An Evaluation of Federated Learning in Training COVID-19 Patient Risk Prediction using Deep Learning Model

Yushen (Ido) Chen
Department of Computer Science
University of Western Ontario
Ontario, Canada
yche2692@uwo.ca

Zifan Zhu
Department of Computer Science
University of Western Ontario
Ontario, Canada
zzhu459@uwo.ca

Min Zhang
Department of Computer Science
University of Western Ontario
Ontario, Canada
mzhan782@uwo.ca

***Abstract*—The COVID-19 pandemic has created a challenge for our healthcare system. Therefore, We have to manage and protect patient's data privacy. Federated Learning is a technique where the training data can be kept at the data source without uploading it to a central server. Due to this characteristic, we used techniques like correlation checks and PCA dimensionality reduction and tried two Federated Learning algorithms – Federated Averaging and Federated Proximal in our project, aiming to determine the most effective way to predict a patient's risk. However, we also need to keep the potential risks and drawbacks of this framework in mind. For example, if a client is infected with a virus, an attacker may upload the wrong model parameters, causing the shared model to be negatively impacted. Plus, there might be differences in accuracy when comparing it to the traditional centralized training approach. In our research, we divided our data into independent and identically distributed data and non-independent and identically distributed data to simulate real-world scenarios. We compared the performance of Federated Averaging and Federated Proximal. Our result shows that the Federated Learning Model can be used in the medical domain with privacy protection, achieving good accuracy. Future work could focus on exploring different models and the refinement of Federated Proximal.**

***Index Terms*—Federated Learning, COVID-19 datasets, privacy**

I. Introduction

After machine learning first came out in 1959, it has developed dramatically, and now it can achieve high accuracy in almost every aspect of our life when given enough data to train. Datasets play an extremely important role in machine learning, as a good dataset can effectively increase the accuracy and the efficiency of a machine learning model. More and more data are being collected thanks to many data-collecting organizations. Nevertheless, in some specific areas, data cannot be easily collected compared to other types. Take the medical field for example, the information can be very personal, and people may not be willing to share such kind of information to others. To them, privacy is more important than upgrading machine learning models. Additionally, as for other kinds of confidential data, even after the data is collected, it is not safe to transfer it to the server.

Considering these factors, Federated Learning (FL) came out and can effectively solve these problems. FL does not ask individuals to provide their personal data. Instead, the training process takes place directly on the individuals' devices, which effectively protects people's privacy. Consequently, the parameters after the training process are sent back to the server to update the model. In this way, the server does not have any access to the data, successfully protecting the privacy of the individuals and the safety of the data.

Our experiment used FL to predict the severity and the mortality risk of COVID-19 patients. We divided the data to simulate independent and identically distributed (iid) data and non-independent and identically distributed (non-iid) data. Also, we performed some feature engineering on the data to make it easier to deal with. We measured the accuracy between centralized learning method and FL. Additionally, we compared the accuracy between Federated Averaging (FedAvg) and Federated Proximal (FedProx) on iid data and non-iid data.

The result of our experiment showed that the FL had almost the same accuracy as centralized learning. And FedProx had better performance than FedAvg when dealing with non-iid data. As for iid data, the difference was very tiny.

Our contribution is to utilize FL on the COVID-19 patient's dataset to predict the severity and mortality risk of COVID-19 patient. Also, we take the device heterogeneity into consideration and use Federated Proximal to address this problem. Moreover, to enhance privacy, we implement encryption on model updates so that the parameters of the model are also safe and will not be negatively affected by any intended actions.

The rest of this paper is organized as follows: Section II describes the background and related work of our paper. Section III provides the methods of our model. The proposed approach and the data preparation are introduced in detail. Section IV presents the results of our research. The performance between iid data and non-iid data is compared. Other comparisons, such as the global model and client model, are also available. Section V is our conclusion and discusses future work that can be done.

II. Background and Related Work

We briefly reviewed studies related to COVID-19 identification, the diagnosis of COVID-19, and its severity based on COVID-19-related data. We also reviewed current studies using federated learning in the medical area and optimization algorithms, including FedAVG and FedProx.

