Differential Privacy Practice on Diagnosis of COVID-19 Radiology Imaging Using EfficientNet

Zümrüt Müftüoğlu
Department of Big Data and Artificial Intelligence Applications
The Presidency of the Republic of Turkey,
The Digital Transformation Office
Ankara, Turkey
zumrut.muftuoglu@cbddo.gov.tr
https://orcid.org/0000-0003-3754-7491

M. Ayyüce Kızrak
Department of Big Data and Artificial Intelligence Applications
The Presidency of the Republic of Turkey,
The Digital Transformation Office
Ankara, Turkey
ayyuce.kizrak@cbddo.gov.tr
https://orcid.org/0000-0001-8545-4586

Tülay Yıldırım
Department of Electronics and Communication Engineering
Yıldız Technical University
Istanbul, Turkey
tulay@yildiz.edu.tr
https://orcid.org/0000-0001-9993-5583

***Abstract*— Medical sciences are an important application area of artificial intelligence. Healthcare requires meticulousness in the whole process from collecting data to processing. It should also be handled in terms of data quality, data size, and data privacy. Various data are used within the scope of the COVID-19 outbreak struggle. Medical and location data collected from mobile phones and wearable devices are used to prevent the spread of the epidemic. In addition to this, artificial intelligence approaches are presented by using medical images in order to identify COVID-19 infected people. However, studies should be carried out by taking care not to endanger the security of the data, people, and countries needed for these useful applications. Therefore, differential privacy (DP) application, which was an interesting research subject, has been included in this study. CXR images have been collected from COVID-19 infected 139 and a total of 373 public data sources were used for a diagnostic concept. It has been trained with EfficientNet-B0, a recent and robust deep learning model, and proposal the possibility of infected with an accuracy of 94.7%. Other evaluation parameters were also discussed in detail. Despite the data constraint, this performance showed that it can be improved by augmenting the dataset. The most important aspect of the study was the proposal of differential privacy practice for such applications to be reliable in real-life use cases. With this view, experiments were repeated with DP applied images and the results obtained were presented. Here, Private Aggregation of Teacher Ensembles (PATE) approach was used to ensure privacy assurance.**

Keywords—: COVID-19, deep learning, EfficientNet, X-Ray, radiology imaging, PATE, differential privacy.

I. INTRODUCTION

Coronavirus, which was uncovered for the first time in Wuhan, China in December 2019, was proclaimed as a new coronavirus by the World Health Organization (WHO-WHO) on 11 February 2020 and named as COVID-19 (2019-nCov) [1]. Governments take various prevention to reduce the extent of the epidemic. Some of them are to close the borders and to recommend social distance limits. However, the number of individuals affected by coronavirus continues to increase in most countries. Clinical trials, medicines, and vaccines need to be developed and implemented to ensure that the importance and challenges are also medically functional. Based on the data published in the process, while waiting for the Polymerase Chain Reaction (PCR) test time required for diagnosis, pioneering estimates help healthcare professionals, accelerate diagnosis by data science and artificial intelligence studies.

New studies are added to the literature for the diagnosis of COVID-19 using medical images. A study examines the clinical features of pneumonia patients infected with coronavirus and influenza virus and emphasizes that it is possible to record the stage of the disease based on chest CT and CXR images and other test results. CT and CXR shots are used as an auxiliary diagnostic parameter recommended by radiologists and other experts [2], [3]. Wang et al. Deep learning model uses 1119 CT images in their studies. However, these images also contain images of patients diagnosed with viral pneumonia. The total accuracy rate is calculated as 79.3%, specificity for test dataset 83%, sensitivity 67%. In this research where modified-Inception was used as the deep learning model, the accuracy of giving correct diagnosis to COVID-19 positive patients was 85.2% [4]. Zhao et al. in their study, use a convolutional neural network-based model by making use of 275 CT images. In this binary classification study, the general accuracy rate is 84.7%, while it reaches 85.3% in F1-score [5].

This study uses artificial intelligence for a successful and rapid diagnostic recommendation using CXR prior to the waiting time for tests due to the density in hospitals. Nevertheless, the focus of the study is on the privacy approaches that should be taken into consideration while producing artificial intelligence solutions by making use of medical data. This paper covers an application to draw attention to data privacy while reaching fast solutions. Differential privacy ensures that data-driven work develops in a reliable environment. It allows governments or technology companies to collect and share information about individuals while protecting the privacy of individual users.

