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

by

Fahim Shahriar Ahmed
22101511
Chowdhury Saima Rawshan
24141280
Sayani Das
22101754
Sadia Habib Nizhum
24241224
Sackline Naien Ridvi
22101631

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2026.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Fahim Shahriar Ahmed

Chowdhury Saima Rawshan

Sayani Das

Sadia Habib Nizhum

Sackline Naien Ridvi

Approval

The thesis titled "A Federated Learning based efficient approach to detect cervical Cancer Using PAP-SMEAR images" submitted by

1. Fahim Shahriar Ahmed (22101511)
2. Chowdhury Saima Rawshan(24141280)
3. Sayani Das(22101754)
4. Sadia Habib Nizhum (24241224)
5. Sackline Naien Ridvi(22101631)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in 31^{st} January, 2026.

Examining Committee:
Supervisor:
(Member)

Dr. Md. Ashraful Alam

Associate Professor
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

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.In the present thesis work, proposes an efficient privacy-preserving federated learning based framework for cervical cancer detection using Pap smear images.Detection of cervical cancer is first done with comparative experimental analysis for pretrained and hybrid lightweight models such as VGG16, ResNet18, ResNet50, DenseNet121, EfficientNetB0, Vision Transformer and MobileNetV2 in iid and non-iid data distribution.In addition, FedPapVisionNet, a lightweight mechanism for federated modelling was introduced which improves classification performance while safeguarding privacy.It merges the efficient convolutional operation and attention mechanism.With only 3,90,215 trainable parameters, it achieves impressive diagnostic accuracy on datasets.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.

Keywords: Federated Learning, Cervical Cancer, PAP-SMEAR Images, Convolutional Neural Networks (CNN), Privacy-Preserving Machine Learning, Decentralized Learning, Medical Image Classification, Data Privacy, Artificial Intelligence in Healthcare, Early Cancer Detection, Non-IID Data Partition, FedAvg, FedProx, Scaffold.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Cervical Cancer is one of the deadliest of Cancers in women.Millions of women are affected every year as followed by WHO.Early identification and adequate and timely attention are vital to decrease morbidity and mortality of the disease.Pap smear has long been the most reliable method for cervical screening and for detection of precancer and cancer.However, manually analyzing pap smear images is a slow,effort-intensive process and vulnerable to inaccuracies caused by human mistakes.Therefore, there is an increasing demand for computer-based objective and efficient diagnostic systems that aid in efficient and correct cervical cancer screening performed by professionals.

Latest innovations in AI and deep learning have proven to be very promising in medical imaging, particularly with the usage of CNNs.Such deep learning models can identify intricate patterns in medical images, and provide an effective way to sensitively and accurately detect the abnormal and cancerous cells in the pap smear slides.

1.2 Motivation

Motivation for this research is that one of the biggest obstacles for integrating AI in healthcare remains in the confidentiality and protection of potentially delicate patient data.Federated Learning (FL) provides an optimistic approach to this issue.FL permits ML model training via collective learning on a large number of isolated devices or institutions, while ensuring the patient's data remains local to it, and is not shared outside of the entity that possesses the data.While ensuring the privacy retention in the setting, only the model updates are communicated, and collaborative learning can be conducted using heterogeneous datasets.

In this paper we offer a solution for efficient and accurate identification of cervical cancer using Pap smear images in which we present Federated Learning-based method.The goal is to devise a system that can enhance the accuracy of diagnosis without compromising patient privacy, by applying deep learning and federated learning techniques.

1.3 Problem Statement

Cervical cancer still constitutes an important global public health problem, and its early detection contributes to diminish its mortality. Pap smear testing is important for screening, but the manual interpretation of such images is slow, with a risk of human error and lacks of capabilities for detecting the diseases in the early stage. Current diagnostic approaches are hampered to early diagnose cervical cancer prior to a large damage is done, making an efficient therapeutic action difficult. While deep learning has the capability to automate image interpretation,Privacy-related challenges make it challenging to adopt AI widely in healthcare. Existing instances face difficulty of preserving the patient privacy authority among the availability of the huge data required to provide proper knowledge to the machine learning models. Efficient, precise, and privacy-preserving methods are urgently needed for this cancer detection with Pap smear images.

1.4 Objective

The objective of this study is to establish 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.

- The purpose of the project is developing a Federated Learning-based system for collaborative training across the disjointed medical institutions without compromising on the privacy of raw patient data.
- To incorporate high-performing deep learning models like CNN models to classify cervical cell Pap smear images.
- To explore the how effectively transfer learning, class balancing and data augmentation works to enhance topical modeling performance on heterogeneous and small data.
- Preserving data privacy and security during training with federated or differential privacy, to be compliant with medical data protection regulations.
- To evaluate how federated learning model performs in IID and non-IID manner regarding real-world disparities among medical institutions.
- To conduct a comparative analysis to decide the advantages and disadvantages of a federated model in comparison with a centralized model through the benchmarking process with the help of popular evaluation metrics such as precision, recall, accuracy and F1-score.
- To prepare a scalable, applicable system for early stage Cervical Cancer diagnosis and increased diagnosis in resource constrained and privacy sensitive medical environments.

1.5 Methodology in brief

Data and Preprocessing of the Data

In this work, we will use the SIPaKMeD Dataset, which is a public dataset for cervical cancer detection. This dataset has 4049 Pap smear images consisting 5 different classes: Dyskeratotic cells, Koilocytotic cells, Metaplastic cells, Parabasal cells and Superficial-Intermediate cells. The size of each image in the data set is 66x66 pixels, corresponding to isolated cells.
In order to train deep learning models, the SIPaKMeD dataset is highly beneficial, with a rich set of annotated Pap smear images for both classifying and detecting cervical cancer. The dataset will undergo some operations like resize every image to the fixed size of 66x66 pixels, augmentation techniques including rotation, flip, zoom, and shear for a data feeding purpose, normalization to make the data more consistent, segmentation to enable the model.

Federated Learning Data Split: The dataset will also be fragmented across multiple simulated clients (each approximately representing a different hospital or healthcare center) that are asked to perform computations locally in a distributed learning environment so as not to violate privacy. All clients will have only part of the data, which guarantees the data privacy, but supports collaboration of model training without sharing data information.

The Federated Learning Framework

The key purpose of the study is to use Federated Learning (FL) which enables decentralized, privacy sensitive training of machine learning models. The processing is as follows for the execution:

- **Client Setup:** The task of mimicking multiple clients (hospitals) with different subscriptions of the global dataset. The dataset will be divided into IID and non-IID distributions in order to evaluate the model in practical real world use.
- **Model Initialization:** We initialize the proposed model globally and share it with all the clients. Local model updates are trained by each client on its own data.
- **Model Training and Aggregation:** Clients train their own models separately and then upload the updated weights to the server. The server then works with the updates according to Federated Averaging (FedAvg) for IID and FedProx and Scaffold algorithm for non-IID data partition, and sends the updated model back to the clients.
- **Federated hyper-parameter optimization:** Optimize hyperparameters to maximize model performance.

Hybrid Model Development

During this stage, the proposed methodology was carried out through the following steps:

- **Best Model Integration for Federated Learning:** Initially, several well-established deep learning architectures were selected based on their proven performance in medical image classification. These included VGG16, MobileNetV2, ResNet18, ResNet50, DenseNet121, EfficientNetB0, and Vision Transformer (ViT). Each of these pretrained models was integrated with a Federated Learning (FL) framework to enable privacy-preserving training across distributed datasets. This phase focused on evaluating their performance, convergence behavior, and computational cost under both IID and non-IID federated settings.

- **Hybrid Models with Federated Learning:** Drawing from the findings from the pre-trained baseline models, hybrid architectures were devised. Enhancing the representation of features while keeping it light.appropriate for dispersed learning. Efficiently combined the architectures of EfficientNet, ResNet, and attention to achieve performance.Blocks involving convolution, residual connection, and attention. Developing.Based on these designs, FedPapVisionNet was the final proposed hybrid architecture.It was created as a lightweight model that preserves discriminative. By decreased model complexity, it is capable making it what.federated and resource-limited settings.

- **Efficient Convolution and Attention Mechanisms:** The hybrid architectures leverage mobile-style convolutions inspired by EfficientNet and MobileNet as well as residual connections to facilitate effective gradient flow in federated training. Moreover, a Convolutional Block Attention Module (CBAM) will be incorporated for channel-wise and spatial-wise attention [20] which will enable the model to look at essential features of cervical cell images.[20].

- **Lightweight Model Optimization for Federated Learning:** The last proposed model, FedPapVisionNet, was optimized for reduced computation and communication cost with competitive classification accuracy. This model is trained with the FedAvg for IID data distribution and FedProx technique and later SCAFFOLD technique for non-IID data distribution, which helps to reduce client drift. This is particularly useful when the data distribution is non-IID.

Training Phase

With the federated learning architecture and hybrid models in place, the training stage will move on as follows:

- **Training the Model on the clients:** Every client will train the hybrid model on their local data. These models will use CNN and Transformer layers to classify the images in the correct categories.

- **Client-specific Model Fine-tuning:** Each client's model will be fine-tuned on its local dataset and the updates are notified to the central server.

- **Federated Aggregation:** The server will aggregate the update models to construct a universal model. This process will iterate on the validation set until desired accuracy for the model is acheived.

Evaluation

Both qualitative and quantitative evaluation approaches will be applied to assess how the model performs such as: Visualizations to create feature maps, Confusion Matrix to analyze model prediction for all the classes and metrics like accuracy, recall, precision, F1-Score to compare the traditional machine learning and other deep learning models' performance.

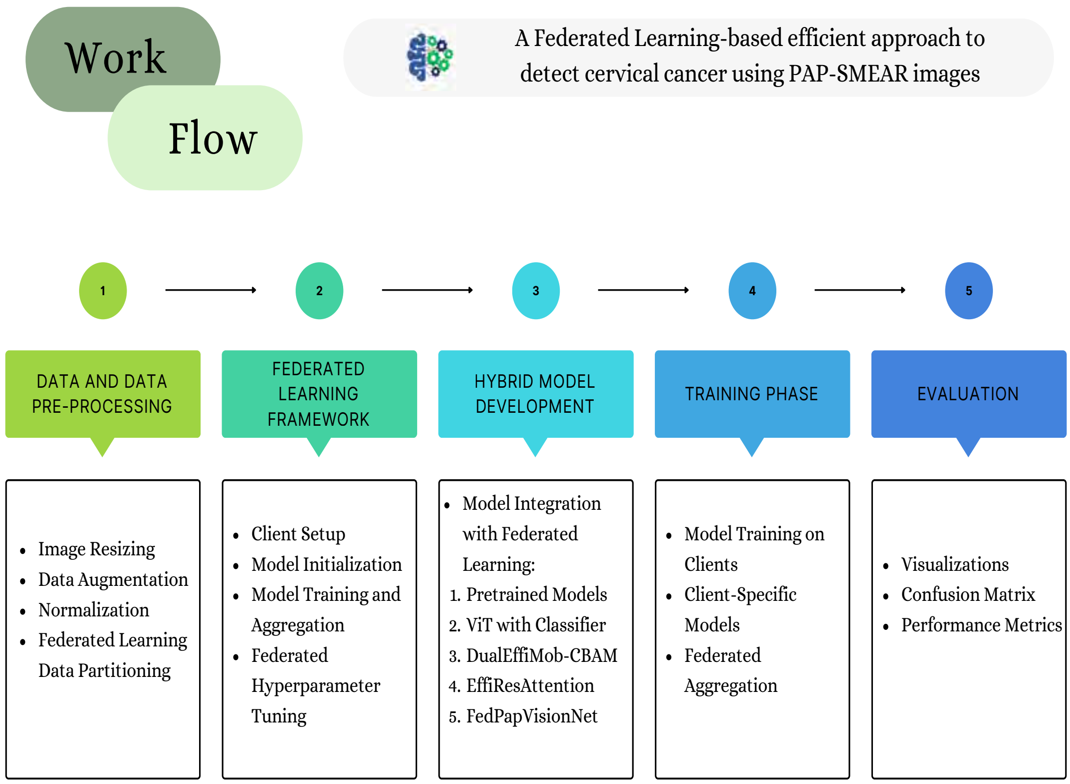


Figure 1.1: Workflow

1.6 Scope and Challenges

This study is dedicated to constructing a privacy-preserving and efficient system using Pap smear images for detecting cervical cancer with a combination usage of Federated Learning (FL) and Deep Learning (DL) models. This involves implementing FL across decentralized datasets, utilizing CNN-based models for image classification, and analyzing performance under IID and non-IID distributions.

However, privacy is one of the most important issues that need to be resolved during federated training, including potential model-updates breaches. Heterogeneity of data across institutions and class imbalance of medical data, may prevent the two-step model to be generalized. It is also non-trivial to get the model converged and deal with the communication overhead in the federated scenario. Also, there is a necessity for using light weight models where resources are limited.

Chapter 2

Literature Review

2.1 Preliminaries

beginsloppypar The following section offers a structured summary of the research's literature review, detailing the structure of its organization, and the theories, concepts, definitions, and models upon which this study is built. The proposed review includes various research works from 2017 to 2025 and mainly based on detection of cervical cancer, deep learning models and federated learning in medical image processing. This timeline shows the progress of research on theoretical and practical applications of deep learning such as automated detection of cervical carcinoma based on Pap smear and colposcopy images.

This article starts with the progress and the issues on the detection of cervical cancer, primarily by the Pap smear test as the general approach of screening. Studies in this field are directed to automate Pap smear image interpretation relying on deep learning techniques (e.g., CNN, ViTs and Autoencoders). Compared the results of performances among the top models of DenseNet, VGG, ResNet, Xception in the way they classify cervical cells and the abnormalities of cervical tissues.

The papers include a comprehensive survey on deep learning models of medical imaging in cervical cancer, too. We present a review of CNNs which have been extensively applied mainly via their capability to self-learn hierarchical features of images and automate the extraction of the features. Furthermore, ViTs and VAEs are assessed to check their potential to deal with medical image datasets and enhancing model performance.

In the second part of the review, we discuss FL in healthcare, with specific focus on its use in medical image classification. The importance of privacy-preserving processing for health data is highlighted in studies with models such as FedAvg, FedProx, and hybrid federated ones to address non-IID data. The ability for federated learning to make use of decentralized data sources in order to train models while ensuring patient privacy is a crucial aspect for healthcare based applications.

In the process of conducting the literature review, different models and their performance indicators are reviewed critically. When using cervical cell classification, cancer based tests as an example and many other, common evaluation metrics such as the recall, F1-score, precision, AUC and accuracy are often used to assess the performance of classifying deep models. The review also points out the dataset diversity, data augmentation, and hyperparameter tuning may help improve this best performance and generalization of the model.

This review combines some of the machine learning concepts including transfer learning, supervised learning and ensemble learning. Transfer learning, in particular, is emerging as a crucial methodology in utilizing pre-trained models to address the issue that large well-annotated datasets are often difficult to obtain in medical imaging.

2.2 Review of Existing Research

2.2.1 Cervical Cancer Detection Using Deep Learning Models

Cervical cancer identification has been studied a lot due to its negative effect on women's health in the world. Pap smears, while effective at identifying precancerous and cancerous cells, struggle with limitations relating to high human dependence, slow throughput, and errors. Detecting cervical cancer from Pap smear images being one of the most intricate yet, rapid advancements in machine learning and deep learning based models, like Convolutional Neural Networks (CNN), Vision Transformers (ViT), Autoencoders etc, have demonstrated remarkable results in automation and accuracy for detecting cervical cancer from Pap smear images. There have been various attempts to overcome the drawbacks of diagnostic tools based on different methods, datasets or model structures. Here are some of the key studies that investigated these strategies.

A study conducted by Sher Lyn Tan et al.[18] explores the difficulties related to the detection of cervical cancer, specifically focusing on automating the classification of cervical cells using deep learning techniques from Pap smear images. The paper evaluates the performance of using different CNN architectures (like VGG-16; ResNet-50; DenseNet-121) applying transfer learning on the open-access Herlev dataset. According to the results, DenseNet-201 attained the best classification accuracy of 87.02% despite some challenges faced by the present study such as small size of dataset and class-imbalance. The study showcases how deep learning models can automate the detection of cervical cancer,without dependence on manual segmentation and feature extraction, a strategy that is error-prone. Nevertheless, the small and imbalanced data set is still a limitation and influences the general model performance.

A related study by Hossein Ashtarian et al. [1] investigates the reasons of participation in cervical cancer screening. The study found that inadequate knowledge, fear and embarrassment of cancer diagnosis was the reason many women cited for not having a Pap smear test. The study underlines that there is a need for education programmes on awareness and participation that can lead to more diverse and representative datasets for training machine learning models. While not specific to machine learning, the results emphasize the need for raising awareness of cervical cancer and the Pap test including to achieve higher levels of participation in screening programmes.

A significant contribution to the field is made by Marina E. Plissiti et al. [3] with the introduction of the SIPaKMeD dataset, an important dataset for evaluating and comparing the performance of classification methods applied to cervical cell images. The dataset consists of 4,049 annotated cervical cell images, which consists of five classes. The authors contrast feature-based cell classification using hand-engineered cell features and CNNs. Results demonstrate that CNN-based approaches are more effective over the other methods, with the classification accuracy of 95.35%. Nonetheless, the paper assumes that the SIPaKMeD dataset is limited in size and it is class-imbalanced and the latter may bias the models. These limitations demonstrate the importance of having bigger and more diverse datasets to enhance the fortitude of the models.

The study by Ishak Pacal et al. [23] applied a hybrid deep learning model combining CNN and Variational Autoencoders (VAE) for classifying cervix cells as normal and abnormal. The DCAVN model performs an excellent classification accuracy of 99.4%, making it superior to traditional machine learning algorithms. The hybrid model's dimensionality reduction and overfitting inhibition are appealing features for automated cervical cancer diagnosis. However, this study is limited to two categories (normal and abnormal) and it is proposed to enlarge to more for a complete diagnosis.

In a study by Aditya Khamparia et al. [7], the authors propose a hybrid model incorporating Convolutional Neural Networks (CNNs) with Variational Autoencoders (VAEs) for classification of cervical cancer. The DCAVN model has been prepared with the Herlev dataset provided a classification accuracy of 99.4%, outperforming the other machine learning approaches. The performance was compared to the previous models, and higher precision and recall were obtained using this hybrid model. A shortcoming of the study is that only two categories were involved (normal and abnormal), and in future research the classification task can be extended to include more detailed categories then a more precise diagnosis can be reached.

Another related research was performed by Milad Rahimi et al. [12] studies machine learning for survival prediction on cervical cancer patients. The analysis comprehensively summaries prediction models for overall survival, disease-free survival, and progression-free survival. It can be seen that the machine learning model, such as Random Forest and Logistic Regression, play an important role in survival prediction, with AUC scores varying from 0.40 to 0.99. But the paper points to a number of hurdles, such as deep learning models' inscrutability, which hampers clinical implementation. The authors recommend that future models should work on greater interpretability so that clinicians can trust them more.

In a novel approach, Tatsuhiro Baba et al. [21] propose a multi-branch deep learning model integrated with Multi-Head Self-Attention (MHSA) and CNN for cervical cancer classification. The system obtained 98.522% accuracy by utilizing the SIPaKMeD dataset that demonstrated the strength of taking advantage of a combination of different deep learning strategies for image classification. The results though promising, show the performance of the model heavily relies on the dataset and also poses question to the generalizability of the model to other image datasets. Furthermore, the computation cost of the multi-branch model is overwhelming, meaning that it can not be effectively applied in resource-constrained environment.

Another significant study is by N. Sompawong et al. [4], which applies Mask R-CNN to automate Pap smear analysis. The model demonstrated a better accuracy of 91.7% and high sensitivity and specificity for individual nuclei detection. However, sensitivity limitations in classifying certain atypical cells are revealed by the study which will lead to misclassification. The authors note that additional studies are called for to optimize the model and increase its sensitivity, particularly in dealing with atypical cell populations.

The study discussed in Yuan et al. [6] works to apply a deep learning model to detect cervical low-grade intraepithelial lesions (LSIL) and high-grade squamous intraepithelial lesions (HSIL) using colposcopy images. The ResNet, U-Net, and Mask R-CNN models with deep learning identified the lesions with a sensitivity of 85.38
A study by Harmanpreet Kaur et al. [14] compares 16 deep transfer learning models for classification of cervical cancer using Pap smear images. The paper concludes that ResNet101 has the best accuracy (95.56%) and transfer learning is a promising method for cervical cancer detection. The study also represents the importance of bigger and diverse datasets as well as to handle better overlapping cells which are the defining factor of real-world clinical application.

Dakshayani Himabindu et al. [22] introduce LSTEDL-CCS (Leveraging Swin Transformer with Ensemble of Deep Learning Model for Cervical Cancer Screening), which aims to enhance the performance of colposcopy images for cervical cancer detection. The approach utilizes a two-stage multi-heads model with the first Swin Transformer as the feature extractor, and the second is an ensemble of CNN architectures (ResNeXt, EfficientNetB0, and MobileNetV2) for local image details features and global image features. Such features are concatenated, are then combined via Canonical Correlation Analysis (CCA) and classified with an Extreme Learning Machine (ELM). Findings indicate high accuracy in detecting cervical cancer and demonstrate the potential to address the limitations of current screening techniques. Yet, the study has several limitations that influence generalisability including reliance on the quality of the dataset and lack of external validation. Additionally, its computational load is another concern, particularly in resource-poor environments. Moreover, the interpretability of deep learning models is still a concern that could affect clinical implementation.

Moreover, its computational burden is another issue, especially in resource-poor conditions. In addition, clinical application may be hampered due to unclear workings of the deep learning algorithms.

