2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Detecting Lung Cancer with Federated and Transfer Learning

G. Mostafa[1], Md Shifat Hamidi[2], and Dewan Md. Farid[1]
[1]Department of Computer Science and Engineering, United International University,
United City, Madani Avenue, Badda, Dhaka 1212, Bangladesh
[2]Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology,
Dhaka, Bangladesh
Email : {gmostafa1210;sifat.hamidi}@gmail.com, dewanfarid@cse.uiu.ac.bd

2023 26th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-5901-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIT60459.2023.10441256

***Abstract*—Lung cancer is a disease that affects and causes abnormalities in the lungs. The current methods to find and treat lung cancer require precise and timely detection to improve patient outcomes and survival rates. However, traditional approaches to lung cancer detection face challenges due to the extensive patient information spread across different medical institutions and research centers. Concerns about data privacy and the need to keep data controlled have prevented the consolidation of this valuable data into a central repository for analysis. As a result, the development of effective and accurate detection models has been limited by restricted access to diverse and comprehensive datasets. To overcome these challenges, federated learning (FL) has emerged as a promising approach in the healthcare field. It has great potential in the healthcare or medical sector where we can get a better Machine learning model while ensuring patient data privacy. This paper presents an FL approach for detecting lung cancer in medical images after setting an initial weight using Transfer Learning. Using this approach, we achieved a break a hand accuracy of 91.03% in detecting lung cancer. This demonstrates the potential of FL for accurate and privacy-preserving medical diagnosis.**

***Index Terms*—Machine Learning, Federated Learning, Medical Data, Data Privacy, Cancer Detection, Lung Cancer, Transfer Learning, VGG16, MobileNetV2, Xception, EfficientNetB4, ResNet15, ImageNet**

I. Introduction

Federated Learning (FL) emerges as an advanced distributed machine learning technique that facilitates multiple participants in training autonomous machine learning models on their own datasets while sustaining the confidentiality of their data [1]. So, instead of moving raw data around, the parties collaborate by sharing model updates (parameters), which are then amalgamated with all other encrypted models' weights to make a global generalizable model over time. This enables organizations to harness the power of machine learning while assuring their data security.

FL holds the potential to transform the healthcare industry through its integration of numerous diverse medical applications by ensuring privacy-preserving analysis of crucial and sensitive medical data [2]. For instance, Federated Averaging (FedAvg), a well-known FL algorithm, can be employed to develop models on distributed clinical information, featuring electronic health records, without undergoing data exchange or centralization [3]. The FL approach can train a deep learning model on an extensive data set of clinical images, such as mammograms or biopsy slides [4]. This approach can improve the model's performance by leveraging the combined data from multiple remote collaborators, while still maintaining the isolation of individual patients' information. Fig. 1 visually represents the basic operational framework.

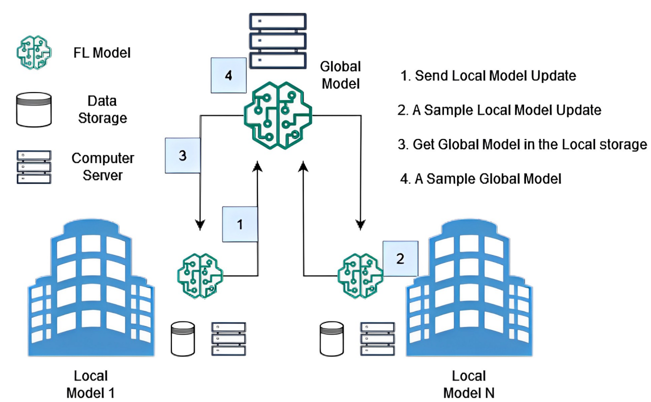


Fig. 1: Workflow of Federated Learning.

An updated algorithm of FedAvg is presented by Li et al. in [5] where they discussed the use of federated optimization in heterogeneous networks. The authors proposed a method that employs FL to construct a global model while considering the heterogeneity attributes of the participating clients. The process focused on a technique called meta-learning to adapt the model updates from each client to the global model, and it allows clients to choose their local best model based on their validation metric. They evaluate the performance of their approach on a real-world data set and demonstrate that it outperformed alternate methods in terms of both convergence speed and prediction accuracy. FL can also address the issue of limited data availability in the medical sector, which can be a significant hurdle for the development of accurate cancer detection models. By aggregating different model updates from multiple parties, FL can improve the performance of machine learning models even with limited data at each side [6].

