A Federated Learning Based Efficient Approach to Detect Cervical Cancer Using PAP-SMEAR Images

Chowdhury Saima Rawshan
Department of Computer Science & Engineering
BRAC University
chowdhury.saima.rawshan@g.bracu.ac.bd

Sadia Habib Nizhum
Department of Computer Science & Engineering
BRAC University
sadia.habib.nizhum@g.bracu.ac.bd

Sackline Naien Ridvi
Department of Computer Science & Engineering
BRAC University
sackline.naien.ridvi@g.bracu.ac.bd

Fahim Shahriar Ahmed
Department of Computer Science & Engineering
BRAC University
fahim.shahriar.ahmed@g.bracu.ac.bd

Sayani Das
Department of Computer Science & Engineering
BRAC University
sayani.das@g.bracu.ac.bd

Abstract—
Cervical cancer is one of the most common malignancies and easily preventable through early diagnosis of pap smear images which results in a lower death rate among women throughout the world. The typical examination of pap smear images is tiring and has a lot of human mistakes. Additionally, having less privacy due to data centralization is a significant concern. This work proposes an efficient privacy-preserving federated learning based framework for cervical cancer detection using Pap smear images. FedPapVisionNet, a lightweight mechanism for federated learning was introduced which improves the accuracy while safeguarding privacy. With only 3,90,215 trainable parameters, it merges the efficient convolutional operation and attention mechanism and achieves impressive diagnostic accuracy on the dataset. Ultimately, the use of federated optimization techniques such as FedAvg, FedProx, and Scaffold allows for the assessment of convergence and robustness in homogeneous and heterogeneous data. Experimental results demonstrate that FedPapVisionNet achieved accuracy of 95.68% in IID distribution, 86% using FedProx algorithm and 90% using SCAFFOLD strategy in non-IID distribution, highlighting its effectiveness for accurate, scalable and privacy-aware cervical cancer detection in real-world healthcare applications.

Index Terms—Federated Learning, Cervical Cancer Detection, Pap Smear Images, Deep Learning, Privacy-Preserving Machine Learning, Non-IID Data Distribution, Medical Image Classification

I. Introduction

Cervical cancer is a significant public health issue and remains one of the most common cancers among women worldwide. Pap smear screening for early detection can significantly reduce the death rate; however, manual screening requires tedious, protocol-driven reviews and is labor-intensive and error-prone. Due to the increasing workload faced by healthcare systems and the limited availability of specialists, there is a growing demand for automated and reliable computer-aided diagnostic systems for cervical cancer screening. Recent advancements in deep learning, especially convolutional neural networks, have shown great promise in medical imaging by learning complex features and classifying diseases accurately.

The use of deep learning based diagnostic models has had great success. However, they cannot be used practically due to data privacy. In centralized learning frameworks, large volumes of patient data get transferred to a central server. This might violate privacy regulations and ethical standards. Federated Learning has emerged as a promising solution that enables trained machine learning models at decentralized institutions without sharing patients' raw data.

A framework based on federated learning is proposed in this paper for the automation of cervical cancer detection via Pap smear images. In this study, we develop a lightweight hybrid architecture called FedPapVisionNet to cope with the computational and communication bottlenecks in federation environments.Our contribution in this paper are :

- Establishing an effective and privacy-preserving diagnostic framework for cervical cancer detection based on Papsmear images by taking advantage of the distributed mechanism of Federated Learning and integrated with powerful Deep Learning models.
- Developing a Federated Learning-based system for collaborative training across the disjointed medical institutions without compromising on the privacy of raw patient data.

- Incorporate high-performing CNN model to classify cervical cell Pap smear images.
- Evaluating how federated learning model performs in IID and non-IID manner regarding real-world disparities among medical institutions.

II. Literature Review

A. Cervical Cancer Detection With Deep Learning Models

Research has been done over the years to detect cervical cancer from images of pap smear to overcome the issue of manual screening. Also, manual screening relies on experts and is slow and error-prone. As a result, several recent studies have utilized machine learning and deep learning to automatically classify cervical cells.

According to Sher Lyn Tan et al. [1], the implementations of transfer learning with CNN architectures VGG-16, ResNet-50 and DenseNet-121 on the Herlev dataset, resulted in DenseNet-201 achieving 87.02% accuracy and limited due to small size of dataset and imbalanced classes. Barriers to screening participation include lack of knowledge, fear, and embarrassment, as identified by Hossein Ashtarian et al. [2]. Awareness programs can address these issues using respective datasets. Marina E. Plissiti et al. [3] create SIPaKMeD dataset with 4,049 labelled images in five classes. It demonstrates that CNN-based methods outperform feature-based methods with accuracy of 95.35%. Notably, they mention concerns with dataset size and class imbalance.

Ishak Pacal et al. [4] and Aditya Khamparia et al. [5] achieved 99.4% accuracy for binary normal vs abnormal classification using Hybridized CNNs with Variational Autoencoders. Thus, multi-class extension of this work is needed. Melad Rahimi et al. [6] presented the review of the survival prediction models – like Random Forest and Logistic Regression – having AUC in range 0.40 to 0.99. They also highlighted that the survival prediction models have interpretability issues for clinical implementation. The multi-branch CNN+MHSA model developed by Tatsuhiro Baba et. al. [7] achieved 98.5228% accuracy on SIPaKMeD while highlighting issues relating to generalizability and complexity.

N. Sompawong and colleagues [8] implemented Mask R-CNN, which, although it yielded a 91.7% success rate, faced challenges in accurately detecting atypical cells. Chunny Yuan et al. [9] achieved 85.38% sensitivity and 84.10% accuracy for the LSIL/HSIL classification with DICE score of 61.64%. Harmanpreet Kaur et al. [10] measured 16 transfer models representing ResNet101 giving 95.56%. Examined segmentation trends with U-Net that attained a recall of 98% and HDFF that attained binary accuracy of 99.85%. Paisit Khanarsa and Satanat Kitsiranuwat [13] presented an ensemble transfer learning method with ResNet152V2 that achieved 95% – 97% accuracy and claimed the need for automation of single-cell extraction.

B. Federated Learning in Medical Imaging

A privacy-preserving framework that facilitates collaborative model training on decentralized data without sharing sensitive patient data has brought much attention to federated learning (FL) in healthcare. Subramanian et al. [14] examined decentralized FL algorithms for cancer classification on Kaggle datasets, evaluating FedAvg, FedProx, and other methods. In non-IID settings, they achieved 83.31% accuracy after 150 communication rounds with Bayesian hyperparameter optimization. However, it should be noted that the dataset was not very diverse and did not evaluate the real-time performance trade-off with privacy.

According to a paper produced by Darzidehkalani et al. [15], heterogeneous data distributions across institutions present challenges to FL applications in medical imaging. Although FedAvg has been recognized as balanced with respect to performance, privacy and communication overhead, the impact of data heterogeneity and privacy mechanisms on convergence and computational efficiency. McMahan et al. [16] proposed a method for training machine learning models on decentralized data silos, called federated averaging, which preserves user privacy. Through model averaging where the local updates are averaged, FedAvg cuts down on communication costs and achieves accuracy similar to that of centralized training.

C. Cervical Cancer Detection with Federated Learning

In cancer detection, FL is fast gaining acceptance. The study of Federated Machine Learning for cervical cancer prediction by Muhammad Umar Nasir [17] achieved an accuracy of 99.26%. The sensitivity and specificity is 99.6%. But, the dataset chosen may not represent the actual scenario. Peta and Koppu [18] implemented FL in breast cancer classification and achieved an accuracy of 95.68% despite communication overhead and encryption challenges. Shehnaz Joynab et al. [19] utilized FL-based CNN models on the SIPaKMeD dataset, achieving an accuracy of 93.36% under IID settings while the same model achieved an accuracy of 78.43% in non-IID settings.