. Equally important study by Betelhem Zewdu Wubineh et al. [19] systematically reviews the techniques for segmentation and classification for Pap smear images used in the detection of cervical cancer. The article discusses limitations of Pap smear testing manually, including Slowness, Human error and outcome reliant on skilled

cytopathologists. As per the results of this work, automated and accurate screening tools are extremely important to mitigate anxiety and stress in high-profile situations. This paper also underlines the importance of using larger and more varied data sets to improve the models generalizability. According to the study, U-Net method gave recall as high as 98
In a related study, Paisit Khanarsa and Satanat Kitsiranuwat [15] introduced an ensemble deep learning method for classifying cervical cancer from conventional Pap smear images. The experiment uses SIPaKMeD dataset and utilizes transfer learning with deep learning models: VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7. Additionally, the authors used ensemble methods like the Maximum Occurrence and Maximum Probability to enhance the classification rate. The results indicate that the ResNet152V2 network attained an accuracy of 95% using the Maximum Occurrence ensemble method and an accuracy of 97% using the Maximum Probability method. The study also indicated the difficulty of classifying multi-cell images because of the overlapped cytoplasm as well as of poor image quality. Nevertheless, the ensemble strategy greatly enhanced the classification accuracy compared to single models. Nevertheless, diversity limitation of SIPaKMeD, and the fact that single-cell images were manually extracted are the main limitations of this study, which indicates that diverse datasets, and automated way of extracting single cells are needed in practical applications.

2.2.2 Federated Learning in Medical Imaging

Federated learning (FL) has shown potential to address the challenges in the privacy and security of patients' data sharing and accessing in the healthcare domain, particularly for medical image analysis. A few studies have employed FL for cervical cancer detection to utilize below is a review of several key studies focusing on the application of FL in medical imaging, specifically for cervical cancer detection.

A study by Subramanian et al.[9] investigates decentralized FL algorithms for cancer classification in the healthcare domain. They performed their experiment using Kaggle datasets, cervical cancer images, and considered FL algorithms such as FedAvg and FedProx. The performance was consistently better than FedAvg, in particular in the case of non-IID learning, where classification accuracy of 83.31% was achieved with 150 communication rounds, as against 81.91% of FedAvg. They employed Bayesian optimization to search for hyper-parameters, and this helped in enhancing model performance. However, the study cited some limitations including low dataset diversity, which was sourced from Kaggle and featured cervical, lung and colon cancer. Also, the fact that no real-time privacy or tradeoff evaluation is available was perceived as a limitation, being a major issue for medical applications, in which patient data confidentiality is critical.

Another relevant comprehensive study by Darzidehkalani et al. [8] present the applications of FL in medical images and the numerous approaches to accommodate the issue of the data privacy and security. They highlight that data are heterogeneous and there are no unified structure in which they are storing, which can make difficult to perform collaborative model training across the institutions. The writers claim that FedAvg is the most balanced FL method, in terms of the trade-off of performance, privacy, and communication overhead. However, they highlight that heterogeneity in institutional data may result in poor model performance and hinder convergence, and that privacy techniques like differential privacy can introduce computational overhead. The work provides insight to the possibilities of FL in medical imaging and emphasizes the necessity for more research to overcome these challenges.

2.2.3 Cervical Cancer Detection with Federated Learning

Muhammad Umar Nasir [17] introduce Federated Machine Learning (FML) for cervical cancer prediction, using federated learning to train models without patient's sensitive data sharing. "Research on these assays shows accuracy as impressive as 99.26%, with sensitivity and specificity rates as high as 99.6%." The implications of privacy-preserving processes and the application of the Internet of Medical Things (IoMT) for efficient collection of data from different patients is stressed by the authors. However, the paper also notes that the data set may not reflect the conditions under which ads are actually seen. The ability of the model to perform well in clinical settings with heterogeneous data quality and availability has not been adequately tested.

In a study by Peta and Koppu [11], FL has been used for breast cancer classification, which shows the strength of Federated Learning in addressing the data privacy while working with the imbalanced dataset. Their research involved federated learning using CNNs and a secure data storage system supported by a framework for Federated Learning Flower (FLF). As you can see the paper shows 95.68% accuracy and 95.6% recall showing that the model performs well on these imbalanced data. But they argue that high communication overhead and data diversity are critical bottlenecks, especially in resource-limited settings. Besides, the requirement of further sophisticated encryption methods for ensuring data security is still a challenge.

In a study focused on cervical cancer detection, Shehnaz Joynab et al. [13] explore the integration of FL with CNN architectures using the SIPaKMeD dataset which includes cervical cell images and is classified into five classes. Experiment shows that CNN-based federated learning models outperformed the traditional machine learning models such as LightGBM, ExtraTrees and SVM. The optimal federated learning model (IL2), trained on an IID configuration reached a 93.36% accuracy at classification with a ROC-AUC score of 98.2%. The model accuracy reduced to 78.43% in non-IID setting which brings forward the problem of non-uniform data in flux. It also emphasizes the requirement of applying methods such as differential privacy to enhance privacy preservation of patient data in federated learning workflows.

The research by Shehnaz Joynab et al. [13] also highlights the necessity of employing high-end pre-trained models and having real-time feedback in actual deployments. Despite high performance in controlled scenarios, extension of the model to clinical environment has not been yet evaluated. The authors advocate combining privacy-preserving methods with model testing using stronger models in order to obtain higher generalizability and robustness in a diverse range of clinical environments.

2.3 Summary of Key Findings

Machine Learning and Federated Learning based Cervical Cancer Detection addresses some major trends and patterns in building models and also address the problem at hand in the medical field. In the studies reviewed during the period 2017–2025, selected models reveal a tendency toward the adoption of more developed deep learning techniques (CNNs, VAEs and ViTs) for the automated detection of cervical cancer in Pap smear and colposcopy images.

The increasing resilience of transfer learning is one of the most notable trends. Several researches, like Kaur et al. [14] and Khanarsa et al. [15], emphasize on the applications of pre-trained models like DenseNet, ResNet, Efficient Net, which were fine-tuned for the detection of cervical cancer. These models tend to achieve high classification accuracy scores which are sometimes over 90% and have shown remarkable performance over conventional learning approaches. Furthermore, it is important to note the work of Tan et al. [18]. They show that other CNN models, such as DenseNet-201, can also classify cervical cells effectively. Moreover, they achieve an accuracy of 87.02
Another well-known trend in the literature is to incorporate federated learning (FL) to overcome the privacy issues associated with healthcare data. The research works of Malliga Subramanian et al. [9] and Muhammad Umar Nasir et al. [17] explore various federated learning algorithms such as FedAvg, FedProx and FedGKT, which enable model training across multiple decentralized data sets, while still preserving patient privacy. This trend is also relevant for the sensitive health data, where privacy-preserving methods are important. Federated learning is becoming popular, but non-IID data and computational costs are the major obstacles to its feasibility in clinical practice. Nonetheless, the fact that promising results are reported in these studies suggests that federated learning, in particular when combined with sophisticated deep learning models, has significant prospects to improve model performance without violating patient privacy.

In terms of model performance, ensemble models and hybrid models are becoming more attractive. For instance, the work by Himabindu et al. [22] integrates both Swin Transformers and CNN ensembles in order to boost feature extraction and classification accuracy with high detection rate for cervical cancer. Ensemble strategies such as Maximum Occurrence and Maximum Probability can significantly improve model robustness and accuracy, particularly in multi-cell and multi-class scenarios, as represented by Khanarsa et al. [15] and Wubineh et al. [19]. The emerging tendency of incorporating models to achieve higher performance can also be observed in experiments that demonstrate how models combining deep features can reach nearly perfect classification accuracy (99.7%) of cervical cancer from colposcopy and Pap smear images.

The evaluation metrics employed in different studies (sensitivity and specificity, F1 score, accuracy) are uniform and all indicate an emphasis on having an efficient method for use in clinics. But there is a variety in the training dataset size, class imbalance, and under-representation of a diverse population in real-world that makes the models difficult to be generalized across populations. Furthermore, interpretability is still a big problem. Most deep learning models from the reviewed papers are “black boxes”, making it difficult to its clinical acceptance. The importance of explainable AI (XAI) is also highlighted particularly by Khanarsa et al. [15], where transparency and confidence in model predictions are paramount to actual use in healthcare settings.

The results show that in the cervical cancer diagnosis, there is a move towards advanced deep learning models and privacy aware. So far, techniques of transfer learning, ensemble methods and hybrid methods, which can also enhance the accuracy of classification, achieved great success. Federated learning removes the ease of issues related to privacy. However, issues like dimension, data interpretability, and the need to set up different data collections for every study forces us to address similar issues. According to reviewed studies, significant improvement in detection of cervical cancer by combined techniques has been established; yet, additional optimization and real-world validation are necessary in order to achieve their widespread use in the clinic.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

The proposed system entitled *"A Federated Learning Based Efficient Approach to Detect Cervical Cancer Using PAP-SMEAR Images"* ensures both technical feasibility and societal non-violation. This chapter outlines the final specifications and requirements, including technical, social, and ethical considerations, along with the project management plan and economic aspects of the proposed solution. The comprehensive details related to these factors are presented in this chapter.

3.1.1 Technical specifications

The proposed system allows for cervical cancer detection using PAP-smear images in a federated learning–based system that trains the model across multiple distributed clients without sharing raw medical data. This solution balances the demand for data privacy and good performance. System's technical requirements must include suitable software, hardware, dataset, and other necessary resource specifications needed for model training, evaluation, and deployment. In order to organize and develop decentralized trainings, deep learning frameworks and federated learning libraries are required and also GPU accelerated hardware is required to handle the imaging task which is computationally intensive. Reliably extracting features and ensuring consistent model performance depend on properly labeled and preprocessed PAP-smear image datasets. Together, these requirements guarantee that the suggested system can be scalable, reliable and practically useful.

3.1.2 Software Requirements

To develop, train, and evaluate the federated learning model effectively, the following software tools and platforms will be required.

- **Python 3.10.19:** The programming language utilizes Python 3.10.19 for model creation and experimentation.
- **PyTorch:** PyTorch is a deep learning framework for developing and training convolutional neural networks.

- **Torch & Torchvision:** Torch and Torchvision neural networks and support for image preprocessing and pretrained models.
- **Flower (flwr):** Flower is a federated learning framework that enables model training across multiple clients.
- **Scikit-learn:** Scikit-learn is utilized to calculate evaluation metrics such as accuracy, precision, etc.
- **Jupyter Notebook:** A Jupyter Notebook is an interactive environment that lets us experiment, visualize and analyze our results. It was used to run some ipynb files.
- **WSL2 (Windows Subsystem for Linux 2):** The windows subprocess for Linux 2 gives a better tuned performance but it provides a Linux-based development environment on windows and resolves Ray Backend issues.
- **CUDA-supported environment:** An environment that supports CUDA is required for training the model in the GPU.

3.1.3 Data Requirements

The model is trained and validated on quality PAP-smear images dataset. Include the data requirements

- PAP-smear images are digitized to label 5 different classes.
- Diverse types of cervical cells were evenly represented.
- Normalization, resizing, and augmentation techniques to input images.
- Distributed dataset partition.

3.1.4 Hardware Requirements

- The GPU that has been used is the NVIDIA RTX 4080 Super that has accelerated the training of deep learning.
- Powerful CPU and adequate RAM (16GB minimum recommended).
- Sufficient storage capacity for image datasets and trained models.

3.2 Social Impact

The utilization of ML for early detection of cervical cancer in a privacy-preserving way can alleviate healthcare problems. Cervical Cancer Mortality May be Reduced with Early Diagnosis; However, Due to Limited Resources, Delay in Diagnosis and Other Factors Regarding the Use of Data, Privacy, and so on. The suggested system detects cervical cancer at an earlier stage using federated machine learning. Health organizations can collaborate while keeping data private with federated learning. The proposed method enables the training of models on PAP-smear images at the local site without transmitting any sensitive patient data to a centralized server. This useful diagnostic tool contributes to the betterment of cervical cancer diagnosis by strengthening efficiency and accuracy. The suggested approach will therefore accelerate clinical decision-making and reduce the burden on medical staff.

3.2.1 Impact on Society

The proposed federated learning approach can assist the hospitals and management agency to share and plan together, without sharing their patient data outside the hospital, as better detection and diagnosis can decrease the death rate due to cervical cancer.

3.2.2 Health Impact

Most medical professionals can confirm this system will ease the screening workload by making diagnosis faster. The benefits of enhanced detection accuracy are early treatment, improved patient outcomes, and lower costs, among others.

3.2.3 Cultural Impact

In many regions of the world especially in the developing countries, cultural stigma and privacy issues hamper cervical cancer screening participation rate. The architecture of federated learning was designed with privacy in mind to help address such issues and encourage uptake.

3.3 Environmental Impact

Additionally, large datasets need not be transferred to a centralized server for training, which helps reduce the data transmission and storage requirements that result in the further reduction of energy consumption and carbon footprint.

- **Sustainability :** This framework has the potential to be a long-lasting healthcare innovation that offers efficient computation with longer usability.
- **Environmental Impact :** Federated learning minimizes the unnecessary storage of data. Also, it lessens the reliance on high-powered servers that are cloud-hosted. Consequently, we achieve more environmentally-friendly AI.

3.4 Ethical Issues

Medical data being one of the most coveted assets, are often utilized by medical AI systems. As health data is sensitive in nature, the use of medical AI systems raises an ethical issue. For example, medical images or clinical information about a patient are highly sensitive medical data. Any usage of this data needs to occur in and through a robust ethical framework. Thus, medical AI systems should have strong data protection safeguards, transparency as to how decisions are reached and safeguards against bias as well as discrimination. Tackling these ethical matters is crucial to ensuring that AI technologies assist healthcare professionals in a responsible manner while safeguarding patient's rights and interests.

- Decentralized data processing guarantees patient privacy protection.
- Protecting From Data Misuse and Unauthorized Access
- Ensure diverse and representative datasets to avoid algorithmic bias.
- Open explanation of model choices to support medical trust.

3.4.1 Professional Responsibility

The researchers and developers of the system must adhere to medical ethics. They need to comply with data protection regulations and AI ethics. This system intends to assist doctors rather than replacing them.

3.5 Project Management Plan

Effective project management ensures timely completion and quality outcomes.

3.5.1 Schedule

The project is divided into the following phases:

- **Month (1-2) :** Literature review and problem analysis
- **Month (3-4) :** Dataset collection and preprocessing
- **Month (5-6) :** Test the pretrained model on the dataset
- **Month (7-8) :** Model development and federated learning implementation
- **Month (9-10) :** Performance evaluation and optimization
- **Month (11-12) :** Documentation and thesis writing

3.5.2 Resource Management

Organizational work is something important that is needed to perform the planning and organizing the resources of the project to successfully complete it. Whether human resources or overall computing resources, each espouses planning and appropriating. A brief overview of the resources has been provided for the completion of this project.
The necessary resources to achieve the project goal are:

1. **Machinery and Tools**
   The hardware, including the CPUs and GPUs in our case, is meant for the successful execution of the project.

2. **Software and tools:**
   "Software and Tools" includes the necessity for various software to develop the project and also operate it.

3. **Dataset:**
   A good dataset is required so that it performs efficiently. The dataset must be such that it is sufficient to carry out learning of the model and understand and create the performance of

3.5.3 Risk Management

The occurrence of any unpleasant event can lead to the chance of any damage or hazard.

Risk

- Limited availability of high-quality labeled medical data
- Model performance inconsistency across federated clients
- compatibility issues of both Hardware or software

Mitigation Strategies

- Usage of data augmentation and preprocessing techniques
- proper tuning of federated learning parameters

3.6 Economic Analysis

The decentralized architecture that has been proposed enables multiple clients to cooperatively learn a shared prediction model while keeping all the training data local. In this way, the economics become much more efficient in comparison to conventional AI solutions which leverage the centralized architecture model. The conventional AI systems store all the collected data in a central agent first in order to process the model for training. However, it does not hold true for the proposed system. Similarly, the cost of data transfer and maintenance is also high in case of traditional systems. In addition, the local computing resource of the institution is employed to train the model at each client node. The hospital or diagnostic center doesn't have to invest in a high-performing server or cloud-based software platform to train the model. As a result traditional systems require to have a huge infrastructure of the organization for data storage and maintenance.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

The goal of our research is to build a privacy preserving deep learning based pap smear image classifiers using federated learning. The overall workflow consists of three stages:

1. Dataset collection & preprocessing
2. Federated implementation based on Flower framework
3. Model development and performance comparison for IID & non-IID data partition

Several pretrained architectures like MobileNetv2, EfficientNet, ResNet18, ResNet50 etc were implemented on the SipakMed dataset.

Afterwards, data augmentation techniques were used in the training process to enhance model robustness and generalization.To ensure privacy preservation, Federated Learning based algorithm was used for both IID & non-IID data partition among 3 clients. For the IID data partition, pretrained models were initially employed using the FedAvg algorithm. Subsequently, various hybrid models were trained on the dataset to assess improvements in accuracy and efficiency. These hybrid models were then applied to the non-IID data partition with FedProx & Scaffold algorithm for further evaluation.

4.2 Design Specifications

4.2.1 Pretrained Models

1. **VGG16:** VGG16 is one of the well-known deep learning models for image classification. It is known for its powerful implementation and reliable feature extraction for visual recognition tasks. The VGG16 model's input image dimension was set at 224×224 pixels in this study. The VGG16 architecture is a simple yet effective architecture which consists of small convolution filters while stacking deeper layers. This approach facilitates the model learning

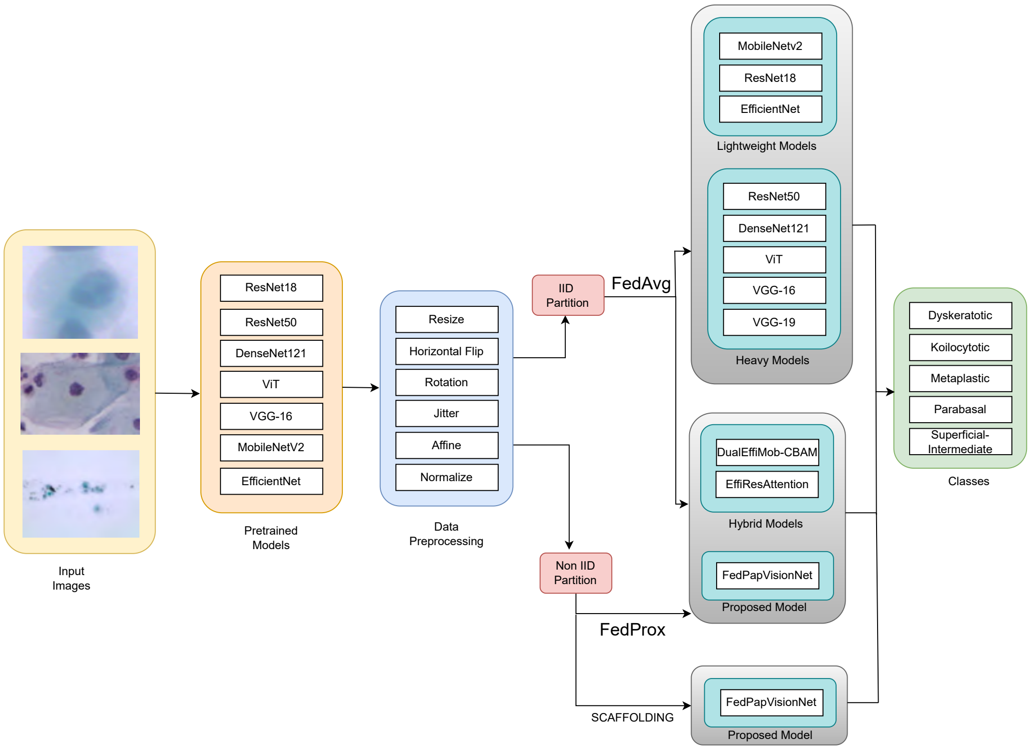


Figure 4.1: Methodology overview of the proposed federated learning framework

complex patterns while keeping the architecture simple. The transfer learning model selected for this work is VGG16 due to its effectiveness and stability.

2. **MobileNetV2:** MobileNetV2 is a CNN architecture for efficient execution on mobile devices. It employs inverted bottleneck blocks and skip connections.

3. **ResNet18:** ResNet18 is a deep neural network that overcomes the vanishing gradient issue through residual learning and skip connections. This allows information to efficiently get across layers for ease of learning.

4. **ResNet50:** ResNet50 is a deeper residual network containing a total of 50 layers that captures more complex images. The use of residual connections improves stability in training and performance on large image classification problems.

5. **ViT:** The ViT is a transformer-based model that treats images as sequences of patches rather than using convolutional layers. It can effectively capture the global relationships and long-range dependencies in images.

6. **DenseNet121:** DenseNet121 connects each layer to every other layer in a feed-forward way using this architecture. Both these features help in reusing any feature and improving the flow of gradients efficiently. This architecture lessens parameter redundancy and enhances feature propagation.

7. **EfficientNetB0:** EfficientNetB0 is one architecture that balances the depth, width, and resolution of a network through compound scaling. It reaches high

accuracy with fewer parameters and a lower computational cost than CNN models.

The main aim of the proposed work was to devise a lightweight deep learning model that can be applied in federated learning situations where computation and communication cost constraints exist. As the paper suggests, the number of trainable parameters used in each architecture was evaluated, as this affects its complexity. To begin with, a baseline hybrid lightweight architecture called EffiResAttention was designed by assembling efficient convolutional layers with residual connections and attention mechanisms consisting of channel and spatial attention using a Convolutional Block Attention Module (CBAM). This model had 16.6 million trainable parameters, making it a reasonable classifier with good performance. Further analysis showed that, under certain conditions, improved feature representation could be achieved while decreasing parameter redundancy.

The architecture was modified into a new hybrid architecture named DualEffiMob-CBAM. It has dual efficient mobile-style convolutional blocks and a Convolutional Block Attention Module (CBAM). With the number of parameters reduced to 8.5 million,this improved model features discrimination using spatial attention and channel-wise attention and is effective in a resource-constrained environment.

Subsequently, a ViT with Classifier model was built with 527,877 parameters to train using the Federated Learning algorithm. It includes temperature scaling for better calibration and utilizes dropout and LayerNorm for enhanced regularization.

Ultimately, FedPapVisionNet is the optimized rendition of the proposed model. FedPapVisionNet adopts efficient convolutional operations, attention mechanisms, and federated learning compatibility, with a total of 390,215 trainable parameters. The architecture finally developed strikes a good balance between being lightweight and classification performance. Additionally, the model requires less computation time, thus corresponding to a lower demand for computational resources in decentralized and privacy-preserving training.