A. COVID-19 related AI studies

Since the 2019 COVID-19 pandemic has brought a lot of risks, including asymptomatic cases, severe illness, and even death, many AI models, especially those using deep learning, have been used to predict patient risk levels and mortality based on different types of data. The study by Kollias et al. (2023) [1] analyzed the detection method of COVID-19. It discusses the severity detection based on a dataset of 3-D CT scan images using baseline approaches, which are deep neural networks based on a common CNN-RNN architecture. This study mainly describes the 3rd COVID-19 competition, where many teams presented their results. A Comparison of these results shows that there are better solutions for the COVID-19 detection challenge, but the best F1 Score for the COVID-19 severity detection challenge is only 73.04%. There is another study by Abdelwhab et al. (2021) [2] that proposed a new deep learning model for COVID-19 identification also using CT scan images. They use two similar CNN to extract features from CT images using the first CNN. Consequently, they extract features from X-ray images using the second CNN. The accuracy of the detection of COVID-19 using their suggested model is 99%.

B. Secure data privacy

Through some studies, it is not hard to see the performance advantage of deep learning networks in COVID-19 identification and severity classification. However, since most of the data needed in this area are real patient information, such as their CT scan images, and this data often requires privacy protection, a new challenge emerges. We now not only need to design a high performance model but also need a model with privacy protection. Majeed et al. (2021) [3] address the challenges and solutions related to protecting patient data privacy. They mention that information privacy refers to securely collecting, storing, analyzing, managing, and publishing personal data. Nevertheless, these are needed to train a centralized deep learning model. It is not possible to achieve the aforementioned privacy requirement using the traditional centralized learning model. Therefore, they also analyze various privacy protection techniques developed during the COVID-19 pandemic period. These techniques include federated learning, blockchain, swarm learning, etc. In the study, federated learning gives good performance as well as provides great privacy protection. Thus, we would like to focus on the federated learning technique and attempt to utilize it into our research.

C. Federated Learning in medical area

Essentially, FL is a machine learning technique that allows decentralized training to protect data privacy. Unlike centralized learning, FL does not require the data to be sent to the central server. It only requires sending the training parameters to the server. Due to this characteristic, FL protects the privacy of the data. Korkmaz et al.(2022) [4] mainly discuss the performance of federated learning technique, with a focus on privacy and security, on medical datasets. The study analyzes the performance of Inception-v3, ResNet50, EfficientNetBo and other deep learning models on the diabetic retinopathy detection dataset, chest x-ray (COVID, pneumonia) dataset, etc., using federated learning technique. Optimization strategies including FedAvg, FedAdam, etc., are also mentioned in the study. These strategies are keys to federated learning's effectiveness in handling decentralized data and protecting data privacy, especially in sensitive domains like healthcare. In our research, we choose FedAvg and FedProx as the FL algorithm to evaluate the performance of our model and compare the performance of these two algorithms, which will be introduced in the following subsection.

D. Federated Learning optimization algorithms

McMahan et al. (2016) [5] proposed the FedAvg algorithm to effectively train deep learning networks based on distributed data on mobile devices. It shows that with FedAvg algorithm, FL exhibits robustness and efficiency with fewer communication rounds. Basically, each participant (e.g., mobile device) trains the model locally and sends the update to a centralized server. The central server averages these updates to represent the global model then sends it back to those participants for further training. FedAvg is good at handling iid data. However, we need to deal with non-iid data in many scenarios such as data in different countries, regions, and clinics. The data might have different sizes or even features. In these cases, using FedAvg to integrate all updates evenly into the global model might affect the performance and accuracy of the model to some extent. So, Li et al. (2018) [6] proposed FedProx framework that can effectively address the challenge of heterogeneity. This algorithm introduces an extra proximal term during the optimization step, acting as a regularizer to limit the deviation of local model updates from the global model. Therefore, it helps to maintain the stability and validity of the model when the data is non-iid or the client performance is uneven. The algorithm of FedProx can be represented using the following formula:

$$w^k = \arg\min_w (\sum_{j=1}^{n} (f_i(w) + \frac{\mu}{2} ||w - w_i^{k-1}||_2^2)) \qquad (1)$$

In the formula, we have a μ, which denotes a hyperparameter controlling the strength of the proximal term. We use this parameter to obtain a simplified FedProx, which we'll introduce in section III.

III. Methods

In this section, we discuss the method of our research, including the research objectives and research methodology. For the research methodology, we divide it into four parts: the proposed approach, dataset and preparation, experiment implementation, and the simplified FedProx.