In this paper, we aim to detect one of the deadliest and most common cancers in this generation by ensuring more robust and accurate diagnosis and treatment with the help of diverse FL methods. The major contributions of this paper are shown below:

- Implementing the FL method involving different clients to train and detect three prominent lung cancer types.
- Performing a rigorous assessment of five different dynamic models used on our diverse dataset.
- Estimating outperformed model which can identify the three types of lung cancer to ensure accurate diagnosis in medical sector .

II. Related Works

In recent years, the healthcare sector has witnessed remarkable advancements in medical diagnosis and image analysis with the integration of cutting-edge technologies like Federated Learning. Xu et al. [7] mentioned some statistical and system challenges and solutions of the FL method on systemic and random biased electronic health records (EHRs) and some implications and potentials in healthcare. A similar approach has been followed by Zhao et al. [8]. In this study, unreliable participants, who have low-quality data, have been identified and reduced their impact on feature capturing and model training. Le et al. in [9] have introduced a privacy-preserving system within a federated environment that enables to identification of similar patients across different institutions. Homomorphic encryption has been applied to patient searches and all this happened with breaching patient privacy.

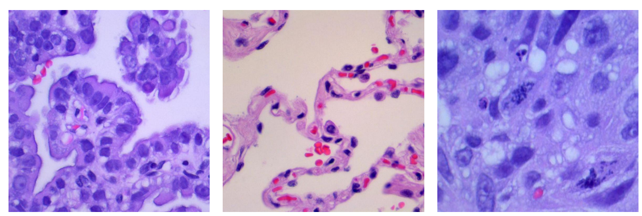
Yaqoob et al. [10] implemented a hybrid classifier-driven FL framework that consists of an MABC-SVM classifier at the client end of health service providers. The MABC works for feature selection, inspired by the intelligent foraging behavior of honeybee colonies, to find relevant patterns iteratively across the different client's data and the addition of SVM helps to classify heart diseases. Dipro et al. [11] focused on decentralized servers to integrate the FL framework for Parkinson's disease detection. Three distinct CNN models have been considered to test and evaluate the performance and among them, the VGG19 model outperformed with 97.3% accuracy. Roth et al. [12] have explored the practical application of FL involving seven clinical institutions across the world. Despite the presence of significant changes in the dataset and without centralizing the information, effectively trained CNN models within the Federated framework. This approach has earned on average 6.3% better results than other equivalents trained on an institute's local data solely.

Holistic exploration in focusing medical domains like cancer prediction and breast cancer classification has appeared through the implementation of FL methods [13], [14]. Jiménez-Sánchez et al. [13] has implemented a method for training a global model using FL while protecting the privacy of individual participants' data. The method uses a deferentially private stochastic gradient descent algorithm and allows clients to select their local best model based on their validation metric. Beguier et al. [14] experimented with a different method that employs a memory-aware curriculum to update a global model. The method enables clients to learn from their data and, it uses a memory buffer to store samples from the global model that enhances learning. Brisimi et al. [15] propose an FL approach for predictive modeling from electronic health records (EHRs). The experiment shows that this approach can improve the performance of machine learning models for predictive modeling from EHRs while preserving the privacy of individual patients. Additionally, it can reduce the communication and computational costs associated with FL. Ishraq R. Rahman et al. in [16] employs gene expression studies and the CuMiDa dataset to enhance cancer categorization by identifying the most suitable classifier model, focusing on lung and bladder cancer datasets while optimizing accuracy through parameter adjustments.

III. Methodology

A. Data Collection and Pre-processing

The dataset used in this study collected from Borkowski et al. [17], consists of 25,000 microscopy images of lung and colon tissue samples that have been labeled as either cancerous or benign. These images were accumulated from multiple institutions ensuring patient privacy and informed consent. The variation in size and resolution make it a well-curated and challenging dataset for the advancement of cancer detection algorithms.



(a) Lung SCC (b) Benign Lung (c) Adenocarcinoma

Fig. 2: Sample of Lung Cancer Classes.