The organization of the paper is as follows. In the second section, details about the relevant COVID-19 medical imaging and dataset used are inscribed. Besides, details about the implemented deep learning and explainability model and experimental results are presented in common with the evaluation metrics. The proposed differential privacy application was introduced in section 3. In section 4, the

results obtained in this study were evaluated from a DP perspective. Section 5 presents results and the discussion of the study to the literature, and in the section 6 the research area is mentioned for the future.

II. DEEP LEARNING APPROACH FOR COVID-19 DIAGNOSIS

At the first stage of the research, COVID-19 CXR and CT images published in international open-sources were investigated and a dataset was created. Accordingly, the deep learning model chosen is explained technically. Performance evaluation is included with the experimental results.

A. COVID-19 Radiology Imaging Dataset

CT imaging is a procedure with cross-sections as a result of computer processing of signals obtained by rotating the patient's body with an X-ray beam [3]. Chest radiography (CXR), which contains information like this but with fewer details, is also used to detect bone tumors, abdominal lesions, and heart disease anomalies. Besides that, CXR is more beneficial for disease progression follow-up due to the per-patient applicability, easy access, and low X-ray exposure [4].

One of the important symptoms of COVID-19 epidemic disease is the intense cough and difficulty breathing. CXR and CT imaging are frequently used for these findings. It demonstrates properties similar to the indications of COVID-19 pneumonia from the MERS and SARS family. However, the asset of discrete nodules and reverse halo increases the likelihood of COVID-19 findings [5]. In recent researches, the relationship between chest images and PCR tests is generally found to be correlated. Moreover, there are typical findings on CT images, and cases that were negative as a result of PCR but became positive after 2-8 days were recorded [5]. Despite this, the clinical images of COVID-19 are still not very clear, and different treatments are also used in diagnosing the disease. The most prominent of these is the examination of chest CXR and CT. It is vital to struggle with the epidemic that these images are easily obtained and quickly appraise by experts [6], [7]. Diagnostic suggestion systems from artificial intelligence-based computer vision methods are already used. CXR and CT images from COVID-19 infected patients further allow data scientists to work with healthcare professionals [8], [9].

This paper utility a dataset of CXR images. The dataset occurs images of 234 healthy and 139 COVID-19 infected individuals who were compiled from the T-Covid group from open source data [10]. In this study, we used most of the data to train to make model training more accurate. This dataset with different dimensions is divided into 80% training, 10% validation, and 10% test sets. Examples from the dataset are shown in Figure 1.

B. Deep Learning Model: EfficientNet-B0

Convolutional neural network-based deep learning models are used in diagnostic offers in medical images. EfficientNet, an up-to-date, cost-efficient, and robust model developed by scaling three parameters such as depth, width, and resolution, is used in this study. It is a group consisting of 8 models, B0-B7. The EfficientNet-B0 model is used by setting the input image size to 224 × 224. Figure 2 shows the scaling architecture that summarizes EfficientNet [11].