A. Research Objective

We have two research objectives to accomplish: (a) using FL in the medical field to provide privacy to patients, and (b) integrating some encryption methods to enhance the safety of the transfer process.

For objective (a), on the one hand, it can provide privacy to the patients, and on the other hand, it can make more people willing to share their information and thereby more data can be collected. This can significantly alleviate the shortage of these kinds of data and build a more accurate model to diagnose diseases. Additionally, more usefully data can contribute significantly to the research field. And it is also not hard to apply in practice. For example, an application can collect the patient information locally, train the local model every day and send the parameters to the server every day. In this way, FL is easily applied in practice.

For objective (b), it takes privacy protection to a new level. Now, not only is the data stored locally in the patients' devices, but also the training parameters are protected. We use the encryption methods to process the data transferring. In this way, our model is suitable not only in the medical scenarios but also capable of dealing with some confidential scenarios.

By combining objective (a) and objective (b), our model is highly capable of protecting the patient's data without lowering the accuracy.

B. Proposed Approach

The framework of our approach can be described in Fig. 1. There are three sections in total.

The first section is the clinic patient information section, which would be stored in devices at clinics. This information can be collected over a certain period and is not accessible by any third party. When the local data is ready, it starts to be trained in the local model, which is our second section. Each clinic represents a client that trains one local model. During this training process, the local training will be performed and the model will generate new parameters.

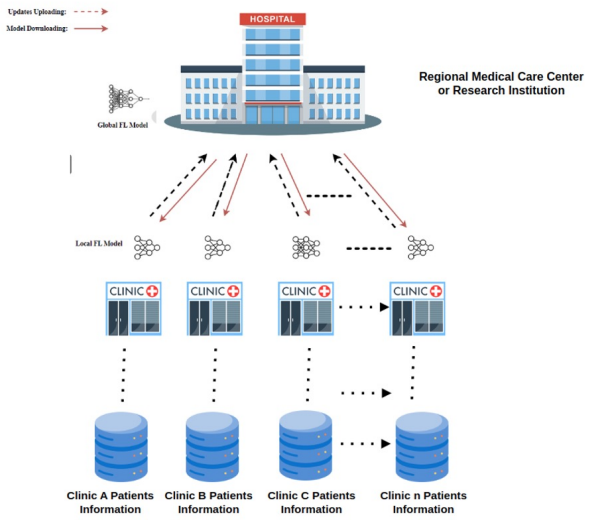

Fig. 1. Federated Learning in Medical Area

Then, the updated parameters are sent to the global FL model section. The global FL model can utilize all the local training updates to form a comprehensive one. Consequently, the global FL model is sent to all the local FL models and updates the parameters. In reality, each client can be a patient, clinic, hospital, region, or even a country. These clients are supposed to be able to train local models concurrently in each round and send the generated parameters to the central server for integration in order of training speed.

In our simulation code, we simply use a FOR loop to create multiple clients and train each of them sequentially instead of many clients at the same time in reality. The reason for this is that our goal is to make comparisons, analyze performance, and validate feasibility rather than create a system that is available for use. However, we also performed tests where multiple clients communicate with a central server using TCP.

But this approach also introduces another data privacy issue: the possibility of receiving third-party attacks during model transmission. Transmission encryption becomes an additional safeguard to protect data privacy. We use an approach that combines symmetric key encryption (AES) algorithm and asymmetric encryption (RSA) algorithm to ensure privacy protection.

RSA is utilized to securely exchange symmetric keys between the client and the central server, and then the shared key is used to encrypt and decrypt the data sent over the TCP connection (i.e., model updates). This approach ensures that the data transmission has strong security and that the shared key transmission is secure.

Specifically in our framework, after exchanging the shared key between the client and the server, the updated data containing the updated model, sensitive information, etc., are encrypted using AES encryption and transmitted to the server. The server then decrypts them using the shared key, integrates them, and encrypts and transmits them back to the individual

clients to complete the subsequent training.

C. Dataset and Preparation

We chose a COVID-19 dataset from Kaggle, which consists of 21 unique features and 1,048,576 COVID-19 patient's data, for our research. This database has 10 usability marks, indicating that it is very easy to use. This dataset provides information such as patients' symptoms, status and medical history. At the initial state, the dataset has many features and contains a great amount of missing data. Therefore, we need to perform some feature engineering first.