III. Methodology

This research proposes a privacy-preserving deep learning framework for automated cervical cancer cell classification using Pap smear images under a federated learning algorithm. The primary objective is to design lightweight yet high performing convolutional and hybrid deep learning models that can be trained collaboratively across distributed clients without sharing sensitive medical data. The overall workflow integrates dataset acquisition

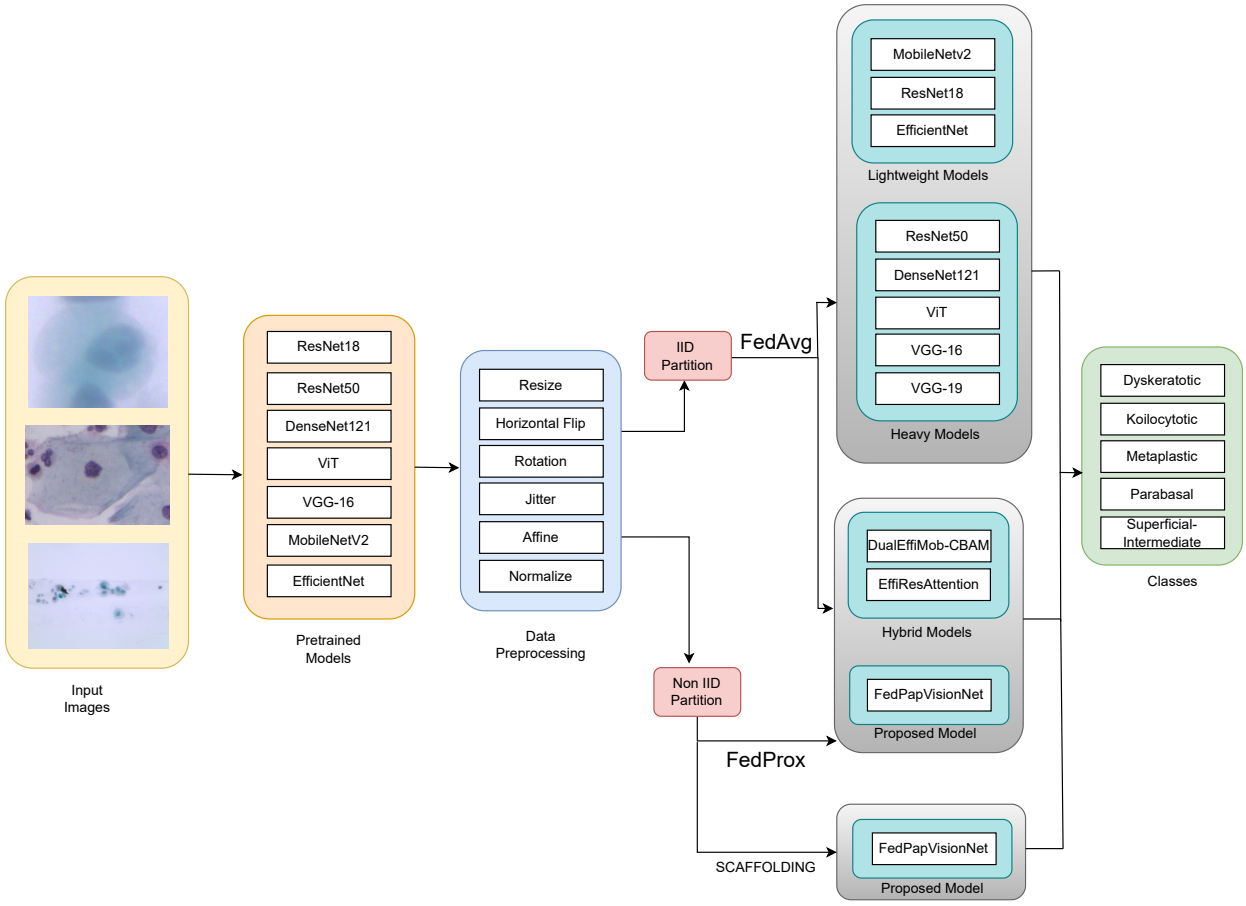


Fig. 1: Methodology overview of the proposed federated learning framework

and preprocessing, federated learning-based training under both IID and non-IID data distributions, and the development of efficient custom and hybrid architectures optimized for decentralized environments. The complete methodological pipeline is illustrated in Fig. 1.

A. Data Acquisition

In this study's experiments, we primarily use the SIPaKMeD (Single-cell Pap Smear) dataset, a well annotated and publicly available cervical cytology image dataset. The database of SIPaKMeD has specification of 4049 isolated single cell images which are isolated by an expert cytopathologist which are used to accurately partition the nucleus and the cytoplasm. The dataset is divided into five cytological classes that are clinically relevant, including Superficial-Intermediate, Parabasal, and others. Cervical cells in these classes represent a wide range of normal, pre-cancerous and abnormal.

B. Data Preprocessing

The SIPaKMeD dataset underwent preprocessing procedures to enhance convergence stability before model training. All images were uniformly resized to 128×128 pixels to satisfy the architectural constraints of convolutional and transformer-based models while preserving essential cellular morphology. Pixel intensity values were normalized to the range $[0, 1]$, followed by channel-wise normalization using a mean of $[0.5, 0.5, 0.5]$ and a standard deviation of $[0.5, 0.5, 0.5]$. This normalization process reduced variability in illumination and staining across samples.

Data augmentation techniques were applied during training to improve robustness and generalization. These included random horizontal flipping with a probability of 0.5 and random rotation within $\pm 15°$. To simulate staining variations, color jittering and small random affine transformations with translation were employed. These augmentation strategies increased dataset diversity and enhanced the model's ability to generalize to real-world acquisition variations.

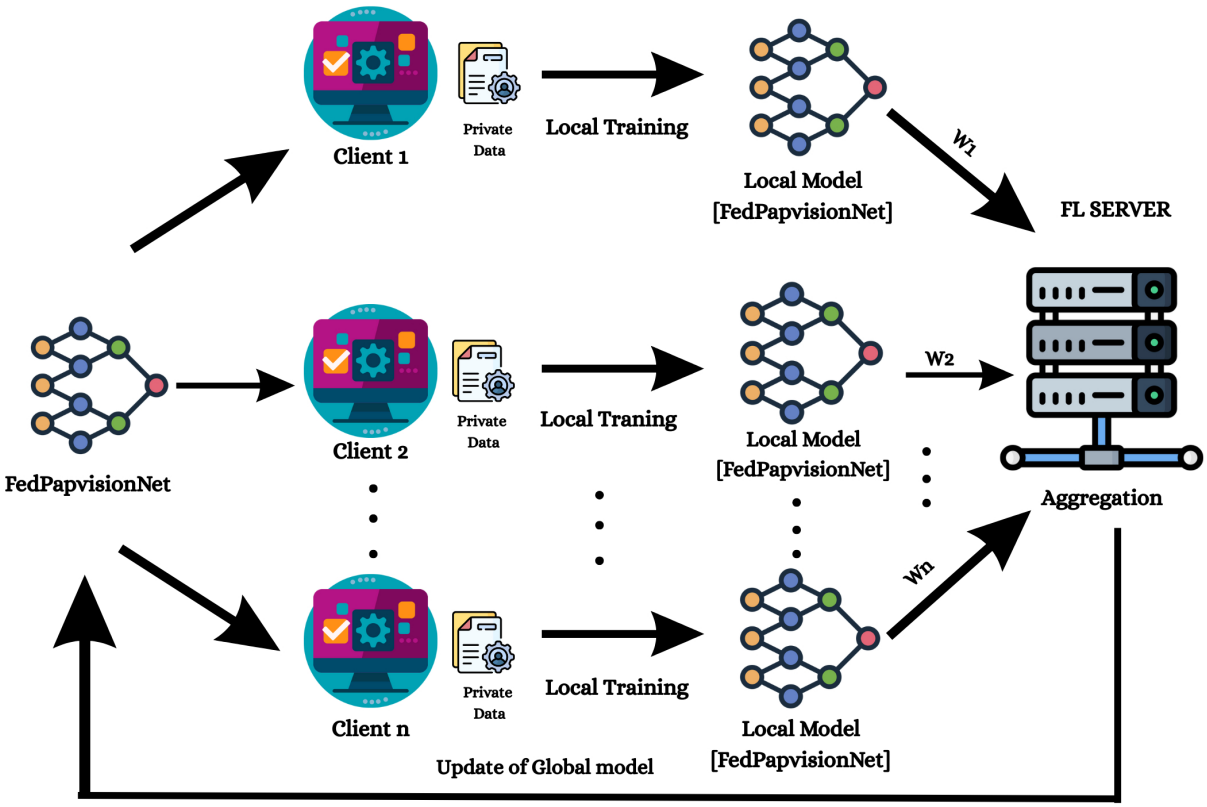


Fig. 2: Federated Learning Workflow

C. Federated Learning Framework

A federated learning framework is adopted as mentioned in 2 to enable decentralized training while preserving data privacy. Instead of aggregating all images at a central server, model training is performed collaboratively across multiple clients, each holding its own local dataset. Only model parameters are exchanged between clients and the central server, ensuring that raw medical images never leave local devices.

Three aggregation strategies are employed. Federated Averaging (FedAvg) serves as the baseline approach, where local model updates are aggregated using weighted averaging. While FedAvg performs well under IID conditions, real-world medical data often exhibit heterogeneity. To address this, Federated Proximal Optimization (FedProx) is applied, introducing a proximal regularization term to constrain local updates and reduce client drift in non-IID settings. Additionally, the SCAFFOLD algorithm is implemented to further mitigate client drift by incorporating control variates that correct biased local updates.

All three proposed models FedPapVisionNet, DualEffiMob-CBAM, and EffiResAttention are trained under FedAvg for IID Data Partition while the models are trained under FedProx and Scaffold Algorithm for non-IID Data Partition.