4.2.2 Custom Models

EffiResAttention

The EffiResAttention design combines the benefits of EfficientNet-B0 and ResNet18 backbones in parallel feature extraction. The two-way framework allows for the concurrent extraction of diverse feature representations by employing the principles of EfficientNet and ResNet. EfficientNet-B0 makes use of depth-wise separable convolution and squeeze-and-excitation block for channel attention. At the same time, ResNet18 requires identity-based skip connections in order to mitigate vanishing gradients in very deep networks. By means of this heterogeneous integration, the model would get the benefit from compound scaling efficiency and residual learning stability.
Considering an image with $\mathbf{X} \in R^{3\times128\times128}$ input, the feature extraction in both paths can be represented as:

$$\mathbf{F}_{\text{eff}} = \Phi_{\text{eff}}(\mathbf{X}) \in R^{C_{\text{eff}} \times h \times w}$$

$$\mathbf{F}_{\text{res}} = \Phi_{\text{res}}(\mathbf{X}) \in R^{C_{\text{res}} \times h' \times w'}$$

where Φ_{eff} is the feature extraction by EfficientNet-B0 and Φ_{res} represents the ResNet18 feature extraction backbone, producing feature maps with channel dimensions C_{eff} and C_{res} at potentially different spatial dimensions.

Residual Learning Pathway in ResNet18: The ResNet18 backbone implements residual learning through sequential residual blocks, where each block learns residual mappings rather than direct transformations. The fundamental residual operation is defined as:

$$\mathbf{R}_{\text{out}} = \text{ReLU}(\mathbf{y}_2 + \text{Shortcut}(\mathbf{x}))$$

where:

$$\mathbf{y}_1 = \text{ReLU}(\text{BN}(\text{Conv}(\mathbf{x}, \mathbf{W}_1)))$$

$$\mathbf{y}_2 = \text{BN}(\text{Conv}(\mathbf{y}_1, \mathbf{W}_2))$$

$$\text{Shortcut}(\mathbf{x}) = \begin{cases} \text{BN}(\text{Conv}_s(\mathbf{x}, \mathbf{W}_s, \text{stride} = 1)) & \text{if dimensions differ} \\ \mathbf{x} & \text{if dimensions match} \end{cases}$$

In these equations, $\mathbf{x}$ represents the input, $\text{Conv}(\mathbf{x}, \mathbf{W}_1)$ denotes the first convolutional layer with weights $\mathbf{W}_1$, BN represents batch normalization, ReLU denotes the Rectified Linear Unit activation function, and $\text{Conv}_s(\mathbf{x}, \mathbf{W}_s, \text{stride} = 1)$ represents the 1×1 convolution for the shortcut connection when input-output dimensions differ. [16]

Spatial Alignment and Feature Concatenation: To ensure compatibility between feature maps from both paths, spatial alignment is performed using bilinear interpolation when dimensions differ:

$$\mathbf{F}_{\text{align}} = \text{Interpolate}(\mathbf{F}_{\text{eff}}, \mathbf{F}_{\text{res}}, \text{mode} = \text{bilinear}, \text{align\_corners} = \text{False})$$

The aligned features are then concatenated along the channel dimension:

$$\mathbf{F}_{\text{concat}} = [\mathbf{F}'_{\text{eff}}; \mathbf{F}'_{\text{res}}] \in R^{C_{\text{total}} \times h'' \times w''}$$

where $[\cdot;\cdot]$ denotes channel-wise concatenation, $C_{\text{total}} = C_{\text{eff}} + C_{\text{res}}$, and $h'' \times w''$ represents the unified spatial dimensions.

Convolutional Block Attention Module (CBAM): The Convolutional Block Attention Module (CBAM) enhances feature representation where it applies attention sequentially in two different dimensions Channel and Spatial attention and makes the network focus on the discriminative part of the feature maps [20].

Channel Attention Module

$$\mathbf{M}_c(\mathbf{F}) = \sigma\left(\mathcal{W}_2\delta\left(\mathcal{W}_1\mathbf{z}_{\text{avg}}\right) + \mathcal{W}_2\delta\left(\mathcal{W}_1\mathbf{z}_{\text{max}}\right)\right)$$

where:

- $\mathbf{z}_{\text{avg}} = \text{GAP}(\mathbf{F}) = \frac{1}{h\times w}\sum_{i=1}^{h}\sum_{j=1}^{w}\mathbf{F}_{:,i,j}$
- $\mathbf{z}_{\text{max}} = \text{GMP}(\mathbf{F}) = \max_{i,j}\mathbf{F}_{:,i,j}$
- $\mathcal{W}_1 \in R^{\frac{C}{r}\times C}, \mathcal{W}_2 \in R^{C\times\frac{C}{r}}$ with reduction factor $r = 16$
- $\delta(\cdot)$ denotes ReLU activation function
- $\sigma(\cdot)$ represents sigmoid activation function

Spatial Attention Module:

$$\mathbf{M}_s(\mathbf{F}) = \sigma\left(f^{7\times7}\left([\mathbf{F}_{\text{avg}}; \mathbf{F}_{\text{max}}]\right)\right)$$

where:

- $\mathbf{F}_{\text{avg}} = \frac{1}{C}\sum_{c=1}^{C}\mathbf{F}_{c,:,:}$
- $\mathbf{F}_{\text{max}} = \max_c \mathbf{F}_{c,:,:}$
- $f^{7\times7}(\cdot)$ denotes a 7×7 convolutional layer with padding
- $[\cdot;\cdot]$ represents channel concatenation

Complete CBAM Operation:

$$\mathbf{F}_{\text{att}} = \mathbf{F} \otimes \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{M}_s(\mathbf{F} \otimes \mathbf{M}_c(\mathbf{F}))$$

where $\otimes$ denotes element-wise multiplication.

Feature Fusion and Dimensionality Reduction:

$$\mathbf{F}_{\text{fused}} = \text{ReLU}\left(\text{BN}\left(\mathbf{W}_{\text{fusion}} * \mathbf{F}_{\text{att}}\right)\right)$$

where $\mathbf{W}_{\text{fusion}} \in R^{512\times C_{\text{total}}\times1\times1}$ and $*$ denotes convolution operation.

Global Average Pooling and Classification:

$$\mathbf{z} = \text{GAP}(\mathbf{F}_{\text{fused}}) = \frac{1}{h''\times w''}\sum_{i=1}^{h''}\sum_{j=1}^{w''}\mathbf{F}_{\text{fused}}(:,i,j) \in R^{512}$$

$$\mathbf{y} = \text{softmax}\left(\frac{\mathbf{W}_3\delta(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{z}))}{T}\right)$$

where T is a learnable temperature parameter initialized to 1.0, and the classifier employs progressive dropout rates (0.5, 0.35, 0.25) for regularization.

Mathematical Formulation of EffiResAttention: The complete forward pass of the EffiResAttention model is represented step-by-step:

$$\mathbf{X}_{\text{input}} \in R^{128\times128\times3} \quad (1)$$

Input size $128 \times 128 \times 3$, where 3 defines RGB channels in Eq. (1).

$$\mathbf{F}_{\text{eff}} = \Phi_{\text{eff}}(\mathbf{X}_{\text{input}}) \in R^{C_{\text{eff}}\times h\times w} \quad (2)$$

$$\mathbf{F}_{\text{res}} = \Phi_{\text{res}}(\mathbf{X}_{\text{input}}) \in R^{C_{\text{res}}\times h'\times w'} \quad (3)$$

EfficientNet-B0 and ResNet18 feature extraction producing C_{eff} and C_{res} channel feature maps in Eqs. (2) and (3).

$$\mathbf{F}_{\text{align}} = \text{Interpolate}(\mathbf{F}_{\text{eff}}, \mathbf{F}_{\text{res}}, \text{size} = (\min(h, h'), \min(w, w'))) \quad (4)$$

Spatial alignment through bilinear interpolation in Eq. (4).

$$\mathbf{F}_{\text{concat}} = [\mathbf{F}'_{\text{eff}}; \mathbf{F}'_{\text{res}}] \in R^{C_{\text{total}}\times h''\times w''} \quad (5)$$

Channel-wise concatenation of aligned feature maps in Eq. (5), where $C_{\text{total}} = C_{\text{eff}} + C_{\text{res}}$.

$$\mathbf{F}_{\text{att}} = \mathbf{F}_{\text{concat}} \otimes \mathbf{M}_c(\mathbf{F}_{\text{concat}}) \otimes \mathbf{M}_s(\mathbf{F}_{\text{concat}} \otimes \mathbf{M}_c(\mathbf{F}_{\text{concat}})) \quad (6)$$

Application of CBAM with channel and spatial attention in Eq. (6).

$$\mathbf{F}_{\text{fused}} = \text{ReLU}\left(\text{BN}\left(\mathbf{W}_{\text{fusion}} * \mathbf{F}_{\text{att}}\right)\right) \in R^{512\times h''\times w''} \quad (7)$$

1×1 convolution for dimensionality reduction to 512 channels in Eq. (7).

$$\mathbf{z} = \text{GAP}(\mathbf{F}_{\text{fused}}) \in R^{512} \quad (8)$$

Global average pooling in Eq. (8).

$$\mathbf{h}_1 = \text{Dropout}(0.5)\left(\text{ReLU}\left(\text{LayerNorm}\left(\mathbf{W}_1\mathbf{z}\right)\right)\right) \in R^{256} \quad (9)$$

First classifier layer with 256 neurons and dropout rate 0.5 in Eq. (9).

$$\mathbf{h}_2 = \text{Dropout}(0.35)\left(\text{ReLU}\left(\text{LayerNorm}\left(\mathbf{W}_2\mathbf{h}_1\right)\right)\right) \in R^{128} \quad (10)$$

Second classifier layer with 128 neurons and dropout rate 0.35 in Eq. (10).

$$\mathbf{y} = \frac{\mathbf{W}_3\mathbf{h}_2}{T} \in R^5 \quad (11)$$

Final output layer with 5 neurons corresponding to number of classes, scaled by temperature parameter T, in Eq. (11).

Architectural Components and Parameter Statistics

- **EfficientNet-B0 Backbone**: 4,007,548 parameters
- **ResNet18 Backbone**: 11,176,512 parameters
- **Attention and Fusion Modules**: 1,320,546 parameters
- **Classification Head**: 165,637 parameters
- **Total Trainable Parameters**: 16.6M parameters

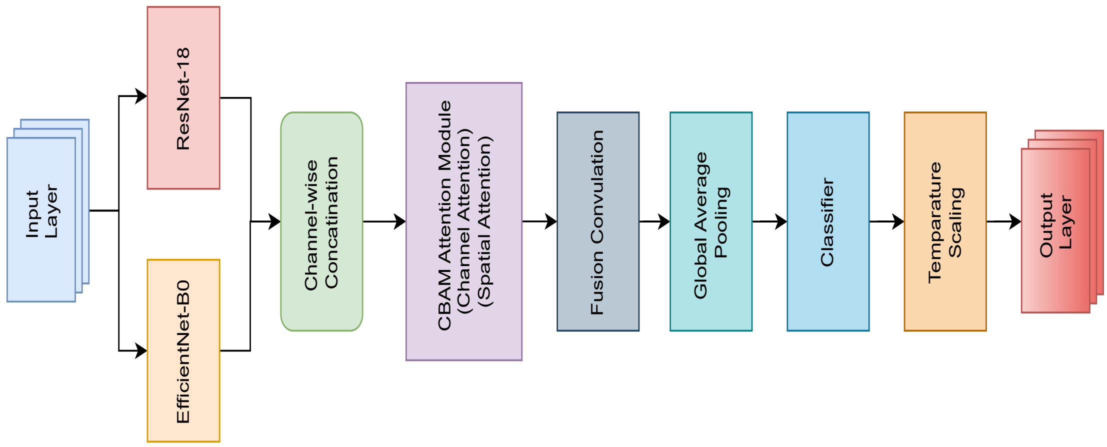


Figure 4.2: Architecture overview of the EffiResAttention model

DualEffiMob-CBAM

The DualEffiMob-CBAM architecture employs a parallel feature extraction strategy that leverages complementary characteristics of two efficient backbone networks: EfficientNet-B0 and MobileNetV2. This dual-path design enables the model to capture diverse feature representations simultaneously, enhancing the overall discriminative capacity while maintaining computational efficiency. EfficientNet-B0 utilizes compound scaling of depth, width, and resolution, while MobileNetV2 employs inverted residual blocks with linear bottlenecks. The parallel processing of these architectures allows the model to benefit from both scaling principles and lightweight design.

Given an input image $\mathbf{X} \in R^{3\times128\times128}$, the feature extraction process in both paths can be represented as:

$$\mathbf{F}_{\text{eff}} = \Phi_{\text{eff}}(\mathbf{X}) \in R^{1280\times h\times w}$$

$$\mathbf{F}_{\text{mob}} = \Phi_{\text{mob}}(\mathbf{X}) \in R^{1280\times h\times w}$$

where Φ_{eff} and Φ_{mob} represent the EfficientNet-B0 and MobileNetV2 feature extraction networks respectively, producing 1280-channel feature maps at reduced spatial dimensions $h \times w$.

Spatial Alignment and Feature Concatenation To ensure compatibility between feature maps from both paths, spatial alignment is performed using bilinear interpolation when dimensions differ:

$$\mathbf{F}_{\text{align}} = \text{Interpolate}(\mathbf{F}_{\text{eff}}, \mathbf{F}_{\text{mob}}, \text{mode} = \text{bilinear})$$

The aligned features are then concatenated along the channel dimension:

$$\mathbf{F}_{\text{concat}} = [\mathbf{F}'_{\text{eff}}; \mathbf{F}'_{\text{mob}}] \in R^{2560 \times h' \times w'}$$

where $[\cdot;\cdot]$ denotes channel-wise concatenation, resulting in 2560 total channels.

Convolutional Block Attention Module (CBAM) The Convolutional Block Attention Module (CBAM) is integrated to enhance feature representation through sequential channel and spatial attention mechanisms. This attention-guided approach enables the model to focus on discriminative features while suppressing irrelevant information **9303478**.

Channel Attention Module The channel attention mechanism computes importance weights across feature channels through both global average and max pooling operations:

$$\mathbf{M}_c(\mathbf{F}) = \sigma\left(\mathcal{W}_2 \delta\left(\mathcal{W}_1 \mathbf{z}_{\text{avg}}\right) + \mathcal{W}_2 \delta\left(\mathcal{W}_1 \mathbf{z}_{\text{max}}\right)\right)$$

where:

- $\mathbf{z}_{\text{avg}} = \text{GAP}(\mathbf{F}) = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} \mathbf{F}_{:,i,j}$
- $\mathbf{z}_{\text{max}} = \text{GMP}(\mathbf{F}) = \max_{i,j} \mathbf{F}_{:,i,j}$
- $\mathcal{W}_1 \in R^{\frac{C}{r} \times C}, \mathcal{W}_2 \in R^{C \times \frac{C}{r}}$ with reduction factor $r = 16$
- $\delta(\cdot)$ denotes ReLU activation function
- $\sigma(\cdot)$ represents sigmoid activation function

Spatial Attention Module The spatial attention mechanism computes importance across spatial locations:

$$\mathbf{M}_s(\mathbf{F}) = \sigma\left(f^{7 \times 7}\left([\mathbf{F}_{\text{avg}}; \mathbf{F}_{\text{max}}]\right)\right)$$

where:

- $\mathbf{F}_{\text{avg}} = \frac{1}{C} \sum_{c=1}^{C} \mathbf{F}_{c,:,:}$
- $\mathbf{F}_{\text{max}} = \max_c \mathbf{F}_{c,:,:}$
- $f^{7 \times 7}(\cdot)$ denotes a 7×7 convolutional layer
- $[\cdot;\cdot]$ represents channel concatenation

Complete CBAM Operation The attention mechanism is applied sequentially:

$$\mathbf{F}_{\text{att}} = \mathbf{F} \otimes \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{M}_s(\mathbf{F} \otimes \mathbf{M}_c(\mathbf{F}))$$

where $\otimes$ denotes element-wise multiplication.

Feature Fusion and Dimensionality Reduction A 1×1 convolution reduces the concatenated feature dimensionality while preserving important information:

$$\mathbf{F}_{\text{fused}} = \text{ReLU}\left(\text{BN}\left(\mathbf{W}_{\text{fusion}} * \mathbf{F}_{\text{att}}\right)\right)$$

where $\mathbf{W}_{\text{fusion}} \in R^{512\times2560\times1\times1}$ and $*$ denotes convolution operation.

Global Average Pooling and Classification Global average pooling generates a fixed-size feature representation:

$$\mathbf{z} = \text{GAP}(\mathbf{F}_{\text{fused}}) = \frac{1}{h' \times w'} \sum_{i=1}^{h'} \sum_{j=1}^{w'} \mathbf{F}_{\text{fused}}(:, i, j) \in R^{512}$$

The classification head employs multiple fully-connected layers with temperature scaling:

$$\mathbf{y} = \text{softmax}\left(\frac{\mathbf{W}_3\delta(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{z}))}{T}\right)$$

where T is a learnable temperature parameter initialized to 1.0, and the classifier employs progressive dropout (0.5, 0.35, 0.25) for regularization.

Mathematical Formulation of DualEffiMob-CBAM The complete forward pass of the DualEffiMob-CBAM model is represented step-by-step:

$$\mathbf{X}_{\text{input}} \in R^{128\times128\times3} \quad (1)$$

Input size 128 × 128 × 3, where 3 defines RGB channels in Eq. (1).

$$\mathbf{F}_{\text{eff}} = \Phi_{\text{eff}}(\mathbf{X}_{\text{input}}) \in R^{1280\times h\times w} \quad (2)$$

$$\mathbf{F}_{\text{mob}} = \Phi_{\text{mob}}(\mathbf{X}_{\text{input}}) \in R^{1280\times h\times w} \quad (3)$$

EfficientNet-B0 and MobileNetV2 feature extraction producing 1280-channel feature maps in Eqs. (2) and (3).

$$\mathbf{F}_{\text{concat}} = [\mathbf{F}'_{\text{eff}}; \mathbf{F}'_{\text{mob}}] \in R^{2560\times h'\times w'} \quad (4)$$

Channel-wise concatenation of aligned feature maps in Eq. (4).

$$\mathbf{F}_{\text{att}} = \mathbf{F}_{\text{concat}} \otimes \mathbf{M}_c(\mathbf{F}_{\text{concat}}) \otimes \mathbf{M}_s(\mathbf{F}_{\text{concat}} \otimes \mathbf{M}_c(\mathbf{F}_{\text{concat}})) \quad (5)$$

Application of CBAM with channel and spatial attention in Eq. (5).

$$\mathbf{F}_{\text{fused}} = \text{ReLU}\left(\text{BN}\left(\mathbf{W}_{\text{fusion}} * \mathbf{F}_{\text{att}}\right)\right) \in R^{512\times h'\times w'} \quad (6)$$

1×1 convolution for dimensionality reduction in Eq. (6).

$$\mathbf{z} = \text{GAP}(\mathbf{F}_{\text{fused}}) \in R^{512} \quad (7)$$

Global average pooling in Eq. (7).

$$\mathbf{h}_1 = \text{Dropout}(0.5)\left(\text{ReLU}\left(\text{LayerNorm}\left(\mathbf{W}_1\mathbf{z}\right)\right)\right) \in R^{256} \quad (8)$$

First classifier layer with 256 neurons and dropout rate 0.5 in Eq. (8).

$$\mathbf{h}_2 = \text{Dropout}(0.35)\left(\text{ReLU}\left(\text{LayerNorm}\left(\mathbf{W}_2\mathbf{h}_1\right)\right)\right) \in R^{128} \quad (9)$$

Second classifier layer with 128 neurons and dropout rate 0.35 in Eq. (9).

$$\mathbf{y} = \frac{\mathbf{W}_3\mathbf{h}_2}{T} \in R^5 \quad (10)$$

Final output layer with 5 neurons corresponding to number of classes, scaled by temperature parameter T, in Eq. (10).

Parameter Statistics

- EfficientNet-B0 backbone: 5,288,548 parameters
- MobileNetV2 backbone: 3,504,872 parameters
- Attention and fusion modules: 229,952 parameters
- Classification head: 105,240 parameters
- **Total trainable parameters**: 8,528,612

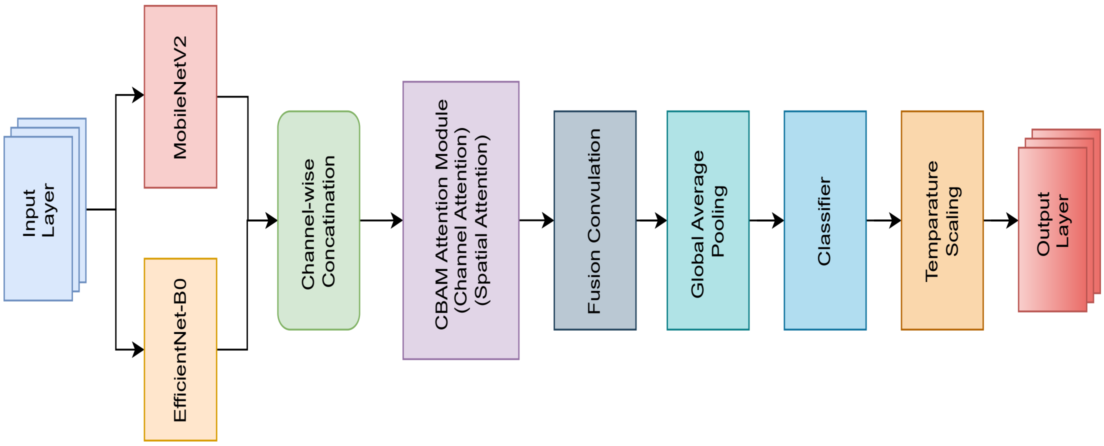


Figure 4.3: Architecture overview of the DualEffiMob-CBAM model

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.

$$R_{\text{out}} = \text{ReLU}(y_2 + \text{Shortcut}(x))$$

Where:

$$y_1 = \text{ReLU}(\text{BN}(\text{Conv}(x, W_1)))$$

$$y_2 = \mathrm{BN}(\mathrm{Conv}(y_1, W_2))$$

$$\mathrm{Shortcut}(x) = \begin{cases} \mathrm{BN}(\mathrm{Conv}_s(x, W_s, \mathrm{stride} = 1)) & \text{if dimensions differ} \\ x & \text{if dimensions match} \end{cases}$$

Here in the Eq. (1), x as input, $\mathrm{Conv}(x, W_1)$ denotes the first convolutional layer with weights W_1, BN denotes batch normalization, ReLU denotes the Rectified Linear Unit activation function and $\mathrm{Conv}_s(x, W_s, \mathrm{stride} = 1)$ denotes the 1×1 convolution for the shortcut connection when the dimensions do not match [16].