For the missing data, we modified the value based on certain regulations. For example, if the pregnancy status for males is missing, we could easily change it to "not pregnant". After addressing the missing data, we attempted to represent the value of every feature as true or false. For other features that cannot be represented as true or false, we used a standard scaler to adjust the values. Then, we did the correlation checks and PCA dimensionality reduction in order to compress the data. After this process, we obtained thirty features to predict the severity of COVID-19 patients and thirty-seven features to predict the mortality risk of COVID-19 patients using iid data. As for none-iid data, we had nineteen features to predict the severity and twenty-six features to predict the mortality.

After the feature engineering, we start to partition the data. In this step, we tried to create iid data and non-iid data by modifying the current whole dataset manually. As for the iid data, we simply shuffle the data and equally divide it into ten parts. Because iid data requires being independent and identically distributed. And the shuffle process is good enough to achieve it. Regarding the non-iid data, we use the medical unit feature to partition the dataset into thirteen different parts. The medical unit feature represents the thirteen types of institutions of the National Health System that provide health care. In this way, we can simulate the data as if it were a real-world circumstance. Essentially, we are modelling different medical unit data as non-iid data.

In the end, we aimed to predict two different objectives: (a) the severity of the COVID-19 patient, and (b) the mortality risk of COVID-19 patients. For objective (a), we had 4 labels, which originally had 7 labels. We decreased the number of the labels because some types of severity were too close to identify. And for objective (b), we had 2 labels, which were alive or dead.

D. Experimental Implementation

Since the dataset is from Kaggle, we were able to refer to what others have done with the data. Combined with our own testing of the data using some classification trees and models, we found that most other people and we had poorer predictions of severity (around 40%-60%). We can, therefore, infer that the data will not perform well in predicting severity. The accuracy increased significantly after we made the prediction result predicting patient mortality. Based on this information, we compare the performance of the FedAvg and FedProx algorithms for the severity prediction part more with the performance of the iid data and non-iid data. Specifically, we use FedAvg and FedProx algorithms to train and test on the iid and non-iid datasets, respectively. As for predicting mortality, we then prefer to test the overall performance of FL when dealing with non-iid data. In addition, we also performed centralized learning on the data to compare the results with those of FL to verify the feasibility of FL application in this domain.

E. Simplified FedProx

In our model, we used a simplified FedProx algorithm instead of a usual one to make a clearer comparison between FedAvg and FedProx. We have 13 sets of non-iid data divided by using the medical unit feature in total, and we use one same μ for all these 13 sets in one experiment. This can help to make the comparison as there are no other differences in each of the different sets. In order to make the experiment more comprehensive, we use 9 different μ to conduct more experiments, repeating the aforementioned process.

IV. Experimental Results

A. FL Experimental Setup

In order to examine the performance of FL in processing COVID-related data using FedAvg and FedProx respectively, we used these algorithms for training iterations for iid data and non-iid data. The experimental parameters for processing iid data are configured as follows: 10 clients for 10 rounds of training, with 5 epochs per round, and a list for comparison of the μ parameter of FedProx $[0.01, 0.04, 0.08, 0.1, 0.15, 0.2, 0.25, 0.3, 0.5]$. The number of clients processing non-iid data is adjusted to 13 (because simulated non-iid data is 13 copies), and other parameters remain the same.

B. Severity Prediction Analysis

In this section, we will focus on comparing the performance of FedAvg and FedProx in processing iid data and non-iid data.

1) IID Scenario Analysis: By comparing Fig. 2 and Fig. 3, it is easy to see that the clients using FedAvg and FedProx perform roughly similarly in handling iid data. This is actually because the simplified version of FedProx we used does not yet reflect the shortcomings in processing iid datasets. In fact, the performance of FedProx in handling iid data is even worse than that of FedAvg due to the influence of different client's proximal parameters. At the same time, Fig. 4 also tells us that there is not much difference between the two algorithms when processing iid datasets and making patient severity predictions.

