2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH) | 978-1-6654-0973-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMARTTECH54121.2022.00055

Securing the Classification of COVID-19 in Chest X-ray Images: A Privacy-Preserving Deep Learning Approach

Wadii Boulila[1,2], Adel Ammar[1], Bilel Benjdira[1], Anis Koubaa[1]
[1]RIOTU Lab, Prince Sultan University, Riyadh, Saudi Arabia
[2]RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba, Tunisia

***Abstract*—Deep learning (DL) is being increasingly utilized in healthcare-related fields due to its outstanding efficiency. However, we have to keep the individual health data used by DL models private and secure. Protecting data and preserving the privacy of individuals has become an increasingly prevalent issue. The gap between the DL and privacy communities must be bridged. In this paper, we propose privacy-preserving deep learning (PPDL)-based approach to secure the classification of Chest X-ray images. This study aims to use Chest X-ray images to their fullest potential without compromising the privacy of the data that it contains. The proposed approach is based on two steps: encrypting the dataset using partially homomorphic encryption and training/testing the DL algorithm over the encrypted images. Experimental results on the COVID-19 Radiography database show that the MobileNetV2 model achieves an accuracy of 94.2% over the plain data and 93.3% over the encrypted data.**

***Index Terms*—Privacy preserving, deep learning, encryption, homomorphic encryption, Paillier scheme, chest X-ray images, COVID-19.**

I. Introduction

The protection of human private data is a persistent challenge. This matter has always been taken seriously in computer science. However, due to the efficiency of deep learning (DL) algorithms, many have used them without any care for the protection of private data. All have been pleased by improving the accuracy of learning and delaying any privacy concerns.

A small interest in this topic has been raised after the deployment of remote DL models over the internet. In fact, a variety of cloud platforms were designed using DL models for both the training and inference phases. The service offered by these platforms was coined later as Deep Learning as a Service (DLaaS) [1]. Hence, private data are transferred over the internet and may also be stored on servers with high vulnerability for leakage. The leakage can be made in an intentional or unintentional way. It can be leaked intentionally through hacking, piracy, or social engineering. It can also be leaked unintentionally by the user himself or the service provider employees. Although being unexpected, the employees are behind 43% of the whole data leakage cases, as affirmed by Intel Security [2]. Around half of these cases were made unintentionally.

The first event that triggered global interest in this problem is the Facebook data privacy scandal [3]. People have discovered that Facebook has sold the personal data of 80 million users to the British firm Cambridge Analytica to train DL algorithms to make political advertising [4]. People have sensed that their private data are threatened and should be protected. DL algorithms should be accompanied during the training and inference by privacy-preserving techniques that prevent any identification of personal data.

The proposed work in this paper focuses on this topic. We designed a data privacy technique to keep private data protected during DL training and inference. The data preserving technique presented in this paper is designed for preserving Chest X-Ray images used to distinguish COVID-19 cases from normal, lung opacity, and pneumonia cases. The DL model does not have access to any identifiable information. It takes only as input encrypted images. The encryption of the Chest X-Ray is done from the user side before feeding them to the DL model. This encryption is homomorphic, which means that the final classification result does not need any decryption of the image. The DL model will only work with these encrypted data without requiring any personally identifiable information. We think this feature is essential, especially in the current COVID-19 pandemic. In fact, several research works are conducted to show the importance of using Chest X-Ray images to detect COVID-19 cases [5]–[10]. Moreover, many people explicitly require that the body details shown in these kinds of images must be kept private and secure [11]. The technique presented in this paper fulfills their requirement at a very low cost (around a 1% decrease in accuracy).

The remainder of this paper is ordered as follows: in Section II, the background section, we will discuss the research works that targeted the topics of Privacy-Preservation Deep Learning (PPDL) and homomorphic encryption. Later, in Section III, we will introduce our PPDL approach for protecting COVID-19 data during the classification task. Then, in Section IV, we will describe the experimental results that defend the utility of the proposed approach. Finally, in Section V, we conclude our work and raise the main limitations that should be targeted in the next research works.