Squeeze-and-Excitation Block FedPapVisionNet includes a Squeeze-and-Excitation (SE) block, which allows for the channel-wise feature representation by explicitly modelling inter-channel dependencies. The SE block allocates importance weights to each individual channel so that the network can highlight informative features and suppress uninformative ones.
The squeeze operation applies global average pooling (GAP) in space to produce a channel-wise descriptor of size 1 × 1 × C1×1×C1×1×C. The channel descriptor goes through two dense layers, with a reduction ratio of r = 16. Following this, a sigmoid activation is performed to get the attention weights. The SE operation is defined as:

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

where $W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{C \times \frac{C}{r}}$ are learnable weight matrices, and σ denotes the sigmoid activation function. This mechanism improves representational capacity without significantly increasing computational complexity [20].

ConvResidualFusion Block The novel ConvResidualFusion block of FedPapVisionNet has parallel features extraction strategies aiding fusion process. It comprises two simultaneous pathways.

- **Convolutional Path:** A standard convolution followed by batch normalization.
- **Residual Path:** A full residual block with identity-based skip connections.

The outcomes from both routes are summed at element level and passed through a ReLU activation. It allows the model to take concurrent advantages of convolutional transformation and residual learning, enriching the feature representation and leading to discriminative features.

Global Average Pooling Subsequent to hierarchical feature extraction, the network utilizes Global Average Pooling (GAP) that reduces each feature map to a scalar. The parameters are significantly reduced by GAP compared to the flattening operations, resulting in a reduced risk of overfitting and improved generalization while training on limited medical datasets. Unlike flattening, which produces a high-dimensional feature vector dependent on spatial resolution, GAP generates a fixed-size representation regardless of input dimensions. This makes the architecture more robust and flexible for varying image sizes.

The pooled characteristics are subsequently directed to a lightweight classifier which includes a fully connected layer with batch normalization, ReLU activation and dropout rate 0.3, and then the final classification layer.
The proposed FedPapVisionNet model mathematically is explained here step by step. The explanation include number of filters and filter size, pooling size, any activation function used, dropout rate and number of neurons in the dense or fully connected layer [16].

$$X_{\text{input}} \in R^{128\times128\times3} \quad (3)$$

Input size 128 × 128 × 3, 3 defines RGB channel in the Eq. (3).

$$X_1 = \text{ReLU}(\text{BatchNorm}(\text{Conv2D}(32, 3\times3, \text{padding} =' \textit{same}')(X_{\text{input}}))) \quad (4)$$

Activation Function "ReLU", number of filter 32 and size (3×3), keeping the padding same in the Eq. (4) [16].

$$X_{\text{SE}} = X_1 \cdot \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(X_1))) \quad (5)$$

Where GAP is the global average pooling operation, $W_1 \in R^{2\times32}$ and $W_2 \in R^{32\times2}$ are weight matrices with reduction ratio 16, and σ represents the sigmoid activation in the Eq. (5).

$$X_2 = \text{ReLU}(\text{ConvPath}(64, 3\times3)(X_{\text{SE}}) + \text{ResPath}(64, 3\times3)(X_{\text{SE}})) \quad (6)$$

Activation Function "ReLU", number of filter for each convolution layer 64 and size (3×3) in the Eq. (6). ConvPath denotes the convolutional pathway and ResPath denotes the residual pathway.

$$X_3 = \text{MaxPooling2D}(2\times2)(X_2) \quad (7)$$

$$X_4 = \text{ReLU}(\text{ConvPath}(128, 3\times3)(X_3) + \text{ResPath}(128, 3\times3)(X_3)) \quad (8)$$

Activation Function "ReLU", number of filter for each convolution layer 128 and size (3×3) in the Eq. (8).

$$X_5 = \text{MaxPooling2D}(2\times2)(X_4) \quad (9)$$

All the maxpooling layers use pool size (2×2) shown in the Eqs. (7) and (9) [16].

$$X_6 = \text{GlobalAveragePooling2D}()(X_5) \quad (10)$$

Global average pooling in the Eq. (10).

$$X_7 = \text{Dropout}(0.5)(\text{ReLU}(\text{BatchNorm}(\text{Dense}(64)(X_6)))) \quad (11)$$

Dense layer with 64 neurons and dropout rate 50% in the Eq. (11).

$$Y = \text{Dense}(5)(X_7) \quad (12)$$

Final output layer with 5 neurons corresponding to number of classes in the dataset. Here, Y denotes output in the Eq. (12).

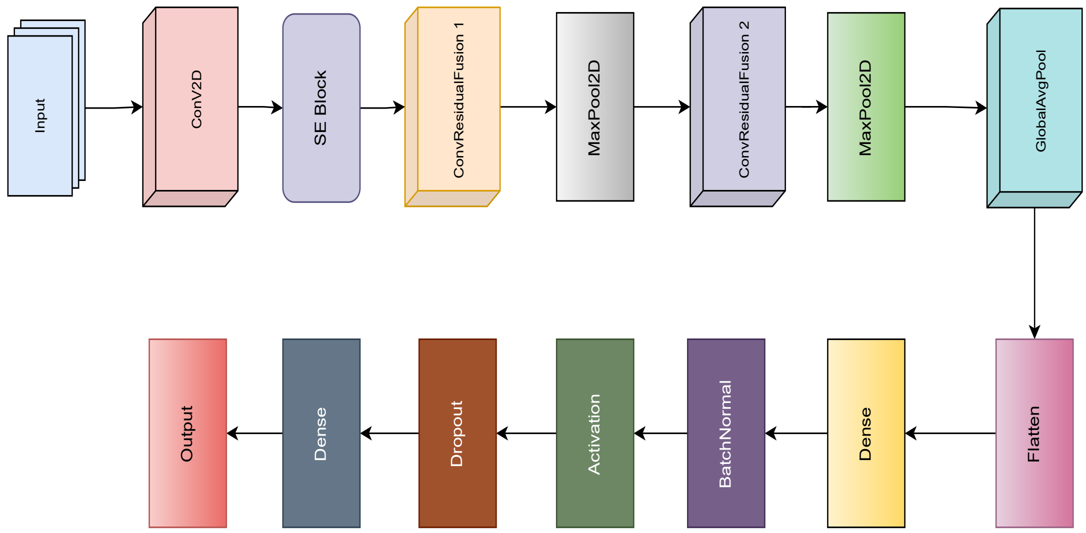
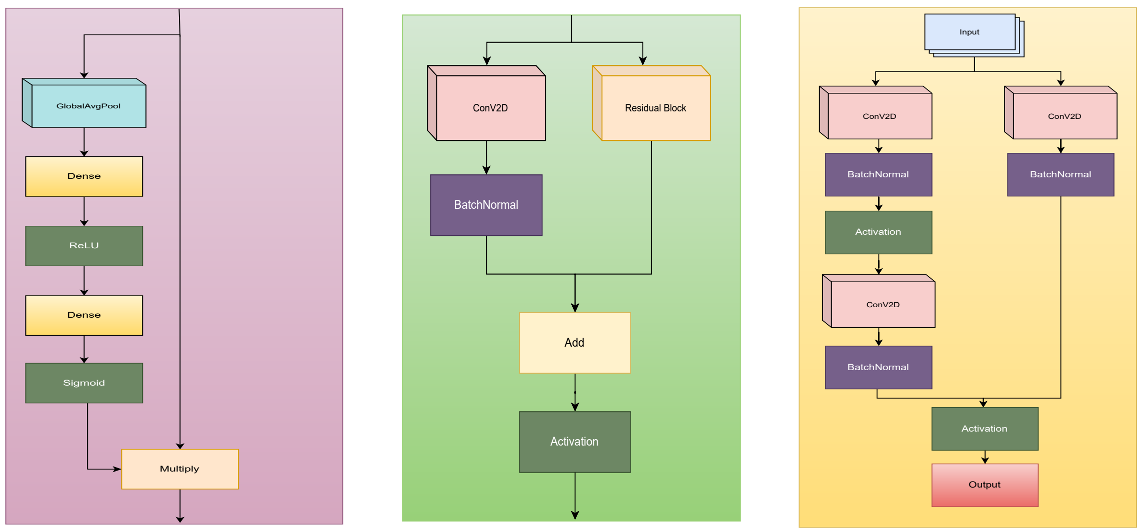


Figure 4.4: Architecture overview of the FedPapVisionNet model



(a) SE Block (b) ConvResidualFusion (c) Residual Block

Figure 4.5: Structural components used in the FedPapVisionNet architecture

4.2.3 Federated Learning Framework

This study adopts a federated learning (FL) approach for the training of deep neural networks in a decentralized manner while keeping privacy intact. Model training is not performed by aggregating all the training samples in the central server but is rather done collaboratively at multiple clients. Each of the clients has its own local dataset. The clients and the central server only exchange model parameters, ensuring that raw medical images never leave the local devices.
Let there be K participating clients indexed by $k \in \{1, 2, \dots, K\}$. Each client k possesses a local dataset $\mathcal{D}_k$ of size n_k, and the total number of samples across all clients is $n = \sum_{k=1}^{K} n_k$. The objective of federated learning is to minimize the global empirical risk:

$$\min_{\mathbf{w}} \ F(\mathbf{w}) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(\mathbf{w})$$

where $\mathbf{w}$ denotes the global model parameters and $F_k(\mathbf{w})$ represents the local objective function of client k.
All three proposed architectures : FedPapVisionNet, DualEffiMob-CBAM, and EffiResAttention are trained under the same federated learning protocol to ensure a fair and consistent comparison.

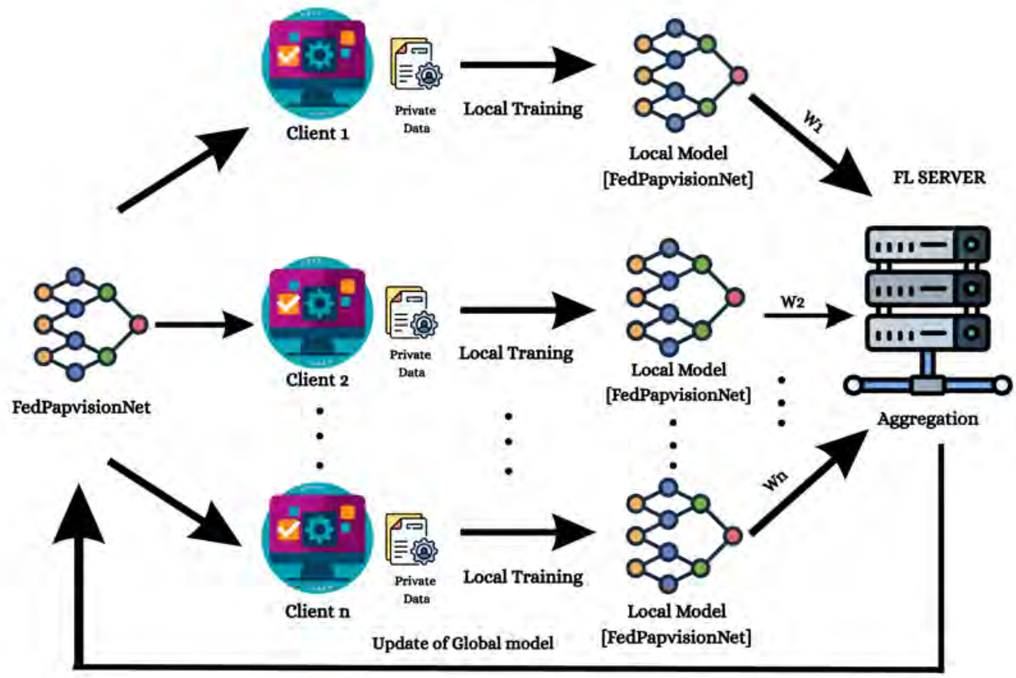


Figure 4.6: Federated Learning Workflow

Federated Averaging (FedAvg)

The baseline aggregation algorithm uses the Federated Averaging (FedAvg). In FedAvg, the central server first initializes a global model $\mathbf{w}^0$ and then transmits it to a subset of participating clients at the start of each communication round.

Local Training At communication round t, each selected client k receives the global model $\mathbf{w}^t$ and performs E epochs of local training based on stochastic gradient descent (AdamW-based optimization). The local objective is characterized as:

$$\mathbf{w}_k^{t+1} = \mathbf{w}^t - \eta \nabla F_k(\mathbf{w}^t)$$

where η denotes the learning rate and $\nabla F_k(\cdot)$ is the gradient computed on client k's local data.[2]

Global Aggregation After performing local training, every individual client then transmits the updated model parameters $\mathbf{w}_k^{t+1}$ to the central server. The server uses a weighted average to aggregate updates: [2]

$$\mathbf{w}^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \mathbf{w}_k^{t+1}$$

In the next communication round, clients get this aggregated model back.
FedAvg works well for independent and identically distributed (IID) data; however, in actual medical settings, client data distributions are often heterogeneous, hence the use of FedProx.

Federated Proximal Optimization (FedProx)

To address the challenges posed by non-IID data distributions, Federated Proximal Optimization (FedProx) is employed as an extension of FedAvg. FedProx introduces a proximal regularization term that constrains local model updates to remain close to the global model.

FedProx Local Objective The local optimization problem at client k is reformulated as:

$$\min_{\mathbf{w}} \; F_k(\mathbf{w}) + \frac{\mu}{2} \|\mathbf{w} - \mathbf{w}^t\|_2^2$$

where :

μ is the proximal coefficient controlling the strength of the regularization term,

$\mathbf{w}^t$ represents the global model parameters at round t.

The corresponding update rule becomes:

$$\mathbf{w}_k^{t+1} = \mathbf{w}^t - \eta \left(\nabla F_k(\mathbf{w}^t) + \mu(\mathbf{w}^t - \mathbf{w}) \right)$$

This proximal term stabilizes local updates and reduces client drift, particularly when local data distributions vary significantly across clients [10].

Aggregation The server aggregation step in FedProx remains identical to FedAvg:

$$\mathbf{w}^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \mathbf{w}_k^{t+1}$$

Stochastic Controlled Averaging (SCAFFOLD)

Stochastic Controlled Averaging (SCAFFOLD) algorithm has been implemented to address the client drift phenomenon in federated learning with heterogeneous data distributions. SCAFFOLD employs control variates to correct local update directions, ensuring convergence stability without requiring hyperparameter tuning for data heterogeneity.

The algorithm maintains dual control variates: a global server variate c and local client variates c_i. These variates estimate gradient directions and are updated iteratively to mitigate divergence between local and global objectives. The correction term $(c - c_i)$ is applied to local gradients, effectively reducing variance across clients with disparate data distributions.
For N clients with local objectives $f_i(x)$, the global minimization problem is defined as:

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

In each communication round r, participating clients $i \in S$ perform K local updates with drift correction [5]:

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

where η_l is the local learning rate. The control variates are updated using the efficient Option II formulation:

$$c_i^{r+1} = c_i^r - c^r + \frac{1}{K\eta_l} \left(x^r - y_i^{r,K} \right)$$

Server aggregation follows the weighted averaging scheme:

$$x^{r+1} = x^r + \frac{\eta_g}{|S|} \sum_{i \in S} \left(y_i^{r,K} - x^r \right)$$

$$c^{r+1} = c^r + \frac{1}{N} \sum_{i \in S} \left(c_i^{r+1} - c_i^r \right)$$

where η_g is the global learning rate. [5]

4.3 Data Collection

The foundation of this research lies in utilizing reliable and well-annotated cervical cell image datasets that enable accurate and automated detection of cervical cancer using deep learning. Data collection in medical image analysis plays a crucial role in building models capable of identifying subtle morphological differences between healthy and abnormal cells. Since this research aims to classify cervical cells into multiple distinct categories while maintaining data privacy in a federated learning setup, it is essential to use a dataset that provides a comprehensive representation of various cell types captured under consistent laboratory conditions.

The Herlev dataset, one of the earliest publicly available Pap smear image collections, was initially considered for this research. It contains 917 single-cell images collected at Herlev University Hospital, Denmark, categorized into seven classes, including both normal (superficial, intermediate, parabasal) and abnormal (mild, moderate, severe dysplasia, and carcinoma in situ) cells. While the dataset has been widely used for binary classification tasks (normal vs. abnormal), its relatively small size and limited variability make it less suitable for robust multiclass classification. These constraints reduce its effectiveness for deep learning-based models that require a larger and more diverse dataset. Therefore, to overcome these limitations and enable a finer differentiation between normal, pre-cancerous, and abnormal cells, the SIPaKMeD dataset was chosen as a more comprehensive alternative.

The SIPaKMeD (Single-cell Pap Smear) dataset, introduced by Plissiti et al. (2018), offers a significantly larger and more diverse collection of 4049 single-cell images obtained from Pap smear slides . Each cell image was manually segmented by expert cytopathologists, ensuring precise delineation of the nucleus and cytoplasm for high-quality feature extraction. The dataset is divided into five cytological classes such as Superficial–Intermediate, Parabasal, Koilocytotic, Dyskeratotic, and Metaplastic representing a wide spectrum of normal and abnormal cervical cell types. Images were captured using an OLYMPUS BX53F optical microscope with a CCD camera (Infinity 1 Lumenera) under the Papanicolaou (Pap) staining technique, enhancing the visual contrast of cellular structures. The dataset's well-balanced distribution, high-resolution imaging, and expert annotations make it ideal for training and evaluating deep learning models in cervical cancer detection. Moreover, being open-source, it complies with ethical and data-sharing standards, allowing for reproducible and privacy-conscious research in federated learning environments [3].

4.4 Data Description

The SIPaKMeD dataset includes images of 4049 isolated cells, which are manually cropped from clustered cell images of Pap slides. The primary feature of this dataset is the classification of samples. There are five cytological classes used for diagnosis, which are described below.

- **Dyskeratotic Cells**: Often indicative of infection, inflammation, or precancerous changes.
- **Koilocytotic Cells**: A common sign of Human Papillomavirus (HPV) infection.
- **Metaplastic Cells**: Transformed squamous cells originating in the cervix that may or may not be cancerous.

- **Parabasal Cells**: The smallest squamous epithelial cells in the body; an elevated presence may indicate abnormality.
- **Superficial–Intermediate Cells**: The most commonly observed cells in Pap tests, which may show morphological alterations.

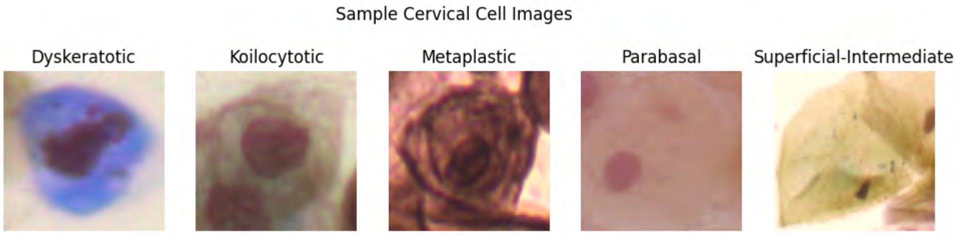


Figure 4.7: Sample Cervical Cell Images

Each image was collected under consistent lighting, magnification, and staining conditions using Papanicolaou (Pap) techniques. Metadata such as patient identifiers were excluded to maintain data privacy. The dataset's structure, with its clear, multi-class labeling, provides a robust ground truth for supervised learning, enabling the development of a nuanced classification model that goes beyond a simple binary (normal/abnormal) diagnosis. For simple binary diagnosis Herlev Dataset is being used. Mostly for training and testing machine learning models to classify individual cervical cells as normal or abnormal.

4.5 Data Analysis

To facilitate the development of a federated learning (FL) system, the SIPaKMeD dataset required careful structuring and partitioning to simulate a real-world, multi institutional collaborative environment. The primary objective was to create a data pipeline that reflects the distributed nature of healthcare data while preserving patient privacy with a core tenet of FL approach. The distribution of images across the five classes is relatively balanced, which helps mitigate potential issues with class imbalance during model training. The original class distribution is as follows:

Class	Number of Images	Percentage
Superficial–Intermediate	825	20.4%
Parabasal	780	19.3%
Koilocytotic	828	20.4%
Dyskeratotic	813	20.1%
Metaplastic	803	19.8%

Table 4.1: Distribution of images across different classes.

