ASEC 2024 Conference

IMPLEMENTATION OF FEDERATED LEARNING FOR PERIPHERAL BLOOD CELL IDENTIFICATION ACROSS DIFFERENT CLINICAL CENTERS

Lorenzo Natali[2], Kevin Barrera[2], Anna Merino[1] and José Rodellar[2]

[1]CORE Laboratory. Biochemistry and Molecular Genetics Department. Biomedical Diagnostic Center. Hospital Clínic. Barcelona. Spain
[2]Department of Mathematics. Technical University of Catalonia. Barcelona. Spain

INTRODUCTION

Morphological analysis of peripheral blood is essential for the diagnosis of 80% of hematological diseases. Although automatic classification systems in morphological analyzers support diagnosis, image variability caused by differences in reagents, sample preparation, and analyzer optics between centers affects their performance. To address this challenge, this study proposes using federated learning, an innovative approach that enables collaborative training of a model, adapting to the specific characteristics of each center and maintaining high performance despite image variability.

METHODS

The federated learning workflow is illustrated in Figure 1. The Core Laboratory of the Hospital Clínic de Barcelona served as the reference center, providing 10 298 images categorized into five leukocyte classes: basophils, eosinophils, lymphocytes, monocytes, and neutrophils. Four public datasets were included as external centers: C1 (14 514 images), C2 (2 513 images), C3 (5 000 images), and C4 (11 353 images). Data from all centers were divided into training, validation, and test sets.

Initially, a VGG16 network was trained with data from the reference center, achieving 99.4% accuracy on its test set. However, when evaluated on external centers, the model's performance significantly decreased. Table 1 presents metrics such as accuracy, precision, specificity, and F1-score, highlighting this drop in performance before applying federated learning (FL).

To improve generalization, an FL approach was implemented. The first three convolutional blocks of the VGG16 network were frozen, and the remaining layers were fine-tuned using the training sets from each center. The FedDyn technique was applied to aggregate the adjusted weights, resulting in a Final Global Model that improved performance across all centers while maintaining privacy.