I. Federated Averaging (FedAvg): Federated Averaging (FedAvg) is employed as the baseline aggregation algorithm. In FedAvg, the central server initializes a global model w^0 and distributes it to a subset of participating clients at each communication round. At round t, each selected client k receives the global model w^t and performs E epochs of local training using stochastic gradient descent with AdamW optimization. The local update is given by:

$$w_k^{t+1} = w^t - \eta \nabla F_k(w^t), \tag{1}$$

where η denotes the learning rate and $\nabla F_k(\cdot)$ represents the gradient computed on the local dataset of client k [16]. After local training, clients transmit their updated parameters to the server, which aggregates them using

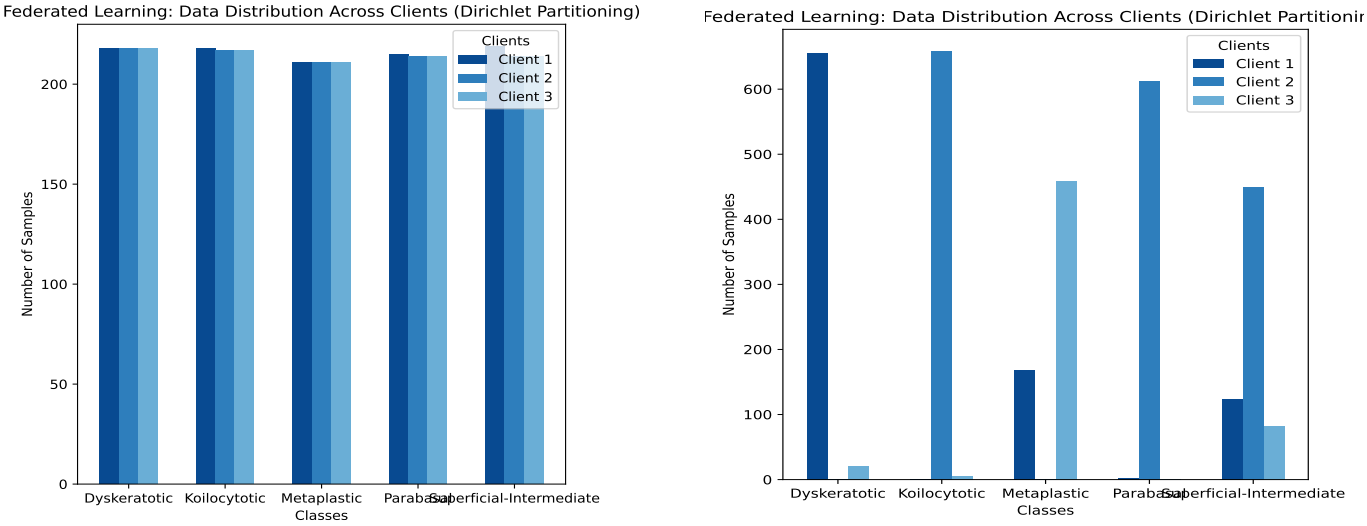


Fig. 3: IID & Non IID Data settings

weighted averaging:

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{t+1}. \tag{2}$$

Although effective for IID data, FedAvg suffers performance degradation under heterogeneous data distributions.

II. Federated Proximal Optimization (FedProx): To mitigate client drift in non-IID settings, Federated Proximal Optimization (FedProx) extends FedAvg by introducing a proximal regularization term. The local optimization objective at client k is formulated as:

$$\min_{w} F_k(w) + \frac{\mu}{2}\|w - w^t\|_2^2, \tag{3}$$

where μ controls the strength of regularization and w^t denotes the global model. The corresponding update rule becomes:

$$w_k^{t+1} = w^t - \eta \nabla F_k(w^t) + \mu(w^t - w). \tag{4}$$

The server aggregation remains identical to FedAvg. [20]

III. Stochastic Controlled Averaging (SCAFFOLD): SCAFFOLD is implemented to further reduce client drift by introducing control variates that correct biased local updates. For N clients with local objectives $f_i(x)$, the global optimization problem is defined as:

$$\min_{x \in \mathbb{R}^d} f(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x). \tag{5}$$

During each communication round r, participating clients perform drift-corrected local updates:

$$y_i^{r,k} = y_i^{r,k-1} - \eta_l \nabla f_i(y_i^{r,k-1}) + c^r - c_i^r, \tag{6}$$

where c^r and c_i^r denote global and local control variates, respectively. Server aggregation follows weighted averaging to update both the global model and control variates [21].

D. Data Partitioning

The SIPaKMeD dataset was partitioned into IID and non-IID configurations to simulate different data distribution scenarios across medical institutions in a federated learning setting. In the IID setup, data were randomly and evenly distributed among clients, ensuring balanced representation of all five cervical cell classes. Conversely, the non-IID setup introduced intentional data imbalance using Dirichlet partitioning algorithm, with each client containing samples dominated by specific classes to reflect real-world institutional heterogeneity.

E. Training Configuration

All models are trained using AdamW optimizer with initial learning rate of 0.001 for IID settings and 0.0005 for non-IID settings. The learning rate follows cosine annealing schedule for IID experiments and step-wise decay (decay rate 0.96) for non-IID experiments. Weight decay is set to 1×10 for IID and 5×10 for non-IID configurations.

For IID experiments, 25-200 communication rounds (model-dependent) were conducted with 3-5 local epochs per round and batch size of 32. For non-IID experiments, we use 50 communication rounds with 5 local epochs and batch size of 10 to allow finer gradient updates under heterogeneous conditions. FedProx employs =0.1 as the proximal coefficient. All experiments are executed on NVIDIA RTX 4080 Super GPU with CUDA support.

F. Model Architectures and Design Specifications

Pretrained Models: A diverse range of pretrained deep learning architectures are selected including VGG16, VGG19, MobileNetV2, ResNet18, ResNet50, DenseNet121, EfficientNetB0 and Vision Transformer (ViT) to train on FedAvg algorithm for IID Data Partition.

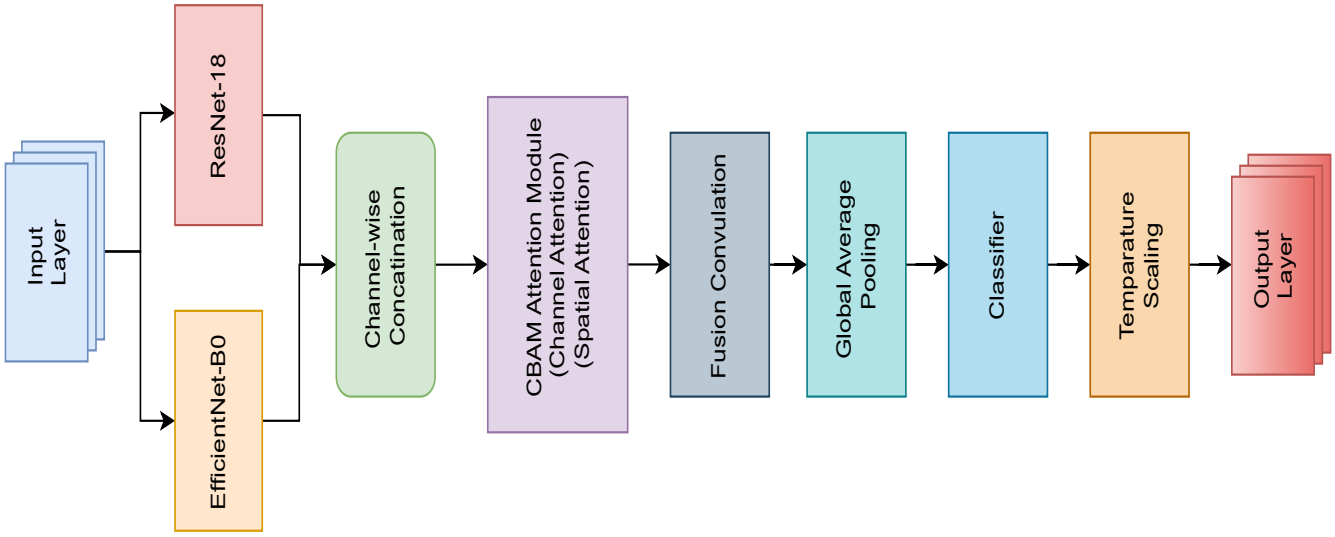


Fig. 4: Architecture overview of the EffiResAttention model

TABLE I: Model Complexity Comparison

Model	Parameters
DualEffiMob-CBAM	8,528,611
EffiResAttention	16,670,243
FedPapVisionNet	390,215

Hybrid Models:

1) I. EffiResAttention: The EffiResAttention structure makes use of EfficientNet-B0 and ResNet18 features in a parallel way. The dual-path design allows the simultaneous learning of diverse feature representations. EfficientNet-B0 uses depthwise separable convolutions and squeeze-and-excitation for compound scaling efficiency. While ResNet18 incorporates residual learning with identity-based skip connections to minimize vanishing gradient problems. The input image undergoes feature extraction simultaneously on both paths independently.