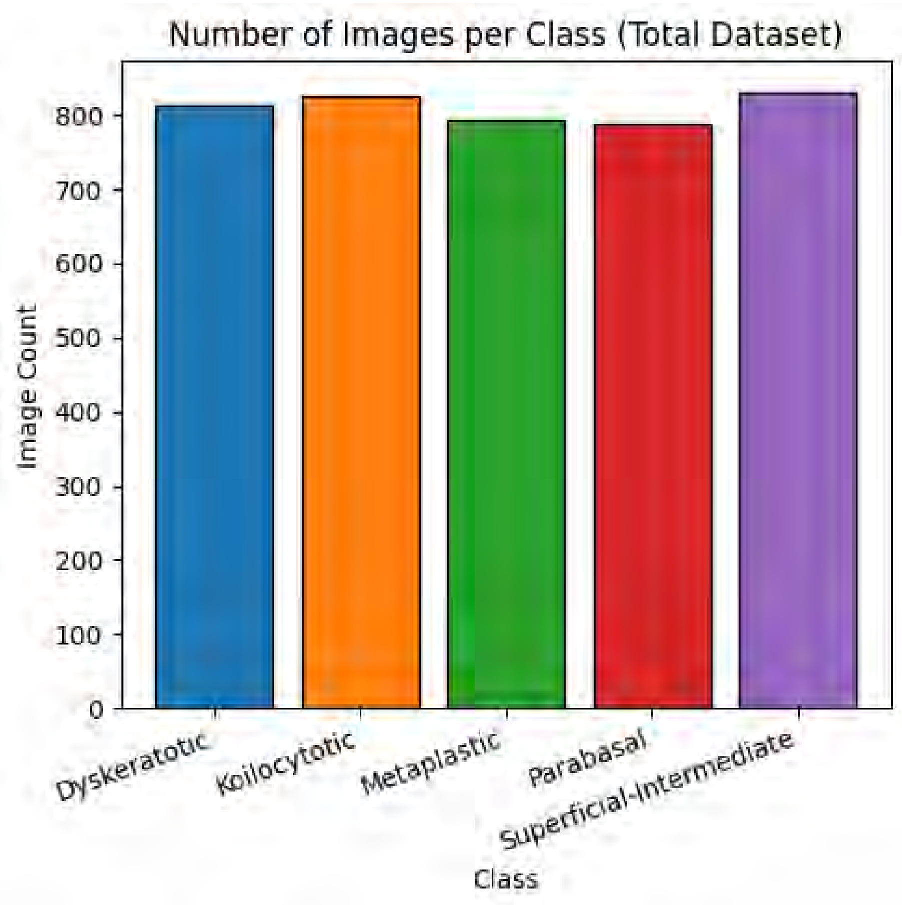


Figure 4.8: Distribution of images across different classes

4.5.1 Data Partition

For this study, the SIPaKMeD dataset is divided into two setups; IID and non-IID to emulate the various distribution of data between different medical institutions in a federated learning environment. In the IID (Independent and Identically Distributed) scenario, the dataset was uniformly partitioned among the clients through random permutation so as to ensure each client gets a balanced and representative subset containing samples of all five classes of cervical cells. This set-up assumes an ideal case, where each participating hospital possesses a varied collection of cell images. On the other hand, the non-IID setup was designed to simulate heterogeneous behavior similar to that of healthcare centers. The distribution of the dataset was imbalanced here, where each client contained images from mostly one class, which introduced more variations in both data and volume. Such intentions of the non-uniformity were to evaluate how stable and robust the federated model could be and also how flexible it would become to heterogeneous data.

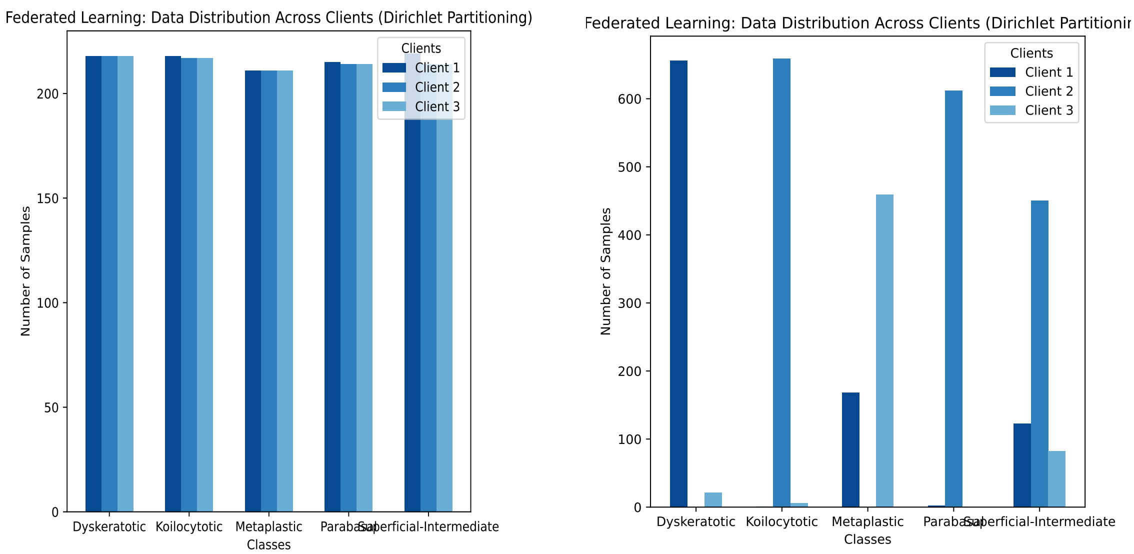


Figure 4.9: IID & Non IID Data settings

4.5.2 Visual Characteristic

The cervical cell images in all five categories were processed with pixel intensity histograms and Heatmap visualizations. The heat map shows morphologies and texture visually where the differences are not easy to see on the raw images.

- **Histogram Analysis**

  For each class, the histograms of intensities are created by converting the image to greyscale and getting the frequency distribution of the intensity of the pixels (0-255). The histograms show that the intensities are clearly different for the normal and abnormal classes:

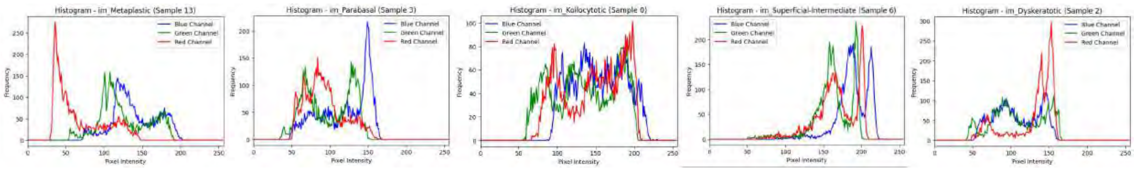


Figure 4.10: Heatmap Visualization of five Classes

  These intensity patterns provide useful cues for texture-sensitive CNN filters during the classification process.

- **Heatmap Visualization**

  To visualize spatial intensity concentration, class-wise heatmaps were produced by averaging pixel values across all images within each category. The resulting heatmaps reveal characteristic structural patterns:

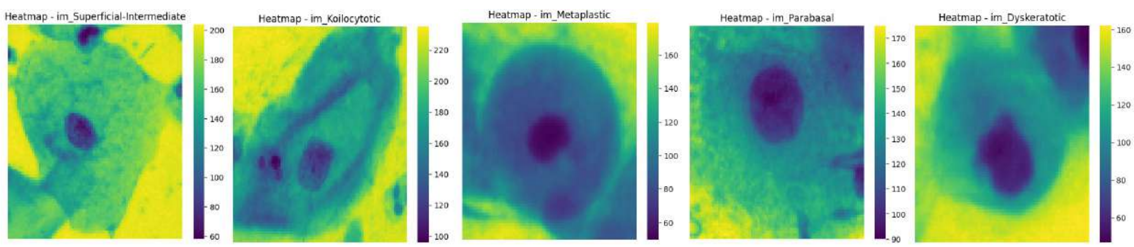


Figure 4.11: Heatmap Visualization of five cell images

  These heatmaps provide a global structural overview for each class, complementing the local pixel-level insights obtained from histograms.

- **Interpretation**

  The combined histogram and heatmap analyses demonstrate consistent visual differences between normal and abnormal cells. Morphological variations are reflected in measurable intensity distributions, which supports the suitability of deep learning models that rely on both local and global intensity variations for accurate classification.

4.6 Data Preprocessing

Several preprocessing steps were implemented on the SIPaKMeD dataset prior to training and evaluating the proposed model to prepare the data and improve classification accuracy. Image preprocessing is essential in medical imaging, as variations in image size, intensity, and background can significantly affect classification performance.

- **Image Size**: The SIPaKMeD dataset images are originally provided in different resolutions. All images were resized to 128×128 pixels. This resizing preserves the spatial structure of cells while satisfying the input requirements for batch processing and the architectural constraints of CNN-based models such as ResNet18, ResNet50, MobileNetV2, ViT, DenseNet121, VGG16, and EfficientNet-B0.
- **Image Normalization**: To ensure training stability and faster convergence, normalization was applied to all images prior to model training. Pixel intensities were first rescaled from the range $[0, 255]$ to $[0, 1]$. Subsequently, channel-wise normalization was performed using a mean of $[0.5, 0.5, 0.5]$ and a standard deviation of $[0.5, 0.5, 0.5]$. This normalization reduces the impact of illumination and staining variations in cytology images and enhances robustness in feature learning.
- **Random Horizontal Flip**: To increase dataset variability, random horizontal flipping was applied with a probability of 0.5. This augmentation enables the model to learn spatial invariance by exposing it to mirrored versions of the same cells.
- **Random Rotation**: Random rotation within a range of $\pm 15^\circ$ was applied to account for minor orientation changes that may occur during real-world image acquisition.
- **Color Jitter**: Color jittering was applied to adjust brightness, contrast, and saturation by a factor of 0.2, and hue by 0.1. This augmentation simulates variations in lighting and staining conditions, improving the model's robustness to illumination changes.
- **Random Affine Transformation**: A random affine transformation without rotation was applied using a translation factor of 0.1 along both spatial axes. These small positional shifts help the model generalize better to translated versions of cell images.

4.7 Implementation of Selected Designs

.

4.7.1 Deep learning models

A total of seven transfer learning models were employed on the proposed dataset, chosen for their proven success in image classification. Fine-tuning was performed on each model to evaluate their effectiveness and comparative performance.

Table 4.2: Experimental and Training Configuration of the Proposed Model

Category	Parameter	Value & Description
Federated Learning Setup	Number of clients	3
	Number of communication rounds	10
	Local epochs per round	3
	Total epochs per client	3
	Batch size	32
	Optimizer	Adam
	Initial learning rate	0.001
Training Configuration	Maximum learning rate	0.002
	Minimum learning rate	0.00001
	Learning rate schedule	Cosine annealing with warm-up
	Warm-up rounds	5
Model Architecture	Number of output classes	5
	Dropout rate	0.05
Input Configuration	Input image size	128 × 128 × 3
Dataset Split	Training data	80%
	Validation data	20%
	Resize	128 × 128
Training Data Augmentation	Random horizontal flip	Probability = 0.5
	Random rotation	±15 degrees
	Color jitter	Brightness, contrast, saturation = 0.2; hue = 0.1
	Random affine transformation	Translation = (0.1, 0.1), rotation = 0
Validation Preprocessing	Resize	128 × 128

4.7.2 Hybrid Models

EffiResAttention

Hyperparameter Configuration: As shown in Table 1, the hyperparameters used in EffiResAttention for medical image classification are employed with an image size of 128 × 128. In addition, the model has been trained for 25 rounds and 3 local epochs for IID and 50 rounds and 5 epochs for Non-IID using a learning rate of 0.001 and a batch size of 32. The model is compiled with the AdamW optimizer using default parameters. The division of data into training and testing takes place with a ratio of 8:2. In other words, the model is trained using 80% of the data and evaluated on the remaining 20%. This ratio was selected after assessing multiple splits and selecting the most suitable one.
The EffiResAttention architecture combines dual-path feature extraction for residual feature groups with attention assembly and high-roll temperature classification to create a robust symmetry. The digressive model is utilized for cervical cell classification in Pap smear image analysis.

DualEffiMob-CBAM

Hyperparameter Configuration The hyperparameters for DualEffiMob-CBAM medical image classification are represented in Table 1. The input images are applied at a training size of 128 $\times$ 128. The experiments consist of 25 rounds and 3 local epochs for IID and 50 rounds and 5 epochs for Non-IID, where the learning rate is 0.001 and the batch size is 32. The AdamW optimizer is utilized and cosine annealing is applied as a learning rate scheduling. An 80-20 split is used for dividing data into training and testing sets. This ratio was adopted based on empirical experiment observations. The model is trained on GPU for all experiments.
The DualEffiMob-CBAM architecture effectively combines dual-path feature extraction, attention-guided fusion, and temperature-scaled classification to create a robust model suitable for medical image analysis, particularly for cervical cell classification tasks in Pap smear image analysis.

FedPapVisionNet

Hyperparameter Configuration The hyperparameter configuration used in the proposed FedPapVisionNet Model Table 1 presents the hyperparameters used in classifying medical images. In this case, the image size is taken as 128 $\times$ 128, and 200 rounds and 5 local epochs for IID and 50 rounds and 5 epochs for Non-IID are utilized. The proposed FedPapVisionNet model is trained utilizing a learning rate of 0.001 and uses AdamW as the optimizer with 32 as the batch size. The Model is executed using the GPU for the model training. To add, a split of training-test is made in the ratio 8:2.
Different values for data splits were employed, and this paper found 8:2 to yield the best results. In this case, 80% is meant for training and 20% for testing. FedPapVisionNet is an architecture that integrates residual learning, channel attention mechanism and parallel fusion pathways making it an effective model for application in federated learning settings for the task of classification of Pap smear images.

4.7.3 Federated Learning Framework

Training Configuration and Optimization Details

To ensure consistency across experiments, all three models are trained using identical federated learning hyperparameters.

Table 4.3: Federated Learning Hyperparameters for EffiResAttention (IID Settings)

Category	**Parameter**	**Value**
Federated Learning Configuration	Number of clients (K)	3
	Client participation per round	100%
	Number of communication rounds	25
	Local epochs (E)	3
	Batch size	32
Optimization Strategy	Optimizer	AdamW
	Initial learning rate	0.001
	Learning rate scheduler	Cosine Annealing
	Scheduler $T_{\max}$	50
	Warmup strategy	Linear
	Weight decay	1×10^{-5}
Activation & Regularization	Activation function	ReLU
	Output activation	Softmax
	Dropout rate	0.5

Table 4.4: Federated Learning Hyperparameters for Dual EffiMob CBAM (IID Settings)

Category	**Parameter**	**Value**
Federated Learning Configuration	Number of clients (K)	3
	Client participation per round	100%
	Number of communication rounds	25
	Local epochs (E)	3
	Batch size	32
Optimization Strategy	Optimizer	AdamW
	Initial learning rate	0.001
	Betas	(0.9, 0.999)
	Learning rate scheduler	Cosine Annealing
	Warmup strategy	Linear
	Weight decay	1×10^{-5}
Activation & Regularization	Activation function	ReLU
	Output activation	Softmax
	Dropout rate	0.5

Table 4.5: Federated Learning Hyperparameters for FedPapsVisionNet (IID Settings)

Category	**Parameter**	**Value**
Federated Learning Configuration	Number of clients (K)	3
	Client participation per round	100%
	Number of communication rounds	200
	Local epochs (E)	5
	Batch size	32
Optimization Strategy	Optimizer	AdamW
	Initial learning rate	0.005
	Betas	(0.9, 0.999)
	Learning rate scheduler	Step-wise Decay
	Weight decay	1×10^{-5}
Activation & Regularization	Activation function	ReLU
	Output activation	Softmax
	Dropout rate	0.5

Table 4.6: Federated Learning Hyperparameters for EffResAttention (Non-IID Settings)

Category	**Parameter**	**Value**
Federated Learning Configuration	Number of clients (K)	3
	Client participation per round	100%
	Number of communication rounds	50
	Local epochs (E)	5
	Batch size	10
	Alpha (α)	0.3
	Mu (μ)	0.1
Optimization Strategy	Optimizer	AdamW
	Initial learning rate	0.0005
	Betas	(0.9, 0.999)
	Epsilon (ϵ)	1×10^{-8}
	Learning rate scheduler	Step-wise Decay
	Weight decay	5×10^{-4}
	Decay rate	0.96
Activation & Regularization	Activation function	ReLU
	Output activation	Softmax
	Dropout rate	0.5

Table 4.7: Federated Learning Hyperparameters for Dual EffiMob-CBAM (Non-IID Settings)

Category	**Parameter**	**Value**
Federated Learning Configuration	Number of clients (K)	3
	Client participation per round	100%
	Number of communication rounds	50
	Local epochs (E)	5
	Batch size	10
	Alpha (α)	0.3
	Mu (μ)	0.1
Optimization Strategy	Optimizer	AdamW
	Initial learning rate	0.0005
	Betas	(0.9, 0.999)
	Epsilon (ϵ)	1×10^{-8}
	Learning rate scheduler	Step-wise Decay
	Weight decay	5×10^{-5}
	Decay rate	0.96
Activation & Regularization	Activation function	ReLU
	Output activation	Softmax
	Dropout rate	0.5

Table 4.8: Federated Learning Hyperparameters for FedPapsVisionNet (Non-IID Settings)

Category	**Parameter**	**Value**
Federated Learning Configuration	Number of clients (K)	3
	Client participation per round	100%
	Number of communication rounds	50
	Local epochs (E)	5
	Batch size	10
	Alpha (α)	0.3
	Mu (μ)	0.1
Optimization Strategy	Optimizer	AdamW
	Initial learning rate	0.0005
	Betas	(0.9, 0.999)
	Epsilon (ϵ)	1×10^{-8}
	Learning rate scheduler	Step-wise Decay
	Weight decay	5×10^{-4}
	Decay rate	0.96
Activation & Regularization	Activation function	ReLU
	Output activation	Softmax
	Dropout rate	0.5

Chapter 5

Result Analysis

5.1 Performance Evaluation

5.1.1 Pretrained Models

As mentioned before, some pre-trained models like MobileNetV2, DenseNet112, ResNet18, ResNet50, VGG16, EfficientNet, ViT were implemented in the SipakMed dataset. Among them, from Fig. 5.1 the lightweight models, EfficientNet achieved the best accuracy of 97.16%, MobileNet achieved 95.93%, and ResNet18 achieved 95.31% accuracy. From the heavyweight pre-trained models, DenseNet121 achieved the highest accuracy of 97.04%, VIT achieved 96.42%, ResNet50 achieved 95.06%, and VGG16 achieved the lowest accuracy of 89.14%.

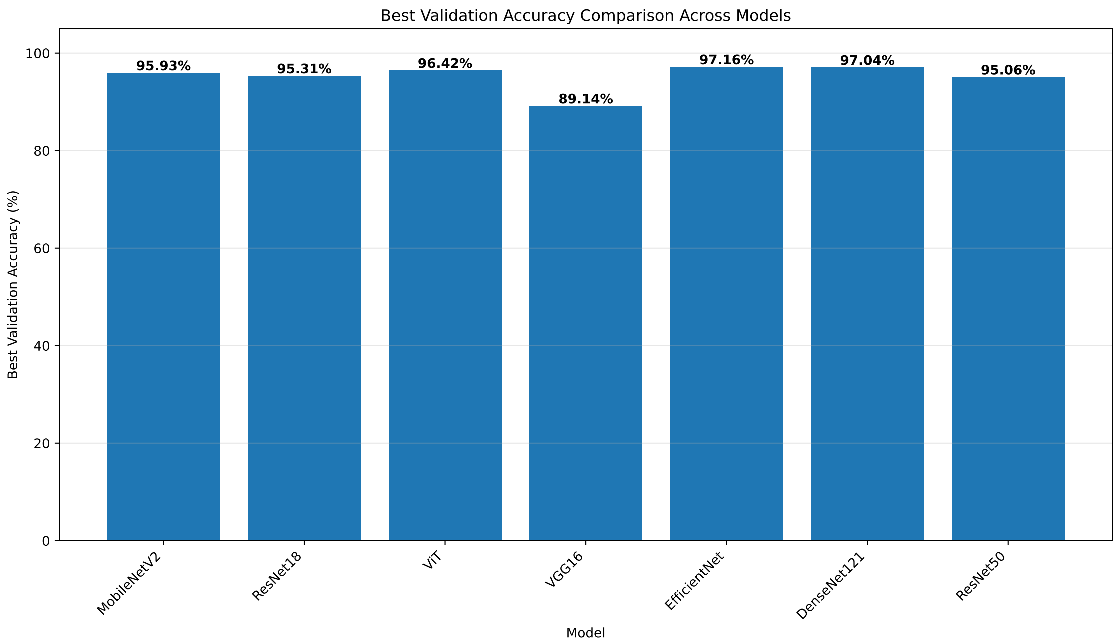
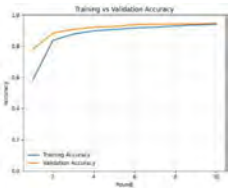


Figure 5.1: Accuracy comparison of pretrained models without federated learning

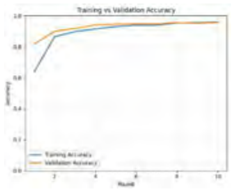
Several data augmentation techniques were used and based on the results some pre-trained models and custom models were implemented using Federated Learning algorithm for IID and Non-IID data partitions where FedAvg algorithm was used for IID partition and FedProx algorithm was used for Non-IID partition.

5.1.2 IID Data Partition

Pretrained Models

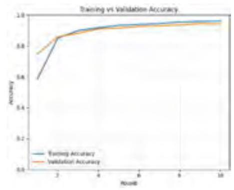


(a) ResNet18

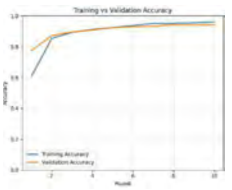


(b) ResNet50

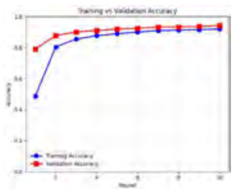
(c) DenseNet112



(d) VGG19

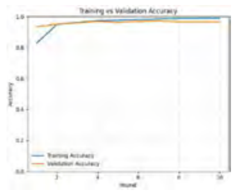


(e) VGG16



(f) MobileNetV2

(g) EfficientNet



(h) ViT

Figure 5.2: Training and validation accuracy curves of different deep learning architectures under IID data partition using the FedAvg algorithm

Table 5.1 summarizes the Performance Metrics which are Precision, Recall, F1-score and Accuracy for all models and classes. The ResNet50 model performed significantly better than all other models, reaching a maximum overall accuracy of 95.56%. ResNet18 may be significantly lower than ResNet50; its accuracy of 94.57% is not insignificant. The VGG19 model achieved an accuracy rate of 94.32%. However, its precision and recall were lower than that of ResNet models.

In terms of individual class performance, the Parabasal class had the maximum precision and recall in most models, achieving a score of (1.0) in several instances. The VGG16 and ViT models provided the best performance in Superficial-Intermediate, achieving F1-scores of 0.9860 and 0.9915, respectively.

