A Federated Learning Aided System for Classifying Cervical Cancer using PAP-SMEAR Images

Nazia Shehnaz Joynab[a], Muhammad Nazrul Islam[a,*], Ramiza Rumaisa Aliya[a], A.S.M. Rakibul Hasan[a], Nafiz Imtiaz Khan[a], Iqbal H. Sarker[b]

[a]*Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka-1216, Bangladesh*
[b]*Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh*

Abstract

Cervical cancer is a significant contributor to female mortality on a global scale, especially in low-income countries where effective screening programs for the detection and treatment of precancerous conditions are lacking. Classification of pap-smear test cervical cell images is crucial as it gives essential information for the diagnosis of malignant or precancerous lesions and thus helps in providing a proper diagnosis. Most of the existing methods require accumulating pap-smear test images of all patients in a centralized location for classification purposes. However, this procedure may hamper the privacy of patient data and create data ownership issues. In this study, different convolutional neural network-based federated learning architectures are introduced to achieve both the objectives of accurate image classification and data privacy in three different experimental settings. In the proposed system, the updates of the locally trained models get aggregated with an initially untrained global model in order to increase its performance. In traditional ML-based systems, the more the train data, the more efficiently the model performs, but in the proposed system, clients can participate remotely to train a robust model even with the disadvantage of possessing a small dataset. The proposed CNN-based FL architecture showed test accuracy of 94.36% and 78.4% in an IID (Independent and Identically Distributed) and a non-IID setting respectively. Thus multiple hospitals across different countries can use the proposed system to train their local models with their private dataset without sharing it centrally, which eventually will help to build the central model of federated learning architecture with diverse datasets.

Keywords: Federated Learning; Healthcare Informatics; Computer Vision; Cervical Cancer

1. Introduction

Cervical cancer originates in cells of the cervix, a lower end of the uterus. This cancer typically remains asymptomatic in the early stages. Severe cervical cancer symptoms include vaginal bleeding after or between periods, watery and odorous discharge, and pelvic pain etc. The human papillomavirus (HPV), a sexually transmitted infection, is accountable for more than 95% cases of cervical cancer. In 2020, there was a global estimation of 604,127 new cases and 341,831 deaths due to cervical cancer [1]. With 569,847 new cases every year, cervical cancer has become the fourth most frequent disease afflicting women globally, after breast, colorectal, and lung cancers [2].

Regular clinical checkups and Pap tests are crucial for the timely detection of cervical cancer in women [3]. The Pap test is a widely used method to collect cells from the cervix and examine them for any anomalies that may suggest the existence of cancer. With early identification of these unusual cells, the Pap smear test can prevent the progression of cervical cancer. In the proposed system, a famous dataset named SIPaKMeD [4] is used where cervical cell images are classified into one of these five categories: Dyskeratotic cells, Koilocytotic cells, Metaplastic cells, Parabasal cells and Superficial-Intermediate cells.

If Dyskeratotic cells are present in a pap smear, it may indicate either infection or inflammation in the cervix, or it may be a sign of precancerous or cancerous cells. These cells are characterized by having vesicular nuclei that are similar to Koilocytotic cells. There are various causes of Dyskeratotic cells in a pap smear, such as Human papillomavirus (HPV) infection, Chlamydia infection, Herpes simplex virus (HSV) infection, Bacterial vaginosis, Yeast infection, use of hormonal birth control, and smoking.

Koilocytes are epithelial cells with a hyperchromatic nucleus that is replaced by perinuclear vacuole(s). Pathologists use these morphological changes to detect HPV-infected epithelial cells in Pap smears. Both low-risk and high-risk HPV infections show Koilocytosis in clinical biopsies. The appearance of Koilocytotic cells in the differentiated layers of the squamous epithelium is a well-known sign of human papillomavirus (HPV) infection [5].

Squamous metaplastic cells possess unique cellular boundaries and may exhibit eccentric nuclei and occasionally contain a significant intracellular vacuole. Depending on their size, they can resemble either small or large parabasal-type cells. There is a discernible contrast between squamous metaplastic cells and parabasal cells. Squamous metaplastic cells exhibit a more in-

*Corresponding author
Email address: nazrul@cse.mist.ac.bd (Muhammad Nazrul Islam)

tense stain in the cytoplasm and display a high level of consistency in terms of their size and shape compared to parabasal cells. These cells in Pap test results don't necessarily indicate malignancy; rather, they indicate that cervical squamous cells have changed but there are no alarming anomalies.

Parabasal cells are the smallest squamous epithelial cells seen on a typical vaginal smear. It is important to note that an elevation in parabasal cells on a vaginal smear doesn't always imply the existence of cancerous cells, yet it might indicate an abnormality. To find out the root of such an abnormality, additional examinations like a colposcopy or biopsy may be necessary. Women are encouraged to undergo regular Pap smear tests to monitor their cervical health and identify any potential problems promptly.

Superficial-intermediate cells make up the vast bulk of the cells detected during a Pap test. They typically have rounded, oval, or polygonal shaped cytoplasm stains. These cells contain a pycnotic nucleus in the center. These cells have recognizable nuclear boundaries (small, pycnotic nuclei in superficial cells, and vesicular nuclei in the intermediate cells), and big, polygonal cytoplasm that is clearly defined. These cells exhibit the morphological alterations caused by more severe lesions.

There have been multiple studies in order to classify cervical cancer using different Machine Learning and Deep Learning approaches [6]. Researchers used Softmax regression (SR), Support vector machine (SVM), and GentleBoost ensemble of decision trees (GEDT) as three different classifiers to identify cervical cancer [7]. The study of Ghoneim et al [8] showed the development of a cervical cancer categorization and detection method based on CNN. Three CNN models and an ELM-based classifier were studied. The shallow CNN model was trained and tested using the 5-fold cross-validation method on the Herlev dataset. They concluded by demonstrating how the ELM-based classifier produced a greater accuracy than all the other methods. In another research [9], the Herlev and SIPaKMeD datasets were integrated for the purpose of detecting cervical cancer. They demonstrated their ability to successfully analyze multi-layer cervical cells and developed a binary and multi-class classification pipeline to identify cancer in Pap smear images.

Deep learning models for the semantic segmentation of images require a huge amount of data. Obtaining enough data in the field of medical imaging is a major challenge. This challenge could be addressed through collaboration between different institutions. Sharing medical data in a centralized location creates various legal, privacy, technical, and data-ownership challenges. Thus for the classification of cervical cancer, a Federated Learning (FL) aided system is proposed in this study. Federated Learning allows individual hospitals to benefit from the rich datasets of multiple non-affiliated hospitals without centralizing the data in one place. Federated Learning utilizes diverse datasets of numerous collaborators to build a strong deep-learning model.

Despite the fact that there have been a number of studies in this area, further study is required to more accurately classify cervical cancer using cutting-edge ML and FL techniques. There hasn't been much research done to classify cervical cancer cells using different Federated Learning approaches. Moreover, the results of existing systems could be improved more with better-performing neural networks. Therefore, the objectives of this research are as follows:

i. To propose different deep learning architectures for the prediction and classification of cervical cancer in two IID settings and one non-IID setting.

ii. To enhance the efficiency and performance of the proposed deep learning architectures integrated with FL using hyper-parameter tuning.

iii. To compare the performance of the proposed FL-based deep learning architectures with the traditional ML models.

This paper is structured into several sections. Section 2 briefly introduces related studies. Section 3 offers an overview of the federated learning framework while Section 4 provides a description of the research methodology. In Section 5, the main development of the CNN-based federated learning architecture is presented along with the results obtained. Section 6 delves into the development of classical ML models. Section 7 provides their performance comparison with CNN-based FL architecture. Finally, the conclusion is presented in Section 8.

2. Literature Review

A number of studies have been carried out in the Medical Imaging domain to determine the application of Deep Learning and Federated Learning in the detection, segmentation, and classification of various types of cancer cells.

An overview of using deep learning and machine-learning algorithms was provided in a study [10], to aid the research of cervical cancer. They explored well-known datasets like Herlev and ISBI and discovered a new publicly accessible database called "SIPaKMeD" for classification purposes. The Herlev dataset, produced by the Herlev Medical University in Denmark, focuses on Pap smear benchmarking, while the ISBI challenge database consists of multi-layered cervical cell volumes. The use of deep learning and machine-learning approaches in differentiating and classifying cervical cytopathology images was highlighted here. Here, the potential for developing more advanced models was suggested to improve accuracy in cervical cytopathology image processing in the future.

The use of hybrid pipelines for the detection and classification of aberrant regions in liquid-based cytology (LBC) pictures using a combination of deep learning (DL) and traditional machine learning (ML) techniques was demonstrated in another study [11]. They made use of a personal database containing 1920 x 2560 pixel photos. They demonstrated how to inspect cervical samples using a RetinaNet model for the detection of aberrant regions. Taha et al. used a pre-trained CNN architecture along with a support vector machine for the detection of Cervical Cancer [12]. For the pre-trained architecture, the AlexNet was used to extract the desired features. They showed how it performs better in test recall, precision, specificity and accuracy scores than other techniques. Again, a focused study

included reviews of the various cervical cancer diagnostic techniques [13]. They have covered state-of-the-art methodologies expressed in major articles on computer-assisted diagnostic systems for cancer detection. The study emphasizes future options for the development of a cost-effective, automated disease classification system.