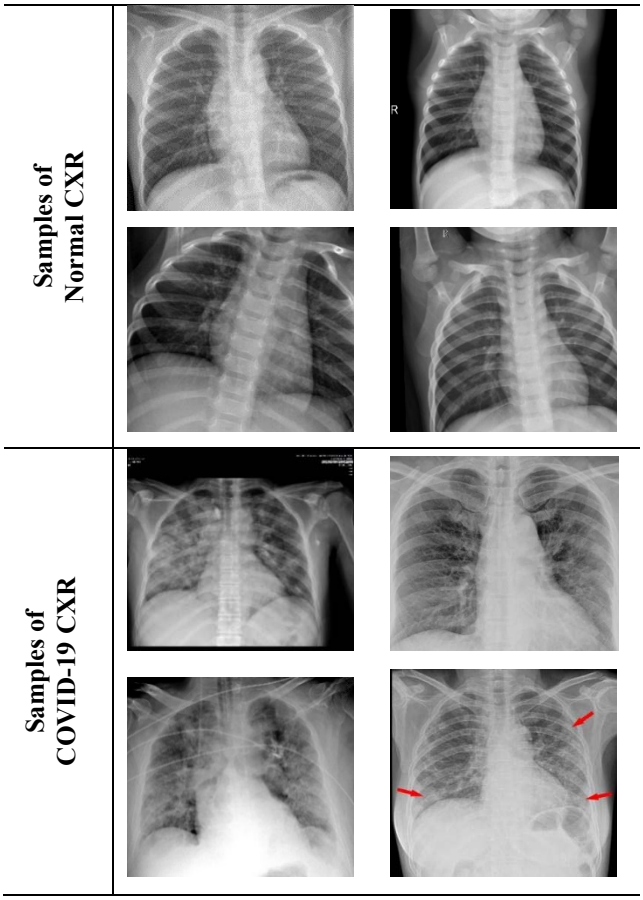


Fig. 1. Samples of the dataset [10].

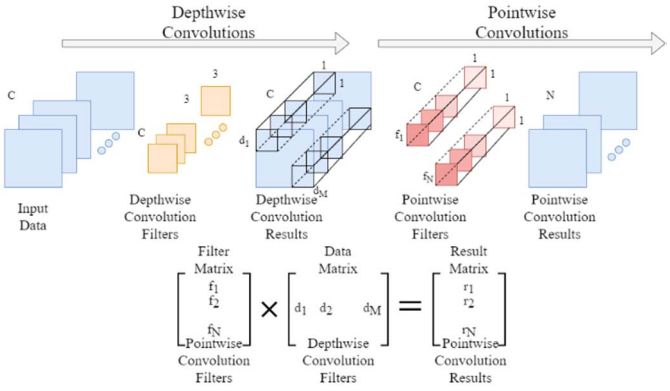


Fig. 2. A basic representation of depthwise and pointwise convolutions.

$\phi = 1$ and grid search for α, β, and to rescale from B0 to B1. α, β, γ set: Accordingly, for scaling from B2 to B7, ϕ is chosen between 2~7. The main building block for EfficientNet is the inverse bottleneck MBConv of MobileNetV2. Using shortcuts between bottlenecks by connecting a much smaller number of channels compared to expansion layers, it was combined with an in-depth separable convolution. Unitization the bottleneck approach, it overcomes the computational complexity k^2 rather than classics methods. In the condition where k expresses the kernel size, it indicates the height and width of the 2-dimensional convolution window.

$$depth \rightarrow d = \alpha^{\varphi} \rightarrow \alpha \geq 1 \quad (1)$$

$$width \rightarrow w = \beta^{\varphi} \rightarrow \beta \geq 1 \quad (2)$$

$$resolution \rightarrow$$

$$r = \gamma^{\varphi} s.t. \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \rightarrow \gamma \geq 1 \quad (3)$$

This paper analyzes the EfficientNet-b0 performance, one of the most sophisticated deep learning models, with CXR images collected from recent COVID-19 cases. Table 1 shows the selected model parameters. The parameters selected in the table are summarized with data augmentation and parameter optimization. The accuracy and loss curves achieved for the model are shown in Figure 3 and Figure 4. In Table II, further the different total validation and test accuracy, it shows the sensitivity, recall and F1 score metrics to be considered when classifying medical images. The performance shows that it is comparable with similar studies. When evaluating the classification results, different methods should be determined according to the domain. In critical research fields such as medical, looking at the accuracy rate can be misleading.

TABLE I. MODEL PARAMETERS.

Optimizer	Adam (β_1= 0.9, β_2= 0.999)
Learning Rate	0.001
Batch Size	32
Epoch	10/120

TABLE II. PERFORMANCE OF THE EFFICIENTNET-B0 MODEL.

Model	Label	Test Precision	Test Recall	Test F1-Score	Validation Accuracy	Test Accuracy
EfficientNet-B0	Normal	0.950	0.950	0.950	0.970	0.947
	COVID-19	0.940	0.940	0.940		

For the implementation of the research, it allows you to improve deep learning applications by using different frameworks such as TensorFlow, Keras, PyTorch. Especially to use this cloud service, it is sufficient to have a Google account. In this work, the results are examined using a data center accelerator with Tesla P100 GPUs, which are accessible in the Colab cloud, to evaluate the performance.