II. Background

A. Privacy-Preserving Deep Learning

Data privacy is an important issue for training and testing DL models, especially in the case of training and inferring

978-1-6654-0973-5/22/$31.00 ©2022 IEEE
DOI 10.1109/SMARTTECH54121.2022.00055

sensitive data (e.g. health records, financial details, location logs, and satellite images) [12], [13]. Many PPDL techniques focused on allowing multiple input parties to train and test DL models without revealing their private data in its original form. These techniques can be subdivided into three groups: cryptographic, perturbation, secure enclaves, and hybrid techniques [14]. Cryptographic methods are used to train and test DL techniques on encrypted data [15]. This category of methods includes Homomorphic Encryption (HE), Secret Sharing (SS), Secure Multi-Party Computation (SMPC), and Garbled Circuit (GC). The perturbation methods aim to alter data values to maintain individual record confidentiality [16]. This category of methods includes Differential Privacy (DP) [17]–[19] and Dimensionality Reduction (DP) [20]. In secure enclaves-based approaches [21] [22], both the prediction model and the data are separately sent by the client and the server to a trusted, secure enclave environment for execution. On the other hand, hybrid methods aim at combining more than one PPDL technique to improve the privacy of data [23]–[26]. A recent survey on privacy-preserving deep learning approaches can be found in [27].

Nevertheless, many of these solutions are not efficient on complex data, but only on simple classification tasks [27], such as MNIST or CIFAR-10. Besides, they often add an important computational cost and communication overhead. Moreover, there is always a trade-off between privacy and model accuracy, due to the use of approximated activation functions.

B. Homomorphic encryption

Most existing encryption algorithms do not allow working on data unless decrypted. However, decrypting the data undermines privacy requirements. Once the data are encrypted by someone, they must first be decrypted before processing, which makes them vulnerable to unauthorized access.

HE removes the need for decrypting the data before usage. In other words, data integrity and privacy are protected while processing the data [28], [29]. HE is a cryptographic technique that has the ability to allow DL techniques to run over encrypted data without losing context. It eliminates the tradeoff between data usability and data privacy and ensures that data remains secure in untrusted environments.

In the case of DL, the algorithm can be trained and tested over encrypted data. If the DL algorithm reaches a good prediction accuracy, it can be deployed, and therefore, in real cases, it will provide a decision over the encrypted data. Using a unique secret key, the user can decrypt the obtained data. Thus, the data's privacy and security are maintained. The HE techniques can be divided mainly into three subcategories, namely Partially Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SWHE), and Fully Homomorphic Encryption (FHE). PHE enables only one form of mathematical operation on the encrypted data. SHE enables all addition and multiplication operations with only a limited range on the encrypted data. FHE enables various types of assessment operations on encrypted data with an unbounded range. PHE schemes are in general more efficient than SHE and FHE, mainly because they are homomorphic with regard to only one type of operation (addition or multiplication). SHE is more general than PHE since it supports more operations; however, it can perform them on only a limited range. The main drawback of FHE is its slow computation speed.

III. Proposed Approach

This section describes the proposed approach for PPDL in the case of Covid-19 classification. The proposed workflow is based on two phases as depicted in Figure 1. The first phase aims to encrypt the COVID-19 dataset using the Paillier method. Then, the encrypted dataset is fed to the DL algorithm for securing training. Phase 2 allows determining the class of encrypted input images based on the trained DL model.

A. HE-based Encryption

In this paper, the Paillier method is used to ensure the encryption of images. Paillier encryption is a PHE satisfying additive homomorphism. In the following, we describe the main concepts used in this study, which are key generation, encryption, and homomorphic addition.

