A Federated Learning Approach to Bone Metastasis Prediction Using Convolutional Neural Network

Conference Paper · December 2023

DOI: 10.1109/ICCIT60459.2023.10441154

Syed Muaz Ali
BRAC University

Salman Ibne Eunus
Genofax

Tasfia Anika Bushra
Daffodil International University

Md Humaion Kabir Mehedi

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

2023 26th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-5901-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIT60459.2023.10441154

A Federated Learning Approach to Bone Metastasis Prediction Using Convolutional Neural Network

Syed Muaz Ali[1], Salman Ibne Eunus[2], Tasfia Anika Bushra[3],
Md Humaion Kabir Mehedi[4], and Annajiat Alim Rasel[5]
[1,2,3,4,5]Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
[1]syed.muaz.ali@g.bracu.ac.bd, [2]salman.ibne.eunus@g.bracu.ac.bd, [3]tasfia.anika.bushra@g.bracu.ac.bd,
[4]humaion.kabir.mehedi@g.bracu.ac.bd, [5]annajiat@gmail.com

***Abstract*—Bone metastasis is a frequently occurring disease and can be a consequence of a number of different cancers such as - prostrate, lung and breast cancers, and predicting them can be really useful for the diagnosis of patients with such diseases. Classifying images of bone scan for bone metastasis prediction requires a huge amount of data to produce a prediction output which is reliable and accurate, but a single medical organization usually do not have access to such amounts of data from other organizations and those organizations are not also ready to share their patients' private data as well due to data security issues. For such scenarios, it is not often possible to train a model with enough data, thus leading to an inaccurate prediction model for bone metastasis. This can be devastating at times due to the occurrence of many false positives or false negatives, if bone metastasis is wrongly classified. In order to find a better solution, so that there is less data protection and privacy issues and therefore more availability of data, we are proposing to use a Federated Learning (FL) based approach for bone metastasis prediction using convolutional neural network. As per our knowledge and background study, we are the first to use federated learning for bone metastasis prediction on the BS-80K dataset. Federated Averaging (FedAvg) strategy was used for implementing the federated learning methodology where different client models were built along with a Global Model.**

***Index Terms*—Federated Learning, CNN, Bone Metastasis, Image Classification**

I. Introduction

Bone Metastasis (BM) is a kind of cancer that occurs in one part of the body in the beginning and later spread to the bones. It commonly occurs in many places of the body such as - spline, thigh, pelvis bones or might occur in any other bones as well. At times, bone metastasis may appear after years or it could be the first sign of cancer. Metastatic cancer in bones is a lethal disease. Detection of normal, benign and malignant tissues for metastatic cancer and thus treating them can be life saving for many [1]. Bone metastases are a prime cause of death, due to serious pain, lack of mobility, pathological fractures, compression in the spinal cord, bone marrow aplasia and hypercalcemia [2]. Therefore, this disease has become a greater concern for physicians, researchers, cancer specialists and many other relevant experts who are working in this field. At present times, there is a widespread use of machine learning in medical diagnosis, due to the rapid growth of deep learning and Artificial Intelligence in healthcare [3]. Deep learning techniques can also lead to outstanding results and can have a great impact in diagnosing bone metastasis. It also requires a vast amount of data, since deep learning is data intensive and need millions of training instances to learn and produce an efficient outcome. As a consequence of the lack of inadequate data, images of histopathology trained on deep learning algorithms from a single hospital may not be able to generalize well when inference with images from different hospitals. Although, powerful computers are available to train such huge quantities of data, but, hospitals and medical organisations have to deal with various data protection rules and regulations, making them unable to share the personal data of patients [4].