From this dataset, we have chosen only lung images and worked on around 3600 images due to resource constraints. The entire dataset is divided into three classes including Lung SCC, Lung Adenocarcinoma, and Benign Lung Tumor (Fig. 2). We manually create 4 clients where each client contains three different lung classes and, around 300 images in each class, and each client has 900 lung data. The images are available in JPEG format and all the images were 780x780 pixels. Image Data Augmentation was applied to pre-process the images and resize them to 128x128x3 formation, The following three channels indicate the RGB image usage for training the CNN models.

B. Proposed Framework

The proposed work employs customized and finely tuned convolutional neural network (CNN) models to extract and comprehend features from lung cancer images effectively. Transfer learning is approached to train on clients (edge devices) model. In this process, the last three layers of each model have been modified to adapt our target domain. The integrated FL layers consist of one flattened layer and two dense layers. The initial dense layer employs ReLU as an activation function and in the final layer, the softmax activation function is applied to facilitate the detection of our multiclass lung cancers. Fig. 4 shows the mentioned structure. The proposed work trains the lung dataset on five distinct deep CNN models.

During the implementation of FL, the fundamental approach has been followed that commences with the distribution of initial model weights by the coordinators to individual clients to inaugurate the model training operation. The initial weights were derived from its own modified CNN model weight. We have addressed the initial weight as global weight and the model's weight as local weight. Throughout each communication round, the recently generated model parameter of each client has been gathered and transmitted back to the coordinator to aggregate those different weights and formulate a new updated global weight by averaging those model parameters. This refined global weight is once again disseminated to each client to enhance the model's robustness and experience to effectively capture the relevant features and complex patterns of lung cancer.

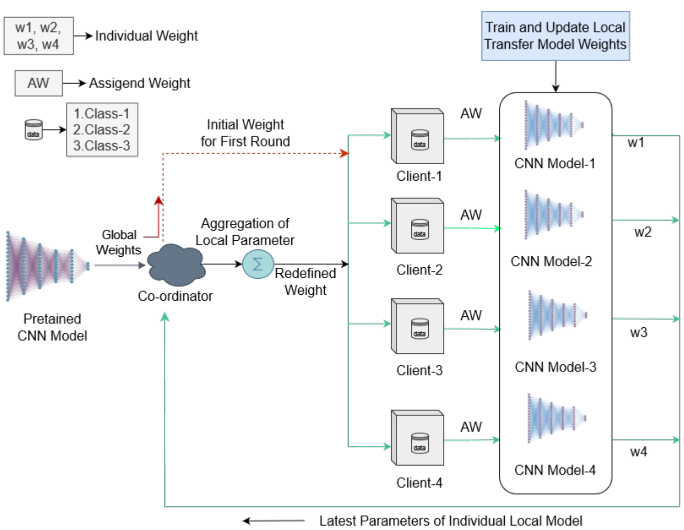


Fig. 3: Applied Federated Learning Flow

In our study, we meticulously executed around a hundred communication rounds, and based on the dataset each client model has undergone approximately thirty epochs each time. Each model of individual clients has been tracked during the end of each communication round along with their performance accuracy, validation loss, and other matrices. Fig. 3 shows the visual representation of our applied FL flow and algorithm 1 provides an idea of our proposed framework. There, w1,w2,w3, and w4 represent the individual local weights of local models and AW stands for the assigned weight from the coordinator to each client. The operation has been split into two segments for weight scaling and aggregating to reduce the computation time.

Let c as the of clients' uses and the sample available from that client is n_c.So, n = $\sum_{c=1}^{C} n_c$ is the total sample size. For scaling the weight in each iteration of the c client,

$$W = [w_1, w_2, \ldots, w_n]$$

W, the list of the original weights of each client k.The scaled final weight list,

$$W_{\text{final}} = [w_{\text{final},1}, w_{\text{final},2}, \ldots, w_{\text{final},n}]$$

Now to scale each client's latest weight,

$$w_{\text{final},i} = \frac{n_c}{n} \times w[i] \tag{1}$$

Here, i ranges from 1 to no of weight available for c client. After that, we summed all the weights of different clients,

$$W_{\text{global}} = \sum_{c=1}^{C} W_{\text{final}} \tag{2}$$

Algorithm 1 Pseudocode for Proposed Framework

Input: total communication rounds, number of clients, global model weights
Output: updated global model weights, performance metrics