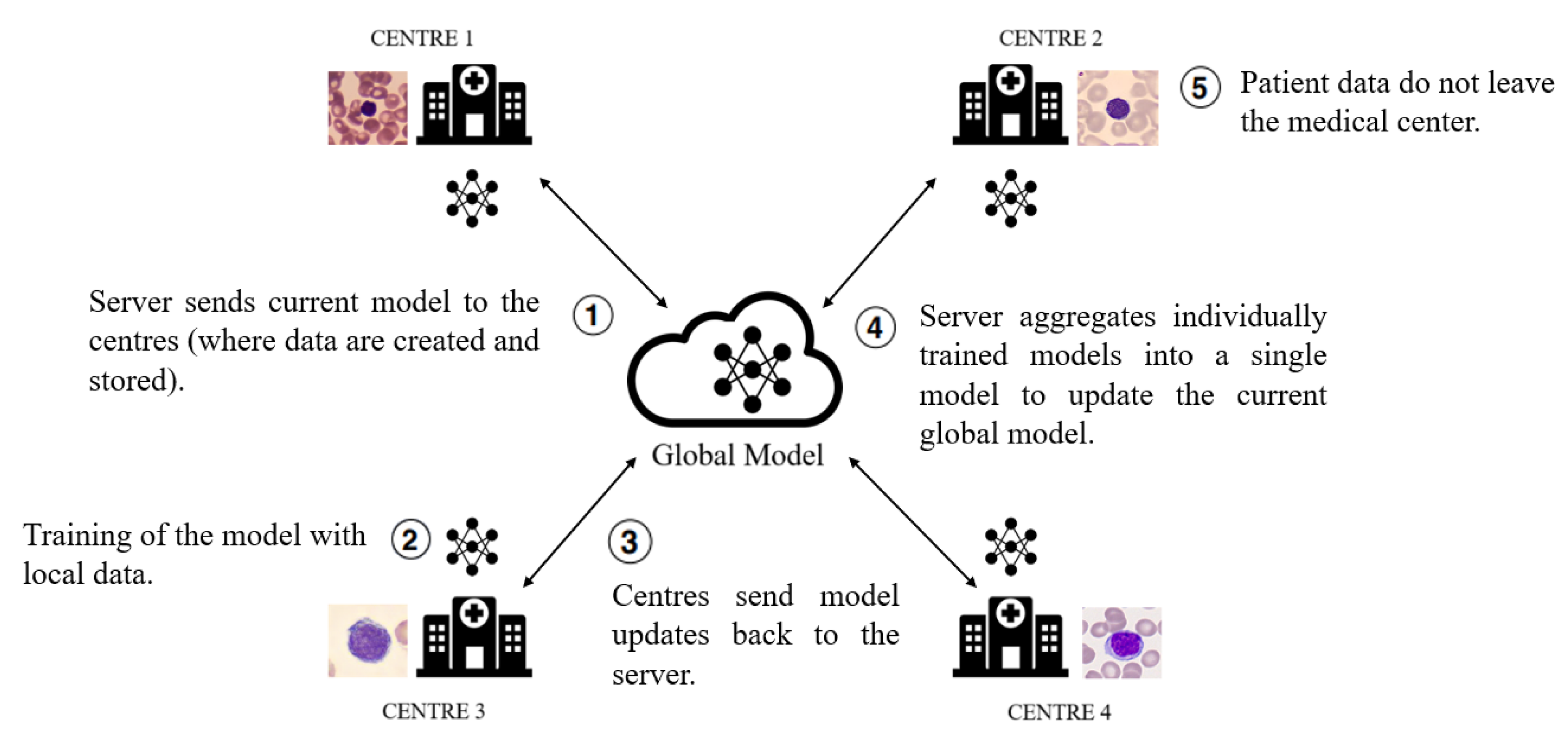

Figure 1.- *Federated learning workflow: 1) The server sends the model to centers, where patient data is stored. 2) Centers train the model locally, ensuring privacy. 3) Updates are sent back to the server. 4) The server refines the global model, enabling collaboration while safeguarding data privacy.*

RESULTS

The Final Global Model obtained through federated learning demonstrated a marked improvement in classification performance across all centers. The test sets for the external centers showed significant gains in accuracy: 96.2% for C1, 99.6% for C2, 99.5% for C3, and 88.8% for C4. These results reflect the ability of the federated learning approach to enhance generalization while addressing data variability between centers.