The performance in precision of DenseNet112 was strong, especially for Dyskeratotic and Superficial-Intermediate classes with F1-scores of 0.9668 and 0.9886, respectively. EfficientNet showed good recall for Parabasal and Superficial-Intermediate classes, but the least precision score for Koilocytotic and Metaplastic classes resulted in overall accuracy of 89.38%, which was least among the models tested. MobileNetV2 and ResNet18 demonstrated a balance of precision, recall and F1-scores, with MobileNetV2 having a slight edge over ResNet18 in the Metaplastic class and Superficial-Intermediate class. The ViT model achieved the highest F1 scores of 0.9817 and 0.9915 for the Dyskeratotic and Superficial-Intermediate classes, respectively, and with an overall accuracy of 96.54% was the second best of all models.

It is evident that application of data augmentation techniques to federated learning models can improve generalization and model robustness when trained on diverse IID data across multiple clients.

Table 5.1: Class-wise performance comparison of Pretrained models in FedAvg Algorithm

Model	Class	Precision	Recall	F1-score	Support	Accuracy
DenseNet112	im_Dyskeratotic	0.9704	0.9632	0.9668	136	0.9531
	im_Koilocytotic	0.8951	0.9119	0.9034	159	
	im_Metaplastic	0.9064	0.9337	0.9199	166	
	im_Parabasal	1.0000	0.9711	0.9853	173	
	im_Superficial-Intermediate	0.9943	0.9830	0.9886	176	
EfficientNet	im_Dyskeratotic	0.7870	0.9779	0.8721	136	0.8938
	im_Koilocytotic	0.8867	0.8365	0.8608	159	
	im_Metaplastic	0.8457	0.8916	0.8680	166	
	im_Parabasal	1.0000	0.7861	0.8803	173	
	im_Superficial-Intermediate	0.9667	0.9886	0.9775	176	
MobileNetV2	im_Dyskeratotic	0.9362	0.9706	0.9531	136	0.9407
	im_Koilocytotic	0.8910	0.8742	0.8825	159	
	im_Metaplastic	0.8882	0.9096	0.8988	166	
	im_Parabasal	0.9940	0.9538	0.9735	173	
	im_Superficial-Intermediate	0.9887	0.9943	0.9915	176	
ResNet18	im_Dyskeratotic	0.9301	0.9779	0.9534	136	0.9457
	im_Koilocytotic	0.9007	0.8553	0.8774	159	
	im_Metaplastic	0.9167	0.9277	0.9222	166	
	im_Parabasal	1.0000	0.9711	0.9853	173	
	im_Superficial-Intermediate	0.9722	0.9943	0.9831	176	
ResNet50	im_Dyskeratotic	0.9301	0.9779	0.9534	136	0.9556
	im_Koilocytotic	0.9006	0.9119	0.9062	159	
	im_Metaplastic	0.9506	0.9277	0.9390	166	
	im_Parabasal	0.9940	0.9653	0.9795	173	
	im_Superficial-Intermediate	0.9943	0.9943	0.9943	176	
VGG16	im_Dyskeratotic	0.9103	0.9706	0.9395	136	0.9407
	im_Koilocytotic	0.8780	0.9057	0.8916	159	
	im_Metaplastic	0.9533	0.8614	0.9051	166	
	im_Parabasal	0.9824	0.9653	0.9738	173	
	im_Superficial-Intermediate	0.9724	1.0000	0.9860	176	
VGG19	im_Dyskeratotic	0.9496	0.9706	0.9600	136	0.9432
	im_Koilocytotic	0.8994	0.8994	0.8994	159	
	im_Metaplastic	0.8857	0.9337	0.9091	166	
	im_Parabasal	1.0000	0.9191	0.9578	173	
	im_Superficial-Intermediate	0.9831	0.9943	0.9887	176	
ViT	im_Dyskeratotic	0.9781	0.9853	0.9817	136	0.9654
	im_Koilocytotic	0.9193	0.9308	0.9250	159	
	im_Metaplastic	0.9394	0.9337	0.9366	166	
	im_Parabasal	1.0000	0.9827	0.9913	173	
	im_Superficial-Intermediate	0.9887	0.9943	0.9915	176	

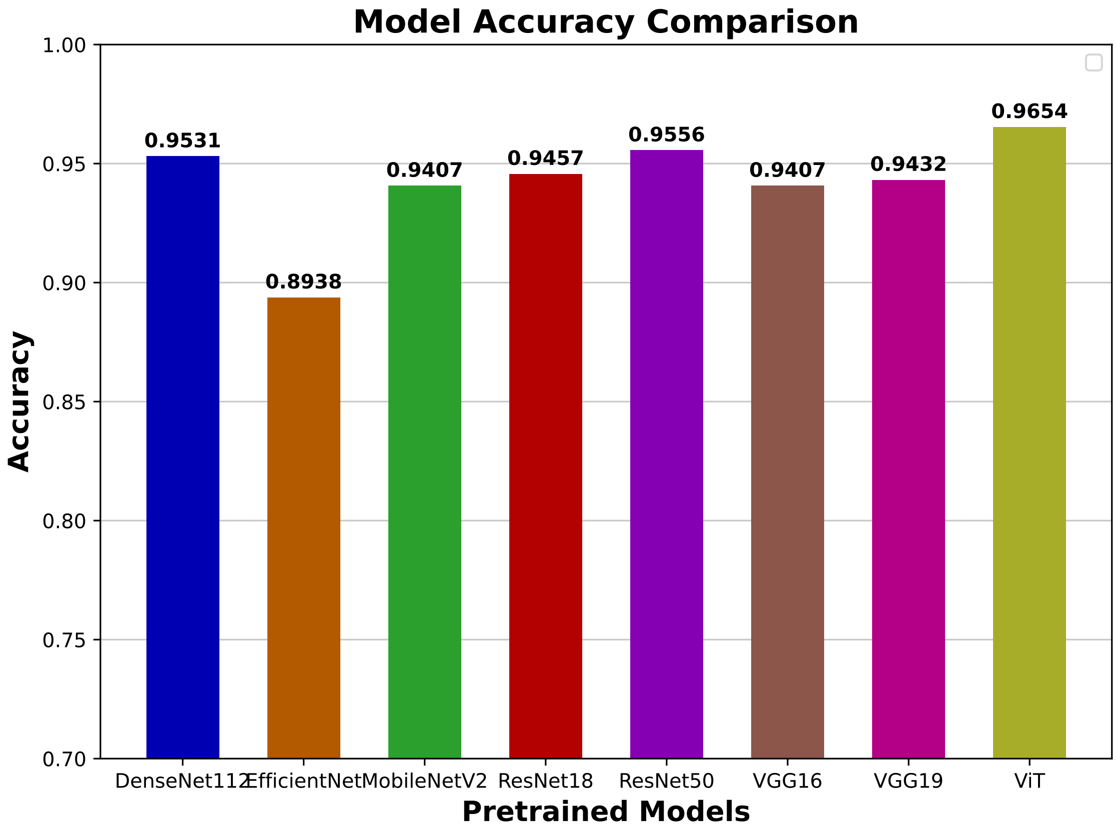


Figure 5.3: Accuracy comparison of pretrained models in federated learning.

Hybrid Models

DualEffiMob-CBAM

The use of FedAvg algorithm on the IID partition of the dataset using custom DualeffiMob-CBAM shows quite a strong classification performance. The classification report shows that all classes have impressive results, with Superficial-Intermediate having the highest precision (0.9887) and recall (0.9943). The F1-score is also well-balanced at 0.9915.

In general, the accuracy of the model was 95.43% indicating its good generalization ability on the dataset. With that, the macro average F1-Score of 0.9532 and weighted average F1-Score of 0.9540 demonstrate consistent performance across classes.

The training and validation loss curve shows a clear and more rapid decrease in training and validation loss. The training loss reached 0.2211 by the completion of training rounds. Thus, model convergence is observed efficiently. The graph on training and validation accuracy shows a steady increase and that the validation accuracy in 95.43. Therefore, the model can fit well to the data.

The DualeffiMob-CBAM model, which has 8,528,611 parameters, achieves an efficient balance between model complexity and performance. This custom model combining efficient architecture and CBAM (Convolutional Block Attention Module) produced good results for federated learning. The medical image classification results show high accuracy and a low number of parameters.

EfiiResAttention

By making use of EffiResAttention custom model on the IID partition of the dataset with the help of FedAvg algorithm, solid results were achieved. It was trained for 25 epochs and 3 rounds. The classification report shows very good results in all classes. The class Superficial-Intermediate has the highest recall and F1 score, scoring 1 and 0.9915 respectively. These results point to a model with good calibration.

So, the overall accuracy of the model was 95.93% with the macro average and weighted average F1-score of 0.9588 and 0.9591 respectively.

As we can see in the above curve representing the training vs validation loss, both training and validation loss have dropped down to negligible amounts quite rapidly especially at the initial rounds.

In other words, the training loss goes down to a final value of 0.2101 while the validation loss hits a final value of 0.1452. The model is successfully learning and converging. The training vs validation accuracy graph shows that the accuracy scores of the training and validation sets are continuously increasing with final validation accuracy of 95.93%. The low gap of 0.0132 between training and validation accuracy finally shows that the model does not overfit but generalizes.

With 16,670,243 parameters, the EffiResAttention model is a good balance of complexity and performance. These results show that this model performs very well in federated learning tasks with good accuracy and optimal number of parameters.

Proposed Model: FedPapVisionNet

The FedAvg algorithm was used to apply the proposed custom model architecture FedPapVisionNet on the IID partition of the data. We employed this model for 200 rounds and 5 epochs per round. The classification report and per-class accuracy indicate that the model produces strong results. With an overall accuracy of 95.68%, it showed excellent performance on all classes. The Superficial-Intermediate class attained the greatest recall (0.9830) and F1-score (0.9858) of all classes, indicating the effectiveness of the model in capturing relevant information. Interestingly, the macro average F1-score and weighted average F1-score was 0.9556 indicating the balance performance across classes.

The training vs validation loss curve shows a drop in the loss in the first few rounds. At the end of the training, this value remained constant. Both training and validation loss converge to a low value. The training versus validation accuracy curve shows that both accuracies improve with the model's validation accuracy getting closer to 1.0. This indicates that the model generalizes well.

The per-class recall is quite impressive. The Dyskeratotic has 97.06%, Koilocytotic has 90.57%, Metaplastics has 93.37%, Parabasal has 98.84% and Superficial-Intermediate has 98.30%. The results show that FedPapVisionNet combines performance and efficiency with 3,90,215 parameters and performs with high accuracy.

Table 5.2: Classification Report For IID Data Partition

Model	Parameters	Class	Precision	Recall	F1-score	Support	Accuracy
DualEffiMob-CBAM	8,528,611	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	
EfiiResAttention	16,670,243	Dyskeratotic	0.964	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	390,215	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	

5.1.3 Non-IID Data Partition(FedProx)

Setting: $\alpha = 0.3$ and $\mu = 0.01$

DualEffiMob-CBAM

After applying the DualEffiMob-CBAM model on the non-IID partition through the FedProx approach, a solid model performance was obtained on the dataset. The classification report demonstrates a generalization of 96%, indicating a good learning process despite the effects of non-IID. The Parabasal class demonstrated outstanding performance as it recorded a perfect precision of 1.00 and a recall of 0.99. Further, the Superficial-Intermediate class also performed well as it recorded a precision of 0.99 and recall of 1.00.

Nonetheless, the performance of the model differed marginally across the other classes. The Dyskeratotic class showed a precision of 0.94 and a recall of 0.97, indicating good performance but a need for improvement in precision. The precision of the Koilocytotic class was somewhat lower than other classes, with values of 0.90 and 0.93 with respect to recall, and an F1 score of 0.92. The Metaplastic class exhibited greater precision of 0.97 and recall of 0.92, but its F1-score of 0.95 was slightly lower than expected. Analysis of the training vs validation loss curve shows the model has converged as evident from the sharp fall of both the training and validation loss in the initial rounds and stabilizing later on. The validation accuracy was consistently high despite a large training to validation accuracy gap, which shows that the validation accuracy is high despite the use of non-IID data. To sum it up, the DualEffiMob-CBAM model implemented with FedProx algorithm showcased good performance on all classes with good overall accuracy. Besides that, the model achieved impressive results in most classes. However, there is still room to improve the precision of the Koilocytotic and Dyskeratotic classes.

EffiResAttention

After implementing the EffiResAttention model on the non-IID partition via the FedProx algorithm, the performance was solid across the dataset. The classification report presents an overall accuracy of 96%, indicating strong generalization despite the fact that the data is not IID. The Superficial-Intermediate class had the highest recall (1.00) and F1-score (0.99). The other two classes Dyskeratotic and Metaplastic also performed well with precision values of 0.95 and 0.96, respectively.

The training vs validation loss curve shows that the model is well-converged as it shows a sharp decline in both training and validation losses. The validation accuracy was very good and remained high through the rounds. The gap between validation & training accuracy is expected in non-IID data. Nevertheless, the model showed consistent performance. In conclusion, the EffiResAttention model with FedProx was able to produce accurate predictions even when the data distributions were non-IID. Further, the accuracy for all the classes of different non-IID datasets was also found to be similar.

Proposed Model: FedPapVisionNet

FedPapVisionNet achieves strong generalization as well as stable convergence when trained with FedProx in Dirichlet non-IID setting. We see that the validation accuracy rises significantly in the early rounds and then stabilizes to 0.858. This indicates that we have successfully mitigated information loss due to client drift. The loss curves suggest that the model is being optimized consistently as the training loss reduces to around 0.22, while the validation loss reduces to around 0.39. With only a small gap between train and validation metrics, there's no huge overfitting in the model as there are low fluctuations in the validation metrics trends after 35-40 rounds of training.

The classification report shows strong performance per class and overall accuracy = 0.86. The highest F1 scores are obtained with Parabasal and Superficial–Intermediate (0.94), while Dyskeratotic remains strong (F1 = 0.91). Koilocytotic (F1 = 0.76) and Metaplastic performance is lower and metaplastic class has high precision (0.94) but low recall (0.57), indicating under-detection and potential class confusion.

Table 5.3: Classification Report for Non-IID Data Partition (FedProx)

Model	**Class**	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
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	

5.1.4 Non-IID Data Partition(SCAFFOLD)

Proposed Model: FedPapVisionNet

FedPapVisionNet gets solid and stable performance with SCAFFOLD after 50 communication rounds and with 3 local epochs per client under a non-IID data partition. Through 810 samples the overall test accuracy comes to 0.90. The control-variates mechanism in SCAFFOLD appears to be effective in mitigating client drift in global updates of the model. According to the scores generated for each class, this model is fairly balanced since we have Dyskeratotic with an F1-score of 0.88 (precision 0.86, recall 0.90) and Koilocytotic with an F1-score of 0.81, which suggests this class is still more difficult under non-IID distributions. Metaplastic has a strong performance (F1 = 0.89), with a high precision (0.92) indicating a reliable positive.

The models with the best F1 scores are the Parabasal (F1 = 0.96) and Superficial-Intermediate (F1 = 0.96), with Superficial-Intermediate boasting a particularly high recall (0.99). This means the model does not miss these cases very often. With a macro average F1 of 0.90, the model performs consistently across classes rather than being dominated by a majority class. All in all, the results indicate that SCAFFOLD effectively trains FedPapVisionNet under realistic, skewed client-data distributions.

Table 5.4: Classification Report for FedPapVisionNet under Non-IID Data Partition (SCAFFOLD)

Class	Precision	Recall	F1-score	Support	Accuracy
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	

5.2 Analysis of Design Solutions

5.2.1 IID Data Partition

Based on computational efficiency performance on the IID partition, the FedPapVisionNet model worked best for federated learning tasks. The accuracy of the two models, EffiResAttention (95.93%) has slightly better accuracy than DualEffiMob-CBAM(95.43%). The proposed FedPapVisionNet achieved an accuracy of 95.68%. However, FedPapVisionNet is a better choice as it has a reasonable number of parameters which is 390,215 when compared to the EffiResAttention model which has 16,670,243 parameters and DualeffiMob-CBAM model which has 85,28,611parameters. Due to its lightweight architecture, FedPapVisionNet is particularly suitable for federated learning since it minimizes server computation for the communication cost of clients.

Federated learning motivates the utilization of lightweight models in federated systems to enable them to scale faster while training and conveying data efficiently. Even though it is a smaller model, FedPapVisionNet achieved an overall accuracy of 95.68%, while the Superficial-Intermediate class reached a recall of 0.9830 and an F1-score of 0.9858, indicating good performance across classes.

Notably, both the training and validation loss curves demonstrate a rapid convergence, with early stabilization indicating efficiency. The curves for the accuracy of training and validation increase steadily. The model's validation accuracy indicates good generalization.

To sum up, although several models perform better than our FedPapVisionNet model in terms of accuracy, FedPapVisionNet is more efficient because of its lightweight architecture. Thus, FedPapVisionNet is recommended for federated learning on IID partitions as it is the most efficient, scalable, and performance-efficient model.

5.2.2 Non-IID Data Partition

The three designs evaluated have different trade-offs with respect to model capacity, convergence behavior and robustness under non-IID client distributions. According to the parameter efficiency perspective, FedPapVisionNet (390,215 parameters) is much more lightweight than DualEffiMob-CBAM (85,28,611), and EffiResAttention (1,66,70,243) which are both heavier. This difference is important in federated learning, where clients have limitations on their compute, memory and communication. By shrinking local training cost and uplink payload size, FedPapVisionNet gets more practical for large-scale or resource-limited deployments.
Considering FedProx with $\alpha = 0.3$ and $\mu = 0.01$, it is evident that both DualEffiMob-CBAM and EffiResAttention achieve an overall accuracy of 0.96. As such, their predictive capacity is relatively strong. Notably, this could be due to their significantly larger representational power. In particular their Parabasal and Superficial–Intermediate class results are consistently high. Nevertheless, this added performance comes at the toll of heavier models that can impact a client-side burden, thus making participation in real-world FL more challenging.
FedPapVisionNet is comparatively lighter, and under the FedProx, it achieved an accuracy of 0.86. This may indicate that lighter models are limited in their performance extensions. Particularly, in Metaplastic, we see high precision but low recall. This indicates under-detection of the class as well as confusion with other classes. Under FedProx, the model may necessitate stronger feature learning, or could benefit from auxiliary techniques such as class-balancing, improved aggregation and/or architectural refinement to match the larger baselines.
Sussequently, SCAFFOLD improves FedPapVisionNet with 0.90 accuracy, meaning that the optimization strategy affects lightweight models under heterogeneity. SCAFFOLD employs control variates that lessen client drift, enhancing class balance and partially bridging the gap with larger networks. With respect to efficiency and deployability practicality, overall, FedPapVisionNet is superior to existing models. At the same time, it falls short of larger models in peak accuracy especially under FedProx hinting at an efficiency-accuracy trade-off rather than a fundamental deficiency.

5.3 Statistical Analysis

5.3.1 Confusion Matrix

IID Data Partition

Across the IID partition, all three confusion matrices are strongly diagonal-dominant, indicating high class separability with limited cross-class leakage. EffiResAttention achieves the highest overall correctness (777/810), corresponding to an accuracy of about 0.959, with particularly clean recognition of Superficial–Intermediate (176/176 correct) and strong Parabasal separation (168/173 correct). The remaining errors are concentrated in clinically similar categories, especially Koilocytotic and Metaplastic, where Koilocytotic is most often predicted as Metaplastic (8 cases) and Metaplastic is most often predicted as Koilocytotic (9 cases). This symmetric confusion suggests feature overlap between these two categories rather than random misclassification.

FedPapVisionNet yield 775/810 and DualEffiMob-CBAM yield 773/810 correct predictions, showing near-parity in aggregate performance while differing in error structure. For FedPapVisionNet, the dominant confusion is Metaplastic to Koilocytotic (10 cases), alongside Koilocytotic to Dyskeratotic (8 cases). This pattern implies that FedPapVisionNet tends to pull Metaplastic samples into the Koilocytotic decision region, which reduces Metaplastic recall relative to EffiResAttention. However, Parabasal and Superficial Intermediate remain highly stable (171/173 and 173/176 correct),meaning the model preserves strong boundary definition for these visually distinctive classes.

DualEffiMob-CBAM shows the strongest Dyskeratotic recall (135/136 correct) but the weakest Koilocytotic recall among the three, driven by Koilocytotic being split across Dyskeratotic (9), Metaplastic (9), and Parabasal (2). In practical terms, EffiResAttention offers the most balanced reduction of the major confusion pair (Koilocytotic and Metaplastic), while the other two models maintain comparable global accuracy but concentrate more errors in that same clinically adjacent region.