ResNet18 utilizes residual learning to facilitate gradient flow and stabilize training in deeper networks. Bilinear interpolation is applied before fusion due to spatial resolution differences. The features that are aligned are concatenated along channel dimensions. To better learn discriminative features, the Convolutional Block Attention Module (CBAM) applies attention on concatenated features through channel and spatial. After tuning in on the address, a 1×1 convolution does feature fusion and dimension reduction. Global average pooling yields a condensed array of relevant characteristics. The classification head uses fully connected layers with increasing dropout and a temperature-scaled output, generating final predictions for the five cervical cell classes.

II. DualEffiMob-CBAM: The parallel extraction approach of model DualEffiMob-CBAM uses complementary features from EfficientNet-B0 and MobileNetV2 backbones. The MobileNetV2 employs inverted residual blocks with linear bottlenecks, and EfficientNet-B0 implements compound scaling across depth, width, and resolution. In distinct branches, 1280-channel feature maps are obtained from the input image via feature extraction. To make the feature maps compatible in case of dimensional mismatches, spatial alignment using bilinear interpolation is used. The features that have been aligned are placed together side by side along the channel dimension.. CBAM is used to learn more discriminative features using attention. The expansion or reduction of dimensionality of feature is accomplished through use of 1×1 convolution that is followed by the feature representation of fixed length by way of global average pooling. The classification head is made up of fully connected layers that apply dropout at progressively higher rates and are subjected to temperature scaling to produce the final predictions over the five target classes.

III. FedPapVisionNet: In the proposed FedPapVisionNet, residual learning is employed as a fundamental design principle to improve feature propagation and stability during training. Each residual block introduces an alternate pathway that allows the input to bypass convolutional layers and be directly added to the block output. This identity mapping ensures that important low-level features are preserved while enabling deeper representations to be learned effectively

The residual operation implemented in FedPapVisionNet is mathematically expressed as:

$$R_{\text{out}} = \text{ReLU}\left(y_2 + \text{Shortcut}(x)\right) \tag{7}$$

where the intermediate transformations are defined as:

$$y_1 = \text{ReLU}\left(\text{BN}\left(\text{Conv}(x, W_1)\right)\right) \tag{8}$$

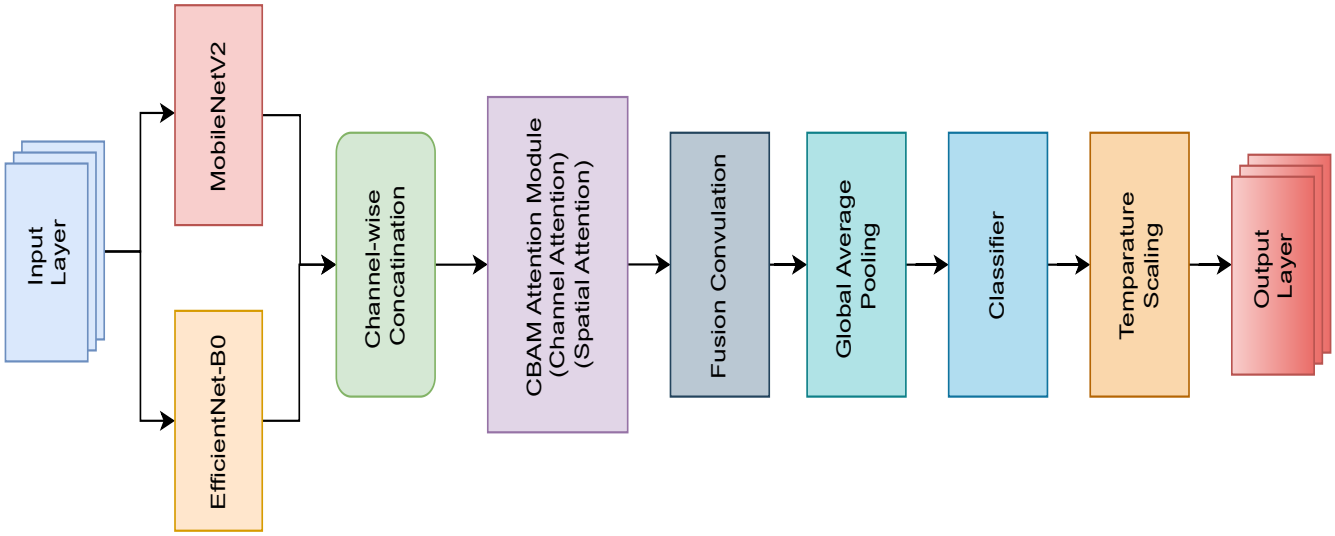


Fig. 5: Architecture overview of the DualEffiMob-CBAM model

$$y_2 = \mathrm{BN}\left(\mathrm{Conv}(y_1, W_2)\right) \tag{9}$$

The shortcut connection is formulated as:

$$\mathrm{Shortcut}(x) = \begin{cases} \mathrm{BN}\left(\mathrm{Conv}_s(x)\right), & \text{if dimensions differ} \\ x, & \text{if dimensions match} \end{cases} \tag{10}$$

2) Squeeze and Excitation Block: FedPapVisionNet model integrates a SE block to enhance channel-wise feature representation by explicitly modeling inter-channel dependencies. The SE mechanism assigns adaptive importance weights to each feature channel, allowing the network to emphasize informative channels while suppressing less relevant ones, thereby improving representational efficiency without substantial computational overhead. The SE operation is mathematically defined as:

$$\mathrm{SE}(x) = x \cdot \sigma\left(W_2 \cdot \mathrm{ReLU}\left(W_1 \cdot \mathrm{GAP}(x)\right)\right) \tag{11}$$

where $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are learnable weight matrices, and $\sigma(\cdot)$ denotes the sigmoid activation function [23].

3) ConvResidualFusion Block: The ConvResidualFusion block is a key architectural component of FedPapVisionNet designed to combine parallel feature extraction strategies. This block consists of two concurrent pathways: a convolutional path composed of standard convolution followed by batch normalization, and a residual path incorporating identity-based skip connections. The outputs from both paths are combined using element-wise summation and passed through a ReLU activation function. This parallel fusion enables the network to benefit simultaneously from direct convolutional transformations and residual learning, resulting in richer and more discriminative feature representations. [23]

4) Global Average Pooling and Classification: After hierarchical feature extraction, Global Average Pooling (GAP) is employed to reduce each feature map to a single scalar value. Compared to flattening operations, GAP significantly reduces the number of trainable parameters, lowers the risk of overfitting, and improves generalization, which is particularly important for medical imaging tasks with limited training data. Furthermore, GAP produces a fixed-dimensional feature vector independent of input resolution, enhancing architectural robustness.

The pooled feature vector is subsequently fed into a lightweight classification head consisting of fully connected layers with batch normalization, ReLU activation, and dropout for regularization.

5) Mathematical Formulation of FedPapVisionNet: The complete forward pass of the FedPapVisionNet model is described step-by-step as follows. The input image is defined as:

$$X_{\text{input}} \in \mathbb{R}^{128 \times 128 \times 3} \tag{12}$$

The initial convolutional feature extraction is given by:

$$\mathrm{Conv2D}(X, W, b) = W * X + b \tag{13}$$

$$X_1 = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv2D}(X_{\text{input}}, W_1, b_1)\right)\right) \tag{14}$$

In Eq. (13), the Conv2D operation uses 32 filters of size 3×3 with same padding, followed by batch normalization and ReLU activation.

Channel-wise recalibration using the SE block is applied as:

$$X_{\text{SE}} = X_1 \cdot \sigma\left(W_2 \cdot \mathrm{ReLU}\left(W_1 \cdot \mathrm{GAP}(X_1)\right)\right) \tag{15}$$