DeepCervix, a hybrid deep feature fusion (HDFF) technique was first used for classification of cervical cytopathology cells [14]. This study presented two distinct steps of data augmentation strategies in dataset pre-processing step. XceptionNet, ResNet50, VGG16, and VGG19 were the four enhanced CNN types introduced to extract complementing features. Prior to the Softmax (SM) layer in each of the DL models, they extracted the features in order to construct feature arrays with 1024 features. To carry out the classification, the feature arrays were then fed into a sequential model that was connected to a dense layer with batch normalization (BN) and a dropout layer in between. This study also used Late Fusion (LF) approach for cervical cancer classification. The performance metrics showed that the HDFF approach performed better than the LF method in terms of classification accuracy.

A novel approach (CVM-Cervix) was put forth in another study [15] that enhanced the overall classification of single-cell cervical cytopathological images by utilizing CNN, Visual Transformer for local and global feature extraction. Moreover, a Multilayer Perceptron module was used to receive the retrieved features for fusion and classification. Eleven different categories of images were obtained from SIPaKMeD and CRIC datasets. Additionally, the model was compressed using a simple post-processing technique.

In [16], an experimental framework was developed to compare the robustness of deep learning techniques in multiscale cell image categorization challenges. This study made use of raw data, gathered from 4049 cell images in the SIPaKMeD public dataset. Next, in order to produce scaled and standardised data, the raw data were pre-processed. Training samples were created by resizing the scaled and standard data to 224×224 pixels. Ultimately, 22 trained deep learning models were utilized to classify unseen test photos. The outcomes showed that the deep learning technique is remarkably resilient to variations in the cell aspect ratio in cervical cytopathological pictures. This was due to the fact that, despite variations in aspect ratios, the nucleus contained the majority of the visual information that set each cell apart. This offered a valid cause for categorizing the cell images.

In [17], a weakly supervised method was proposed for identifying images of cervical cancer nests using Conjugated Attention Mechanism and Visual Transformer (CAM-VT), which could efficiently and accurately examine pap slides. Conjugated attention mechanism and visual transformer modules were used for local and global feature extraction, respectively. An ensemble learning module was then designed to enhance the identification capacity. They performed comparison studies on their datasets to arrive at a plausible interpretation. The average accuracy of the validation set of three repeated trials using this framework was 88.92%, which outperformed 22 well-known deep learning models.

An extensive research on deep learning algorithms for image segmentation and classification in cervical cytopathology was reviewed in another study [18]. Additionally, key ideas in deep learning were described together with the prevalent architectures for them. The review revealed that there is growing interest in the field of deep learning for image processing related to cervical cytopathology. The study found that the majority of cutting-edge techniques for classification and segmentation have been used on the same dataset. So, it was easy to determine which algorithm was superior to the others.

For the first time, federated learning was presented in a study on the modality of cardiovascular magnetic resonance (CMR) [19]. They modified a 3D-CNN network pretrained on action detection and investigated two distinct methods of adding shape prior knowledge into the model, as well as four different data augmentation setups, systematically examining their impact on the various collaborative learning options. They demonstrated that, despite the limited amount of the data (180 people from four centers), privacy-preserving federated learning produced promising results that are competitive with standard centralized learning.

Federated learning enabled deep learning model was used on multimodal brain scans [20]. The quantitative results show that federated semantic segmentation models perform similarly to models trained by sharing data. The comparison was shown among federated learning and two other collaborative learning methods and a conclusion was drawn that the performance of these methods fall short than that of federated learning.

A Federated learning-based cancer diagnosis model was proposed where six first-level impact indicators were identified, as well as historical case data from cancer patients were collected[21]. Various physical examination indicators of patients were used as input in this framework. An auxiliary diagnostic model was built using patients' recurrence time and location, and comparison algorithms included linear regression, support vector regression, Bayesian regression, gradient ascending tree, and multilayer perceptrons neural network. CNN's federated prediction model outperformed single modeling machine learning tree model, linear model, and neural network in terms of accuracy.

In a study conducted by Pati et al. [22], the most comprehensive attempt was taken to develop a precise and generalizable machine learning model for detecting glioblastoma subcompartment boundaries in real-world settings. FL provided unprecedented access to the most common and fatal adult brain tumor dataset, as well as meaningful ML training to ensure model generalizability across out-of-sample data. As Federated Learning enabled large and diverse data, the final consensus model outperformed the public initial model against both the collaborators' local validation data and the entire out-of-sample data.

In another study, the feasibility of using differential-privacy techniques was investigated to protect patient data in a federated learning setup [23]. They developed and tested practical federated learning systems for brain tumor segmentation on the BraTS dataset. The experimental results revealed that there is a tradeoff between model performance and privacy protection

costs. How data dispersion affects FL performance was an outcome of the research conducted by Adnan et al. [24]. The authors explored the possibility of learning from distributed medical data via differentially private federated learning. They mainly showed how FL might be utilized in clinical contexts to guarantee data privacy while also ensuring minimal performance reduction. One study gave an outline of how federated artificial intelligence can be used in medical imaging applications while maintaining security and privacy [25]. They talked about how AI has changed the area of medicine, what is needed for the best privacy preservation, and the privacy and security concerns with medical imaging.

To maintain patient data privacy, Federated Learning algorithms were used for the diagnosis of three types of cancers, namely cervical, lung, and colon cancer through pictures from CT and MRI scans in the study [26]. Here, two FL algorithms, namely FedAvg and FedProx were used to assess the effectiveness of Federated Learning algorithms in diagnosing cancer cells. To ensure enhanced Performance, Bayesian optimization was used to tune the hyperparameters for both local and global models. The dataset includes CT/MRI images of the categories of cervical, lung, and colon cancer lesions, collected from Kaggle. They were able to demonstrate that federated learning models surpass conventional deep learning models in terms of prediction accuracy.

In addressing diagnostic challenges in gynecological malignancies, particularly lymph node metastasis (LNM), Zhijun et al. [27] pioneered an integral approach using federated learning and a multimodal evaluation framework. Traditional imaging techniques, such as CT, MRI, and PET/CT, may face limitations in LNM detection quite often. By combining text and image data to ensure enhanced prediction accuracy, they introduced a composite neural network model within a federated-learning environment. The federated-learning model, without the use of image data, achieved a sensitivity of approximately 92.31%, which increased to 94.12% with MRI integration. They showed how LNM detection may be transformed by emphasizing patient privacy and raising diagnostic standards across various medical centers.

Again, a study conducted by Moshawrab et. al. [28] provides a comprehensive overview of federated learning (FL) aggregation algorithms, discussing their functionalities, drawbacks, and future possibilities. They presented the current state of aggregation algorithms in FL and provided a way to categorize them. By presenting the importance of aggregation in FL, they addressed the privacy-preserving nature of federated machine learning that helped to provide insight for researchers aiming to improve FL aggregation methods and aided in the understanding and development of privacy-centric machine learning technologies.

To summarize, there are some discoveries which were made from the literature review. Firstly, except some recent works, very few such instances of incorporation of Federated Learning in the classification of cervical cancer cells are available. Secondly, most existing systems do not show concern for the development of a personalized Convolutional Neural Network model for the purpose of classification.

3. Federated Learning Framework

Traditional machine-learning approaches require a centralized location i.e. a single server to aggregate user data and carry out the complete training process, whereas federated learning allows continuous training on edge devices while ensuring no sensitive information exits the device. In the classical machine learning model, it is quite common to assume that the data are independent and identically distributed. On the contrary, federated learning assumes the data to be non-IID because in real-world circumstances, various users have different datatypes and the expected and actual number of actors varies.

Federated Learning is a collaborative machine learning technique where multiple edge devices(clients) participate remotely to train a common, robust machine learning model (server) without exchanging/sharing their local data samples with the centralized location or server which helps organizations to make better decisions with AI, addressing critical issues of data privacy, security, access rights and access to heterogeneous data.

There are 3 types of data partitioning systems in federated learning: horizontal data partitioning, vertical data partitioning and federated transfer learning system. In horizontal partition, each edge device has overlapping features with different observations. For instance, if multiple hospitals in various countries collect data on breast cancer patients but have little to no overlapping of patients, that distribution is horizontal partitioning [29]. In the vertical partition, each device has different features with overlapping observations. For example, if a hospital suggests patients to a specific surgeon, both the hospital and the surgeon collect different kinds of data but will have many common patients [29]. If there are a few similar samples with few similar features, but also samples and features do not overlap, then federated transfer learning can be applied in this situation [29].

Depending on real-life scenarios, limitations of available datasets and problem statements, different kinds of ML models are decided to be used as the foundation model of federated learning architecture. Neural networks, decision trees, and even linear models are used on basis of use cases. This base model serves as a global model which is initially distributed as an untrained or pre-trained model from a central server to the local clients. The clients then collaboratively train the global model by continuously improving it in every communication round between the central server and the local clients.

