming: application to medical diagnosis. Comput. Oper. Res. **37**(8), 1389–1394 (2010)
15. D.L. Goroff, Balancing privacy versus accuracy in research protocols. Science **347**(6221), 479–480 (2015)
16. X. Liu, R.H. Deng, K.K.R. Choo, Y. Yang, Privacy-preserving reinforcement learning design for patient-centric dynamic treatment regimes. IEEE Trans. Emerg. Top. Comput. **9**(1), 456–470 (2019)
17. H. Ma, X. Guo, Y. Ping, B. Wang, Y. Yang, Z. Zhang, J. Zhou, PPCD: privacy-preserving clinical decision with cloud support. PLoS ONE **14**(5), e0217349 (2019)
18. M.A. Almaiah, A. Ali, F. Hajjej, M.F. Pasha, M.A. Alohali, A lightweight hybrid deep learning privacy preserving model for FC-based industrial internet of medical things. Sensors **22**(6), 2112 (2022)
19. N. Gomathi, N.P. Karlekar, Ontology and hybrid optimization based SVNN for privacy preserved medical data classification in cloud. Int. J. Artif. Intell. Tools **28**(03), 1950009 (2019)