The first ConvResidualFusion block is expressed as:

$$X_{\text{conv}} = \mathrm{ConvPath}(64, 3 \times 3)(X_{\text{SE}}) \tag{16}$$

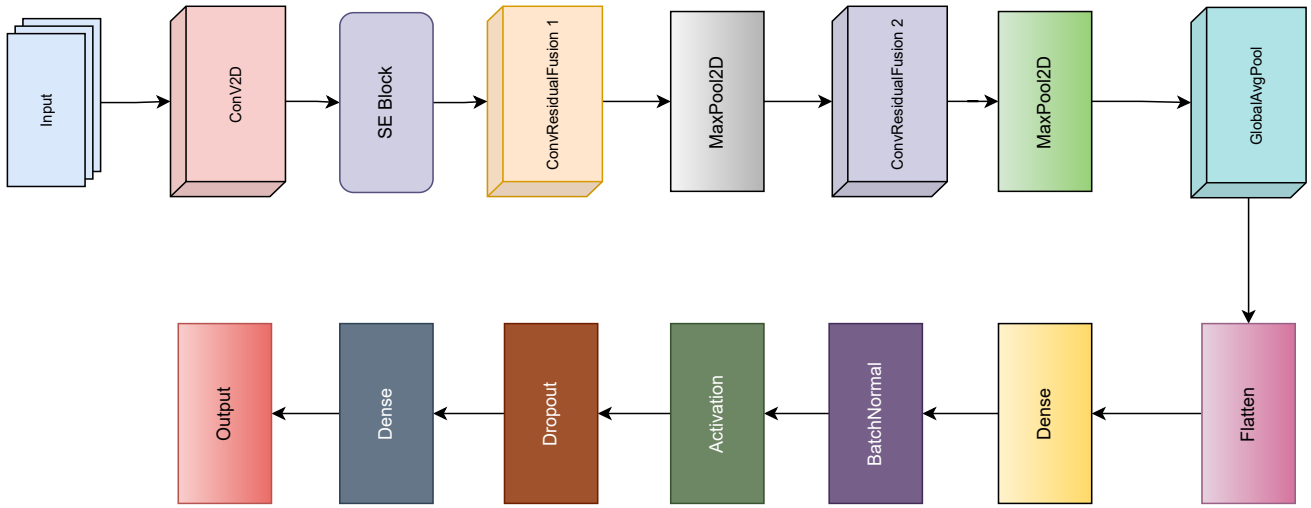


Fig. 6: Architecture overview of the FedPapVisionNet model

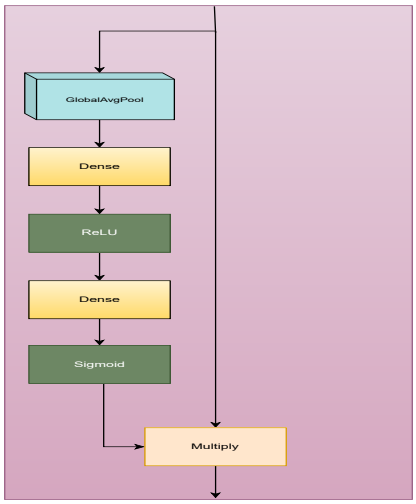


Fig. 7: (a) SE Block

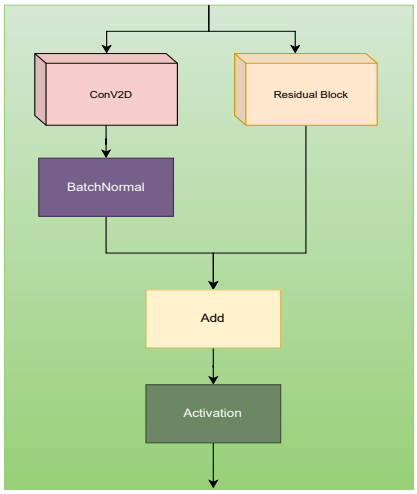


Fig. 8: (b) ConvResidualFusion

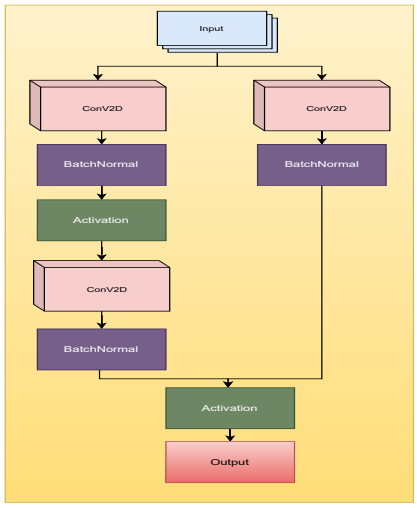


Fig. 9: (c) Residual Block

Fig. 10: Structural components used in the FedPapVisionNet architecture

$$X_{\text{res}} = \text{ResPath}(64, 3 \times 3)(X_{\text{SE}}) \tag{17}$$

$$X_2 = \text{ReLU}(X_{\text{conv}} + X_{\text{res}}) \tag{18}$$

This is followed by max pooling:

$$X_3 = \text{MaxPooling2D}(2 \times 2)(X_2) \tag{19}$$

A second ConvResidualFusion block is defined as:

$$X^{(4)}_{\text{conv}} = \text{ConvPath}\,(128, 3 \times 3)\,(X_3) \tag{20}$$

$$X^{(4)}_{\text{res}} = \text{ResPath}\,(128, 3 \times 3)\,(X_3) \tag{21}$$

$$X_4 = \text{ReLU}\left(X^{(4)}_{\text{conv}} + X^{(4)}_{\text{res}}\right) \tag{22}$$

followed by max pooling:

$$X_5 = \text{MaxPooling2D}(2 \times 2)(X_4) \tag{23}$$

Global average pooling produces:

$$X_6 = \text{GlobalAveragePooling2D}(X_5) \tag{24}$$

The fully connected classification layers are given by:

$$X_7 = \text{Dropout}(0.5)\,(\text{ReLU}\,(\text{BatchNorm}\,(\text{Dense}(64)(X_6)))) \tag{25}$$

Finally, the output layer is defined as:

$$Y = \text{Dense}(5)(X_7) \tag{26}$$

where Y represents the predicted class logits corresponding to the five cervical cell categories.

IV. Results & Discussion

A. Performance Under IID Settings

Under IID data distribution with FedAvg optimization, FedPapVisionNet achieves 95.68% overall accuracy,

TABLE II: Unified Classification Report under IID and Non-IID Data Partitions

Data	Approach	Model	Class	Precision	Recall	F1-score	Support	Accuracy
IID	FedAvg	DualEffiMob-CBAM	Dyskeratotic	0.9375	0.9926	0.9643	136	0.9543
			Koilocytotic	0.9205	0.8742	0.8968	159	
			Metaplastic	0.9277	0.9277	0.9277	166	
			Parabasal	0.9884	0.9827	0.9855	173	
			Superficial–Intermediate	0.9887	0.9943	0.9915	176	
		EffiResAttention	Dyskeratotic	0.9640	0.9853	0.9745	136	0.9593
			Koilocytotic	0.9351	0.9057	0.9201	159	
			Metaplastic	0.9281	0.9337	0.9309	166	
			Parabasal	0.9825	0.9711	0.9767	173	
			Superficial–Intermediate	0.9832	1.0000	0.9915	176	
		FedPapVisionNet	Dyskeratotic	0.9362	0.9706	0.9531	136	0.9568
			Koilocytotic	0.9000	0.9057	0.9028	159	
			Metaplastic	0.9509	0.9337	0.9422	166	
			Parabasal	1.0000	0.9884	0.9942	173	
			Superficial–Intermediate	0.9886	0.9830	0.9858	176	
Non-IID	FedProx	DualEffiMob-CBAM	Dyskeratotic	0.94	0.97	0.96	136	0.96
			Koilocytotic	0.90	0.93	0.92	159	
			Metaplastic	0.97	0.92	0.95	166	
			Parabasal	1.00	0.99	0.99	173	
			Superficial–Intermediate	0.99	1.00	0.99	176	
		EffiResAttention	Dyskeratotic	0.95	0.96	0.95	136	0.96
			Koilocytotic	0.89	0.92	0.90	159	
			Metaplastic	0.96	0.91	0.93	166	
			Parabasal	0.99	0.99	0.99	173	
			Superficial–Intermediate	0.98	1.00	0.99	176	
		FedPapVisionNet	Dyskeratotic	0.90	0.92	0.91	136	0.86
			Koilocytotic	0.69	0.86	0.76	159	
			Metaplastic	0.94	0.57	0.71	166	
			Parabasal	0.92	0.95	0.94	173	
			Superficial–Intermediate	0.90	0.99	0.94	176	
Non-IID	SCAFFOLD	FedPapVisionNet	Dyskeratotic	0.86	0.90	0.88	136	0.90
			Koilocytotic	0.81	0.80	0.81	159	
			Metaplastic	0.92	0.86	0.89	166	
			Parabasal	0.98	0.95	0.96	173	
			Superficial–Intermediate	0.93	0.99	0.96	176	

demonstrating strong performance competitive with larger architectures while using significantly fewer parameters. The class-wise analysis reveals balanced performance as shown in table Comparative evaluation with pretrained architectures shows that while Vision Transformer achieves the highest accuracy (96.54%), and EffiResAttention reaches 95.93%, DualEffiMob-CBAM achieves the accuracy of 95.43% whereas FedPapVisionNet achieves 95.68% accuracy with fewer parameters (390K vs. 8.5M). Among lightweight pretrained models, EfficientNetB0 achieves 97.16% and MobileNetV2 achieves 95.93% accuracy, but these require 5.3M and 3.5M parameters respectively—substantially larger than FedPapVisionNet. The confusion matrix analysis reveals that the primary source of errors occurs between the clinically adjacent Koilocytotic and Metaplastic classes (10 misclassifications in each direction), which exhibit similar morphological features. This pattern is consistent across all evaluated models, suggesting inherent ambiguity in distinguishing these cell types rather than a model-specific limitation. Critically, the clinically stable classes (Parabasal and Superficial-Intermediate) maintain near-perfect classification (171/173 and 173/176 correct respectively), indicating reliable performance on diagnostically clear cases.