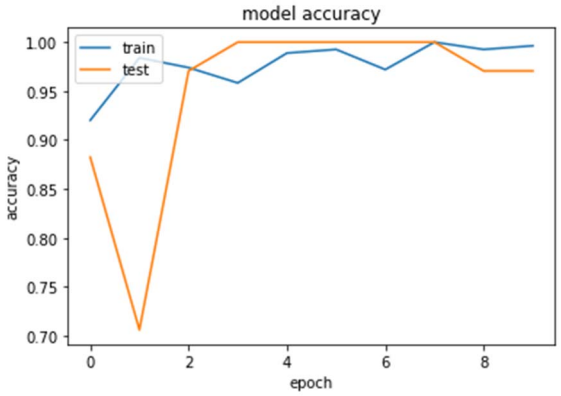


Fig. 3. Result of EfficientNet-B0 model accuracy.

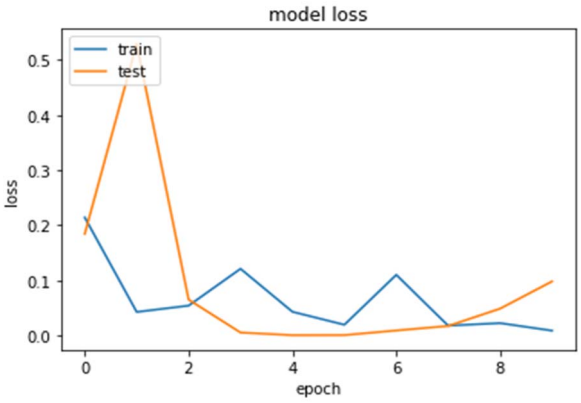
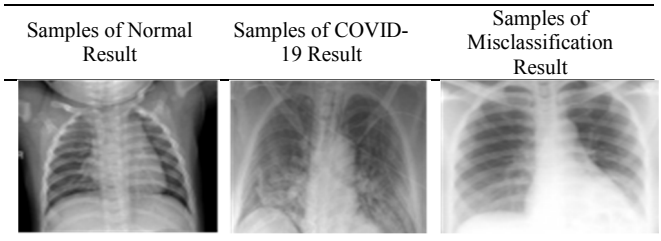


Fig. 4. Result of EfficientNet-B0 model loss.

First, creating a confusion matrix and four possible outcomes should be calculated. Besides, true positive and false positive errors in the confusion matrix are critical. Table III also shows the confusion matrix valuation. Examples of the classification result in the images in the test group obtained are shown in Figure 5.

TABLE III. CONFUSION MATRIX OF THE EFFICIENTNET-B0 MODEL.

		Predicted	
		COVID-19	Normal
Actual	COVID-19	0.941	0.058
	Normal	0.047	0.952



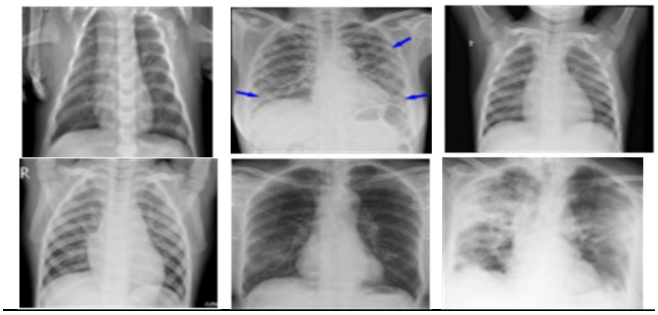


Fig. 5. Samples of after classification to normal and COVID-19 CXR.

At this phase, despite the limited CXR image for detecting COVID-19 infected CXR images, a robust deep learning approach has achieved better accuracy than current studies in the literature. However, all these studies still have many limitations and difficulties. Data quantity, data quality, and reliability are concepts that directly affect the impact of the research. Moreover, patency, accessibility, and privacy should be addressed.

It is also a major criterion that the results obtained can be explained. At the second stage of the research, explainability of the deep learning classifier used with the Grad-CAM approach can be tested. It is also used to identify possible bias in classification. Gradients take the value 1 for the respective class and other possible classes are shown in blue on the heat map. To obtain guided Grad-CAM visualizations, the dot product process is carried out by propagation back the directed heat map [12]. Figure 6 shows the EfficientNet-B0 model used for the diagnosis of COVID-19 from CXR images.