Set initial weights using Transfer Learning
Communication Rounds :
for each communication in total communications **do**
  Get the current global model weights
  Initialize list to store scaled local weights
  Client Updates :
  for each client in the total number of clients **do**
    Set the local model's parameters to global weights
    Calculate steps per epoch for training and validation
    Train local model for one epoch using client's data
    Scale the local model's weights and store them
  end for
  Compute the average of scaled local weights to update the global model
  Assess the global model's accuracy on test data
end for

C. Fine Tuning Parameters

The proposed training configuration uses a batch size of 24, which indicates the number of data samples processed in each training iteration. With 720 training images for each client, the entire dataset is iterated through in batches, and this process repeats 30 times per client in each communication. Here communication round means after training the model locally we send the updated weight to the client server and aggregate them. Validation, performed over 180 images per client with 7 steps after each epoch, assesses model performance while updating weights. The learning rate, set at 0.00001, controls the size of weight adjustments during optimization. Over 100 communication rounds, the FL process involves interactions between the central server and clients, with optimization driven by Stochastic Gradient Descent (SGD) as the chosen optimizer. After that send the updated weight to the local client again. These parameters shape the training model, influencing the duration, intensity, and optimization strategy of the machine learning model. The parameters Setup is shown in table I.

TABLE I: Parameters Setup for Proposed Model

Parameters	Value
Batch Size	24
Training Data for Each Client	720
Steps per Epoch	30
Validation Steps	7
Learning Rate	0.00001
Communication Round	100
Optimizer	Stochastic Gradient Descent

D. *Model Specification*

Due to the diverse attributes, and complex patterns present in the lung dataset, Transfer Learning has been employed in our study. The following model including VGG16, MobileNetV2, Xception, EfficientNetB4, and ResNet15 have been used which is trained on the ImageNet dataset. ImageNet [18] is a large dataset that contains over 14 million images organized into over 20,000 categories, and is often used to train and evaluate machine learning models for image classification, object detection, and other vision tasks. Deep learning models have played a crucial part in the progress and enhancement of several fields. Moreover, they have been widely employed as a standard for evaluating and comparing the efficacy of different models.

1) Transfer Learning: Transfer learning is a technique that exploits the knowledge earned from previous tasks to flourish the generalization of others [19]. It is an optimization that enables rapid progress or improved performance and is related to involving in multi-task learning. In transfer learning, like other neural networks, early layers try to detect edges and middle layers are for shape-capturing, and mainly latter layers are customized and retrained with the target domain to leverage the labeled data of the task it was initially trained on. Fig. 4 gives us an idea of how transfer learning works along with the information of the used model in this research work.

2) VGG16: VGG16 is a convolutional neural network (CNN) model that mostly participated in object detection and classification due to its uniform architecture and easy-to-achieve transfer learning behavior [19]. The model is trained on the ImageNet dataset to update the humongous parameters for better performance. The VGG16 architecture consists of a total of 21 layers, which include 3 dense layers, 13 convolutional layers, and 5 max-pooling layers. The model has often been used as a benchmark for evaluating the performance of other image classification models.

3) MobileNetV2: MobileNetV2 is an efficient and lightweight CNN model and the main components of this architecture are inverted residual blocks and linear bottlenecks [20]. These blocks comprise pointwise convolution(1x1