There are two types of FL communication architecture: centralized and decentralized. Both types of architecture work similarly; the distinction between them is in client-server communication. In a centralized federated learning system, a single central server is used, so there is only one possible point of failure [30]. In a centralized federated learning environment, a central server is utilized to manage all the participating nodes and orchestrate the various steps of the algorithms. The server is in charge of selecting the nodes at the start of the training process and aggregating the received model parameter updates. The server can end up being the system's bottleneck because each of the chosen nodes must communicate updates to the server

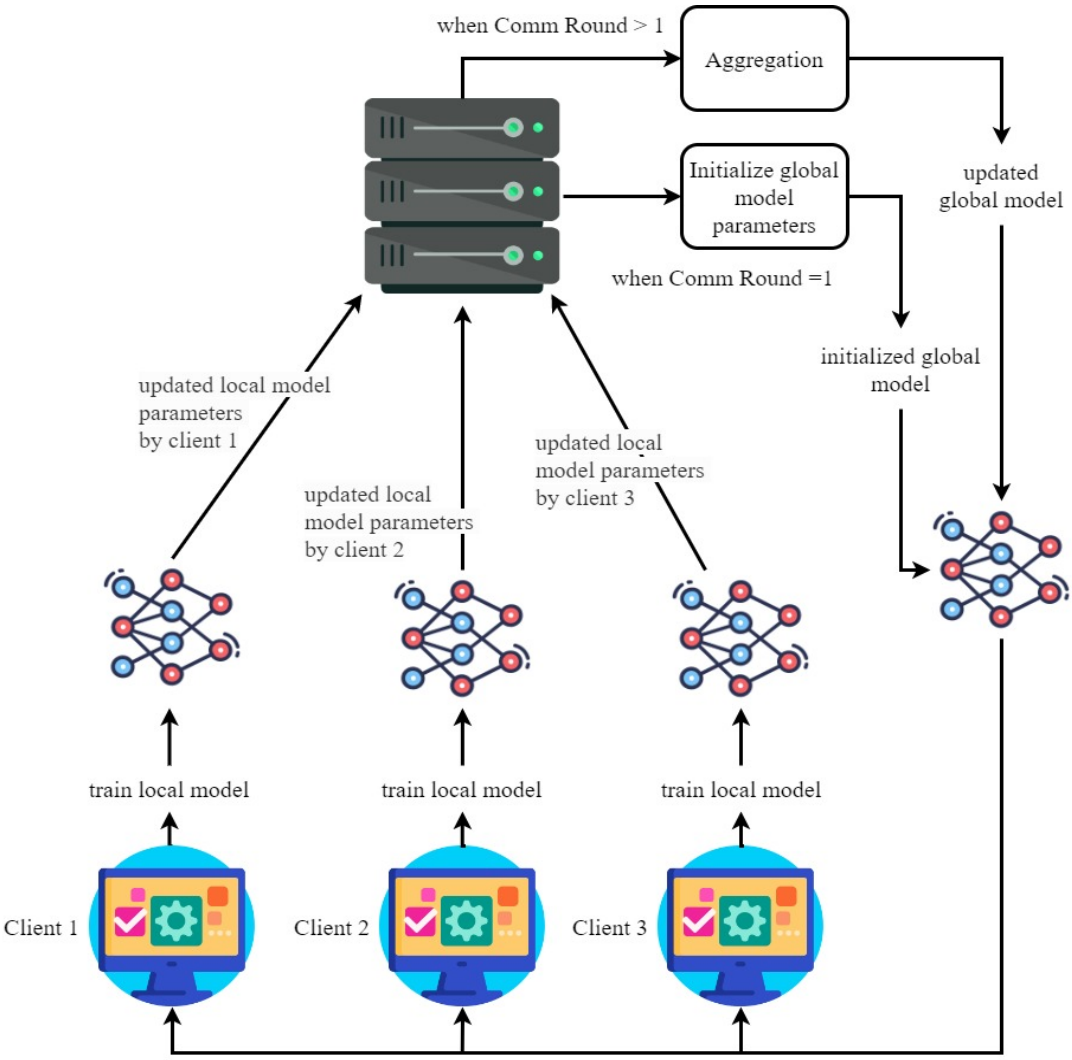

Figure 1: Architecture of Federated Learning

[29]. Decentralized federated learning does not rely on a single central server to provide updates, in contrast to centralized federated learning [30]. In this architecture, the nodes can collaborate among themselves to produce the global model. As the model updates are solely transferred between linked nodes without the coordination of a central server, this configuration avoids single-point failures.

Figure 1 shows the overview of a federated learning framework. In a single federated round, the following steps are carried out: at first, the client devices obtain the initialized global model from the cloud server. Next, they train the model using the local datasets and generate the most recent local model update (model parameters). Then, they send the updated weights of their locally trained model to a central server. After that, the cloud server collects various local update parameters and aggregates the model weights received from all of the client devices by averaging them together. The averaged weights are then sent back to the client devices, which use them to train their local models further. Finally, this process is repeated until the models on the client devices reach satisfactory accuracy.

Stochastic Gradient Descent (SGD) can be applied to as a base to the Federated Learning training algorithm. A single batch gradient calculation is done on each round of communication. Although this approach is computationally efficient, it requires a large number of training rounds in order to produce good models. One approach of FedSGD is that, the client computes the gradient, updates the model and sends it to the server. Now, if the model is updated multiple times before being sent to the server for aggregation, then the method is called FederatedAVG (or FedAVG). Here, the computation is kept in these parameters: w_t (model weights on communication round t), w_t^k (model weights on communication round t on client k), C (Fraction of clients participating in that round), E (No. of training passes each client makes over its local dataset each round), B (Local minibatch size used for client updates), K (number of clients indexed by k), Φ_k (set of data points on client k), n_k (number of datapoints on client k), η (learning rate). The pseudo code for FedAVG algorithm is given below:

Algorithm 1 Pseudo code of FedAVG

Server executes:

initialize w_0
for each round $t = 0, 1, \ldots$ do
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ random set of m clients
  for each client $k \in S_t$ in parallel do
    $w_{t+1}^k \leftarrow$ ClientUpdate (k, w_t)
  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate (k, w) : // Run on client k
  $\mathcal{B} \leftarrow$ (split Φ_k into batches of size B)
  for each local epoch i from 1 to E do
    for batch $b \in \mathcal{B}$ do
      $w \leftarrow w - \eta \nabla \ell(w; b)$
  return w to server

4. Methodology

The methodology adopted in this study in presented in Figure 3. Defining the research objectives through reviewing related literature is the first step of the applied research methodology. Then the dataset of pap smear test cervical cell images SIPaKMeD Dataset [4] is acquired. This dataset is the largest cervical cancer dataset available containing a total of 4049 images of isolated cells (each of size 66X66), taken manually from the cluster cell images of Pap smear slides, and used for the development of classic ML models and the base model of FL architecture. The cell images are divided into five categories including 813 images of Dyskeratotic cells, 825 images of Koilocytotic cells, 793 images of Metaplastic cells, 787 images of Parabasal cells, and the remaining 831 images of Superficial-Intermediate cells. Figure 2 depicts various cell images of this dataset.

Two approaches are followed for the classification of those images. In the first phase, various CNN models are developed and integrated with FL architecture in three different experimental settings. Hyperparameter tuning is carried out to figure out the best possible architecture for a specific setting. In the second phase, different traditional Machine Learning models are developed to determine which would provide an overall increased accuracy. Subsequently, the performance of these models are compared with that of CNN based FL architecture.

5. Development of FL Architecture

5.1. Data Partitioning

The quality and distribution of data have a considerable impact on the methods that are required to be employed for cancer cell classification. To correctly evaluate the performance of a model in machine learning (ML), the data must be dispersed independently and identically (IID). For data to be IID, all random variables must have the same probability distribution and be mutually independent. The real-world data distributions are hardly IID. Deep learning algorithms often fail to generalize well when applied to the medical data of a different hospital. However, FL helps to overcome a certain institution's biases by aggregating models from different institutions. To test the validity of Federated Learning for the classification of Cervical Cancer, three different experimental settings were created. Among them, two of the settings have IID distribution of data and the remaining setting has non-IID distribution of data. Data distribution for three different settings is shown below in Table 1, 2 and 3.

Table 1: Division of images in Experimental Setup 1 (IID)

Class Name	Client_1	Client_2	Client_3	Test
Dyskeratotic	230	230	230	123
Koilocytotic	230	230	228	137
Metaplastic	215	215	215	148
Parabasal	215	215	215	142
Superficial-Intermediate	230	235	231	135

Table 2: Division of images in Experimental Setup 2 (IID)

Class Name	Client_1	Client_2	Client_3	Valid	Test
Dyskeratotic	195	218	237	81	82
Koilocytotic	228	220	212	82	83
Metaplastic	219	199	216	79	80
Parabasal	200	217	212	79	79
Superficial-Intermediate	237	225	202	83	84

Table 3: Division of images in Experimental Setup 3 (non-IID)

Class Name	Client_1	Client_2	Client_3	Valid	Test
Dyskeratotic	650	0	0	81	82
Koilocytotic	0	660	0	82	83
Metaplastic	0	634	0	79	80
Parabasal	0	0	629	79	79
Superficial-Intermediate	0	0	664	83	84

5.2. FL on Experimental Setting 1

5.2.1. Data Preprocessing

The Keras ImageDataGenerator class was used for a few geometrical transformations, including zooming, shearing, rescaling and horizontal flips. Keras model for cervical cell classification was developed using Convolutional Neural Networks. The input_shape argument for the 2 CNN Model was set to (66, 66, 3). This specifies the shape of the input images that the model has been trained on. The input images were expected to have a width of 66 pixels, a height of 66 pixels, and 3 color channels (corresponding to the Red, Green, and Blue color channels). The final output shape of the Keras model was five as in the number of cell classes. These five values were output by the last layer of the model, which has a sigmoid activation function.