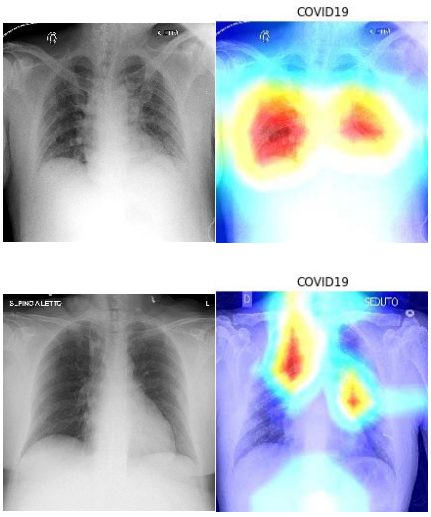


Fig. 6. Original CXR and after EfficientNet-B0 classification feature visualization.

III. DIFFERENTIAL PRIVACY PRACTICE OF COVID-19 DIAGNOSIS

AI approaches preferred in many new-generation studies are sensitive to attack types. Access to training and model parameters and possible privacy leaks for images used in the face recognition system may result in the disclosure of sensitive data [13], [14], [16]. Avoiding this is a separate research topic. Appropriate deep learning models used in this area are shown in Table IV.

TABLE IV. DIFFERENTIALLY PRIVATE DEEP LEARNING METHODS [16].

Related Work	Adversarial Setting	System Setting	Privacy Guarantee Method
Shokri et al. [17]	Additional capabilities	Distributed system	Differentially private Stochastic Gradient Descent (SGD) algorithm with convex objective functions
Adabi et al. [15]	Additional capabilities	Centralized system	Differentially private SGD algorithm with non-convex objective functions
Phah et al. [16]	General capabilities	Centralized system	Objective function perturbation of deep auto-encoder

The methods of guaranteeing privacy are divided into two classes. The first is obtained by adding noise in the optimization algorithm. The second is achieved by adding particular noises to the objective functions before learning. Differential privacy (DP) is a privacy mechanism that provides a security guarantee based on information theory.

Even if the data needed by artificial intelligence systems do not contain health data, these systems can make predictions about the health information of individuals through their inferences. In terms of protecting data privacy, linking the health records itself and metadata about them and making predictions based on cross-correlation are among the most important problems. It is easier to define identity at a personal level based on the data processed with cross-links. This raises ethical, legal, and social concerns such as confidentiality, responsibility, and bias. DP is a mathematical definition for strengthening confidentiality in data-based statistics and machine learning approaches. Accordingly, DP is an acceptable measure of privacy protection to analyze sensitive personal information [17].

Differential Privacy (Ɛ, δ): M denotes a random mechanism in this equation. S represents each output and shows the difference for D and D' datasets when S and M are appropriate. In this way (Ɛ, δ) protects confidentiality [18].

$$Pr[M(D) \in S] \leq exp(\text{Ɛ}).Pr[M(D') \in S] + \delta \qquad (4)$$

where Ɛ is the privacy budget and δ is the probability of error. The ratio of the two possibilities is limited by $e^{\text{Ɛ}}$. If the randomized mechanism M, δ = 0, it simply provides Ɛ-differential confidentiality. Under the condition Δ = 0, stronger differential confidentiality is obtained. (Ɛ, δ)-DP retains the latitude of breaking the differential confidentiality for some low probability events [19].

The following definition is called privacy loss:

$$\int_{M(D)||M(D\prime)} = ln\frac{Pr[M(D)\in S]}{Pr[M(D\prime)\in S]} \qquad (5)$$

The way to attain Ɛ-DP and (Ɛ, δ) -DP to add noise sampled from Laplace and Gauss distributions. The noise is

commensurate with the sensitivity of the mechanism M [20]. A smaller Ɛ stands for more robust privacy [21].

Sensitivity states how much complexity is due in the mechanism. For instance, when we publish a specified query f of dataset D, the sensitivity will calibrate the required volume of noise for $f(D)$. There are two types of sensitivity in the literature of DP: the global sensitivity and the local sensitivity.

IV. EXPERIMENTAL RESULT AND DISCUSSION

With the widespread use of machine learning and deep learning models, privacy concerns have increased, especially for models working on sensitive medical records containing personal data records. In this paper we use differential privacy which is one of the most robust definitions of privacy. It bases on the principle using random noise to provide that the public record doesn't change if one record in the dataset changes.