In such circumstances, a relatively new concept called Federated Learning (FL), comes to the rescue. FL is a machine learning algorithm that trains model on decentralized end-user devices. It allows different data providers to build a model without sharing data which makes it great for protecting the privacy of data. Using federated learning for medical imaging, different medical organizations can share a patient's data without compromising the patient's privacy. Through federated learning, a highly accurate machine learning model can be built if different medical organizations collaborate and thus all organizations can be benefited as the data privacy is protected but they can have the model parameters which can be used to produce an updated and efficient model. The following outlines our significant contributions in this paper:

- Federated learning is applied for bone metastasis prediction to ensure data privacy and security of patients which allows more availability of data to create a robust and accurate model.
- A CNN based deep learning algorithm has been suggested to perform experimentation to compare the results from the proposed Federated Learning (FL) approach in both clients and the server model. We are using a convolutional neural network based approach in our paper as it is robust and easy to implement.
- In this research work, our main focus is to demonstrate the usefulness of Federated Learning (FL) in Bone Metastasis prediction or in medical imaging diagnosis to ensure data privacy and security of patients. For this reason, we focused more on how FL can be implemented in such use cases and less on the complexity of different machine

learning models.

The remaining paper is organized as follows: Section II contains the literature of some relevant research papers on our topic and some background knowledge on the algorithm that previous researchers working on bone metastasis prediction has implemented. Section III contains the proposed algorithm or methodology of the paper. The Section IV displays the experimental results section containing the visualizations of the observations in terms of each model's accuracies and losses while the conclusion is given in Section V.

II. Related Work

In the recent past, Federated Learning has been used by many researchers for medical image classification scenarios. In [5], federated learning has been used for identifying COVID-19 from chest X-ray images. They have compared the performances of four state-of-the-art models with and without the federated learning framework. Moreover, some papers, like [6], have also discussed the certain qualities and limitations of federated learning and provided an overview of current approaches and directions for future work to safeguard patient data. A recent paper has also worked with curriculum-based federated learning for breast cancer classification, which is memory-aware as well. Curriculum-based learning ameliorates breast cancer classification on high-resolution mammograms in a federated learning environment, as shown in [7]. On the other hand, in some of the papers, such as in [8], they have proposed some variations of federated learning called FedDropoutAvg, which is a generalized federated learning approach used for histopathology image classification. FedDropoutAvg performs better than already available FL approaches as they takes advantage of randomness, both in client selection and also in federated averaging process.

We have also found some researchers who are using machine learning and deep learning for bone metastasis prediction. The authors in [9] used machine learning for the detection of bone metastasis in patients who have been recently diagnosed with thyroid cancer. The authors have developed a random forest (RF) model to train on demographic and clinicopathologic data of thyroid cancer patients with an accuracy of 90.4 percent. Furthermore, the researchers in [10] studied how machine learning can be applied to predict bone metastasis in patients with Prostrate Cancer. They have used many different algorithms which includes - Decision Tree (DT), Random Forest (RF), Multi Layer Perceptron (MLP), Logistic Regression (LR), etc, to build prediction model. Among the experimented models Extreme Gradient Boosting (XGB) performed best with an Area Under the Curve (AUC) score of 0.962 and an accuracy of 0.884 using tabular data. The researchers in [11] have also used machine learning to predict bone metastasis in breast infiltrating ductal carcinoma patients. This paper also trained tabular data with multiple linear models like - Logistic Regression (LR), Extreme Gradient Boosting (XGB),Random Forest (RF), etc where XGB showed the best performance among all other models. Also, some researchers have used Whole-body Bone Scintigraphy (WBS) images for the automatic identification of bone metastatic lesions as in [12]. They have used the deep learning algorithm commonly known as Convolutional Neural Network (CNN) to predict bone metastasis. In comparison to the other state-of-the-art models like - VGG16, InceptionV3, DenseNet169, their proposed model performed better in classifying benign and malignant bone lesions. lastly, we have also found another paper which have used Convolutional Neural Network (CNN) for bone metastasis classification using whole body scan images from prostrate cancer patients. This paper in [13] have compared the testing accuracy of their proposed CNN based model with popular CNN architectures like - VGG16, ResNet50, GoogleNet, etc and has confirmed that their proposed model has shown the best performance.

III. Proposed Methodology

In this work,the BS-80K dataset which contains 82,544 bone scan images from 3247 patients was used for bone metastasis prediction. Furthermore, we used the Flower[14] framework for federated learning, along with PyTorch.