5.2.2. Proposed CNN Architectures

A convolutional neural network(CNN) is made up of an input layer, an output layer, and numerous hidden layers in between. Convolution, activation or ReLU, and pooling are three

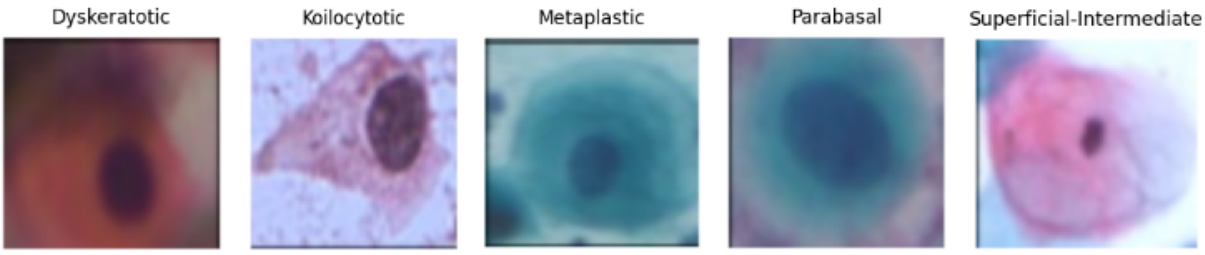

Figure 2: Visualization of the SIPAKMED dataset

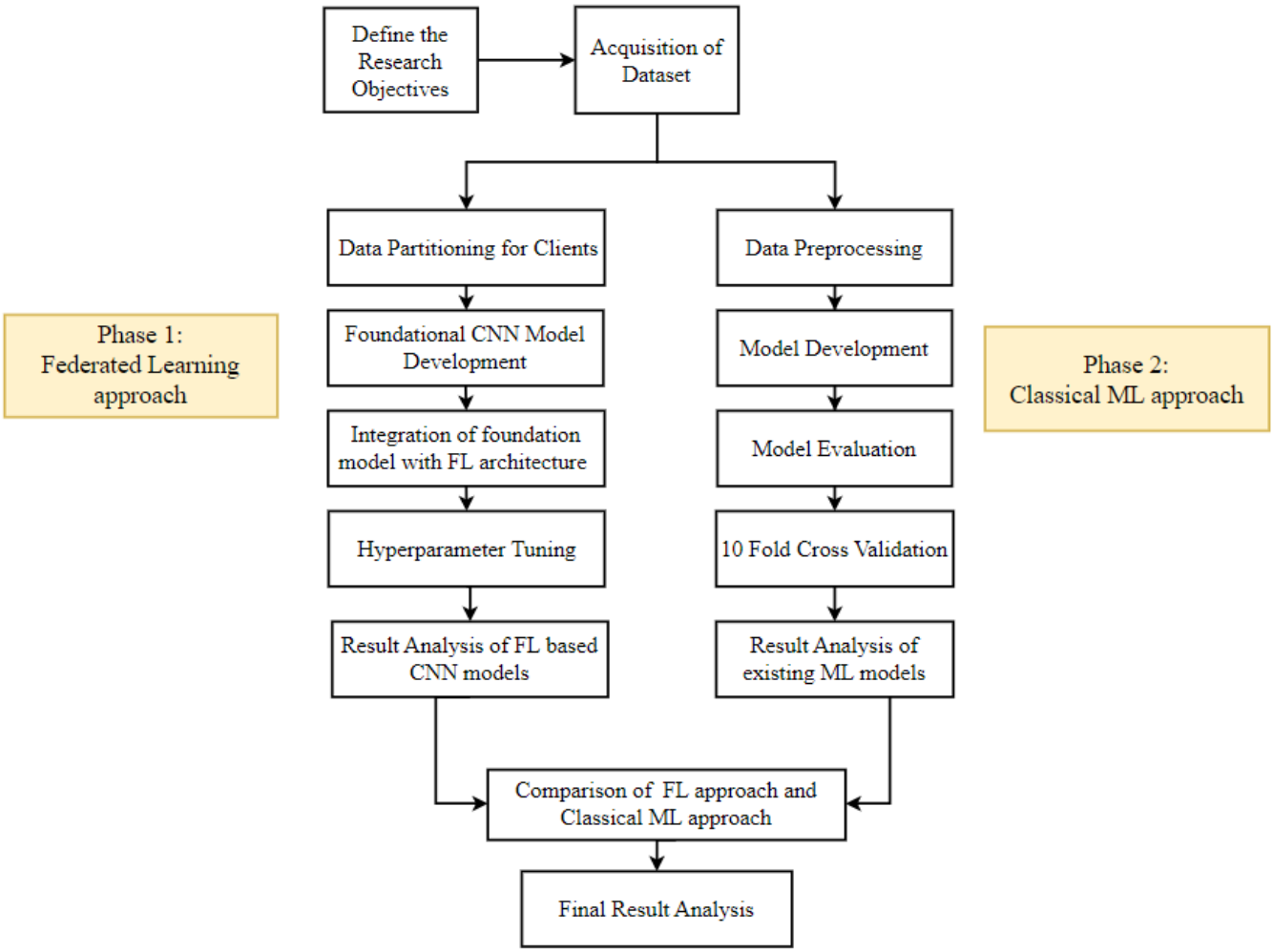

Figure 3: Methodological Outline

of the most used layers of a CNN. In order to extract features from the input image, a convolution layer is used. The image is convolved using a kernel (or filter) in this process. A kernel is a little matrix having a height and width smaller than the picture to be convolved. It is often referred to as a convolution matrix or convolution mask. For convolution, the number of channels in the kernel must match the number of channels in the input image. After the convolution layer, max pooling layer is used to decrease the spatial dimensions of the feature maps created by the convolution layer while preserving the most crucial features. The convolution neural network can operate more quickly and effectively by controlling its computational complexity with the aid of the decreased dimentions of these feature maps. After the convolutional or fully connected layer, a batch normalization layer should be included. This layer normalizes the input values of the layer to which it is attached. The “batch”

in batch normalization refers to the fact that the normalization is carried out on a batch of input data. The training of CNNs can be sped up by batch normalization. By introducing noise to the input data, it can also assist in lowering overfitting. Additionally, it can enhance CNNs' generalization abilities. As a result, the network is better equipped to identify unfamiliar images. The max-pooling layer sends the feature map to the flatten layer, which transforms it from a multi-dimensional array into a one-dimensional array that the dense layers can understand. To avoid model overfitting, dropouts are utilized after the dense layers of the network. In this research, several CNN architectures were developed to integrate and test with Federated Learning Architecture. After hyper-parameter tuning, two architectures stood out in this particular setting, as their test accuracy went above 85%. CNN architecture 1 and 2 is shown in Table 4 and 5 respectively. Hyper-parameter set 1 represented in Table 6 was used for CNN model 1 and Hyper-parameter set 2 and 3 represented in Table 7 and Table 8 was used for CNN model 2.

Table 4: CNN Architecture 1

Layer Type	**Output Shape**
Conv2D	(None, 64, 64, 16)
Conv2D	(None, 62, 62, 32)
MaxPooling2D	(None, 31, 31, 32)
Conv2D	(None, 29, 29, 64)
MaxPooling2D	(None, 14, 14, 64)
Conv2D	(None, 12, 12, 128)
MaxPooling2D	(None, 6, 6, 128)
Dropout	(None, 6, 6, 128)
Flatten	(None, 4608)
Dense	(None, 64)
Dropout	(None, 64)
Dense	(None, 5)

Table 5: CNN Architecture 2

Layer Type	**Output Shape**
Conv2D	(None, 64, 64, 16)
Conv2D	(None, 62, 62, 32)
MaxPooling2D	(None, 31, 31, 32)
BatchNormalization	(None, 31, 31, 32)
Conv2D	(None, 29, 29, 64)
MaxPooling2D	(None, 14, 14, 64)
Conv2D	(None, 12, 12, 128)
MaxPooling2D	(None, 6, 6, 128)
Dropout	(None, 6, 6, 128)
Flatten	(None, 4608)
Dense	(None, 256)
Dense	(None, 128)
Dropout	(None, 128)
Dense	(None, 5)

Table 6: Hyperparameter set 1 used for CNN Model 1

Parameters Used	**Values**
Number of Conv2D layers	4
Number of MaxPooling2D layers	3
Number of Dense layers	2
Number of Dropouts	2
Activation functions used	ReLU for Conv2D layers and Dense layers, and sigmoid for the output layer
Loss function	categorical_crossentropy
Optimizer	Adam
Input image size	66 x 66 x 3 (RGB image with 66 height, 66 width, and 3 channels)
Batch Size	100
Total Communication Round	10
Steps Per Epoch for Local Train	3
Number of Epochs for Local Train	150
Learning Rate	0.013