- **Key generation:** The public key is (n,g) and the private key is (λ,μ) computed as follows:
  1) Compute two random and independent large-prime numbers p and q, where $\gcd(pq,(p-1)(q-1)) = 1$
  2) Compute n and λ, where $n = pq$ and $\lambda = \operatorname{lcm}(p-1, q-1)$. lcm denotes the least common multiple.
  3) Select a random integer g.
  4) Compute n, where $\mu = (L(g^{\lambda} \bmod n^2))^{-1} \bmod n$ and $L(x) = \frac{x-1}{n}$.
- **Encryption:** Compute the ciphertext c, where $c = g^m \cdot r^n \bmod n^2$ and r is a random number ($0 < r < n$ and $gcd(r, n) = 1$)
- **Homomorphic addition:** The product of two ciphertexts will be decrypted to the sum of their corresponding plaintexts: $D(E(m_1, r_1) \cdot E(m_2, r_2) \bmod n^2) = (m_1 + m_2) \bmod n$, and the product of a ciphertext with a plaintext will be decrypted to the sum of the corresponding plaintexts: $D(E(m_1, r_1) \cdot g^{m_2} \bmod n^2) = (m_1 + m_2) \bmod n$.

B. Secure Training

The goal of the secure training is to run the DL algorithm on the encrypted dataset and be able to achieve good performance on identifying the different classes.

The purpose of the presented work is to evaluate the performance of DL algorithms over Paillier-encrypted images. The main challenge is to find a trade-off between the accuracy of the DL algorithm in identifying COVID-19 classes and the security of images. To assess the first point, a good alternative will be to check the performance of the selected DL algorithms on plain and encrypted data. In case of a small difference between the accuracy of the DL algorithm on plain and encrypted data, we can assert that it performs well and can be adopted in real cases. Otherwise, the encryption method

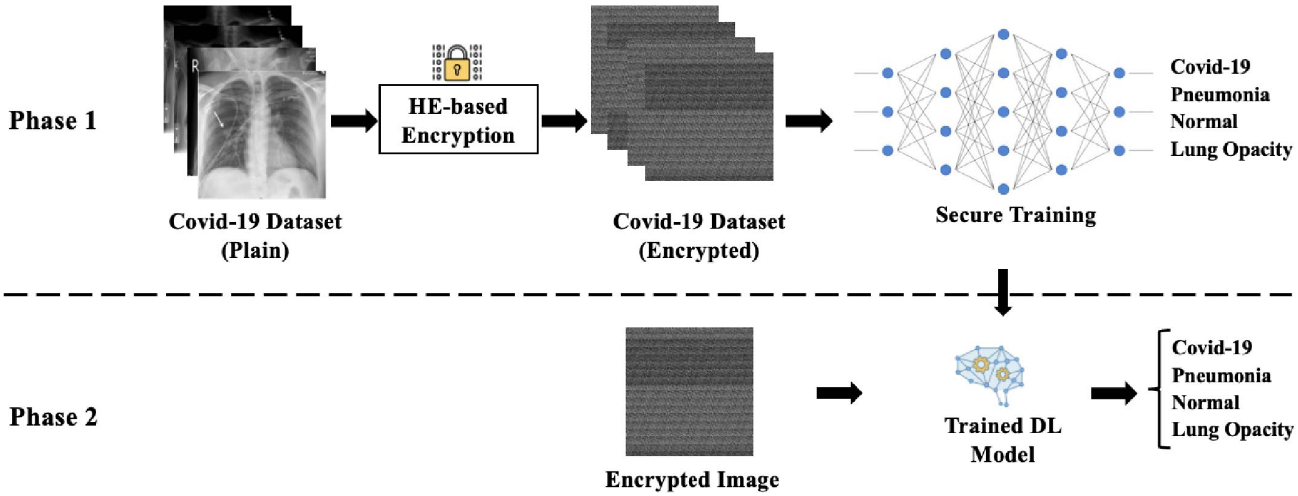

Fig. 1. Proposed approach.

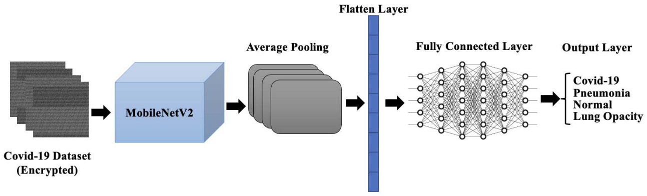

Fig. 2. Architecture of the secure training algorithm.

used for the PPDL is not adequate since it doesn't allow the DL algorithm to learn from encrypted images.

In this paper, we have used MobileNetV2, which is a CNN composed of 53 layers. MobileNetV2 algorithm can be replaced by any other transfer learning algorithm. Figure 2 depicts the architecture of the secure training algorithm used in this study.