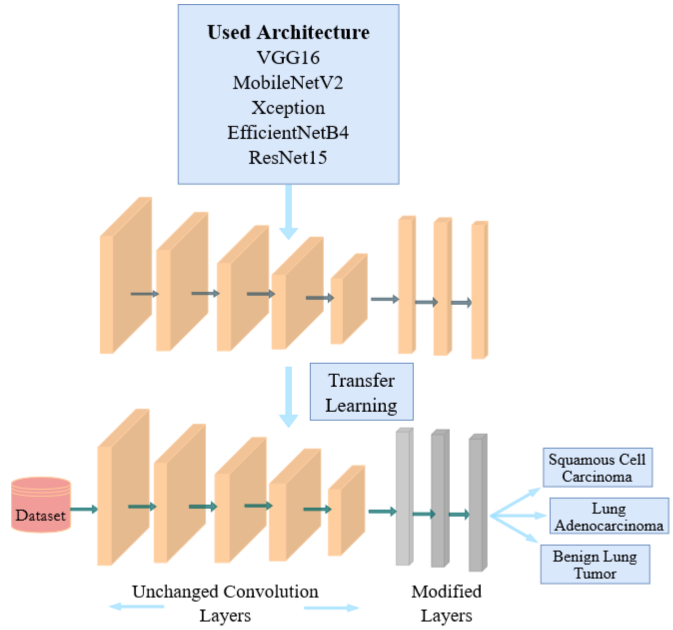


Fig. 4: Modifying CNN Architecture for Applications

filter), depthwise separable convolution(3x3 filter), and another pointwise convolution layer sequentially. The unique formation facilitates to reduction of computational costs while maintaining accuracy and introduces two distinct hyperparameters cited as width multiplier and resolution multiplier to reduce noise and adjust channels.

4) Xception: Xception is an advanced version of the inceptionv3 model with modified depthwise separable convolution [21]. The main objective of Xception is to enhance the efficiency and efficacy of the Inception model by reversing its workflow. To achieve this, the architecture breaks down standard convolutions into two distinct operations: depthwise convolutions and pointwise convolutions.

5) EfficientNetB4: EfficientNetB4 belongs to the EfficientNet series and aims to enhance accuracy by increasing the size of the model while preserving efficiency [22]. The architecture is based on a scaling method that adjusts the depth, width, and resolution of the network consistently. The depth scaling involves including more layers in the network architecture, which improves the model's ability to represent information.

6) ResNet152: ResNet152 stands out due to its unique approach of using residual blocks that incorporate shortcut connections, also known as skip connections, to bypass particular layers [23]. These bypass connections help the network to learn residual functions rather than directly learning the desired mapping. As a result, deep ResNet models like ResNet152 can keep the gradients from disappearing during training by passing on the residual error.

IV. Results and Discussion

Table II encapsulates the performance evaluation of several deep learning models across a specific task. Each model's accuracy, F1 score, precision, recall, AUC, and loss are

TABLE II: Performance Matrices

Models	Accuracy (%)	F1 Score (%)	Precision (%)	Recall (%)	AUC (%)	Loss
VGG16	91.03	90.88	91.23	89.44	98.55	0.315
MobileNetV2	82.77	82.84	83.25	81.77	94.71	0.469
Xception	81.33	87.79	81.52	80.88	94.86	0.419
EfficientNetB4	77.99	77.82	78.03	77.77	91.71	0.674
ResNet152	78.88	81.08	80.44	76.33	93.50	0.478

systematically detailed, shedding light on their respective capabilities. VGG16 emerges as the leader in accuracy, boasting an impressive 91.00%, showcasing its adeptness at making correct predictions. Close behind, MobileNetV2 demonstrates a commendable 82.77% accuracy, while Xception and other models follow suit with their distinct performances. Looking beyond accuracy, the F1 score serves as a key metric for the balance between precision and recall. VGG16 and MobileNetV2 shine with high F1 scores of 90.88% and 82.84%, respectively, indicating a balance between these two critical factors. The models' performance varies in terms of precision and recall, with Xception placing more emphasis on precision which is 81.52%, and EfficientNetB4 showing a balanced trade-off. Moreover, the AUC metric highlights the discriminative power of the models. VGG16 and Xception excel with AUC values of 98.55% and 94.86%, underlining their exceptional ability to distinguish between classes. Considering the loss metric, VGG16 emerges as the frontrunner with the lowest value of 0.315, indicating efficient convergence during training. Conversely, EfficientNetB4 records the highest loss (0.674), suggesting potential room for optimization. Overall, the presented results reveal the strengths and trade-offs of each model, enabling a comprehensive assessment for selecting the most suitable model based on the requirements of the application.