In the final stage of the work, DP is implemented on EfficientNet-B0 model which is used for the diagnosis of COVID-19 from CXR images. During the study, to ensure privacy assurance, we handle Private Aggregation of Teacher Ensembles (PATE) approach as the differential privacy method which was proposed by Papernot et. al. [22]. As illustrated in Figure 7, the PATE framework includes the aggregation of teacher models, an aggregation mechanism, and a student model.

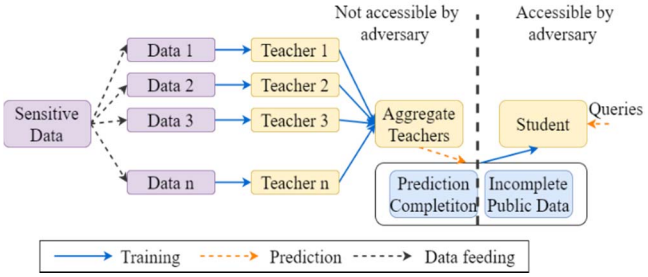


Fig. 7. Overview of the approach: (1) an ensemble of teachers is trained on disjoint subsets of the sensitive data, (2) a student model is trained on public data labeled using the ensemble [23].

In the scope of the PATE framework, we structured the Radiology Imaging Dataset which includes images of healthy and COVID-19 infected individuals as "teacher" and "student" datasets to create "teacher" and "student" models. While the training set is used as "teacher" dataset, the test set is used as "student" dataset. The training of each teacher model is run independently, and outputs are aggregated noisily by adding Laplacian Noise to implement DP using PATE analysis. At the end of this process, obtained private labels are trained with the student model to finalize DP implementation.

We use Pysyft (powered by Pytorch) [24] to perform PATE Analysis. According to the analysis results, the analysis, the data-dependent epsilon is obtained as 5.9782 (for Ɛ= 0.1). The dependent epsilon value, expected to be low, gives a tight bound of privacy loss of the aggregated data and depends on the amount of agreeing for teacher models with each other which makes leaking and tracking sensitive information difficult.

To compare, DP implemented model reaches %71 accuracy rate while the original model has %94.7 accuracy. Even though the accuracy lowered as expected, the results are promising for future work.

V. CONCLUSION

EfficientNet, a robust deep learning model, is recommended by utilizing CXR images to make fast and effective decisions to the healthcare personnel in the PCR test during the COVID-19 epidemic operate. Withal proposed technique, a diagnostic support system can be designed with 94.7% test accuracy, 94% precision, F1-score, and recall accuracy. In addition, one of the important issues encountered in medical data is the explainability of the systems. Therefore, it is explained with the Grad-Cam approach in which regions of the image the EfficientNet-B0 model uses the features in CXR images for the diagnosis of COVID-19. Further, it is emphasized in the artificial intelligence research carried out in the field of medicine that requires particularity in terms of data and model confidentiality and robustness. Therefore, by implementing the DP approximation which has studies on CNN model in the literature has been applied for the first time on EfficientNet model, individuals are encouraged to use artificial intelligence models in healthcare services while providing data security.

VI. FUTURE WORK

In AI studies administer in the domain of health and medicine, it is underlined that it requires fastidiousness in terms of data and model privacy and robustness. Furthermore, studies will be useful for researchers to try several different privacy practices and compare different mechanisms in deep learning-based medical image classification models. Thus, an equitable and secure network can be obtained while providing AI and digitalization in healthcare.

In future work, experiments will be performed for CT images for COVID-19, and it is also aimed to optimize the deep learning diagnostic model with various DP metrics to increase accuracy. By optimizing observed trade-off points, the results of study are promising for preserving data privacy in medical applications for future works. A fair and secure network can be achieved while providing artificial intelligence and digitalization in healthcare without sacrificing performance of model.

ACKNOWLEDGMENT

We wish to acknowledge Yavuz Kömeçoğlu, a machine learning engineer who worked with us for modeling. We are grateful to Radiologist Dr. Nevit Dilmen who shared information about the use of CT and CXR images for the diagnosis of COVID-19. We would also like to thank T-Covid powered by Turkish AI start-up T-Fashion, who created the dataset and shared it for this study.