Table 7: Hyperparameter set 2 used for CNN Model 2

Parameters Used	**Values**
Number of Conv2D layers	4
Number of MaxPooling2D layers	3
Number of Dense layers	3
Number of Dropouts	2
No of BatchNormalization Layer	1
Activation functions used	ReLU for Conv2D layers and Dense layers, and sigmoid for the output layer
Loss function	categorical_crossentropy
Optimizer	Adam
Input image size	66 x 66 x 3 (RGB image with 66 height, 66 width, and 3 channels)
Batch Size	10
Total Communication Round	20
Steps Per Epochs for Local Train	60
Number of Epochs for Local Train	30
Learning Rate	0.001

Table 8: Hyperparameter set 3 used for CNN Model 2

Parameters Used	**Values**
Number of Conv2D layers	4
Number of MaxPooling2D layers	3
Number of Dense layers	3
Number of Dropouts	2
No of BatchNormalization Layer	1
Activation functions used	ReLU for Conv2D layers and Dense layers, and sigmoid for the output layer
Loss function	categorical_crossentropy
Optimizer	Adam
Input image size	66 x 66 x 3 (RGB image with 66 height, 66 width, and 3 channels)
Batch Size	10
Total Communication Round	30
Steps Per Epoch for Local Train	100
Number of Epochs for Local Train	40
Learning Rate	0.001

Table 9: Evaluation Metrics on Test Data

Model	**Hyper-Parameter Set**	**Accuracy**	**Precision**	**Recall**	**F1-score**	**ROC-AUC Score**
CNN 02	3	88.46%	89.58%	88.47%	87.83%	93%
CNN 02	2	87.00%	86.38%	84.47%	86.83%	92%
CNN 01	1	82.18%	82.59%	83.47%	82.62%	89%

From Table 9, it is evident that the test results obtained from CNN architecture 02 and hyper-parameter 3 is the best among all combinations of CNN architecture and hyper-parameter sets

in this setting.

5.2.3. Result Analysis of CNN-based FL

An analysis of a machine learning model's performance on a set of test data is summarized by a confusion matrix. It is frequently used to assess how well classifier models work. These models try to predict a categorical label for each input event. The matrix shows how many true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) the model generated using the test data. If there are n classes, the matrix will have a shape of nxn.

The confusion matrix along with One-Vs-Rest Multiclass ROC curve on test data for the three configurations: CNN architecture 01 + Hyper-parameter Set 01, CNN architecture 02 + Hyper-parameter Set 02, and CNN architecture 02 + Hyper-parameter Set 03 are presented in Figure 4, Figure 5, and Figure 6 respectively. The last configuration performs best among all, as displayed by the One-Vs-Rest Multiclass ROC curve. Computing a ROC curve for each of the n classes constitutes the One-vs-the-Rest (OvR) multiclass technique, sometimes called one-vs-all. A particular class is considered the positive class in each phase, and the remaining classes are collectively considered the negative class. It is evident from the plots that the AUC for each class in the Configuration 3 ROC curve is higher than the Configuration 1 and 2 ROC curves.

5.3. FL on Experimental Setting 2

5.3.1. Data Preprocessing

The torchvision.transforms module provides standard picture transformations. They can be chained together with Compose. Randomized transformations apply the same transformation to all photos in a particular batch, but yield different transformations across calls. In this setting, in built modules of Pytorch were used for image augmentation: RandomRotation, RandomResizedCrop, RandomHorizontalFlip, ToTensor and Normalize. *RandomRotation* rotates an image by an angle. *RandomResizedCrop* selects a random portion of an image and resizes it to a square with a side length of 28 pixels. This crop taken from the original image has a random area (H * W) and a random aspect ratio. The crop is then scaled to a specific size. The *RandomHorizontalFlip* function helps to flip a picture horizontally at random with a specified probability. The *ToTensor* function converts a PIL Image or numpy.ndarray (H x W x C) in the range [0, 255] to a torch.FloatTensor of shape (C x H x W) in the range [0.0, 1.0]. The *Normalize* function is used to normalize a tensor image by calculating its mean and standard deviation. It also helps to obtain data within a certain range and eliminates skewness, allowing the model to learn more quickly and effectively.

5.3.2. Proposed CNN Architecture

In this setting similar to IID distribution, after thoroughly evaluating several CNN architectures through hyper-parameter tuning, the final test accuracy received was 94.36%. The CNN architecture which is used for both experimental setting 2 and 3, is shown in Table 10. Hyper-parameters used for this architecture is shown in Table 11.

Table 10: CNN Architecture 3

Layer Type	**Output Shape**
Conv2D	[-1, 32, 28, 28]
ReLU	[-1, 32, 28, 28]
MaxPooling2D	[-1, 32, 14, 14]
Conv2D	[-1, 64, 14, 14]
ReLU	[-1, 64, 14, 14]
MaxPooling2D	[-1, 64, 7, 7]
Linear	[-1, 128]
ReLU	[-1, 128]
Linear	[-1, 5]

Table 11: Hyperparameters Set 4 used for CNN Model 3 in Experimental Setting 2

Parameters Used	**Values**
Number of Conv2D layers	2
Number of MaxPooling2D layers	2
Activation functions used	ReLU(Conv2D layers)
Loss function	Categorical Cross-entropy
Optimizer	SGD
Input image size	28 x 28
Batch Size	5
Total Communication Round	20
Number of Epochs for Local Train	10
Learning Rate	0.01

5.3.3. Result Analysis of CNN-based FL

In Table 12, the classification report of FedAvg on SIPAKMED dataset is presented for this setting. In an IID distribution of data, 94.36% test accuracy was obtained with a roc-auc score of 98.2%. Curve for training-Vs-validation accuracy was shown in Figure 7. While confusion matrix on test data is presented in Figure 8. And also the One-Vs-Rest Multiclass ROC curve is presented in the Figure 9.

Table 12: Classification Metrics for Experimental Setting 2

Class	Metrics			ROC AUC Score
	Precision	Recall	F1-Score	
Dyskeratotic	0.95	0.94	0.94	0.9817
Koilocytotic	0.88	0.90	0.89	0.9080
Metaplastic	0.89	0.90	0.89	0.9395
Parabasal	0.96	0.99	0.97	0.9858
Superficial-Intermediate	0.97	0.93	0.95	0.9713
Test accuracy				0.943
ROC_AUC Score				0.982