IV. Experimental Results

This section is divided into three parts: dataset description, image encryption results, and CNN classification performances.

A. Dataset Description

In this paper, the COVID-19 Radiography database is used [30], [31]. It is a public dataset that contains four classes, namely COVID-19, viral pneumonia, normal, and lung opacity. The dataset is randomly split into 80% for training, 10% for validation, and 10% for testing. Table I shows the number of images used for training, validation, and testing for each class.

B. Implementation Details

The experiments are carried out using a server with the following configuration properties: an x64-based processor, an Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz, and a 512 GB RAM, with 8 GPUs (NVIDIA Quadro RTX 8000, 48 GB), running on Ubuntu 18.04 . The DL networks are programmed under Python 3.7 programming language. We used both the Keras 2.6 library and the TensorFlow-GPU 2.3 backend.

TABLE I
Dataset Partition

Class	Training	Validation	Testing
COVID-19	2801	350	350
Lung Opacity	4801	601	601
Normal	5403	675	676
Viral pneumonia	963	120	121

C. Results

The first step is to encrypt the dataset using the Paillier encryption scheme. Figure 3 depicts a sample of image encryption for the four classes COVID-19, viral pneumonia, normal, and lung opacity.

Then, the proposed DL architecture is applied to both plain and encrypted datasets. The training and validation accuracy of the MobileNetV2 with plain and encrypted data are depicted in Figures 4 and 5, respectively. The maximum validation accuracy for the plain dataset is 95.04% obtained at the epoch 74. Whereas, the maximum validation accuracy for the encrypted dataset is 94.97% obtained at the epoch 77.

In order to assess more robustly the performance on each dataset, we calculated the confusion matrix on the testing part which did not participate in any way in the training procedure nor in the selection of the best weights. Figures 6 and 7 show the resulting confusion matrices for the plain and encrypted datasets, respectively. In both cases, most misclassified images (76% and 67% of all misclassifications for the plain and encrypted datasets, respectively) are between the 'Normal' and 'Lung opacity' classes. Tables II and III show the detailed results, in terms of precision, recall, and f1-score, on the testing set of the plain and encrypted data, respectively. We

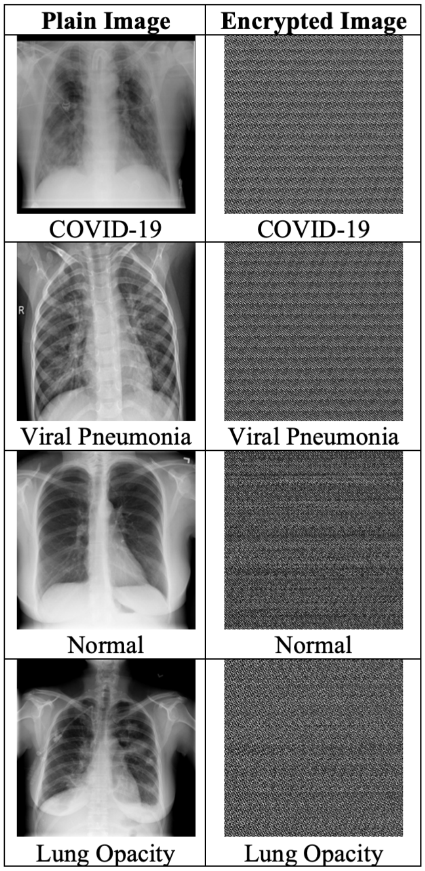

Fig. 3. Sample of plain and encrypted images for each class.

TABLE II
CLASSIFICATION RESULTS ON THE TESTING SET OF THE PLAIN DATA.

	precision	recall	f1-score	support
Covid	0.994	0.949	0.971	350
Lung_Opacity	0.953	0.903	0.927	601
Normal	0.902	0.967	0.934	676
Viral_Pneumonia	0.992	0.975	0.983	121
Accuracy			0.942	1748
Macro avg	0.96	0.949	0.954	1748
Weighted avg	0.944	0.942	0.942	1748

notice that the ’Viral Pneumonia’ class is the most affected by the encryption process, with an f1-sore decreasing from 98.3% to 87.6%, due to the reduced number of images for this class (7% of the dataset) while the three other classes were much less affected. The difference in terms of overall classification accuracy between encrypted (94.22%) and plain (93.25%) data is less than 1% on the testing set. Hence, we can conclude that the Paillier encryption scheme ensures the privacy of data while also maintaining a highly accurate prediction of the class type in the COVID-19 Radiography dataset.