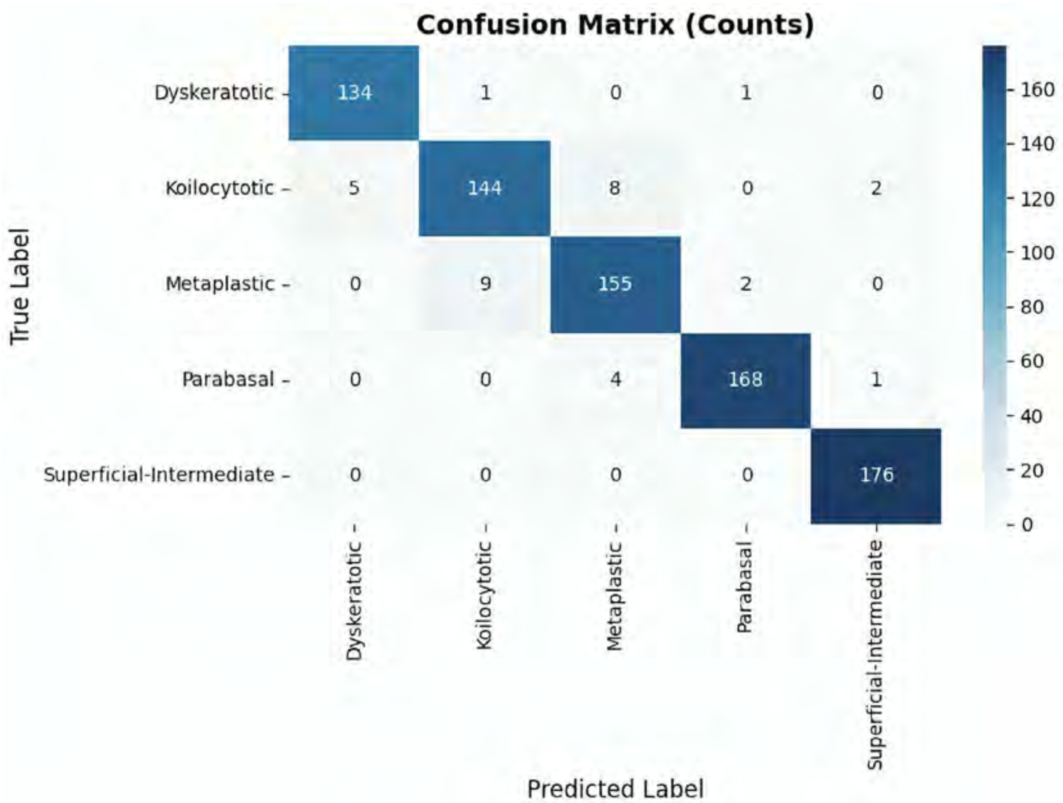


Figure 5.4: Confusion matrix of EffiResAttention under IID settings

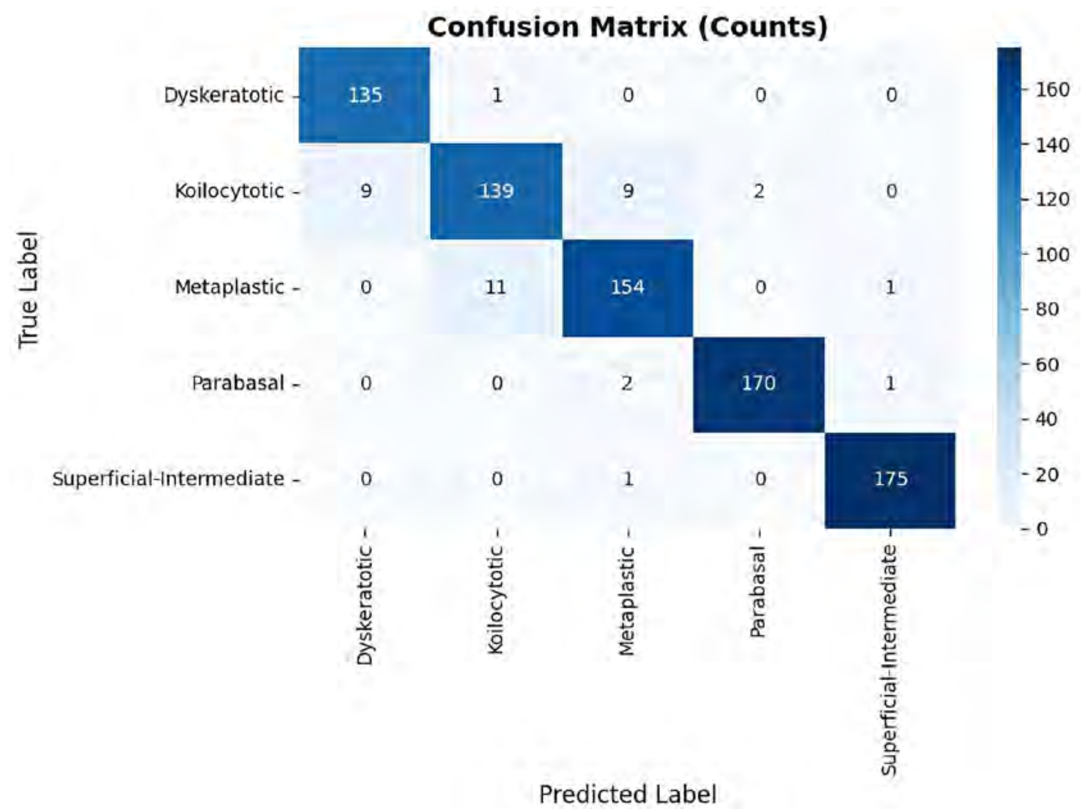


Figure 5.5: Confusion matrix of DualEffiMob-CBAM under IID settings

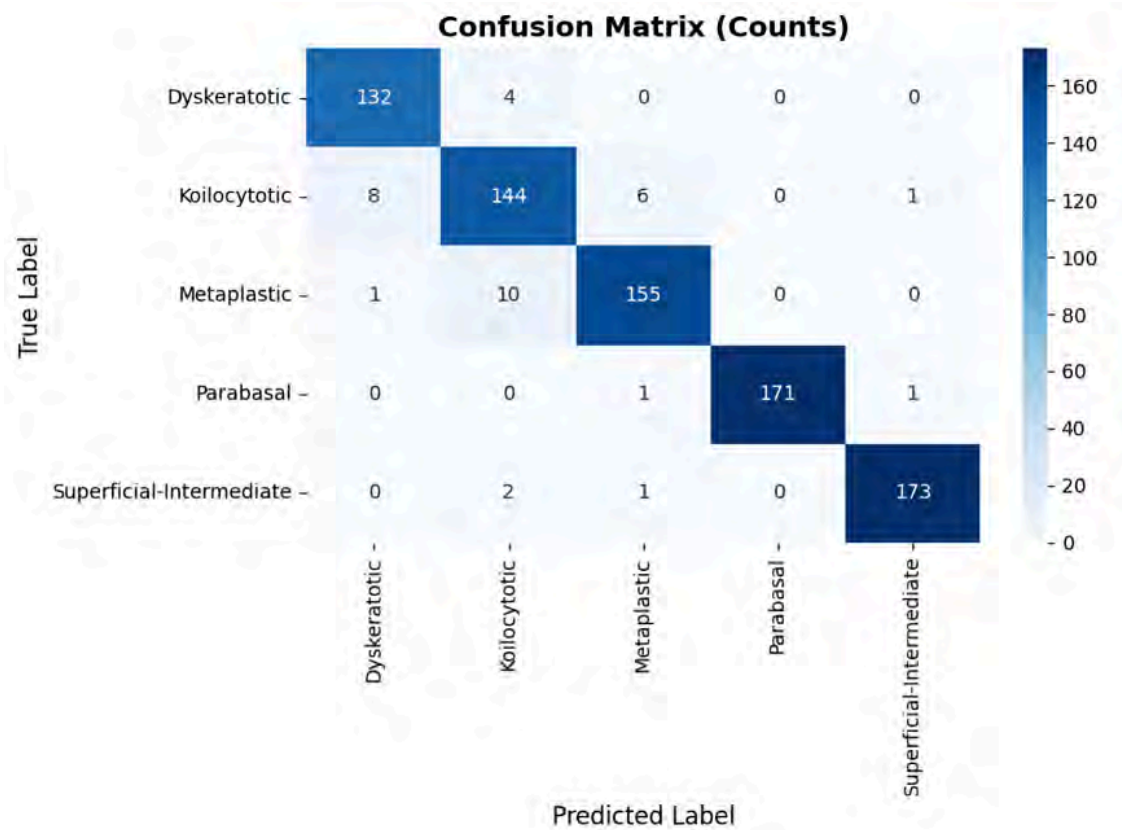


Figure 5.6: Confusion matrix of FedPapVisionNet under IID settings

Non-IID Data Partition

Across the non-IID ($\alpha = 0.3$) setting, the three FedProx-trained confusion matrices show clear differences in diagonal strength and error concentration. DualEffiMob-CBAM is the most diagonal dominant with 780/810 correct,followed by EffiResAttention at 774/810, while FedPapVisionNet (FedProx) drops to 695/810 .

In both heavier models, the last two classes remain extremely stable: Superficial–Intermediate is 176/176 correct for both, and Parabasal is also near-perfect (171/173 in both), indicating strong separability even under client heterogeneity. Error structure is the key differentiator. For EffiResAttention and DualEffiMob-CBAM, the dominant confusion is the clinically adjacent pair Koilocytotic & Metaplastic; Koilocytotic leaks into Metaplastic (4 and 3 cases) and Metaplastic leaks into Koilocytotic (13 and 12 cases), but remains controlled.

FedPapVisionNet (FedProx) shows a much stronger collapse towards this pair, especially Metaplastic to Koilocytotic (48 cases) and Metaplastic to Parabasal, which sharply lowers Metaplastic recall to 95/166. It also exhibits broader dispersion (e.g., Koilocytotic to Dyskeratotic 11 and to Superficial–Intermediate 7), suggesting noisier boundaries under non-IID drift.

Training the same FedPapVisionNet model with SCAFFOLD improves overall correctness to 732/810 and, most importantly, repairs the weakest class: Metaplastic recall rises to 143/166 by reducing Metaplastic to Koilocytotic from 48 to 13. This gain comes with a trade-off: Koilocytotic becomes less stable (recall 127/159) due to increased Koilocytotic to Dyskeratotic (18) and to Metaplastic (9) errors, indicating SCAFFOLD redistributes confusion rather than eliminating it. Overall, the matrices suggest that FedProx favors the larger-capacity backbones, while SCAFFOLD substantially strengthens FedPapVisionNet's class balance, especially for Metaplastic, under the same heterogeneous partition.

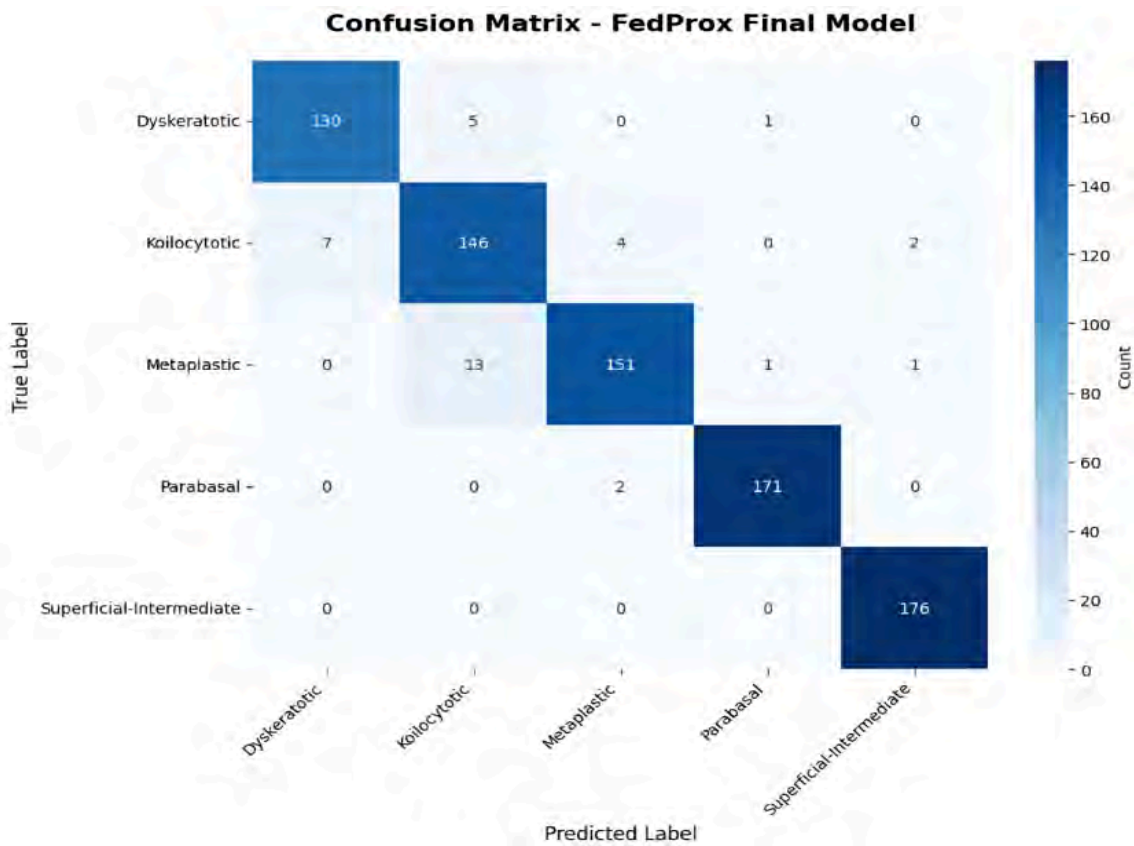


Figure 5.7: Confusion matrix of EffiResAttention non-IID(FedProx)

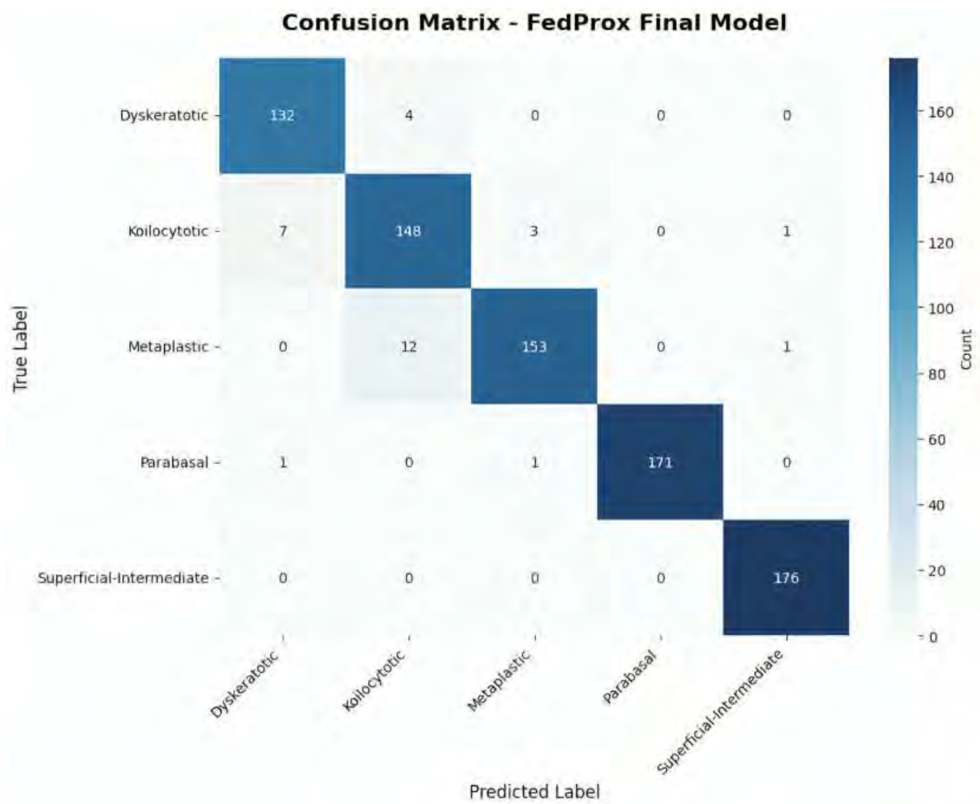


Figure 5.8: Confusion matrix of DualEffiMob-CBAM non-IID(FedProx)

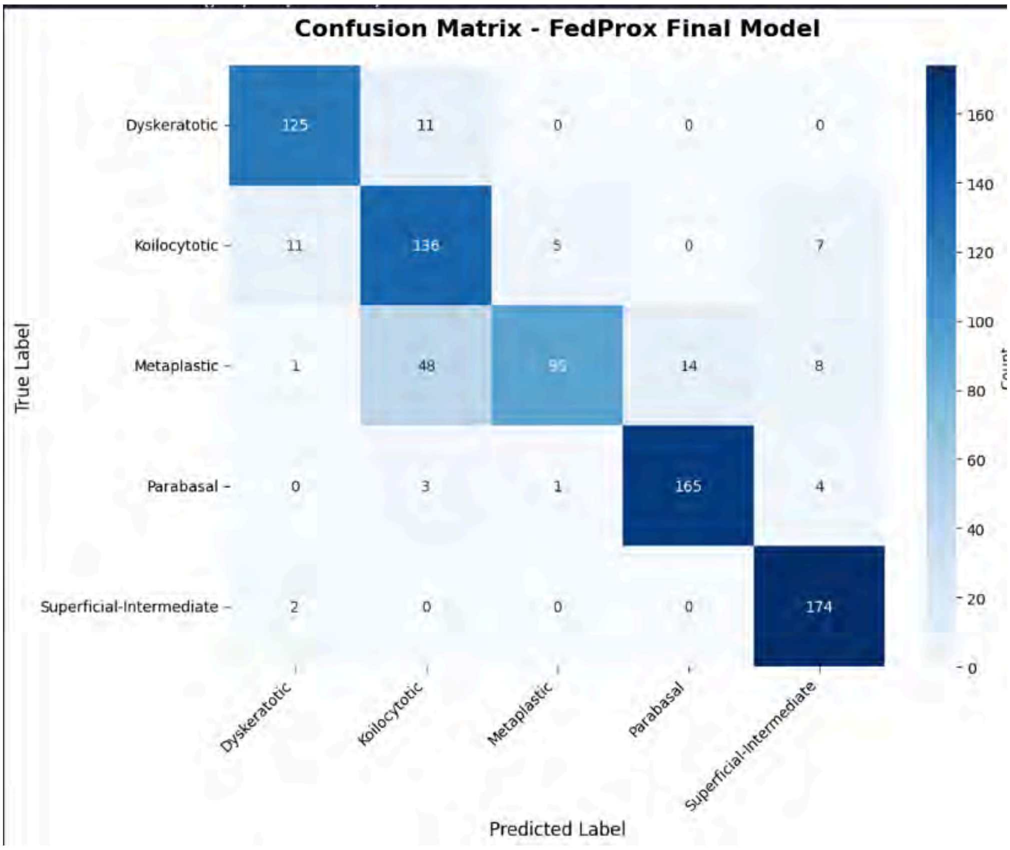


Figure 5.9: Confusion matrix of FedPapVisionNet under non-IID(FedProx)

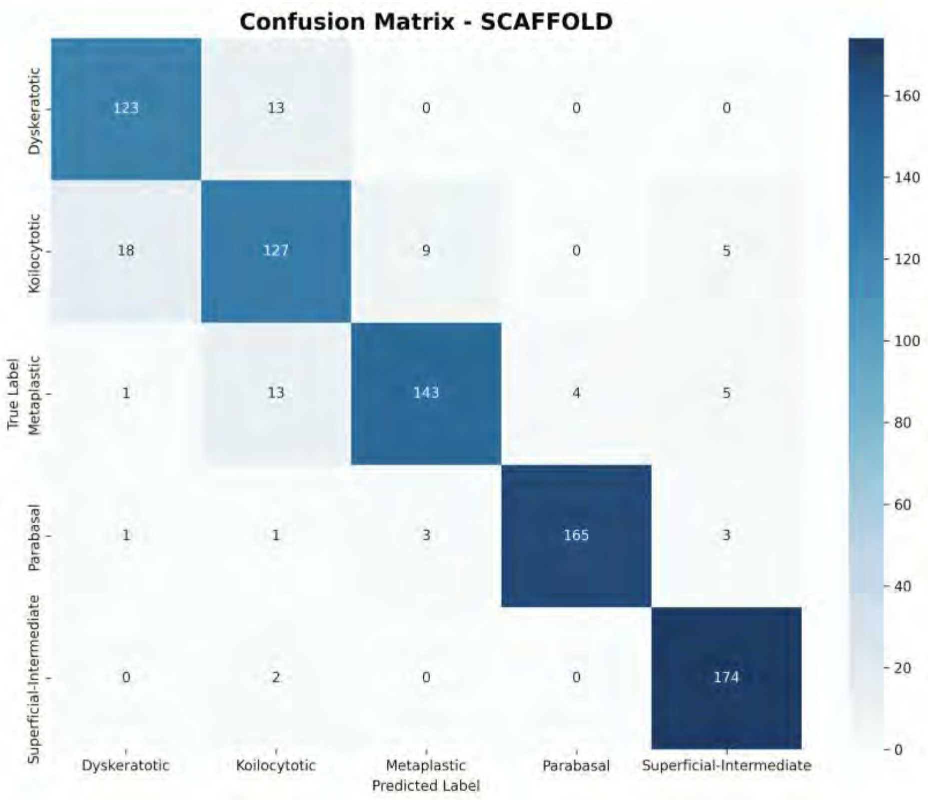
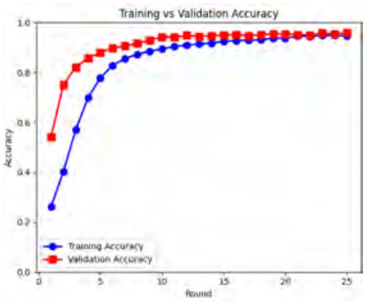


Figure 5.10: Confusion matrix of FedPapVisionNet non-IID(Scaffold)

5.3.2 Loss & Accuracy Curve

IID Data Partition



(a) EffiResAttention

(b) DualEffiMob-CBAM

(c) FedPapVisionNet

(d) EffiResAttention

(e) DualEffiMob-CBAM

(f) FedPapVisionNet

Figure 5.11: Training and validation accuracy (top row) and loss (bottom row) curves for different models under IID data partition using the FedAvg algorithm

Under the IID FedAvg setting, the three models exhibit broadly similar convergence shapes rapid early improvements followed by a slower plateau but differ in how quickly they stabilize and how tightly training and validation track each other. In the accuracy curves, all models reach the high-0.9 range by later rounds, indicating strong learnability under IID partitioning. DualEffiMob-CBAM shows a very fast rise in validation accuracy (reaching 0.9 within the early portion of training) and then a smooth plateau near 0.95, while training accuracy catches up more gradually, leaving a small but visible gap for much of the run. EffiResAttention follows a similar trajectory but with a clearer, more stepwise progression,converging with training accuracy near 0.94–0.95 and maintaining a relatively consistent separation between the curves.FedPapVisionNet reaches its plateau quickly as well and sustains validation slightly above training across most rounds, suggesting stable generalization with limited overfitting under IID data.

The loss curves reinforce these observations. DualEffiMob-CBAM begins with higher training loss and shows a steep early decline, with validation loss dropping quickly and stabilizing at a relatively low level; the parallel downward trends suggest steady optimization without late-stage instability. EffiResAttention displays a similarly sharp initial loss reduction for both training and validation, and the two curves remain close as they flatten, implying controlled generalization error as learning saturates. FedPapVisionNet shows the most abrupt early loss compression, with

both training and validation loss rapidly decreasing into a low regime and then flattening with minimal divergence, indicating strong convergence efficiency and a small generalization gap.
Statistically, the key differences are in convergence speed and gap behavior rather than final performance. Although, DualEffiMob-CBAM and EffiResAttention appear to reach a comparable final accuracy band, whereas FedPapVisionNet achieves similar stability with tighter loss coupling, which is advantageous for consistent deployment under IID conditions where FedAvg aggregation is well-aligned with data homogeneity.