B. Performance Under non-IID Settings

Under non-IID condition ($\alpha = 0.3$) optimization with FedProx, FedPapVisionNet is able to achieve an accuracy of 86%. This shows a performance drop of 9.43% as compared to the IID condition. The Metaplastic class shows most degradation having a low recall of 0.57 as there are 48 instances misclassified as Koilocytotic. The class-wise distribution. Dyskeratotic (0.90, 0.92, 0.91), Koilocytic (0.69, 0.86, 0.76), Metaplastic (0.94, 0.57, 0.71), Parabasal((0.92, 0.95, 0.94) Superficial-Intermediate (0.90, 0.99, 0.94).

On the other hand, larger capacity hybrid ones, such as the DualEffiMob-CBAM and EffiResAttention, can achieve 96% accuracy under the same non-IID conditions with FedProx and show minimal degradation in performance. Due to their superior representational capacity, they can handle heterogeneous client distributions well. Nevertheless, this benefit requires an increase in parameters amounting to 21.8 times and 42.7 times, which causes significant overhead due to communication in federated settings. Utilizing SCAFFOLD on FedPapVisionNet with the same $\alpha = 0.3$ non-IID partition greatly enhances performance to 90% accuracy. The control variate method effectively reduces client drift, with a particularly beneficial impact on the previously difficult Metaplastic class (recall

TABLE III: Accuracy Comparison: Proposed Methods vs. Baseline

Method	IID (%)	Non-IID (%)
Baseline	94.36 (FedAvg)	78.40 (FedAvg)
FedPapVisionNet	95.68 (FedAvg)	90.00 (Scaffold, $\alpha = 0.3$)
		80.00 (Scaffold, $\alpha = 0.1$)

improves from 0.57 to 0.86). In class-wise metrics obtained under Scaffold, Dyskeratotic (0.86, 0.90, 0.88), Koilocytotic (0.81, 0.80, 0.81), Metaplastic (0.92, 0.86, 0.89), Parabasal (0.98, 0.95, 0.96) and Superficial-Intermediate (0.93, 0.99, 0.96). As such, algorithmic optimization could mitigate the effects of model capacity reduction, as these results show.

C. Extreme Heterogeneity ($\alpha = 0.1$)

To stress-test the proposed framework under severe data heterogeneity, we evaluated FedPapVisionNet with SCAFFOLD at $\alpha = 0.1$, representing an extreme scenario where client data distributions are highly skewed. Despite this challenging setting, the model achieves 80% accuracy, surpassing the baseline federated approach by Joynab et al. [19] which reported 78.43% accuracy under non-IID conditions. The training and validation curves demonstrate stable convergence without divergence, validating the robustness of the drift-correction mechanism. This result confirms that FedPapVisionNet with SCAFFOLD can maintain reasonable performance even under the most heterogeneous data distributions encountered in multi-institutional medical collaborations.

D. Convergence Analysis

In terms of optimization dynamics by looking at the loss curves, when employing FedAvg under IID settings, all three architectures deliver a high impact in the beginning, displaying a rapid loss reduction. After achieving the loss drop, all models attain a smooth convergence phase. The final training and validation loss for FedPapVisionNet lie closely at 0.21 and 0.14 respectively, indicating little overfitting and good generalization despite its small size. Non-IID conditions of FedProx lead to larger models with near zero training loss (0.01) and low validation loss (0.13-0.16). The loss of the FedPapVisionNet is bigger (training: 0.23, validation: 0.39) indicating that it is more challenging to fit heterogeneous data with limited capacity. The accuracy curves exhibit a similar behaviour: the DualEffiMob-CBAM and EffiResAttention networks quickly achieve 0.99 training accuracy as well as 0.96 validation accuracy. Conversely, FedPapVisionNet achieves training accuracy of 0.89 and validation accuracy of 0.86 and plateaus out. Using SCAFFOLD changes the convergence patterns of FedPapVisionNet significantly. The decrease in validation loss is faster and stabilizes at lower value compared to FedProx. Similarly, the validation accuracy steadily climbs to 0.90. The improvement comes from SCAFFOLD's explicit correction of local gradient bias. The local gradient bias correction prevents the model from being too pulled towards client-specific optima that do not generalize to the global validation distribution.

E. Parameter Drift Analysis

Parameter drift, measured as the L2 distance between local and global model weights, quantifies client divergence during federated training. Under non-IID conditions with FedProx, DualEffiMob-CBAM exhibits the lowest drift (79.5), indicating strong alignment between local updates and the global model. EffiResAttention shows moderate drift (82-83), while FedPapVisionNet displays the highest and most volatile drift, peaking at 83.4 around rounds 18-20 before settling near 82.2. This non-monotonic behavior reveals that lightweight models are more susceptible to conflicting gradient directions from heterogeneous clients, particularly during the critical middle phase of training when the model is transitioning from initial feature learning to fine-grained classification refinement.

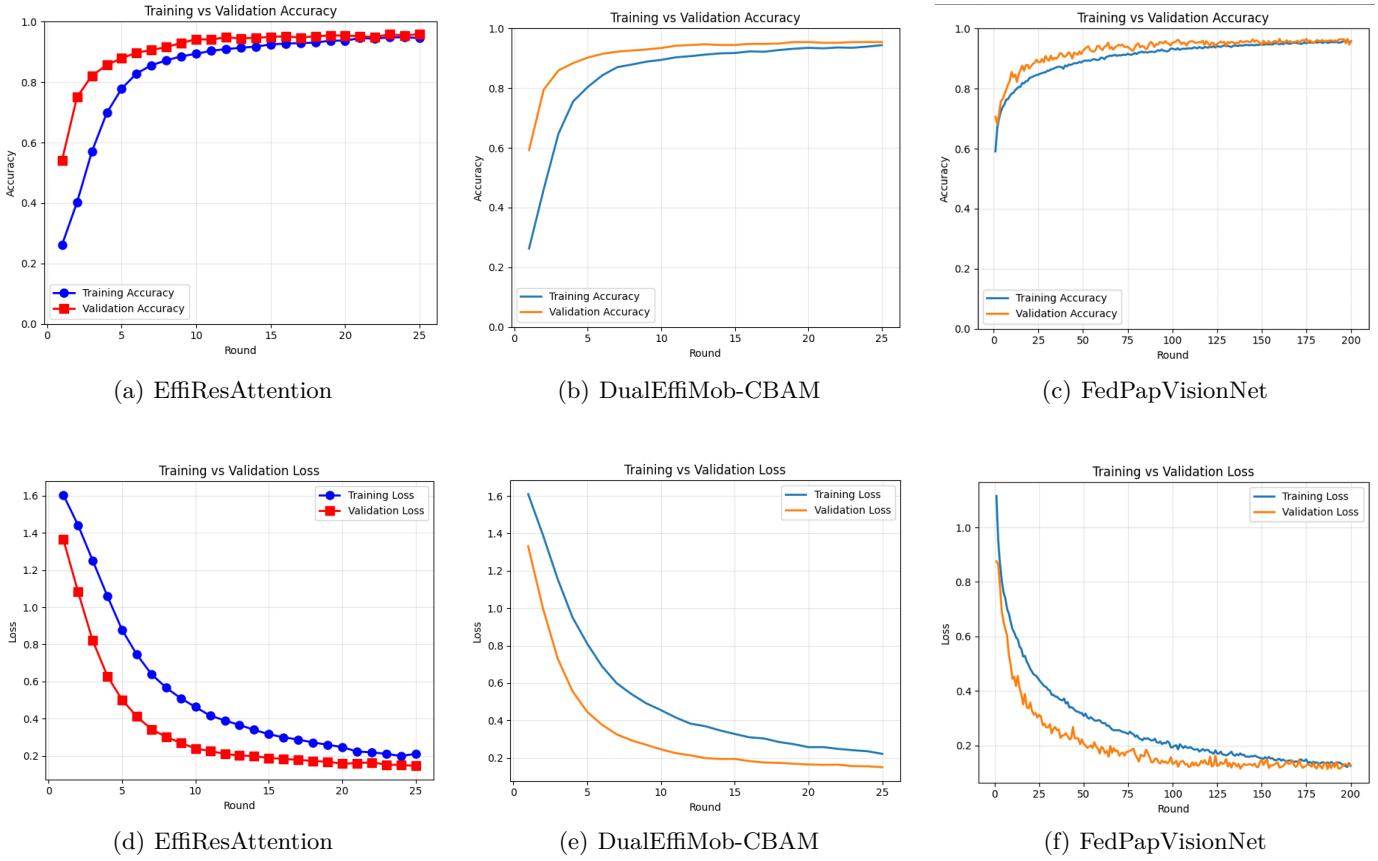
F. Computational Efficiency

The parameter efficiency of FedPapVisionNet (390K) compared to DualEffiMob-CBAM (8.5M) and EffiResAttention (16.6M) translates directly to communication efficiency in federated learning. Each communication round requires transmitting model updates proportional to the number of parameters. With FedPapVisionNet, the communication payload is reduced by 95.4% compared to DualEffiMob-CBAM and 97.6% compared to EffiResAttention. Over 50 communication rounds in a non-IID setting, this represents substantial bandwidth savings—critical for deployment in resource-constrained clinical environments with limited network infrastructure. The modest accuracy trade-off (6-10% under non-IID) is justifiable in scenarios where communication costs, client computational resources, or privacy constraints are primary concerns.FedPapVisionNet is more efficient in parameters than DualEffiMob-CBAM and EffiResAttention, which contributes to its communication efficiency in federated learning. In each communication round, model updates proportional to the number of parameters are transmitted. The communication payload is reduced by 95.4% compared to DualEffiMob-CBAM and 97.6% compared to EffiResAttention for FedPapVisionNet. In a non-IID setup, 50 rounds of communication reflect important savings in bandwidth which is necessary for deploying in a clinical setting.