V. CONCLUSION

HE-based encryption has been used in many research works due to its significant privacy benefits. The increasing need of

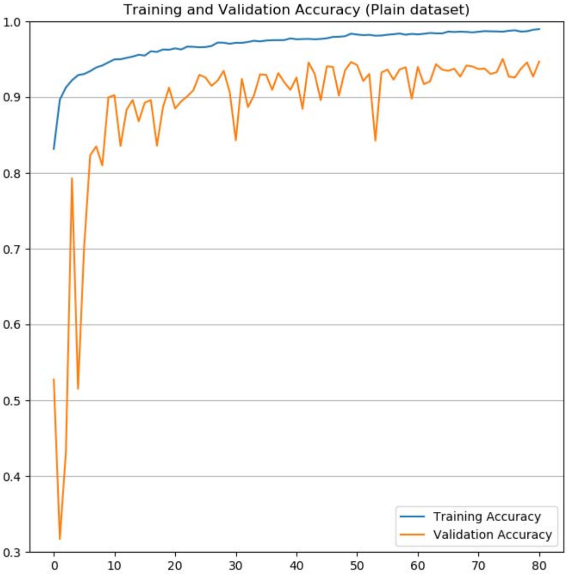

Fig. 4. Training and validation accuracy of the plain dataset.

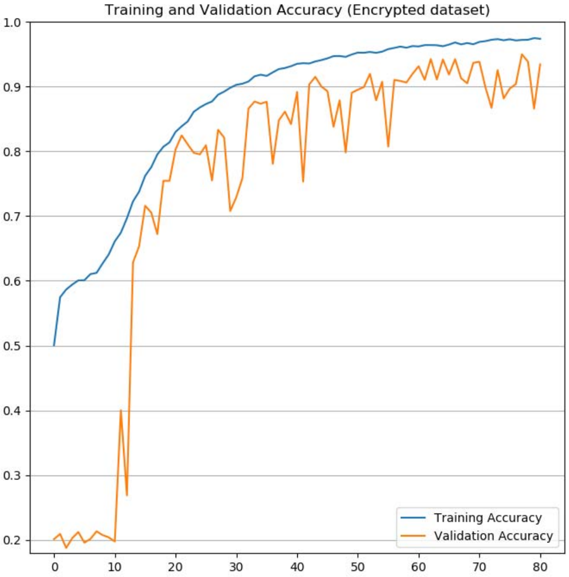

Fig. 5. Training and validation accuracy of the encrypted dataset.

TABLE III
CLASSIFICATION RESULTS ON THE TESTING SET OF THE ENCRYPTED DATA.

	precision	recall	f1-score	support
Covid	0.986	0.980	0.983	350
Lung_Opacity	0.921	0.912	0.916	601
Normal	0.908	0.953	0.930	676
Viral_Pneumonia	0.990	0.785	0.876	121
Accuracy			0.932	1748
Macro avg	0.951	0.907	0.926	1748
Weighted avg	0.934	0.932	0.932	1748

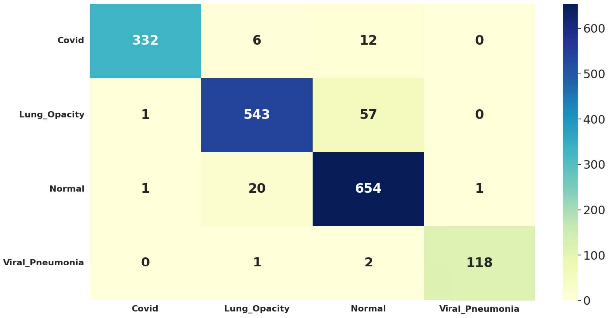

Fig. 6. Confusion matrix on the test part of the dataset of plain images.