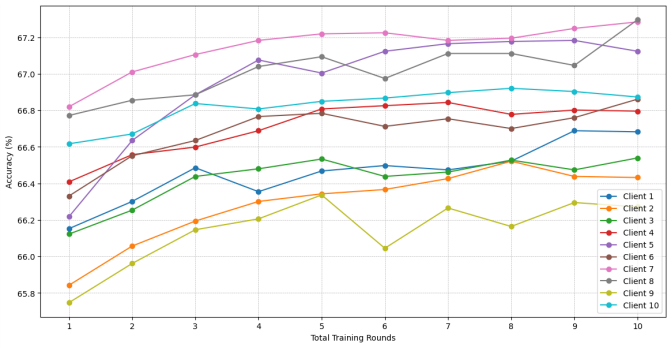

Fig. 2. Client performance with iid data using FedAvg

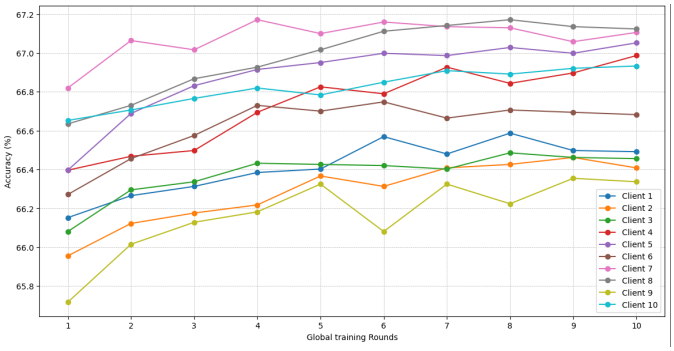

Fig. 3. Client performance with iid data using FedProx (μ = 0.01)

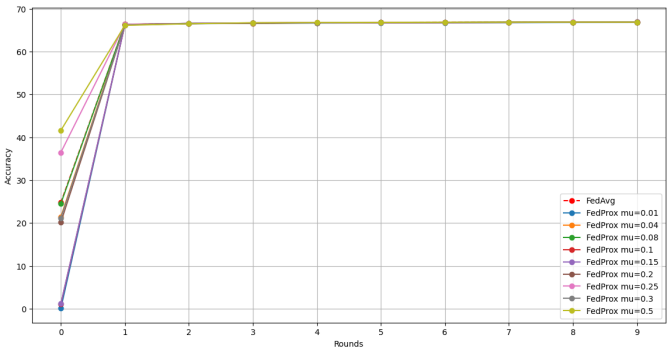

Fig. 4. Global model performance with iid data

2) Non-IID Scenario Analysis: Now we compare the two algorithms using non-iid data. Fig. 5 and Fig. 6 still show that there is not a big difference between clients using FedAvg and FedProx for most of the medical unit. However, if we observe carefully, we find that FedProx can converge slightly faster. And the result using FedProx appears to be more stable. For the worst case, which is the blue curve, FedProx performs much better than FedAvg. Using FedProx, this case has more than 70% accuracy after fourth epoch while FedAvg only has 12% at the same time. In addition, the result of this worst-case scenario is assumed to be caused by the small size of that dataset. Focusing on Fig. 7, in this comparison of global model performance, it is easy to see that this figure illustrates that FedAvg does not perform as well as FedProx when dealing with non-iid datasets. Therefore, FedProx becomes a better solution to deal with non-iid datasets. As for the accuracy, since we created the simulated non-iid data manually and the dataset itself does not perform well for severity prediction, the accuracy stays at a low level.