REFERENCES

[1] S. N. Mali, A.P. Pratap, and B. R. Thorat, "The Rise of New Coronavirus Infection-(COVID-19): A Recent Update", Eurasian Journal of Medicine and Oncology (EJMO), 4(1):35–41, 2020.

[2] COVID19 Global Online Hackathon, Available: https://covid-global-hackathon.devpost.com/, 2020.

[3] M. Y. Ng, et al., "Imaging Profile of the COVID 19 Infection: Radiologic Findings and Literature Review", Radiology: Cardiothoracic Imaging 2.1, e200034, 2020.

[4] T. Ai and Z. Yang, et al., "Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases", 2019.

[5] S. Salehi, and A. Abedi, et al., "Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients", American Journal of Roentgenology, Diagnosis Imaging and Related Science, 1-7. 10.2214/AJR.20.23034, 2019.

[6] Q. Ding, and P. Lu, et al., "The Clinical Characteristics of Pneumonia Patients Co-Infected with 2019 Novel Coronavirus and Influenza Virus in Wuhan China", Journal of Medical Virology, Available: https://doi.org/10.1002/jmv.25781, 2020.

[7] X. Li, and M. Liu, et al., "Preliminary Recommendations for Lung Surgery during the 2019 Novel Coronavirus Disease (COVID-19) Epidemic Period", 20;23(3):133-135, DOI: 10.3779/j.issn.1009-3419.2020.03.01, 2020.

[8] S. Wang, and B. Kang, et al., "A Deep Learning Algorithm Using CT Images to Screen for Corona Virus Disease (COVID-19)", Available: https://doi.org/10.1101/2020.02.14.20023028, 2020.

[9] J. Zhao, and Y. Zhang, et al., "COVID-CT-Dataset: A CT Scan Dataset about COVID-19", Available: https://arxiv.org/abs/2003.13865, 2020.

[10] T-Covid, "A Fast COVID-19 Diagnosis Tool powered by AI", Available: https://covid.tfashion.ai/, 2020.

[11] M. Tan, and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", Thirty-sixth International Conference on Machine Learning (ICML), CA, USA, 2019.

[12] R. R. Selvaraju, and M. Cogswell, et. al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", International Journal of Computer Vision volume 128, pages 336–359, 2020.

[13] M. Fredrikson, S. Jha, and T. Ristenpart, "Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures", In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pages 1322–1333, New York, NY, USA, ACM, 2015.

[14] M. Abadi, A. Chu, I. J. Goodfellow, et. al., "Deep Learning with Differential Privacy", In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, pages 308–318, 2016.

[15] N. Phan, Y. Wang, et. al., "Differential Privacy Preservation for Deep Autoencoders: An Application of Human Behavior Prediction". In AAAI, pages 1309–1316, 2016.

[16] R. Shokri, and V. Shmatikov, "Privacy-Preserving Deep Learning", In SIGSAC, pages 1310–1321, 2015.

[17] C. Dwork, "Differential Privacy: A Survey of Results", In International Conference on Theory and Applications of Models of Computation, 2018.

[18] C. Dwork, "A Firm Foundation for Private Data Analysis", Commun. ACM, 54(1):86–95, 2011.

[19] A. Beimel, K. Nissim, and S. Stemmer, "Private Learning and Sanitization: Pure vs. Approximate Differential Privacy", CoRR, abs/1407.2674, 2014.

[20] B. Jayaraman, D. Evans, "Evaluating Differentially Private Machine Learning in Practice", In 28th USENIX Security Symposium, Santa Clara, CA, USA, 2019.

[21] A. Haeberlen, B. C. Pierce, and A. Narayan, "Differential Privacy Under Fire", SEC'11: Proceedings of the 20th USENIX conference on Security, Pages 33, August 2011.

[22] N. Papernot, M. Abadi, and Ú. Erlingsson, et. al., "Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data", Conference on Learning Representations (ICLR), 2017.

[23] N. Papernot, S. Song, and I. Mironov et. al., "Scalable Private Learning with PATE", Sixth International Conference on Learning Representations (ICLR), 2018.

[24] T. Ryffel, A. Trask, M. Dahl, B. Wagner, J. Mancuso, et. al., "A Generic Framework for Privacy Preserving Deep Learning", Available: https://arxiv.org/abs/1811.04017, 2018.