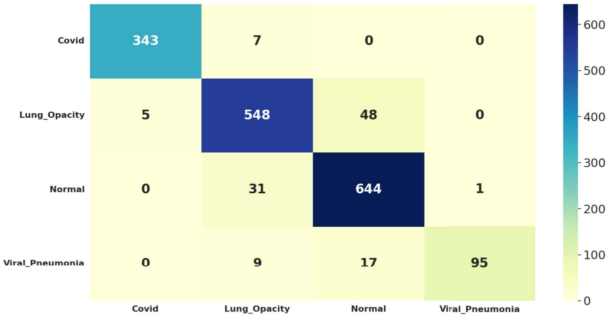

Fig. 7. Confusion matrix on the test part of the dataset of encrypted images.

ensuring privacy-preserving of data while using DL techniques makes HE a very attractive topic to the research community. In this paper, we propose using the PHE-based Paillier algorithm to tackle the problem of the privacy of sensitive healthcare data when using DL algorithms. The Paillier encryption enables securing the data and preserving good classification accuracy. Experiments conducted on the COVID-19 Radiography database show an accuracy of 94.2% for plain data and 93.3% for encrypted data. In this sense, the proposed research can serve as a tool to classify encrypted data by non-trustworthy third parties without disclosing confidentiality. However, the main limitation of HE encryption is its slow computation, which remains the major problem of this technique. Further studies will focus on proposing hybrid encryption techniques such as FHE encryption and SSC that can be used in conjunction with one another.

References

[1] S. Latif, M. Driss, W. Boulila, Z. e. Huma, S. S. Jamal, Z. Idrees, and J. Ahmad, "Deep learning for the industrial internet of things (iiot): A comprehensive survey of techniques, implementation frameworks, potential applications, and future directions," *Sensors*, vol. 21, no. 22, p. 7518, 2021.

[2] F. Mireshghallah, M. Taram, P. Vepakomma, A. Singh, R. Raskar, and H. Esmaeilzadeh, "Privacy in deep learning: A survey," *arXiv preprint arXiv:2004.12254*, 2020.

[3] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *arXiv preprint arXiv:2011.11819*, 2020.

[4] S. Meredith, "Facebook-cambridge analytica: A timeline of the data hijacking scandal," ttps://www.cnbc.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html, 2018, accessed: 2021-11-18.

[5] W. Taylor, Q. H. Abbasi, K. Dashtipour, S. Ansari, S. A. Shah, A. Khalid, and M. A. Imran, "A review of the state of the art in non-contact sensing for covid-19," *Sensors*, vol. 20, no. 19, p. 5665, 2020.

[6] S. B. Atitallah, M. Driss, W. Boulila, and H. B. Ghézala, "Randomly initialized convolutional neural network for the recognition of covid-19 using x-ray images," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 55–73, 2022.

[7] S. Ben Atitallah, M. Driss, W. Boulila, A. Koubaa, and H. Ben Ghézala, "Fusion of convolutional neural networks based on dempster–shafer theory for automatic pneumonia detection from chest x-ray images," *International Journal of Imaging Systems and Technology*, 2021.

[8] M. Rehman, R. A. Shah, M. B. Khan, N. A. A. Ali, A. A. Alotaibi, T. Althobaiti, N. Ramzan, S. A. Shaha, X. Yang, A. Alomainy *et al.*, "Contactless small-scale movement monitoring system using software defined radio for early diagnosis of covid-19," *IEEE Sensors Journal*, 2021.

[9] S. Guefrechi, M. B. Jabra, A. Ammar, A. Koubaa, and H. Hamam, "Deep learning based detection of covid-19 from chest x-ray images," *Multimedia Tools and Applications*, vol. 80, no. 21, pp. 31 803–31 820, 2021.