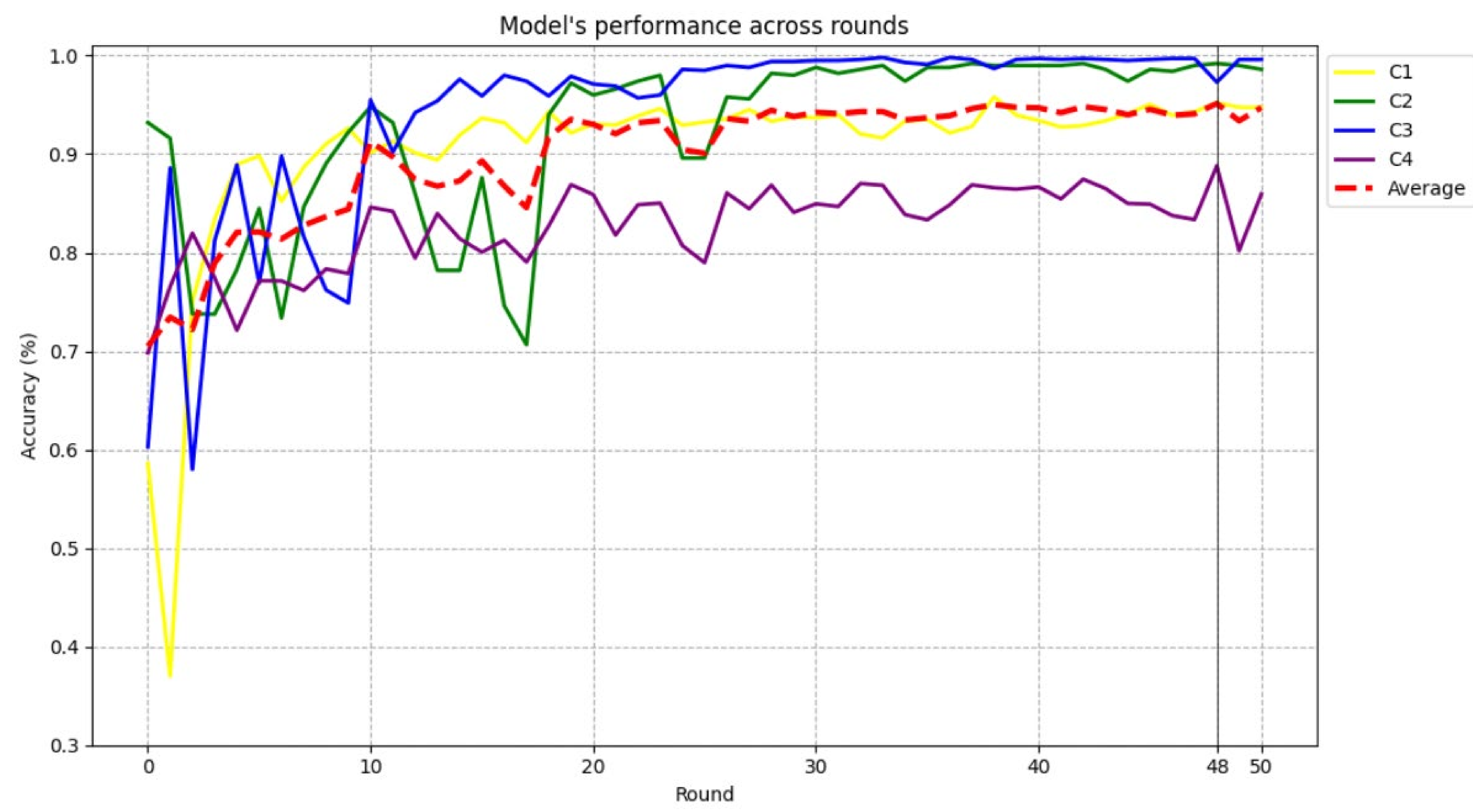

Figure 2.- *Federated learning process: The graph shows accuracy trends during training rounds, while the table outlines improvements in accuracy, precision, specificity, and F1 score across centers after federated learning.*

Figure 2 shows the federated learning process, illustrating accuracy trends over training rounds and highlighting the improvement compared to the initial global model. Table 2 presents the classification metrics after applying the federated learning strategy, demonstrating enhanced model performance across all centers.

	Accuracy	Precision	Specificity	F1 Score
Center C1	0.586	0.736	0.942	0.657
Center C2	0.932	0.944	0.983	0.931
Center C3	0.603	0.803	0.901	0.562
Center C4	0.698	0.711	0.952	0.785

Table 1.- *Classification performance metrics of the VGG16 model before applying federated learning across external centers: accuracy, precision, specificity, and F1-score.*

	Accuracy	Precision	Specificity	F1 Score
Center C1	0.952	0.964	0.992	0.958
Center C2	0.992	0.992	0.998	0.992
Center C3	0.973	0.974	0.993	0.973
Center C4	0.888	0.884	0.992	0.885

Table 2.- *Classification performance metrics of the Final Global Model after applying federated learning across centers: accuracy, precision, specificity, and F1 score*

CONCLUSION

This study demonstrates that federated learning can effectively fine-tune classifiers in multicenter settings, making the model more robust to the variability between different datasets. This approach shows potential as a tool for automatic recognition in multicenter contexts.