A. Work Plan

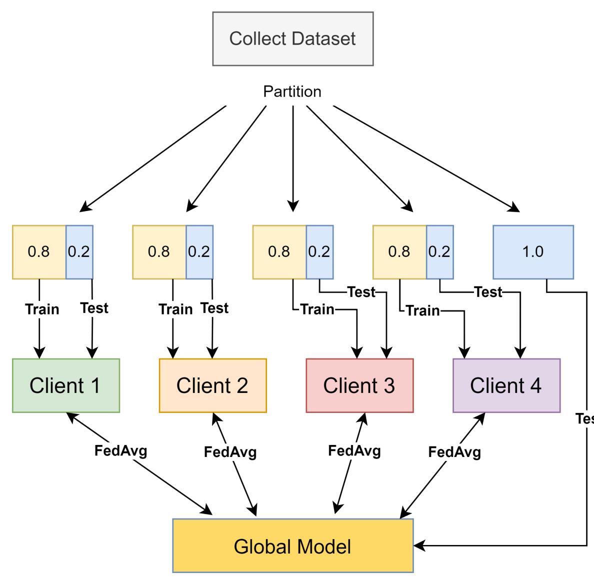

Fig. 1. Work Plan

Figure 1 shows the work plan of our paper where it shows an overview of the paper. Initially, we collected the dataset and then partitioned the dataset into five equal parts. From four of the partitions, we split the dataset into training and testing which were used for training the four clients and also for evaluation. Furthermore, we initialized a Global Model for training the clients with Federated Averaging strategy and one of the five partitions were used for testing the Global Model.

B. Dataset

On this paper, we used the BS-80K[15] dataset for bone metastasis prediction. The dataset contains 82,544 bone scan images gathered from West China Hospital from 3,247 patients. The dataset contains bone scan images of ankle, chest, elbow, head, knee, pelvis, shoulder, vertebra and whole body in left or right with posterior or anterior views where the images are labelled as normal or abnormal. On this paper, we used all the images except for whole body for binary classification.

C. Federated Learning

Unlike classical machine learning where the dataset is stored centrally on a server, on federated learning[16], the data is distributed across different clients and there is one global server for training the models. For implementing Federated Learning, we used the Flower framework with Pytorch. The Flower library is able to create servers and clients for an environment suitable to federated learning. Inititally, we created a global model with random weights. Then we created 5 different clients where each of the clients had one distinct partition from the dataset as local data. We used the Federated Averaging[16] method to average all the models running on the clients for the final model.

D. Convolutional Neural Network

Convolutional Neural Network, or CNN, is a popular deep learning algorithm used in computer vision applications. CNN uses images as input. Inspired by the structure and organization of the visual cortex, CNN's architecture resembles the pattern of connectivity between neurons in the human brain. Each object in the image has a distinct level of importance due to the learnable parameters (weights and biases) that are assigned to it, enabling it to be distinguished from other objects. CNN requires a lot less pre-processing than other classification algorithms. Convolutional, pooling, and fully connected layers are among the layers that make up CNN. The pooling layer down-samples the image to save computation, the convolutional layer applies filters to the input image to extract features, and the fully connected layer makes the final prediction. The neural network uses gradient descent and back-propagation to determine which filters are best. A collection of learnable filters, or kernels, with small widths, heights, and input volume-matched depths make up convolutional layers.

E. Federated Averaging

An algorithm that saves communication when conducting distributed training with a large number of clients is called federated averaging. Clients store their data locally in Federated Averaging to protect their privacy. Individual clients communicate with one another via a central parameter server. Every client receives the parameters from this central server, which also gathers the most recent parameters from the clients. Federated Averaging is typically thought of in a centralized fashion, requiring a great deal of communication between the clients and the central server and frequently resulting in potential channel blocking. The privacy of the entire system may be compromised if an attack targets the central server. By calculating the weighted average of each model update, the Federated Averaging algorithm generates the aggregated model.