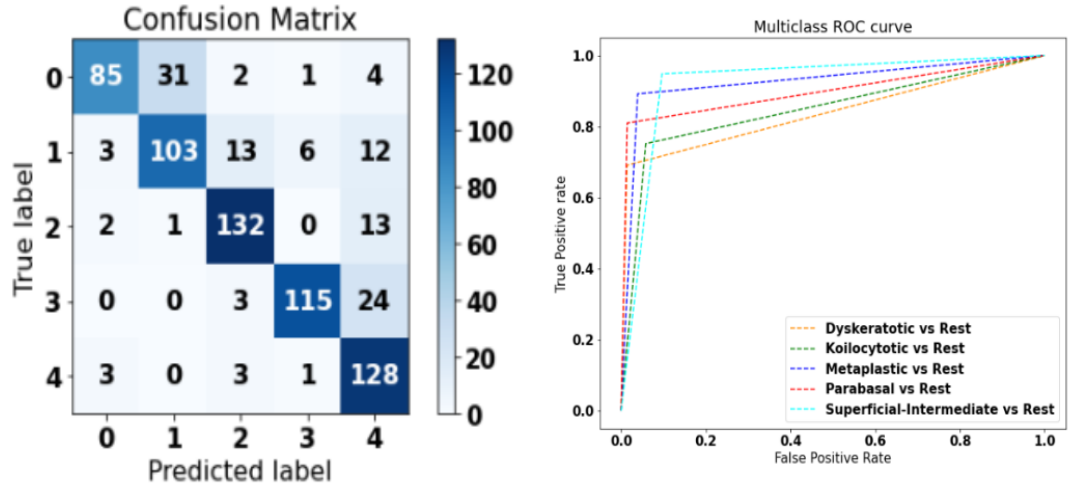

Figure 4: Confusion Matrix and ROC_AUC_Curve with CNN 1 + Hyper-parameter Set 1

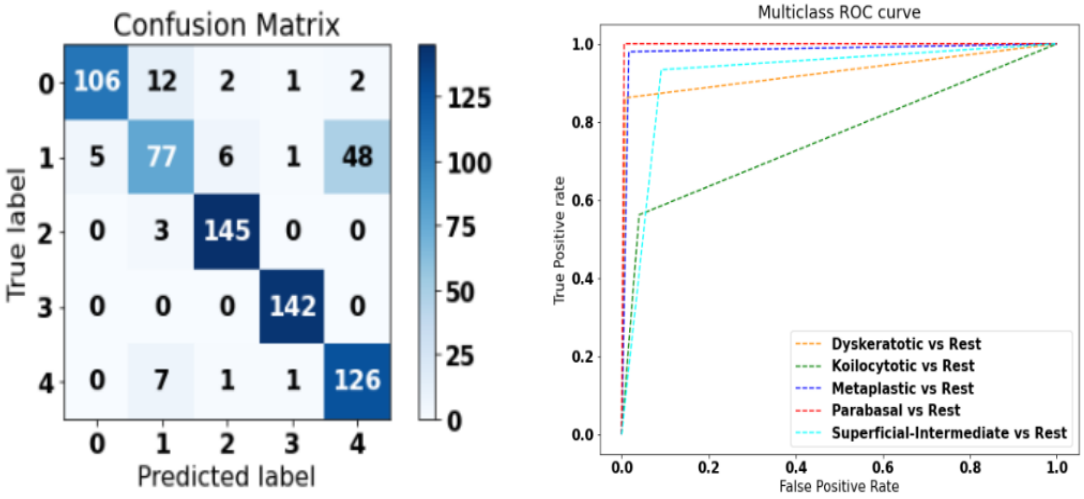

Figure 5: Confusion Matrix and ROC_AUC_Curve with CNN 2 + Hyper-parameter Set 2

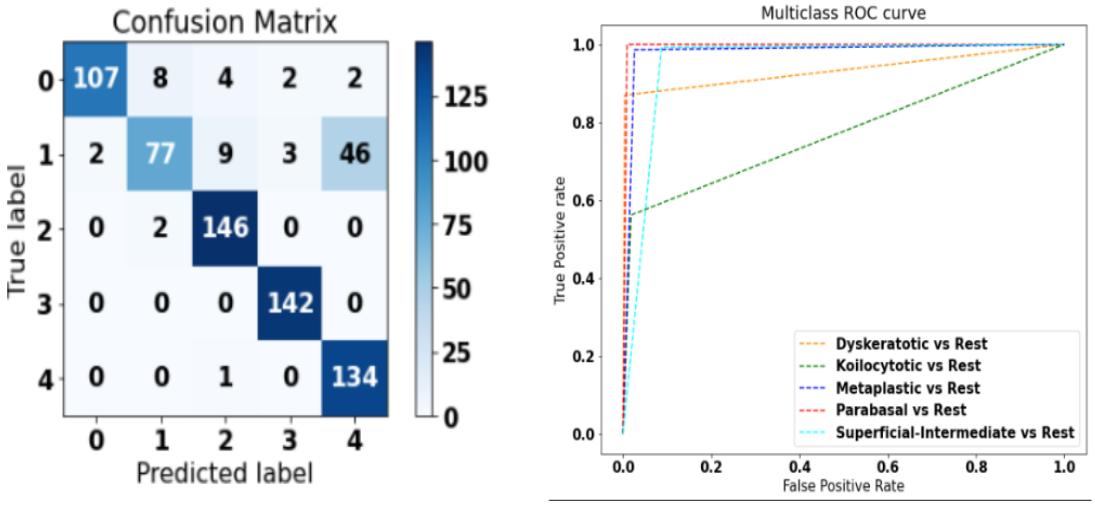

Figure 6: Confusion Matrix and ROC_AUC_Curve with CNN 2 + Hyper-parameter Set 3

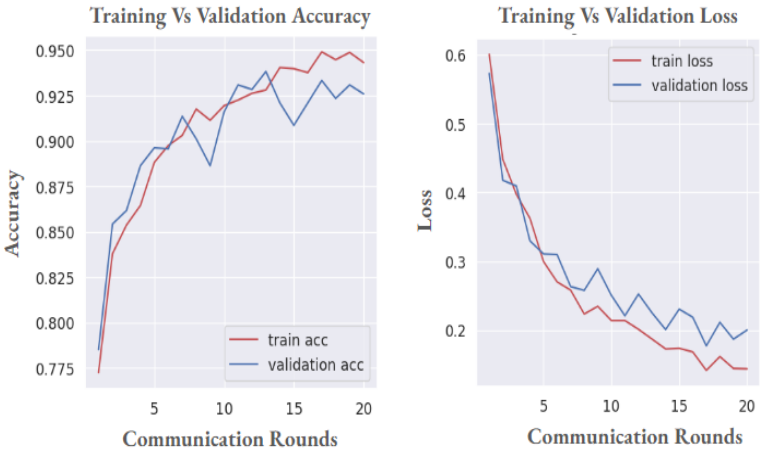

Figure 7: Train Vs. Validation Accuracy and Loss

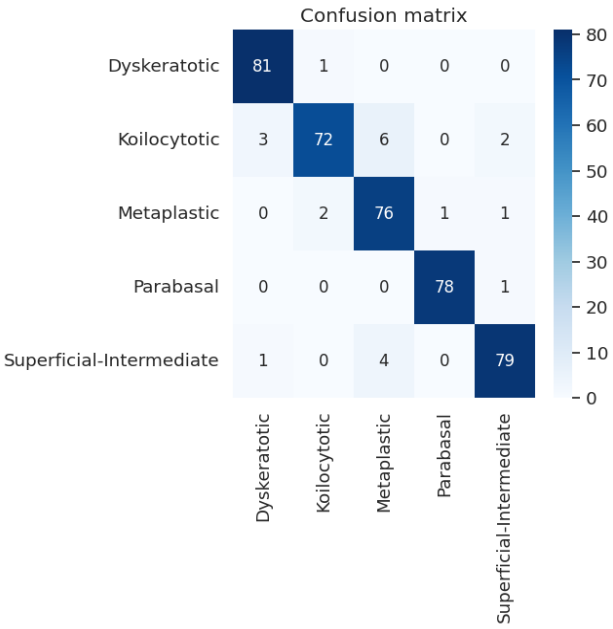

Figure 8: Confusion Matrix

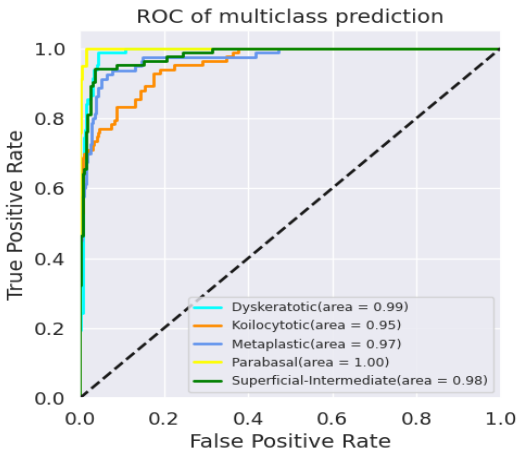

Figure 9: ROC Curve for test data

5.4. FL on Experimental Setting 3

5.4.1. Data Preprocessing

In this setting, in built modules of Pytorch were used for image augmentation: RandomRotation, RandomResizedCrop, RandomHorizontalFlip, RandomVerticalFlip, ToTensor and Normalize. *RandomVerticalFlip* helps to flip an image vertically.

5.4.2. Proposed CNN Architecture

CNN architecture described in Table 10 was used in this setting as well. The final test accuracy received was 78.43%. Hyper-parameters used for this setting is shown below in Table 13.

Table 13: Hyperparameters Set 5 used for CNN Model 3 in Experimental Setting 3

Parameters Used	**Values**
Number of Conv2D layers	2
Number of MaxPooling2D layers	2
Activation functions used	ReLU(Conv2D layers)
Loss function	Categorical Cross-entropy
Optimizer	SGD
Input image size	28 x 28
Batch Size	5
Total Communication Round	250
Number of Epochs for Local Train	15
Learning Rate	0.001

5.4.3. Result Analysis of CNN-based FL

In Table 14, the classification report of FedAvg on SIPAKMED dataset is presented for non-IID setting. In a non IID distribution of data, 78.43% test accuracy was obtained with a roc-auc score of 89.9%.

Figure 10 shows a comparison curve for accuracy and loss of training and validation data on total communication rounds. The confusion matrix on test data for non-IID setting is presented in Figure 11. The One-Vs-Rest Multiclass ROC curve was also presented in the Figure 12. To compare the performance of CNN based FL architectures with classical ML models, the development of existing ML models is shown in section 6.

Table 14: Classification Metrics for Experimental Setting 3

Class	Metrics			ROC AUC Score
	Precision	Recall	F1-Score	
Dyskeratotic	0.93	0.82	0.87	0.90
Koilocytotic	0.70	0.69	0.70	0.76
Metaplastic	0.78	0.80	0.79	0.88
Parabasal	0.95	0.91	0.93	0.97
Superficial-Intermediate	0.63	0.73	0.67	0.80
Test accuracy				0.784
ROC_AUC Score				0.899