Non-IID Data Partition

FedProx

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

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

Figure 5.12: Training and validation accuracy (top row) and loss (bottom row) curves for different models under Non-IID data partition using the FedProx algorithm

In the non-IID FedProx setting ($\alpha = 0.3$ and $\mu = 0.01$), all three models show clear differences in their convergence behavior for both accuracy and loss. DualEffiMob-CBAM achieves the best performance in training and quickest convergence as seen in the accuracy curves. It experiences a fast rise of training accuracy from the high-0.8s and a saturation at 0.99, with a validation accuracy saturating at approximately 0.95–0.96 yet exhibiting strong oscillations in round-to-round transfer. Differences in client distributions can lead to overfitting and stronger local specialisation even with FedProx regularisation, and this interactions may be the cause. The training accuracy of EffiResAttention reaches approximately 0.99, close to saturation; and the validation accuracy varies smoothly in a band around 0.95–0.96, which is slightly smoother than that of DualEffiMob-CBAM, suggesting a marginally better generalization.

EffiResAttention shows strong optimization efficiency training loss sharply decays to almost near-zero (0.01). Validation loss quickly drops from around 0.73 to 0.13-0.16 and then remains in a narrow band with small noise consistent with stable generalization. DualEffiMob-CBAM experiences a similar loss collapse at the outset. The plot has been cropped but it can be seen that it shows the same steep fall with the validation-loss and subsequent flattening as before seen with FedProx with well-converged models. On the other hand, FedPapVisionNet experiences significantly higher losses: the training loss drops from around 0.52 to 0.23 while validation loss reduces from around 1.33 to 0.39, which means a higher absolute generalization error. Moreover, it indicates slower and/or less complete fitting under non-IID conditions.

The FedPapVisionNet display is different as it starts from substantially lower validation accuracy (0.46) and climbs steeply into the curve before plateauing at 0.86, while training accuracy climbs more gradually to 0.89. As compared to the other two, its lower ceiling suggests that the lightweight model suffers from more statistical heterogeneity, while later-stage validation oscillations indicate remaining client drift effects that are only partially suppressed by FedProx.
According to the statistics, the heavier backbones DualEffiMob-CBAM & EffiResAttention converge faster, yield a higher validation plateau, and have lower loss floors. On the other hand, FedPapVisionNet sacrifices peak accuracy for efficiency and is more sensitive to heterogeneous client distributions.

Scaffold

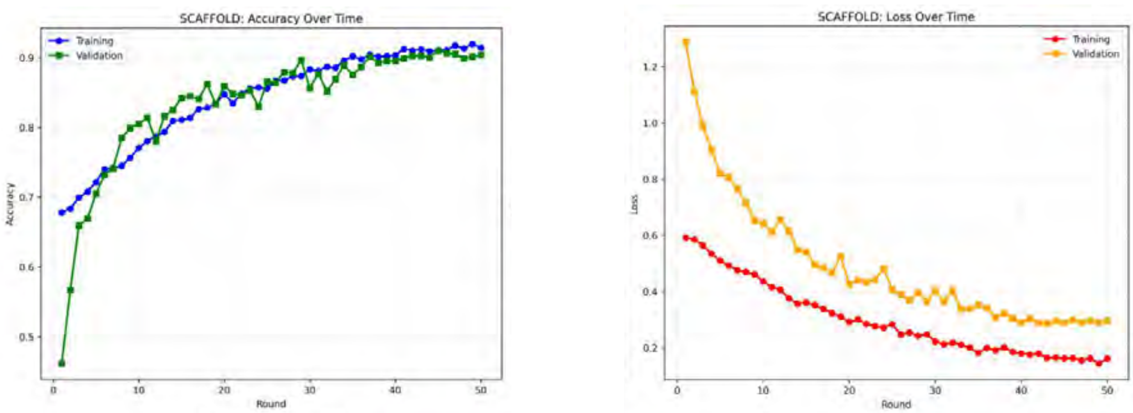


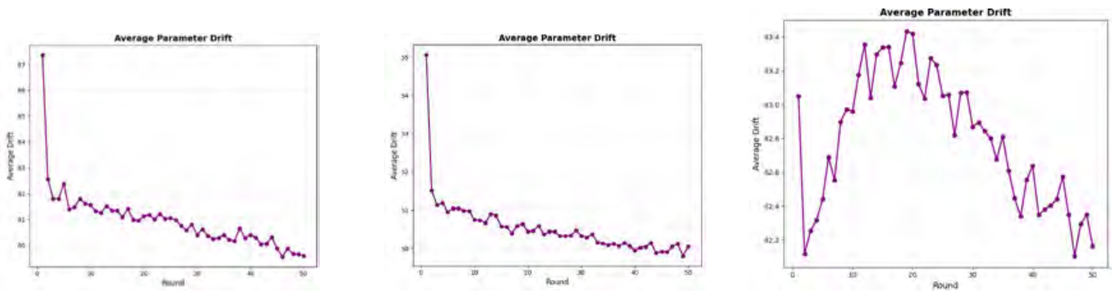
Figure 5.13: Training and validation accuracy & loss curves for FedPapVisionNet under Non-IID data partition using the SCAFFOLD algorithm

To mitigate the limitations of FedProx algorithm on FedPapVisionNet, SCAFFOLD algorithm was applied on the proposed FedPapVisionNet algorithm. SCAFFOLD demonstrates better generalization and convergence stability compared to FedProx-trained FedPapVisionNet. FedProx accuracy curves peak quickly in value but remain lower, while SCAFFOLD continues to climb to settle to 0.90 with a smaller training-validation gap.

The control variates used by SCAFFOLD prevent client drift and enable the global model to benefit more reliably from heterogeneous client updates.

The loss curves display a similar trend. Loss during training remains quite higher for FedProx around the end and validation loss settles around 0.39, thus the model does not fit completely and still retains some heterogeneity effects. The model on SCAFFOLD performs better than the baseline model on all datasets as validation loss drops faster at the start and becomes more stable at a lower value across rounds. SCAFFOLD provides FedPapVisionNet with a statistically significant uplift under non-IID training, by improving the validation plateau and loss floor above other protocols.

5.3.3 Parameter Drift



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

Figure 5.14: Parameter Drift Curve for different models under Non-IID data partition using the FedProx algorithm

Each of the three FedProx (non-IID) drift curves shows a transient behaviour of all models and then much slower stabilisation. However, the magnitude and smoothness of drift is very different.” The Paper presents a summary of results from experiments that employed the dualEffiMob model to understand how model drift can be reduced to near-zero with ease of parameter tuning. The observed behaviour indicates a strong alignment between the local updates of individual clients and the direction of the global model. This suggests that FedProx is effectively preventing the divergence of clients from the main model for this backbone.

EffiResAttention starts off at a higher drift level of around 87 but falls quickly into the 82–83 range after the first few rounds. This is followed by a slow decline to around 79.5–80 at the end. The curve is mostly decreasing, with small oscillations. This means that while FedProx dampens the heterogeneity in our model, there is still moderate client-to-client inconsistency in the later rounds. Compared to DualEffiMob-CBAM, the higher drift magnitude indicates higher sensitivity to non-IID variation even if the convergence direction is stable.

FedPapVisionNet displays a irregular drift dynamics. The initial dip to near 82.1 is followed by upward drifting across early-to-mid training, peaking at 83.4 at near rounds 18-20. The sender subsequently settles back down to values of near 82.2 by round 50. This non-monotonic behaviour shows that the drift of the client is lengthy, during training instance that learning is most critical. The local objectives pull the updates in different competing directions before eventually settling down to one direction. This statistically means that, under non-IID, the early alignment with each client is weaker. This can lead to slower improvement in validation metrics and increased class-wise confusion. In general, DualEffiMob-CBAM achieves the best performance on drift control, EffiResAttention performs moderate in alignment yet steadily improves; while FedPapVisionNet has the most unstable drift trajectory, thereby motivating SCAFFOLD and other drift-correction methods for the proposed model.

5.4 Final Design Adjustments

To push the proposed FedPapVisionNet to its limits under even harsher statistical heterogeneity, we decreased the Dirichlet parameter to $\alpha = 0.1$ and trained the model using SCAFFOLD for 40 communication rounds with 5 local epochs. This setting embodies an extreme non-IID scenario where the label distributions across clients are highly skewed, and there is an anticipation of severe client drift. Despite this, SCAFFOLD stabilizes training by correcting the client update bias using control variates, the results show. The training accuracy rises in a smooth and consistent manner towards a near saturation level, indicating strong local optimization even with harsh data imbalance. Validation accuracy settles down around 0.80, lower and noisier than in the $\alpha = 0.3$ case. This is reflective of the difficulty of the task, not optimization failure. The validation loss oscillates in a controlled manner rather than diverging shows that To push the proposed FedPapVisionNet to its limits under even harsher statistical heterogeneity, we decreased the Dirichlet parameter to $\alpha = 0.1$ and trained the model using SCAFFOLD for 40 communication rounds with 5 local epochs. This setting embodies an extreme non-IID scenario where the label distributions across clients are highly skewed, and there is an anticipation of severe client drift. Despite this, SCAFFOLD stabilizes training by correcting the client update bias using control variates, the results show. The training accuracy rises in a smooth and consistent manner towards a near saturation level, indicating strong local optimization even with harsh data imbalance. Validation accuracy settles down around 0.80, lower and noisier than in the $\alpha = 0.3$ case. This is reflective of the difficulty of the task, not optimization failure. The validation loss oscillates in a controlled manner rather than diverging shows that To push the proposed FedPapVisionNet to its limits under even harsher statistical heterogeneity, we decreased the Dirichlet parameter to $\alpha = 0.1$ and trained the model using SCAFFOLD for 40 communication rounds with 5 local epochs. This setting embodies an extreme non-IID scenario where the label distributions across clients are highly skewed, and there is an anticipation of severe client drift. Despite this, SCAFFOLD stabilizes training by correcting the client update bias using control variates, the results show. The training accuracy rises in a smooth and consistent manner towards a near saturation level, indicating strong local optimization even with harsh data imbalance. Validation accuracy settles down around 0.80, lower and noisier than in the $\alpha = 0.3$ case. This is reflective of the difficulty of the task, not optimization failure. The validation loss oscillates in a controlled manner rather than diverging shows that To push the proposed FedPapVisionNet to its limits under even harsher statistical heterogeneity, we decreased the Dirichlet parameter to $\alpha = 0.1$ and trained the model using SCAFFOLD for 40 communication rounds with 5 local epochs. This setting embodies an extreme non-IID scenario where the label distributions across clients are highly skewed, and there is an anticipation of severe client drift. Despite this, SCAFFOLD stabilizes training by correcting the client update bias using control variates, the results show. The training accuracy rises in a smooth and consistent manner towards a near saturation level, indicating strong local optimization even with harsh data imbalance. Validation accuracy settles down around 0.80, lower and noisier than in the $\alpha = 0.3$ case. This is reflective of the difficulty of the task, not optimization failure. The validation loss oscillates in a controlled manner rather than diverging shows that SCAFFOLD maintains the desired level of client drift. In general, the results confirm that the proposed model combined with SCAFFOLD achieves convergence and reasonable generalization even under extreme non-IID conditions, thus validating the robustness of the last design choice for privacy-preserving deployment in the real world.

5.5 Comparisons and Relationships

Joynab et al. (2024) [13] studied the same cervical-cell (Pap smear) classification problem and reported a best IID FedAvg test accuracy of 94.36% and a non-IID FedAvg accuracy of 78.43%, showing a clear performance drop once heterogeneity increases. In this experiment, FedPapVisionNet achieves 95.43% IID accuracy (773/810 correct), which exceeds the paper's IID ceiling while retaining strong diagonal dominance in the confusion matrix—meaning class separation is consistently high. Practically, our classification report behavior is also stronger in the clinically important stable categories. Parabasal and Superficial–Intermediate show very high precision or recall,and most remaining errors are concentrated in the clinically adjacent Koilocytotic & Metaplastic pair rather than being spread randomly across classes. This is a good sign that it implies the model is learning meaningful features and only struggles where genuine visual overlap exists.Finally, our training/validation curves plateau smoothly in the high 0.9 band with a tight generalization gap, indicating steadier convergence under IID than the variability typically seen when models overfit early.

Table 5.5: 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, $\alpha = 0.1$)

Simultaneously, in their non-IID experimental setting using FedAvg, the model achieves 78.43% test accuracy with a ROC-AUC score of 0.899. Their non-IID classification report also shows uneven class-wise behavior, for example, Koilocytotic drops to Precision = 0.69 / Recall = 0.70, while Parabasal remains strong at Precision = 0.91 / Recall = 0.93, meaning the non-IID split particularly hurts the clinically confusing categories. Against that baseline, our proposed FedPapVisionNet (3,90,215 parameters) is stronger under non-IID in multiple ways. First, even with $\alpha = 0.3$, our FedProx result establishes the non-IID sensitivity profile (695/810), and switching to SCAFFOLD raises performance further (732/810) while specifically repairing class balance;most notably boosting Metaplastic recall (143/166) by sharply reducing the earlier Metaplastic to Koilocytotic collapse. Second, to stress-test robustness under extreme heterogeneity, we reduced the Dirichlet parameter to = 0.1 and still achieved 80% accuracy using SCAFFOLD, surpassing Joynab et al.'s 78.43% despite the harsher non-IID partition.

5.6 Discussions

Prior research in automated cervical cancer detection has shown that deep learning models outperform traditional handcrafted methods under IID conditions, but such centralized approaches assume homogeneous data access that is unrealistic in clinical practice. Federated learning addresses privacy concerns; however, studies based on FedAvg consistently report significant performance degradation under non-IID data distributions. Nazia Shehnaz Joynab et al. highlight this limitation by demonstrating that non-IID federated cervical cell classification achieves only 78.43% accuracy, with pronounced class-wise imbalance in clinically confusable categories such as Koilocytotic and Metaplastic. These findings confirm that client drift and statistical heterogeneity remain critical challenges in federated medical imaging and motivate the need for drift-resilient, lightweight architectures.

The research will present FedPapVisionNet an efficient CNN architecture having only 390215 parameters and will systematically evaluate the drift-aware optimization strategies. Results demonstrate the positive impact of explicit drift correction methods on both convergence stability and class-wise reliability. Notably, FedPapVisionNet together with SCAFFOLD achieves 90% accuracy under non-IID (=0.3) and maintains 80% accuracy under more severe heterogeneity (= 0.1), outperforming FedAvg-based benchmarks. Notably, these gains are also reflected in overall accuracy, recall, and precision across classes that are clinically difficult to classify.

The significance of the proposed method is that it can achieve these three favorable outcomes, that is, privacy, robustness, and efficiency, at the same time from a real-world perspective. Due to its lightweight design, it is suitable for resource-restricted deployment in a healthcare setting with low communication and computation overheads. This research advances federated learning from a theoretical privacy solution to a practical and deployable framework through stable and balanced performance under realistic non-IID conditions. May improve cervical cancer screening use in a real clinical setting.

Chapter 6

Conclusion

6.1 Summary of Findings

This paper studies the complete privacy-preserving cervical cancer detection using federated learning and further focuses on the robustness of the model when tested on IID and non-IID data. FedPapVisionNet is a lightweight and expressive neural architecture for federated environments proposed in this thesis, based on the gaps in prior studies, especially the severe degradation of FedAvg performance under non-IID scenarios. With only 390,215 trainable parameters, the model is meant to improve diagnostic accuracy without sacrificing communication and deployability in resource-limited clinical environments.

Intensive experiments on the SIPaKMeD Pap-smear dataset ensure FedPapVisionNet attains accurate IID settings. Besides, it matches heavy hybrid baselines with much fewer parameters. Under non-IID conditions, naïve aggregation (FedAvg, FedProx) leads to class-specific collapse between Koilocytotic and Metaplastic cells, which is a challenging clinical pair. Incorporating SCAFFOLD, the proposed schema improves non-IID accuracy to 90%. Moreover, metaplastic recall (143/166) has been significantly restored under harsh data skew ($\alpha = 0.1$ and 0.3).

An analysis of the confusion matrix and convergence shows that the benefits do. They are not just aggregates; they are clinically relevant as well. Confusing categories instead of facilitating simple classes. In conclusion, the primary contribution.One goal of this work is to show that drift-aware optimization can enable more lightweight execution.Architectural design can deliver stability, security of private life and reliability in diagnosis.Both within federated learning systems. The results indicate that deployment is possible.Federated AI on cervical cancer screening across intrusive hospitals without data.Centralization which depicts the real-world constraints of contemporary health systems.

6.2 Contributions to the Field

The research has the potential to impact healthcare work in the future. Detection through Federated learning techniques. A new model FedPapVisionNet. Cervical cancer has been proposed using Pap smear images detection. This skeleton. It

addresses the important privacy challenge in the healthcare industry as patient data will not get centralized.Failure to understand AI solutions and their limitations poses a major barrier to the adoption of AI in healthcare.
FedPapVisionNet has been designed to work well on IID and non-IID data distributions, which are common in medical scenarios. The results show that classic methods such as FedAvg and FedProx suffer from class-specific collapse between Koilocytotic and Metaplastic cells, which are clinically similar classes. Both metrics indicate the significant recall and precision improvements with FedPapVisionNet and SCAFFOLD.
The effective performance of FedPapVisionNet over SCAFFOLD facilitates better accuracy and class balance specifically in the hard classes when the data is non-IID in nature. Consequently, we acquired a robust and reliable system. Further, this is beneficial in real-world clinical deployment where one requires privacy, accuracy, and generalization.
According to the findings, working with lightweight federated architectures,such as FedPapVisionNet, could offer great potential for large-scale privacy-protected and federated artificial intelligence applications in health care. The results make for an easier use of federated artificial intelligence in sensitive health domains without any loss of performance.

6.3 Recommendations for Future Work

In future work, the performance of federated learning systems in non-IID settings, particularly cervical cancer detection and other medical applications, can be enhanced with more heterogeneous datasets. The research was however able to curb the setback of drift through SCAFFOLD and achieve a 90% accuracy on non-IID, and future scope could be to lessen further the drift between local and global models. An approach could be to look at dynamic control variates as well as advanced drift correction techniques that take any extreme heterogeneous nature of data into account. In addition, further trials on datasets with greater class imbalance and skew could broaden the model's generalisability and bring it closer to real-world application. Through the architectural lightness design and drift-aware optimization, stability and accuracy have been assured. However, a lot needs to be done to strengthen the design, especially, in resource-limited environments like the health-care system.

Bibliography

[1] 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," *International journal of community based nursing and midwifery*, vol. 5, no. 2, p. 188, 2017.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, PMLR, 2017, pp. 1273–1282.

[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 *2018 25th IEEE international conference on image processing (ICIP)*, IEEE, 2018, pp. 3144–3148.

[4] N. Sompawong et al., "Automated pap smear cervical cancer screening using deep learning," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 7044–7048.

[5] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International conference on machine learning*, PMLR, 2020, pp. 5132–5143.

[6] C. Yuan et al., "The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images," *Scientific reports*, vol. 10, no. 1, p. 11 639, 2020.

[7] A. Khamparia, D. Gupta, J. J. Rodrigues, and V. H. C. de Albuquerque, "Dcavn: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network," *Multimedia Tools and Applications*, vol. 80, pp. 30 399–30 415, 2021.

[8] 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.

[9] M. Subramanian, V. Rajasekar, S. VE, K. Shanmugavadivel, and P. Nandhini, "Effectiveness of decentralized federated learning algorithms in healthcare: A case study on cancer classification," *Electronics*, vol. 11, no. 24, p. 4117, 2022.

[10] X. Yuan and P. Li, "On convergence of fedprox: Local dissimilarity invariant bounds, non-smoothness and beyond," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10 752–10 765, 2022.

[11] J. Peta and S. Koppu, "Enhancing breast cancer classification in histopathological images through federated learning framework," *IEEE Access*, vol. 11, pp. 61 866–61 880, 2023.

[12] 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.

[13] 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," *Informatics in Medicine Unlocked*, vol. 47, p. 101 496, 2024.

[14] 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.

[15] P. Khanarsa and S. Kitsiranuwat, "Deep learning-based ensemble approach for conventional pap smear image classification," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 18, no. 1, pp. 101–111, 2024.

[16] M. H. K. Mehedi, M. Khandaker, S. Ara, M. A. Alam, M. Mridha, and Z. Aung, "A lightweight deep learning method to identify different types of cervical cancer," *Scientific Reports*, vol. 14, no. 1, p. 29 446, 2024.

[17] M. U. Nasir, O. K. Khalil, K. Ateeq, B. S. Allah Almogadwy, M. Khan, and K. M. Adnan, "Cervical cancer prediction empowered with federated machine learning.," *Computers, Materials & Continua*, vol. 79, no. 1, 2024.

[18] 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," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 16, no. 1, pp. 16–38, 2024.

[19] 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.

[20] Z. Zhang, J. Gao, S. Li, and H. Wang, "Rmcnet: A liver cancer segmentation network based on 3d multi-scale convolution, attention, and residual path," *Bioengineering*, vol. 11, no. 11, 2024, ISSN: 2306-5354. DOI: 10.3390/bioengineering11111073 [Online]. Available: https://www.mdpi.com/2306-5354/11/11/1073

[21] 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.

[22] D. D. Himabindu, E. L. Lydia, M. Rajesh, M. A. Ahmed, and M. K. Ishak, "Leveraging swin transformer with ensemble of deep learning model for cervical cancer screening using colposcopy images," *Scientific Reports*, vol. 15, no. 1, p. 7900, 2025.

[23] I. Pacal, "Investigating deep learning approaches for cervical cancer diagnosis: A focus on modern image-based models.," *European Journal of Gynaecological Oncology*, vol. 46, no. 1, 2025.