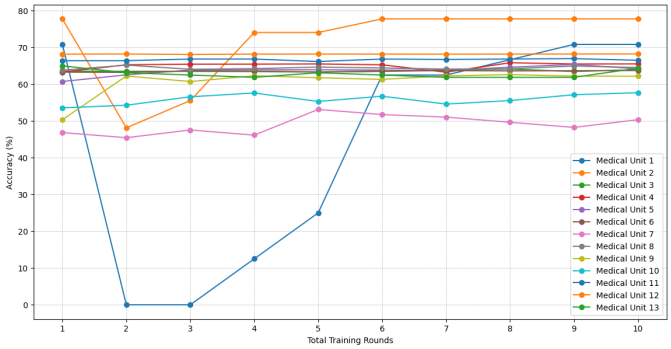

Fig. 5. Client performance with non-iid data using FedAvg

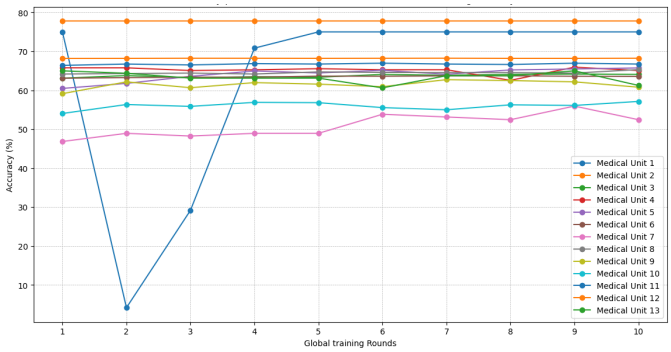

Fig. 6. Client performance with non-iid data using FedProx (μ = 0.25)

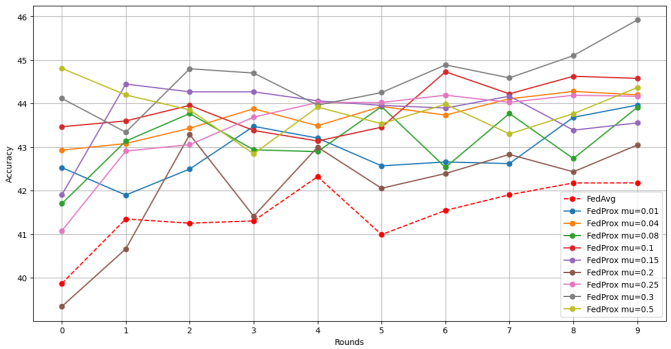

Fig. 7. Global model performance with iid data

C. Mortality Prediction Analysis with Non-IID data

Now we know that FedProx is good with non-IID data. We will show the results of using non-IID data and the FedProx algorithm to predict patient mortality.

Fig. 8 shows how the client performs in the training rounds using FedProx with $\mu = 0.5$. Fig. 9 illustrates that there is not

much difference in the performance of FedAvg and FedProx in predicting patient mortality even with non-iid datasets. This is because mortality prediction is actually a simple binary classification problem, so they both perform well. However, we can still see that some global models using FedProx are slightly better than FedAvg.

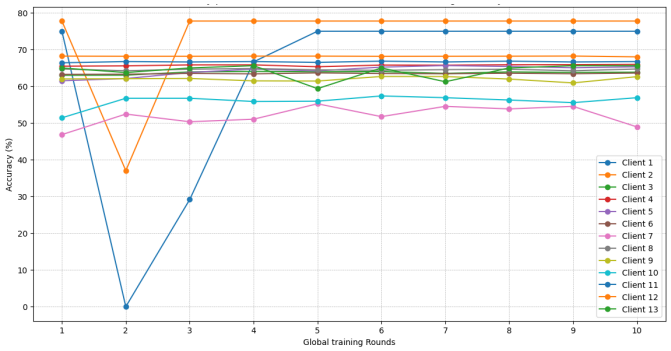

Fig. 8. Client performance with non-iid data using FedProx (μ = 0.5)

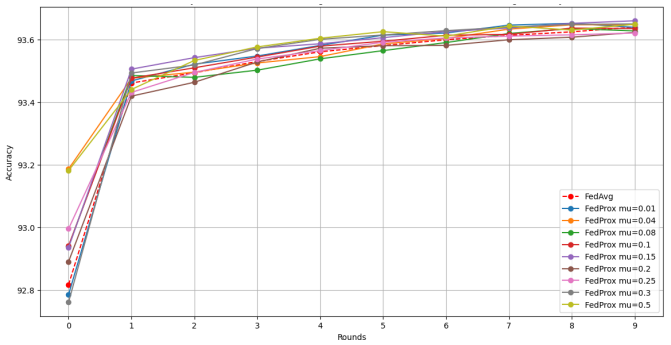

Fig. 9. Global model performance with non-iid data

D. Model Accuracy Comparison

Fig. 10 and Fig. 11 illustrate the performance of the centralized learning model and the FL model in predicting COVID-19 patient mortality risk and severity. The FL model performs better than the centralized learning model in predicting COVID-19 patient mortality. However, the federated learning model performs worse than the centralized learning model in predicting the severity of COVID-19 patients. In most cases in real life, decentralized training adversely affects performance. However, we can see that the performance is still excellent in the mortality prediction case.