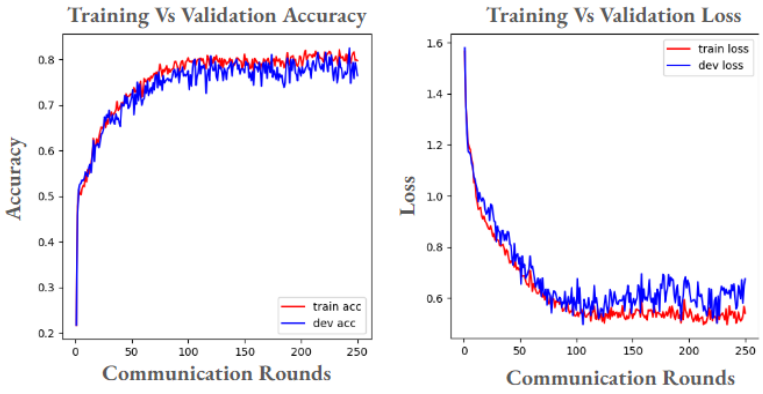

Figure 10: Train Vs. Validation Accuracy and Loss

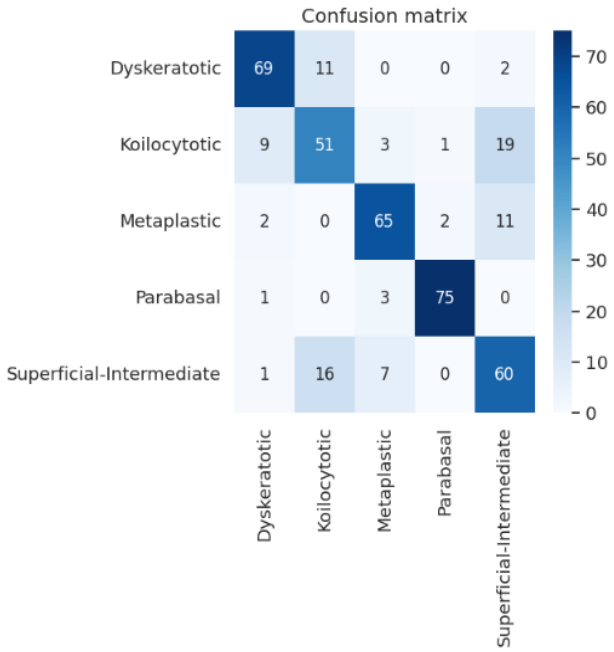

Figure 11: Confusion Matrix

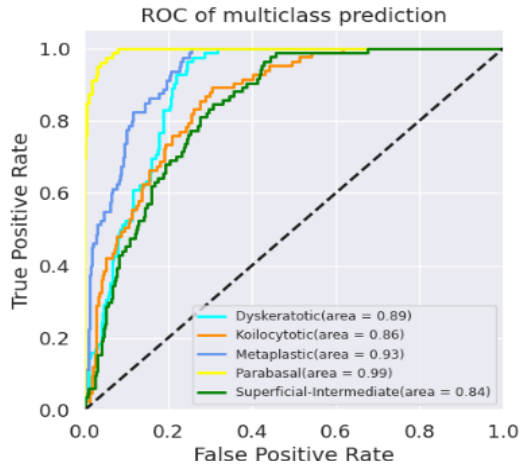

Figure 12: ROC Curve for test data

6. Development of Classical ML Models

6.1. Data Pre-processing

Gray-scale conversion of images was done to remove the difficulties associated with computational requirements. In order to feed information from a 1-D array into a classification model, a technique known as flattening is utilized to transform multidimensional arrays into 1-D arrays. The images are resized into 28x28 pixels, and a data frame of 784 pixels for each image was created. Each image was re-scaled to the range of 0 to 255. The dataset was divided into train and test images with a test ratio of 20%. The Synthetic Minority Oversampling Technique (SMOTE) was used for handling imbalanced data. Figure 13 shows the effects of using smote.

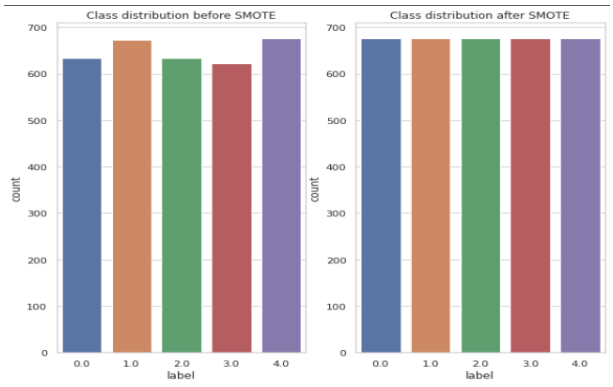

Figure 13: Class distribution before and after applying SMOTE

6.2. Result Analysis of the ML Models

Table 15 shows the value of these performance metrics: accuracy, precision, recall, and f1-score for train data as well as test data for classical ML models. Figure 14,15,16 shows the confusion matrices for test data for the top 3 models: LightGBM, Histogram Gradient Boosting, and Extratrees. The LightGBM model is found to perform better than the other models.

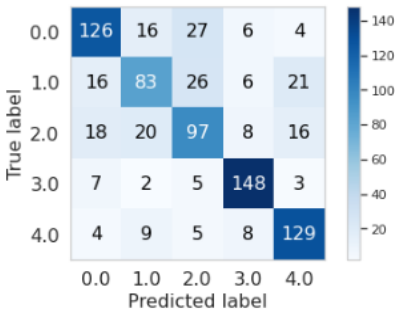

Figure 14: Confusion matrices for LightGBM

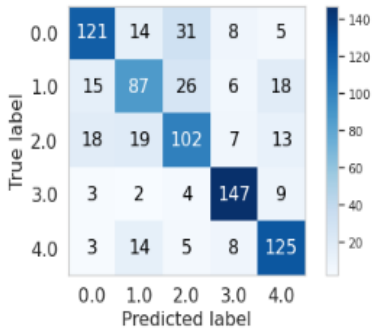

Figure 15: Confusion matrices for Histogram Gradient Boosting

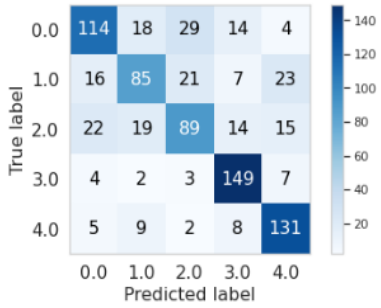

Figure 16: Confusion matrices for Extra Trees

6.3. Cross Validation Results

Cross-validation is a resampling method that trains and tests a model using different portions of the data on different iterations. The number of groups into which a specific data sample is to be divided is indicated by the procedure's sole parameter, k. The process is hence frequently referred to as k-fold cross-validation. In this study, 10-fold cross validation is applied on the entire dataset. Here the top 2 models which performs best on test data are: LightGBM and Extreme Gradient Boosting. In Table 16, the results are shown for LightGBM model and in Table 17, the results are shown for Extreme Gradient Boosting model.

7. Performance Comparison

In this section, a brief comparison is shown between the performance metrics of two metholodogies. The performance metrics selected for the comparison are: test accuracy, test precision, test recall and test f1-score. From Table 18, it can be concluded that CNN-based FL-architecture performs best compared to the traditional ML models.
Furthermore, only one study [26] was found that showed a test accuracy of 81.91% with FedAvg and 83.313% with FedProx algorithm on a non-IID setting where each client had a unique class of data. But their acquired dataset also contained lung and colon cancer classes. As a result, the performance metrics of this research cannot be directly compared with this study. As such, a conclusion can be drawn that the efficacy and validity of Federated Learning is proved via different experimental settings through this research.

8. Discussion & Conclusion

This study has attained a standard test accuracy by harnessing the benefits of integrating Federated Learning with deep learning algorithms. The proposed CNN-based FL architecture showed test accuracy of 94.36% and 78.43% on an IID and a non-IID setting respectively. A brief comparison was drawn between the proposed architectures and classical ML models which proved the significant performance of federated learning over classical ML models.

Although there have been numerous studies for the classification of cancerous cervical cells, further research is required to accurately classify these cells through advanced Deep Learning and Federated Learning techniques. There has been limited research conducted on the classification of cervical cancer cells through multiple Federated Learning approaches. To overcome data sharing challenges faced in related studies, this study in cervical cancer classification will be a significant initiation for more relevant research in the future. The findings of this study can be used as a benchmark for an unbiased assessment of a computer-assisted diagnostic tool that would be selected by physicians in a real world setting.

It is important to acknowledge that privacy is a valid concern that the current work does not fully address. However, employing techniques like differential privacy and secure multi-party computation can help extend this research more and tackle all issues related to data security. As a result, future work for this research will focus on introducing differential privacy, which would help to give results as accurate as possible while addressing data privacy concerns by enabling confidentiality of a database. Moreover, with more diverse datasets the scope of this research can be more broadened. The system can be upgraded through continuous user feedback by deploying in the real world. Different pre-trained models can also be tested to determine which may result in better accuracy when integrated with federated learning.

Data Availability

The datasets selected for analysis in this study were obtained from https://paperswithcode.com/dataset/sipakmed .

Research and Ethics Committee

We confirm that ethical approval was applied for conducting this research. The research was conducted using the open source data. Therefore, the ethical committee headed by the Research & Development Wing of Military Institute of Science and Technology (MIST) decided that it is not required to have formal approval.