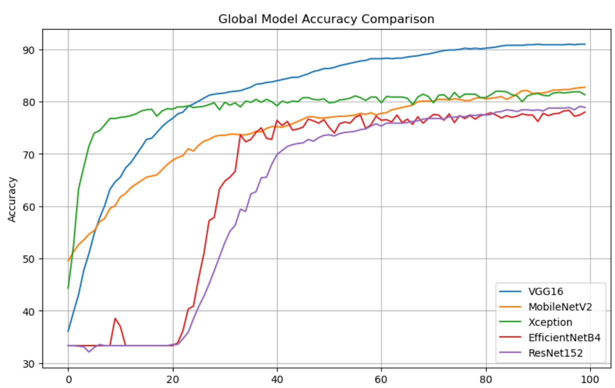


Fig. 5: Global Model Accuracy Comparison Graph.

Fig. 5 represents accuracy evolution across the communication rounds of diverse CNN models. Initially, the majority of models exhibit substantial improvement over the communication round except for efficientnetB4 and Resnet152, which lag up to a particular round. Approximately after forty rounds, CNN models demonstrated a tendency to reach saturation point and displayed moderate growth thereafter. Remarkably, VGG16 maintains consistent improvement throughout

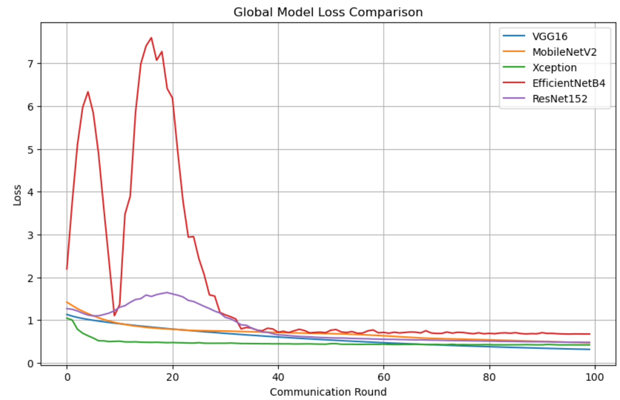


Fig. 6: Global Model Loss Comparison Graph.

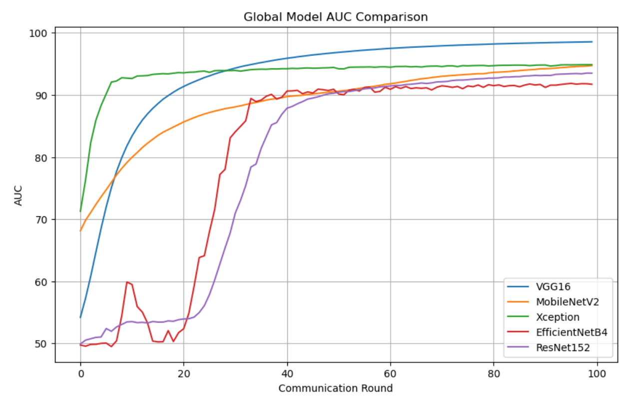


Fig. 7: Global Model AUC Comparison Graph.