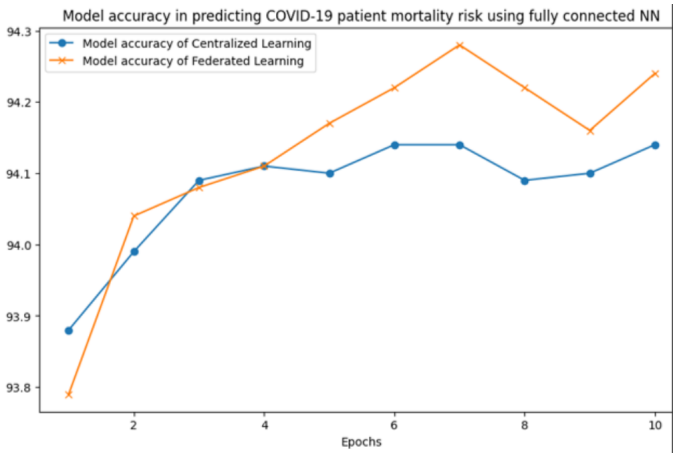

Fig. 10. Model accuracy comparison between Centralized Learning and Federated Learning in predicting mortality

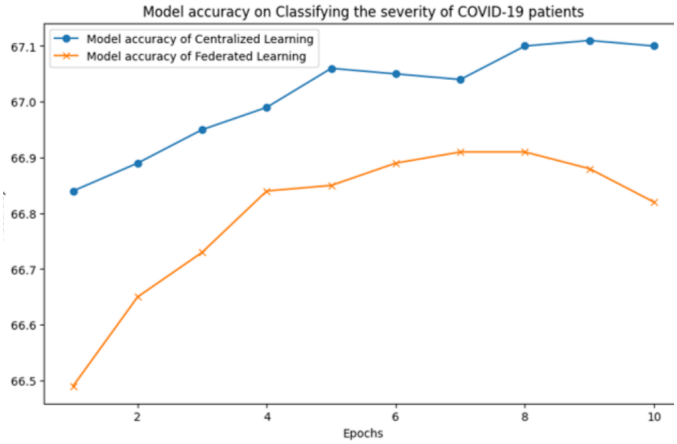

Fig. 11. Model accuracy comparison between Centralized Learning and Federated Learning in predicting severity

Therefore, we can use FL for prediction in the medical domain as long as we have the right data, or use the appropriate optimization algorithms, and strategies to adjust the client model updates and the global model integration, such as using FedProx to deal with non-iid data.

V. Conclusion and Future Work

In our research, we utilized FL to conduct the patient risk prediction using the COVID-19 dataset. Regarding the optimization method, FedProx outperforms FedAvg in non-iid data. For iid data, the two algorithms perform almost the same. Also, when comparing FL with centralized learning, FL performs well and only has 0.2% performance deficiency in the worst case. Since FL can provide privacy protection and, therefore, is expected to collect more data, it can outperform centralized learning when given more data.Moreover, we used encryption to protect the transmission process. Our research provides a novice way for the medical industry to protect patient privacy. In the future, we would like to try more different deep learning models to find the most optimal one

and try to utilize the method to medical images. Also, focus should be paid to the modified FedProx, which can be refined further.

REFERENCES

[1] D. Kollias, A. Arsenos and S. Kollias, "AI-Enabled Analysis of 3-D CT Scans for Diagnosis of COVID-19 & its Severity," *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, Rhodes Island, Greece, 2023, pp. 1-5.

[2] O. Abdelwhab and B. Fatima, "A New Deep Learning Model for COVID-19 Identification using Chest X-ray and CT Scan Images," in *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, El Oued, Algeria, 2021, pp. 1-4.

[3] A. Majeed and S. O. Hwang, "A Comprehensive Analysis of Privacy Protection Techniques Developed for COVID-19 Pandemic," *IEEE Access*, vol. 9, 2021, pp. 164159-164187.

[4] A. Korkmaz, A. Alhonainy, and P. Rao, "An Evaluation of Federated Learning Techniques for Secure and Privacy-Preserving Machine Learning on Medical Datasets," in *2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, DC, USA, 2022, pp. 1-7.

[5] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017*, Journal of Machine Learning Research: Workshop and Conference Proceedings, vol. 54, 2017.

[6] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, "Federated Optimization in Heterogeneous Networks," in *Proceedings of MLSys 2020*.