V. Discussion

A. Implications for Clinical Deployment

According to experimental results, cervical cancer screening using federated learning can obtain clinically acceptable performance without compromising patient privacy. With a 95.68% IID accuracy and 90% non-IID accuracy (with SCAFFOLD), our results surpass the regular error rates (10-15%) of human cytopathologists



(a) EffiResAttention (b) DualEffiMob-CBAM (c) FedPapVisionNet

(d) EffiResAttention (e) DualEffiMob-CBAM (f) FedPapVisionNet

Fig. 11: Training and validation accuracy (top row) and loss (bottom row) curves for different models under IID data partition using the FedAvg algorithm

and gets close to the performance levels observed in centralized studies. Most notably, clinically stable categories (Parabasal, Superficial-Intermediate) classified with near perfect performance show that our model is capable of reliably identifying definitively normal cells or negative false rates. Thus, it can reduce unnecessary follow-up. The remaining challenges, were highlighted as those, faced in distinguishing between Koilocytotic and Metaplastic cells, which known manual cytopathology presents, where similar type usually requires additional clinical context for diagnosis. In practice, the model was potentially implemented as a firstline screening tool to flag potentially abnormal samples for expert review, rather than provide fully autonomous diagnosis.

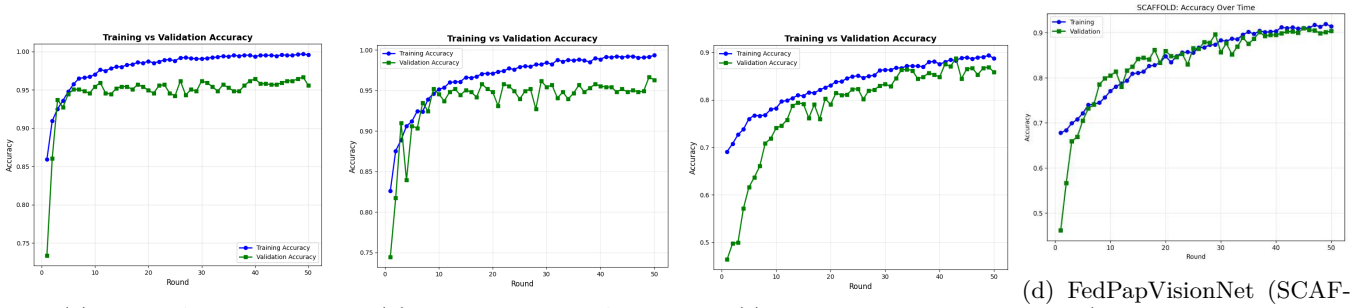
B. Privacy-Utility Trade-offs

The federated learning framework successfully maintains data privacy by keeping raw images within each institution. However, privacy guarantees rely on the assumption that model updates do not leak sensitive information—a limitation not addressed in this study. Future work should incorporate differential privacy mechanisms to provide formal privacy guarantees, though this typically introduces additional accuracy degradation. The current 5-10% non-IID accuracy gap leaves headroom for such privacy-enhancing techniques while still maintaining clinically useful performance above 80-85%.

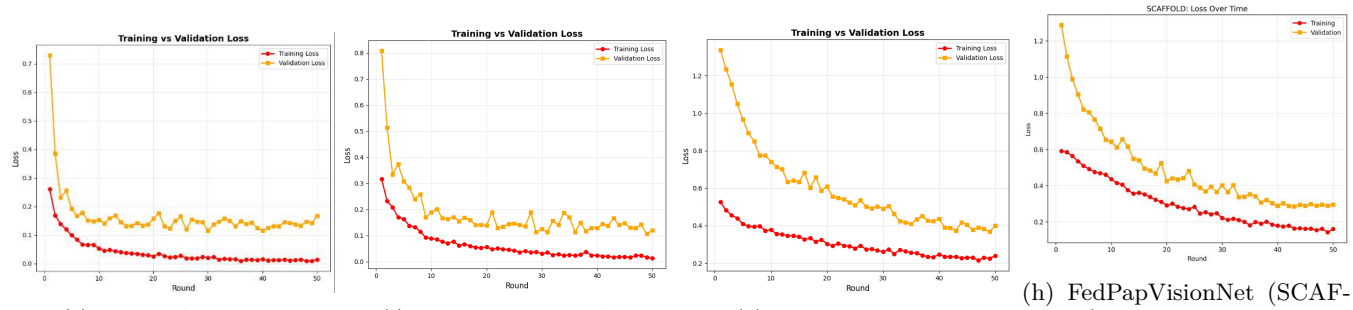
C. Limitations

The findings have limited generalizability due to many limitations. To begin with, experiments conducted on a single dataset SIPaKMeD with only 4049 images may not reflect the morphological variation seen in large-scale clinical practice. The three clients simulated federated setting does not represent real world scenario complexities like varied client availability, network latency, device variety, or malicious participants. The third assumption is that communication rounds are synchronized, whereas real-world federated systems often require asynchronous aggregation because not all clients will always connect.

Furthermore, the assessment solely examines individual images from a single cell, as clinical Pap smear slides contain multiple overlapping cells. Moving from single-cell classification to classification of the whole slide analysis incurs additional technical difficulties not addressed here. Ultimately, the research does not assess performance on external data or real clinical workflows and therefore limits conclusions about deployment readiness.



(a) EffiResAttention (b) DualEffiMob-CBAM (c) FedPapVisionNet (d) FedPapVisionNet (SCAFFOLD)



(e) EffiResAttention (f) DualEffiMob-CBAM (g) FedPapVisionNet (h) FedPapVisionNet (SCAFFOLD)

Fig. 12: Training and validation accuracy (top row) and loss (bottom row) curves for different models under Non-IID data partition

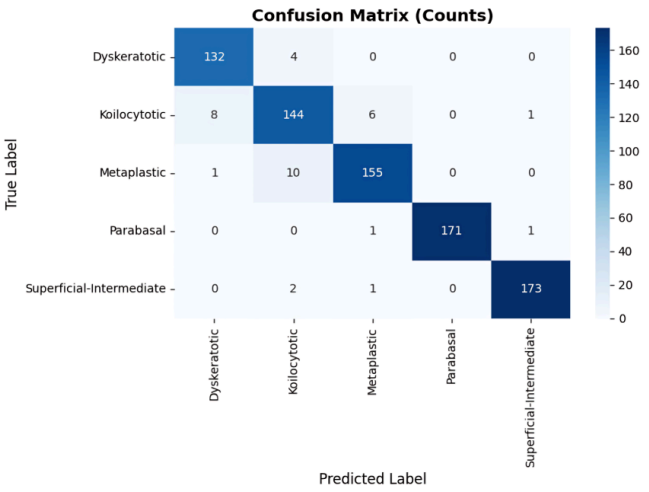
VI. CONCLUSION AND FUTURE WORK

This study presents FedPapVisionNet, a lightweight federated learning framework that uses Pap smear images for cervical cancer detection. With only 390,215 parameters, the proposed architecture reaches competitive accuracy requirements of 95.68% (IID), and 90% (non-IID with SCAFFOLD), showing good trade-off characteristics. Analysis of federated optimization techniques through benchmark tasks revealed that drift-awareness SCAFFOLD substantially improves performance in the presence of data heterogeneity as the recall of the Metaplastic class improves from 57% to 86%. The lightweight designs were shown to have practical advantages in resource-constrained federated settings through extensive comparison to pretrained and hybrid models.