F. CNN Architecture

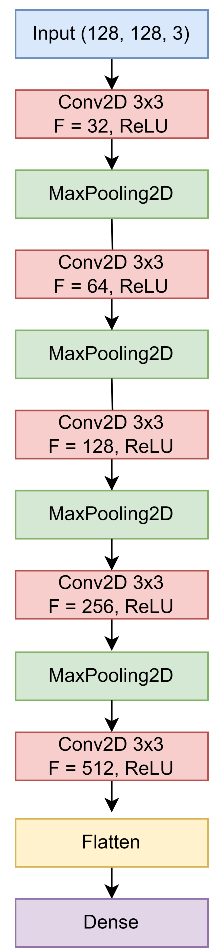

Fig. 2. Proposed Convolutional Neural Network architecture

Figure 2 shows the convolutional neural network architecture used on this paper. The convolutional neural network architecture takes an input shape of 128x128x3 and performs multiple convolution2D operation of kernel size 3x3 where ReLU [17] activation function has been applied on the generated feature maps and also MaxPooling2D operation has been carried out multiple times to down-scale the features. Finally, a Flatten operation is performed on the features generated by the final Conv2D layer and passed to a Dense layer for final classification.

G. Data Pre-processing

In this research paper, we have used images from the BS-80K dataset, while excluding the ones for the whole-body scans. There were total 76,050 images left after excluding the whole-body scans where only 3,560 images were of 'Abnormal' class and the rest 72,490 images were of 'Normal' class which made the dataset of selected images highly imbalanced as around 5% of the images belong to only one class out of two. In order to tackle the imbalance problem, we have reduced the number of 'Normal' samples to same number as the 'Abnormal' samples, making the total number of samples to 7,120. Furthermore, we partitioned the datasets into 5 different partitions where each had 1,424 images. The 'Normal' images were set to Class 0 and 'Abnormal' images were set to Class 1. Again, four of the partitions were split into 80% training and 20% testing for training and testing the four clients and one entire partition was used for testing the Global Model. During training the models, the images were normalized and also Data Augmentation[18] was applied where we randomly flipped the images horizontally and vertically. Those randomly flipped images where also used in the training to increase the variability of the data.

IV. Experiments and Results

We initially created the global model with arbitrary weights. Using the Flower framework, we employed a simulation strategy where we set number of clients to 4. The simulation was set to run for 50 rounds. On each of the rounds, a minimum of two clients would be selected for training and also for evaluating their performance on their local test sets. The models were trained using Adam [19] optimizer and the loss function used was cross entropy. The batch size used was 32 and models were run for 10 epochs on rounds they were trained. Figure 3 shows the training loss of the four clients and the Figure 4 shows the training accuracy of the four clients where two of them were selected each of the 50 rounds and ran for 10 epochs.

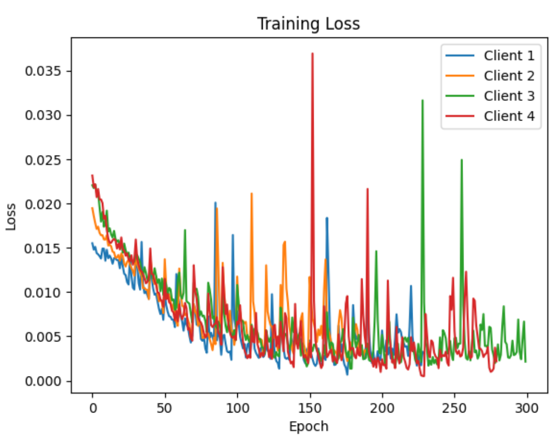

Fig. 3. Training Loss of four clients

Furthermore, in terms of testing set, the Figure 5 shows the testing accuracy of the four clients where 'Evaluation' indicates the number of attempt when the clients were selected