Funding statement

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Table 15: Performance metrics for the developed models

Model	Train				Test			
	Accuracy	**Precision**	**Recall**	**F1 score**	**Accuracy**	**Precision**	**Recall**	**F1 score**
LightGBM	99.9%	99.9%	99.9%	99.9%	72.0%	71.4%	71.8%	71.4%
HGB	100.0%	100.0%	100.0%	100.0%	71.9%	71.5%	71.7%	71.5%
Extratrees	100.0%	100.0%	100.0%	100.0%	70.1%	69.4%	70.1%	69.4%
SVM	69.4%	69.0%	69.4%	68.9%	69.8%	68.9%	69.5%	68.8%
Gradient Boosting	84.6%	84.7%	84.6%	84.5%	68.6%	67.8%	68.5%	67.8%
XGBoost	77.0%	77.5%	77.0%	76.8%	68.3%	67.8%	68.1%	67.4%
MLP	70.5%	71.6%	70.5%	70.4%	67.3%	68.1%	67.0%	66.8%
HGB + XGBoost	76.8%	77.1%	76.8%	76.6%	66.9%	66.2%	66.7%	66.0%

Table 16: Cross validation result for developed LightGBM model

Model	Evaluation Metric	Datafolds										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean**
LightGBM	Train Accuracy	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Train Precision	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Train Recall	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Train F1-Score	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Test Accuracy	65.9%	71.3%	72.6%	69.1%	67.9%	70.1%	71.8%	74.3%	69.8%	75%	70.8%
	Test Precision	64.9%	71.1%	72.4%	69.1%	67.8%	70.0%	71.7%	74.1%	69.6%	74.8%	70.6%
	Test Recall	65.9%	71.4%	72.6%	69.2%	67.9%	70.1%	71.9%	74.4%	69.8%	75%	70.9%
	Test F1 Score	65.3%	71.1%	72.5%	68.9%	67.6	70.1%	71.7%	74.1%	69.6%	74.8%	70.6%

Table 17: Cross validation result for developed Extreme Gradient Boosting model

Model	Evaluation Metric	Datafolds										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean**
XGBoost	Train Accuracy	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Train Precision	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Train Recall	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Train F1-Score	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Test Accuracy	67.4%	70.1%	72.3%	68.4%	68.8%	68.6%	70.4%	70.4%	72.6%	73.7%	70.3%
	Test Precision	66.6%	69.7%	72.0%	68.3%	69.3%	68.4%	70.2%	70.1%	70.2%	73.8%	70.0%
	Test Recall	67.5%	70.1%	72.4%	68.4%	68.9%	68.7%	70.4%	70.4%	72.5%	73.8%	70.3%
	Test F1 Score	66.7%	69.7%	72%	68.1%	68.7%	68.4%	70.2%	70.1%	72.1%	73.5%	69.9%

Table 18: Comparison between the performance metrics of two methodologies on Test data

Methodology	Model	Accuracy	Precision	Recall	F1-score
Phase 1	CNN 03 + Hyper-parameter Set 4 (IID)	94.36%	94.54%	94.36%	94.32%
	CNN 02 + Hyper-parameter Set 3 (IID)	88.46%	89.58%	88.47%	87.83%
	CNN 02 + Hyper-parameter Set 2 (IID)	87.00%	86.38%	84.47%	86.83%
	CNN 01 + Hyper-parameter Set 1 (IID)	82.18%	82.59%	83.47%	82.62%
	CNN 03 + Hyper-parameter Set 5 (non-IID)	79.60%	80.17%	79.66%	78.77%
Phase 2	LightGBM	72.0%	71.4%	71.8%	71.4%
	Histogram Gradient Boosting	71.9%	71.5%	71.7%	71.5%
	Extratrees	70.1%	69.4%	70.1%	69.4%
	SVM	69.8%	68.9%	69.5%	68.8%
	Gradient Boosting	68.6%	67.8%	68.5%	67.8%

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments The authors would like to thank the

anonymous reviewers for their invaluable comments and suggestions, which helped improve the quality of the paper.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a cancer journal for clinicians 71 (2021) 209–249.

[2] S. Pimple, G. Mishra, Cancer cervix: Epidemiology and disease burden, Cytojournal 19 (2022).

[3] M. A. Razzak, M. N. Islam, T. Broti, E. S. Kamal, S. Zahan, Exploring usability problems of mhealth applications developed for cervical cancer: An empirical study, SAGE Open Medicine 11 (2023) 20503121231180413.

[4] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, 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.

[5] E. Krawczyk, F. A. Suprynowicz, X. Liu, Y. Dai, D. P. Hartmann, J. Hanover, R. Schlegel, Koilocytosis: a cooperative interaction between the human papillomavirus e5 and e6 oncoproteins, The American journal of pathology 173 (2008) 682–688.

[6] M. A. Razzak, M. N. Islam, M. S. Aadeeb, T. Tasnim, Digital health interventions for cervical cancer care: A systematic review and future research opportunities, PloS one 18 (2023) e0296015.

[7] A.-u. Rehman, N. Ali, I. A. Taj, M. Sajid, K. S. Karimov, An automatic mass screening system for cervical cancer detection based on convolutional neural network, Mathematical Problems in Engineering 2020 (2020) 1–14.

[8] A. Ghoneim, G. Muhammad, M. S. Hossain, Cervical cancer classification using convolutional neural networks and extreme learning machines, Future Generation Computer Systems 102 (2020) 643–649.

[9] A. R. Bhatt, A. Ganatra, K. Kotecha, Cervical cancer detection in pap smear whole slide images using convnet with transfer learning and progressive resizing, PeerJ Computer Science 7 (2021) e348.

[10] R. Surendiran, M. Thangamani, M. S, R. P, Exploring the cervical cancer prediction by machine learning and deep learning with artificial intelligence approaches, International Journal of Engineering Trends and Technology 70 (2022) 94–107. doi:10.14445/22315381/IJETT-V70I7P211.

[11] E. L. Silva, A. F. Sampaio, L. F. Teixeira, M. J. M. Vasconcelos, Cervical cancer detection and classification in cytology images using a hybrid approach, in: Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part II, Springer, 2021, pp. 299–312.

[12] B. Taha, J. Dias, N. Werghi, Classification of cervical-cancer using pap-smear images: a convolutional neural network approach, in: Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21, Springer, 2017, pp. 261–272.

[13] P. Shanthi, K. Hareesha, R. Kudva, Automated detection and classification of cervical cancer using pap smear microscopic images: A comprehensive review and future perspectives, Engineered Science 19 (2022) 20–41.

[14] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, Q. Wang, Deepcervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques, Computers in Biology and Medicine 136 (2021) 104649.

[15] W. Liu, C. Li, N. Xu, T. Jiang, M. M. Rahaman, H. Sun, X. Wu, W. Hu, H. Chen, C. Sun, et al., Cvm-cervix: A hybrid cervical pap-smear image classification framework using cnn, visual transformer and multilayer perceptron, Pattern Recognition 130 (2022) 108829.

[16] W. Liu, C. Li, M. M. Rahaman, T. Jiang, H. Sun, X. Wu, W. Hu, H. Chen, C. Sun, Y. Yao, et al., Is the aspect ratio of cells important in deep learning? a robust comparison of deep learning methods for multi-scale cytopathology cell image classification: From convolutional neural networks to visual transformers, Computers in biology and medicine 141 (2022) 105026.

[17] Z. Fan, X. Wu, C. Li, H. Chen, W. Liu, Y. Zheng, J. Chen, X. Li, H. Sun, T. Jiang, et al., Cam-vt: A weakly supervised cervical cancer nest image identification approach using conjugated attention mechanism and visual transformer, Computers in Biology and Medicine 162 (2023) 107070.

[18] M. M. Rahaman, C. Li, X. Wu, Y. Yao, Z. Hu, T. Jiang, X. Li, S. Qi, A survey for cervical cytopathology image analysis using deep learning, IEEE Access 8 (2020) 61687–61710.

[19] A. Linardos, K. Kushibar, S. Walsh, P. Gkontra, K. Lekadir, Federated learning for multi-center imaging diagnostics: a simulation study in cardiovascular disease, Scientific Reports 12 (2022) 3551.

[20] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, S. Bakas, Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4, Springer, 2019, pp. 92–104.

[21] Z. Ma, M. Zhang, J. Liu, A. Yang, H. Li, J. Wang, D. Hua, M. Li, An assisted diagnosis model for cancer patients based on federated learning, Frontiers in Oncology 12 (2022) 860532.

[22] S. Pati, U. Baid, B. Edwards, M. Sheller, S.-H. Wang, G. A. Reina, P. Foley, A. Gruzdev, D. Karkada, C. Davatzikos, et al., Federated learning enables big data for rare cancer boundary detection, Nature communications 13 (2022) 7346.

[23] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso, et al., Privacy-preserving federated brain tumour segmentation, in: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10, Springer, 2019, pp. 133–141.

[24] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, H. R. Tizhoosh, Federated learning and differential privacy for medical image analysis, Scientific reports 12 (2022) 1953.

[25] G. A. Kaissis, M. R. Makowski, D. Rückert, R. F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nature Machine Intelligence 2 (2020) 305–311.

[26] M. Subramanian, V. Rajasekar, S. VE, K. Shanmugavadivel, P. Nandhini, Effectiveness of decentralized federated learning algorithms in healthcare: A case study on cancer classification, Electronics 11 (2022) 4117.

[27] Z. Hu, L. Ma, Y. Ding, X. Zhao, X. Shi, H. Lu, K. Liu, Enhancing the accuracy of lymph-node-metastasis prediction in gynecologic malignancies using multimodal federated learning: Integrating ct, mri, and pet/ct, Cancers 15 (2023) 5281.

[28] M. Moshawrab, M. Adda, A. Bouzouane, H. Ibrahim, A. Raad, Reviewing federated learning aggregation algorithms; strategies, contributions, limitations and future perspectives, Electronics 12 (2023) 2287.

[29] B. Pfitzner, N. Steckhan, B. Arnrich, Federated learning in a medical context: a systematic literature review, ACM Transactions on Internet Technology (TOIT) 21 (2021) 1–31.

[30] C. Korkmaz, H. E. Kocas, A. Uysal, A. Masry, O. Ozkasap, B. Akgun, Chain fl: Decentralized federated machine learning via blockchain, in: 2020 Second international conference on blockchain computing and applications (BCCA), IEEE, 2020, pp. 140–146.