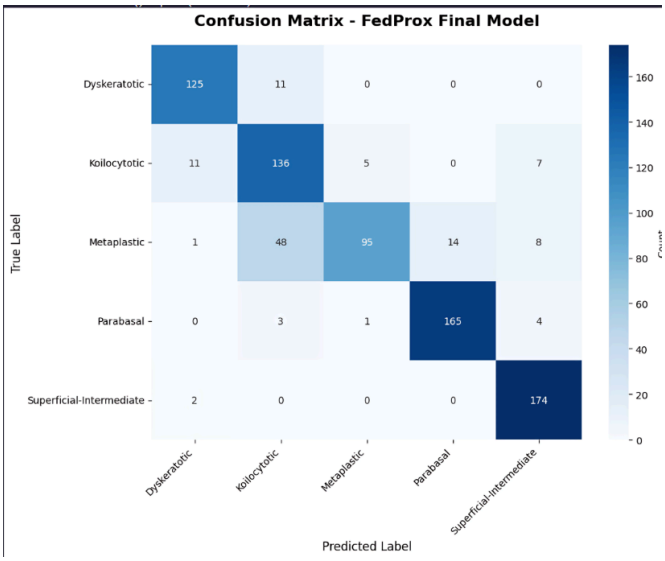
Future research will focus on incorporating differential privacy to attain formal privacy guarantees with bounded accuracy degradation, extending the framework to whole-slide image analysis along with automated cell segmentation, measuring performance on external datasets and multi-institutional real-world deployments, investigating asynchronous federated optimization to address availability-constrained clients, developing explainable AI techniques to improve clinical trust and interpretability, and examining semi-supervised/self-supervised learning and its applicability to federated learning with unlabeled data. Thanks to these advances, federated medical imaging can move from the confined walls of experiment-controlled settings to actual clinical use. This will enable privacy-preserving collaborative learning to improve worldwide cervical cancer screening.

VII. Acknowledgement

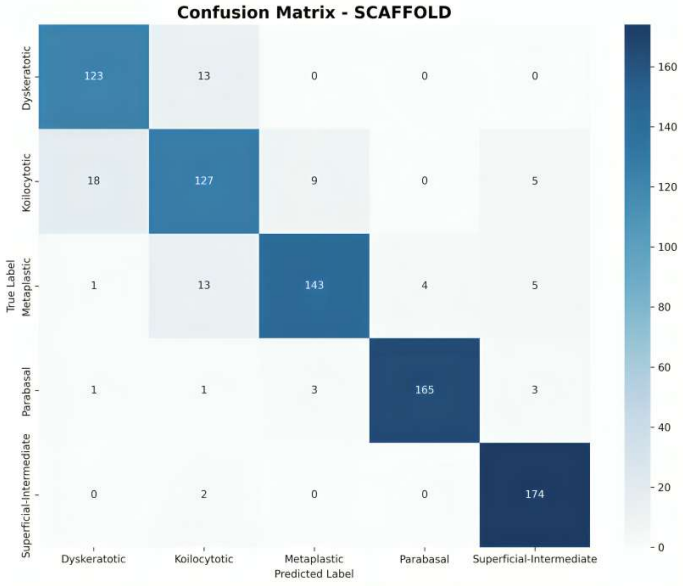
The contribution of the Department of Computer Science and Engineering of BRAC University with computational resources and research is highly acknowledged in this work. The authors would also like to express gratitude to Dr. Md. Ashraful Alam for supervising and guiding throughout the research.



(a) FedPapVisionNet under IID data partition.



(b) FedPapVisionNet under Non-IID data partition.



(c) Performance of SCAFFOLD-based federated model under Non-IID setting.

Fig. 13: Confusion matrices comparing different models under Non-IID data distribution.

References

[1] S. L. Tan, G. Selvachandran, W. Ding, R. Paramesran, and K. Kotecha, "Cervical cancer classification from pap smear images using deep convolutional neural network models," Interdiscip. Sci. Comput. Life Sci., vol. 16, no. 1, pp. 16–38, 2024.

[2] H. Ashtarian, E. Mirzabeigi, E. Mahmoodi, and M. Khezeli, "Knowledge about cervical cancer and pap smear and the factors influencing the pap test screening among women," Int. J. Community Based Nurs. Midwifery, vol. 5, no. 2, p. 188, 2017.

[3] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images," in Proc. 25th IEEE Int. Conf. Image Process. (ICIP), 2018, pp. 3144–3148.

[4] I. Pacal, "Investigating deep learning approaches for cervical cancer diagnosis: a focus on modern image-based models," Eur. J. Gynaecol. Oncol., vol. 46, no. 1, 2025.

[5] A. Khamparia, D. Gupta, J. J. P. C. Rodrigues, and V. H. C. de Albuquerque, "DCAVN: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network," Multimedia Tools Appl., vol. 80, pp. 30399–30415, 2021.

[6] M. Rahimi, A. Akbari, F. Asadi, and H. Emami, "Cervical cancer survival prediction by machine learning algorithms: a systematic review," BMC Cancer, vol. 23, no. 1, p. 341, 2023.

[7] P. Chatterjee, S. Siddiqui, R. S. A. Kareem, and S. R. Rao, "Multi-modal graph neural networks for colposcopy data classification and visualization," Cancers, vol. 17, no. 9, p. 1521, 2025.

[8] N. Sompawong, J. Mopan, P. Pooprasert, W. Himakhun, K. Suwannarurk, J. Ngamvirojcharoen, T. Vachiramon, and C. Tantibundhit, "Automated pap smear cervical cancer screening using deep learning," in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), 2019, pp. 7044–7048.
[9] C. Yuan, Y. Yao, B. Cheng, Y. Cheng, Y. Li, Y. Li, X. Liu, X. Cheng, X. Xie, J. Wu et al., "The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images," Sci. Rep., vol. 10, no. 1, p. 11639, 2020.
[10] H. Kaur, R. Sharma, and J. Kaur, "Classification of cervical cancer from pap smear images using deep learning: A comparison of transfer learning models," 2024.
[11] D. D. Himabindu, E. L. Lydia, M. V. Rajesh, M. A. Ahmed, and M. K. Ishak, "Leveraging swin transformer with ensemble of deep learning model for cervical cancer screening using colposcopy images," Sci. Rep., vol. 15, no. 1, p. 7900, 2025.
[12] B. Z. Wubineh, A. Rusiecki, and K. Halawa, "Segmentation and classification techniques for pap smear images in detecting cervical cancer: A systematic review," IEEE Access, 2024.
[13] P. Khanarsa and S. Kitsiranuwat, "Deep learning-based ensemble approach for conventional pap smear image classification," ECTI Trans. Comput. Inf. Technol. (ECTI-CIT), vol. 18, no. 1, pp. 101–111, 2024.
[14] M. Subramanian, V. Rajasekar, S. VE, K. Shanmugavadivel, and P. S. Nandhini, "Effectiveness of decentralized federated learning algorithms in healthcare: a case study on cancer classification," Electronics, vol. 11, no. 24, p. 4117, 2022.
[15] E. Darzidehkalani, M. Ghasemi-Rad, and P. M. A. Van Ooijen, "Federated learning in medical imaging: part II: methods, challenges, and considerations," J. Am. Coll. Radiol., vol. 19, no. 8, pp. 975–982, 2022.
[16] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. Artif. Intell. Statist., 2017, pp. 1273–1282.
[17] M. U. Nasir, O. K. Khalil, K. Ateeq, B. S. A. Almogadwy, M. A. Khan, and K. M. Adnan, "Cervical cancer prediction empowered with federated machine learning," Comput. Mater. Continua, vol. 79, no. 1, 2024.
[18] J. Peta and S. Koppu, "Enhancing breast cancer classification in histopathological images through federated learning framework," IEEE Access, vol. 11, pp. 61866–61880, 2023.
[19] N. S. Joynab, M. N. Islam, R. R. Aliya, A. R. Hasan, N. I. Khan, and I. H. Sarker, "A federated learning aided system for classifying cervical cancer using pap-smear images," Inform. Med. Unlocked, vol. 47, p. 101496, 2024.
[20] X. Yuan and P. Li, "On convergence of fedprox: Local dissimilarity invariant bounds, non-smoothness and beyond," in Proc. Adv. Neural Inf. Process. Syst., vol. 35, 2022, pp. 10752–10765.
[21] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in Proc. Int. Conf. Mach. Learn., 2020, pp. 5132–5143.
[22] T. S. B. Ahmed, T. Rahman, S. Biswas, S. R. Sabuj, M. B. Bhuian, M. A. Moni, and M. A. Alam, "A vision transformer-based hybrid neural architecture for automated handwritten Bangla character recognition and braille conversion," Knowl.-Based Syst., vol. 114546, 2025.'
[23] M. H. K. Mehedi, M. Khandaker, S. Ara, M. A. Alam, M. F. Mridha, and Z. Aung, "A lightweight deep learning method to identify different types of cervical cancer," Sci. Rep., vol. 14, no. 1, p. 29446, 2024.