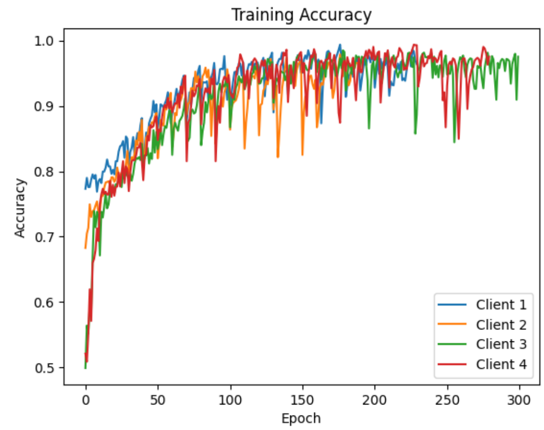

Fig. 4. Training Accuracy of four clients

to evaluate on their local test set. Similarly, the Figure 6 shows the testing loss of the clients on their local test set. Furthermore, Figure 7 shows the accuracy of the global model.

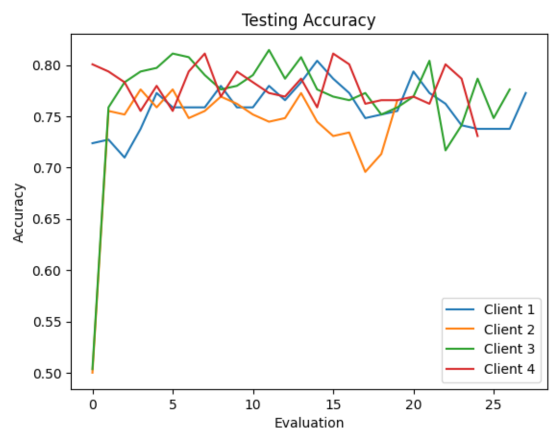

Fig. 5. Testing Accuracy of four clients

Client 1 managed to reach the lowest amount of loss of 0.01543 on the fifth time it was selected to train and achieved the highest accuracy of 0.80419 on the 15th time it was selected to evaluate and finally reached an accuracy of 0.77272 on the last time it was evaluated. For Client 2, the lowest loss of 0.01638 was obtained on the second time it was selected to evaluate and also reached the highest accuracy of 0.77622 on the fourth time. Although, during the last time it was selected for evaluation, it reached a final accuracy of 0.76573 with a loss of 0.02716. Client 3 reached a maximum accuracy of 0.81468 on the 12th time it was evaluated and reached the lowest loss of 0.01403 on the 7th time and on the last time it was evaluated and reached an accuracy of 0.77622 and loss of 0.02093. Furthermore, Client 4 reached a maximum accuracy of 0.81118 on the 8th time it was evaluated and reached lowest loss of 0.01361 on second time it was evaluated. However, on the final time it was evaluated, it reached an accuracy of 0.73076 and a loss of 0.02751. Table I shows the classification

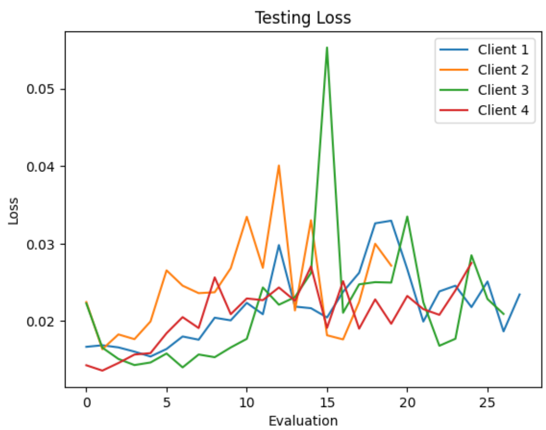

Fig. 6. Testing Loss of four clients

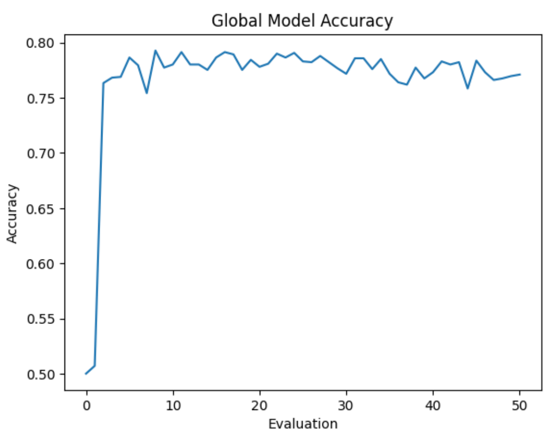

Fig. 7. Accuracy over evaluations on the Global Model

report of Client 1 on their test on the final evaluation set where it achieved a Precision of 0.81 on class 0 and 0.74 on Class 1, Sensitivity of 0.71 on Class 0 and 0.83 on Class 1 and F1 Score of 0.76 on Class 0 and 0.79 on Class 1.