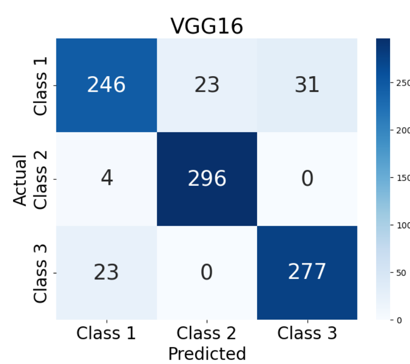


Fig. 8: Confusion Matrix of VGG16

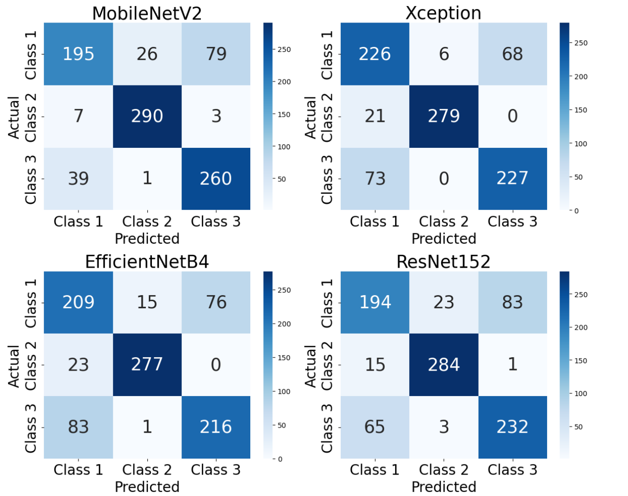


Fig. 9: Confusion Matrix of MobileNetV2, Xception, EfficientNetB4, ResNet152

the communication process. Fig. 6 depicts that, all models show substantial loss reduction with effective convergence over communication periods, except EfficientB4. Fig 7, VGG-16, MobilenetV2, and Xception consistently display effective convergence with significant AUC improvement.EfficientnetB4 and Resnet152 initially step behind but catch up to other models to effectively distinguish lung instances. Fig. 8 and 9 show the confusion matrix of the achieved model.

V. Conclusion

In this paper, we have tried to effectively detect different lung cancers by leveraging decentralized data sources while preserving individual data ownership and privacy. Five distinct CNN models with the help of transfer learning have been conducted to achieve superior performance. The results presented in this research paper align with our ideology and demonstrate precise accuracy in lung cancer detection. Nevertheless, it is essential to recognize the fundamental challenges of FL execution, such as communication overhead, model aggregation complication, and potential bias by distinct data dispersion. FL demonstrates a promising scope in the dynamic sector of healthcare in the future. The instinctive nature of distributed and privacy concerns enables multidimensional clinical data integration and in-depth image analysis. Moreover, for chronic diseases, it could be applied as continuous monitoring of the different patients and appear as a personalized treatment planner.

References

[1] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," *arXiv e-prints*, pp. arXiv–1602, 2016.

[2] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.

[3] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020.

[4] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, "Federated learning and differential privacy for medical image analysis," *Scientific Reports, Nature Machine Intelligence*, vol. 12, no. 1, p. 1953, 2022.

[5] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.

[6] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[7] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021.

[8] L. Zhao, Q. Wang, Q. Zou, Y. Zhang, and Y. Chen, "Privacy-preserving collaborative deep learning with unreliable participants," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1486–1500, 2019.

[9] J. Lee, J. Sun, F. Wang, S. Wang, C.-H. Jun, X. Jiang *et al.*, "Privacy-preserving patient similarity learning in a federated environment: development and analysis," *JMIR medical informatics*, vol. 6, no. 2, p. e7744, 2018.

[10] M. M. Yaqoob, M. Nazir, M. A. Khan, S. Qureshi, and A. Al-Rasheed, "Hybrid classifier-based federated learning in health service providers for cardiovascular disease prediction," *Applied Sciences*, vol. 13, no. 3, p. 1911, 2023.

[11] S. H. Dipro, M. Islam, A. Al Nahian, M. S. Azad, A. Chakrabarty, and T. Reza, "A federated learning based privacy preserving approach for detecting parkinson's disease using deep learning," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 139–144.

[12] H. R. Roth, K. Chang, P. Singh, N. Neumark, W. Li, V. Gupta, S. Gupta, L. Qu, A. Ihsani, B. C. Bizzo *et al.*, "Federated learning for breast density classification: A real-world implementation," in *Domain adaptation and representation transfer, and distributed and collaborative learning*. Springer, 2020, pp. 181–191.

[13] A. Jiménez-Sánchez, M. Tardy, M. A. G. Ballester, D. Mateus, and G. Piella, "Memory-aware curriculum federated learning for breast cancer classification," *arXiv preprint arXiv:2107.02504*, 2021.

[14] C. Beguier, J. O. d. Terrail, I. Meah, M. Andreux, and E. W. Tramel, "Differentially private federated learning for cancer prediction," *arXiv preprint arXiv:2101.02997*, 2021.

[15] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," *International journal of medical informatics*, vol. 112, pp. 59–67, 2018.

[16] I. R. Rahman, S. B. Soumma, and F. B. Ashraf, "Machine learning approaches to metastasis bladder and secondary pulmonary cancer classification using gene expression data," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 430–435.

[17] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (lc25000)," *arXiv preprint arXiv:1912.12142*, 2019.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[19] S. K. Rahut, R. Sharmin, and R. Tabassum, "Bengali abusive speech classification: A transfer learning approach using vgg-16," in *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*. IEEE, 2020, pp. 1–6.

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[22] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.