[10] M. Ben Jabra, A. Koubaa, B. Benjdira, A. Ammar, and H. Hamam, "Covid-19 diagnosis in chest x-rays using deep learning and majority voting," *Applied Sciences*, vol. 11, no. 6, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/6/2884

[11] W. Boulila, S. A. Shah, J. Ahmad, M. Driss, H. Ghandorh, A. Alsaeedi, M. Al-Sarem, and F. Saeed, "Noninvasive detection of respiratory disorder due to covid-19 at the early stages in saudi arabia," *Electronics*, vol. 10, no. 21, p. 2701, 2021.

[12] W. Boulila, H. Ghandorh, M. A. Khan, F. Ahmed, and J. Ahmad, "A novel cnn-lstm-based approach to predict urban expansion," *Ecological Informatics*, vol. 64, p. 101325, 2021.

[13] M. Jemmali, L. K. B. Melhim, A. Alourani, and M. M. Alam, "Equity distribution of quality evaluation reports to doctors in health care organizations," *PeerJ Computer Science*, vol. 8, p. e819, 2022.

[14] J. Liu and X. Meng, "Survey on privacy-preserving machine learning," *Journal of Computer Research and Development*, vol. 57, no. 2, p. 346, 2020.

[15] M. Alkhelaiwi, W. Boulila, J. Ahmad, A. Koubaa, and M. Driss, "An efficient approach based on privacy-preserving deep learning for satellite image classification," *Remote Sensing*, vol. 13, no. 11, p. 2221, 2021.

[16] M. A. P. Chamikara, P. Bertók, D. Liu, S. Camtepe, and I. Khalil, "Efficient data perturbation for privacy preserving and accurate data stream mining," *Pervasive and Mobile Computing*, vol. 48, pp. 1–19, 2018.

[17] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy." *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.

[18] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," *arXiv preprint arXiv:1610.05755*, 2016.

[19] Z. Bu, J. Dong, Q. Long, and W. J. Su, "Deep learning with gaussian differential privacy," *Harvard data science review*, vol. 2020, no. 23, 2020.

[20] T. Chanyaswad, J. M. Chang, and S.-Y. Kung, "A compressive multi-kernel method for privacy-preserving machine learning," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 4079–4086.

[21] F. Tramer and D. Boneh, "Slalom: Fast, verifiable and private execution of neural networks in trusted hardware," *arXiv preprint arXiv:1806.03287*, 2018.

[22] O. Ohrimenko, F. Schuster, C. Fournet, A. Mehta, S. Nowozin, K. Vaswani, and M. Costa, "Oblivious multi-party machine learning on trusted processors," in *25th {USENIX} Security Symposium ({USENIX} Security 16)*, 2016, pp. 619–636.

[23] N. Kumar, M. Rathee, N. Chandran, D. Gupta, A. Rastogi, and R. Sharma, "Cryptflow: Secure tensorflow inference," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 336–353.

[24] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, “A hybrid approach to privacy-preserving federated learning,” in *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, 2019, pp. 1–11.

[25] C. Juvekar, V. Vaikuntanathan, and A. Chandrakasan, “GAZELLE: A low latency framework for secure neural network inference,” in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 1651–1669.

[26] M. Chase, R. Gilad-Bachrach, K. Laine, K. Lauter, and P. Rindal, “Private collaborative neural network learning,” *Cryptology ePrint Archive*, 2017.

[27] H. C. Tanuwidjaja, R. Choi, S. Baek, and K. Kim, “Privacy-preserving deep learning on machine learning as a service—a comprehensive survey,” *IEEE Access*, vol. 8, pp. 167 425–167 447, 2020.

[28] P. V. Parmar, S. B. Padhar, S. N. Patel, N. I. Bhatt, and R. H. Jhaveri, “Survey of various homomorphic encryption algorithms and schemes,” *International Journal of Computer Applications*, vol. 91, no. 8, 2014.

[29] F. Armknecht, C. Boyd, C. Carr, K. Gjøsteen, A. Jäschke, C. A. Reuter, and M. Strand, “A guide to fully homomorphic encryption.” *IACR Cryptol. ePrint Arch.*, vol. 2015, p. 1192, 2015.

[30] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al Emadi *et al.*, “Can ai help in screening viral and covid-19 pneumonia?” *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.

[31] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan *et al.*, “Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images,” *Computers in biology and medicine*, vol. 132, p. 104319, 2021.