	Client 1		
Class	**Precison**	**Sensitivity**	**F1 Score**
0	0.81	0.71	0.76
1	0.74	0.83	0.79

TABLE I: Classification Report of Client 1 on final evaluation

Table II shows the classification report of Client 2 where it achieved Precision of 0.77 and 0.76 on class 0 and 1, Sensitivity of 0.76 and 0.77 on class 0 and 1 and finally F1 Score of 0.76 and 0.77 on Class 0 and 1.

	Client 2		
Class	**Precision**	**Sensitivity**	**F1 Score**
0	0.77	0.76	0.76
1	0.76	0.77	0.77

TABLE II: Classification Report of Client 2 on final evaluation

Furthermore, the Table III shows the classification report of Client 3 where it achieved Precision of 0.82 and 0.74, Sensitivity of 0.71 and 0.85, F1 Score of 0.76 and 0.79 on class 0 and 1.

	Client 3		
Class	**Precision**	**Sensitivity**	**F1 Score**
0	0.82	0.71	0.76
1	0.74	0.85	0.79

TABLE III: Classification Report of Client 3 on final evaluation

Similarly, Table IV shows the classification report of Client where it achieved Precision, Sensitivity and F1 Score of 0.80, 0.72 and 0.76 on Class 0 and 0.75, 0.82 and 0.78 on Class 1.

	Client 4		
Class	**Precision**	**Sensitivity**	**F1 Score**
0	0.77	0.66	0.71
1	0.70	0.80	0.75

TABLE IV: Classification Report of Client 4 on final evaluation

Moreover, Table V shows the classification report of the Global Model where it achieved precision of 0.8, sensitivity of 0.72 and F1 Score of 0.76 on Class 0 and Precision of 0.75, Sensitivity of 0.82 and F1 Score of 0.78 on Class 1.

	Global Model		
Class	**Precision**	**Sensitivity**	**F1 Score**
0	0.80	0.72	0.76
1	0.75	0.82	0.78

TABLE V: Classification Report of Global Model on final evaluation

Finally, the maximum and final testing accuracies are shown on Table VI.

	Max Accuracy	**Final Accuracy**
Client 1	0.80419	0.77272
Client 2	0.77622	0.76573
Client 3	0.81468	0.77622
Client 4	0.81118	0.73076
Global Model	0.79283	0.77106

TABLE VI: Maximum and Final Accuracies of four clients and the Global Model

V. Conclusion and Future Works

In this research work, we have used bone scan images to train and predict bone metastasis. We have used a federated learning approach to protect private data and a unique CNN based approach. In this paper, our main focus was to show the efficacy of Federated Learning in medical imaging use cases and how we can use FL to ensure data security and privacy. In future, we look forward to use our proposed model and federated learning based approach to experiment with other similar datasets as well. In addition, we also plan to experiment and validate our results with other CNN based transfer learning models as well.

REFERENCES

[1] N. I. Papandrianos, E. I. Papageorgiou, A. Anagnostis, and A. Feleki, "A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans," *Applied Sciences*, 2020.

[2] F. Macedo, K. Ladeira, F. Pinho, *et al.*, "Bone metastases: An overview," *Oncology Reviews*, May 2017. DOI: 10.4081/oncol.2017.321. [Online]. Available: https://doi.org/10.4081/oncol.2017.321.

[3] E. Darzidehkalani, M. Ghasemi-rad, and P. van Ooijen, "Federated learning in medical imaging: Part ii: Methods, challenges, and considerations," *Journal of the American College of Radiology*, vol. 19, no. 8, pp. 975–982, 2022, ISSN: 1546-1440. DOI: https://doi.org/10.1016/j.jacr.2022.03.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1546144022002812.

[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*, vol. 12, no. 1, Feb. 2022. DOI: 10.1038/s41598-022-05539-7. [Online]. Available: https://doi.org/10.1038/s41598-022-05539-7.

[5] B. Liu, B. Yan, Y. Zhou, Y. Yang, and Y. Zhang, *Experiments of federated learning for covid-19 chest x-ray images*, 2020. DOI: 10.48550/ARXIV.2007.05592. [Online]. Available: https://arxiv.org/abs/2007.05592.

[6] G. Singh, V. Violi, and M. Fisichella, "Federated learning to safeguard patients data: A medical image retrieval case," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 18, Jan. 2023. DOI: 10.3390/bdcc7010018. [Online]. Available: https://doi.org/10.3390/bdcc7010018.

[7] A. Jiménez-Sánchez, M. Tardy, M. A. G. Ballester, D. Mateus, and G. Piella, "Memory-aware curriculum federated learning for breast cancer classification," *Computer Methods and Programs in Biomedicine*, vol. 229, p. 107318, Feb. 2023. DOI: 10.1016/j.cmpb.2022.107318. [Online]. Available: https://doi.org/10.1016/j.cmpb.2022.107318.

[8] G. N. Gunesli, M. Bilal, S. E. A. Raza, and N. M. Rajpoot, "Feddropoutavg: Generalizable federated learning for histopathology image classification," *ArXiv*, vol. abs/2111.13230, 2021.

[9] W.-C. Liu, Z.-Q. Li, Z.-W. Luo, W.-J. Liao, Z.-L. Liu, and J.-M. Liu, "Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer," *Cancer Medicine*, vol. 10, no. 8, pp. 2802–2811, Mar. 2021. DOI: 10.1002/cam4.3776. [Online]. Available: https://doi.org/10.1002/cam4.3776.

[10] W.-C. Liu, M.-X. Li, W.-X. Qian, *et al.*, "Application of machine learning techniques to predict bone metastasis in patients with prostate cancer," *Cancer Management and Research*, vol. Volume 13, pp. 8723–8736, Nov. 2021. DOI: 10.2147/cmar.s330591. [Online]. Available: https://doi.org/10.2147/cmar.s330591.

[11] W.-C. Liu, M.-X. Li, S.-N. Wu, *et al.*, "Using machine learning methods to predict bone metastases in breast infiltrating ductal carcinoma patients," *Frontiers in Public Health*, vol. 10, Jul. 2022. DOI: 10.3389/fpubh.2022.922510. [Online]. Available: https://doi.org/10.3389/fpubh.2022.922510.

[12] Y. Liu, P. Yang, Y. Pi, *et al.*, "Automatic identification of suspicious bone metastatic lesions in bone scintigraphy using convolutional neural network," *BMC Medical Imaging*, vol. 21, no. 1, Sep. 2021. DOI: 10.1186/s12880-021-00662-9. [Online]. Available: https://doi.org/10.1186/s12880-021-00662-9.

[13] N. Papandrianos, E. Papageorgiou, A. Anagnostis, and K. Papageorgiou, "Bone metastasis classification using whole body images from prostate cancer patients based on convolutional neural networks application," *PLOS ONE*, vol. 15, no. 8, J. Gwak, Ed., e0237213, Aug. 2020. DOI: 10.1371/journal.pone.0237213. [Online]. Available: https://doi.org/10.1371/journal.pone.0237213.

[14] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, and N. D. Lane, "Flower: A friendly federated learning research framework," *ArXiv*, vol. abs/2007.14390, 2020.

[15] Z. Huang, X. Pu, G. Tang, *et al.*, "BS-80k: The first large open-access dataset of bone scan images," *Computers in Biology and Medicine*, vol. 151, p. 106221, Dec. 2022. DOI: 10.1016/j.compbiomed.2022.106221. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2022.106221.

[16] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *International Conference on Artificial Intelligence and Statistics*, 2016.

[17] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

[18] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.