Article

A Collaborative Federated Learning Framework for Lung and Colon Cancer Classifications

Md. Munawar Hossain [1,*] , **Md. Robiul Islam** [1] , **Md. Faysal Ahamed** [1] , **Mominul Ahsan** [2] and **Julfikar Haider** [3,*]

1 Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh; robiulruet00@gmail.com (M.R.I.); faysalahamedjishan@gmail.com (M.F.A.)
2 Department of Computer Science, University of York, Deramore Lane, Heslington YO10 5GH, UK; mominul.ahsan2@gmail.com
3 Department of Engineering, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK
* Correspondence: munawark7@gmail.com (M.M.H.); j.haider@mmu.ac.uk (J.H.)

Abstract: Lung and colon cancers are common types of cancer with significant fatality rates. Early identification considerably improves the odds of survival for those suffering from these diseases. Histopathological image analysis is crucial for detecting cancer by identifying morphological anomalies in tissue samples. Regulations such as the HIPAA and GDPR impose considerable restrictions on the sharing of sensitive patient data, mostly because of privacy concerns. Federated learning (FL) is a promising technique that allows the training of strong models while maintaining data privacy. The use of a federated learning strategy has been suggested in this study to address privacy concerns in cancer categorization. To classify histopathological images of lung and colon cancers, this methodology uses local models with an Inception-V3 backbone. The global model is then updated on the basis of the local weights. The images were obtained from the LC25000 dataset, which consists of five separate classes. Separate analyses were performed for lung cancer, colon cancer, and their combined classification. The implemented model successfully classified lung cancer images into three separate classes with a classification accuracy of 99.867%. The classification of colon cancer images was achieved with 100% accuracy. More significantly, for the lung and colon cancers combined, the accuracy reached an impressive 99.720%. Compared with other current approaches, the proposed framework showed an improved performance. A heatmap, visual saliency map, and GradCAM were generated to pinpoint the crucial areas in the histopathology pictures of the test set where the models focused in particular during cancer class predictions. This approach demonstrates the potential of federated learning to enhance collaborative efforts in automated disease diagnosis through medical image analysis while ensuring patient data privacy.

Keywords: lung cancer; colon cancer; histopathological image analysis; image classification; decentralized machine learning; federated learning; privacy preservation; explainability

1. Introduction

Cancer ranks as the second most prevalent cause of mortality globally, following cardiovascular ailments. In 2022, there were 9.74 million cancer-related deaths and almost 20 million new cases reported globally [1]. Lung cancer is responsible for the highest mortality rate and the second highest number of incidents worldwide. On the basis of the latest data from the World Health Organization (WHO), 2.20 million new cases and 1.79 million deaths were reported in 2020. Figure 1 shows that lung cancer has the highest number of cases, accounting for 12.4%, narrowly surpassing breast cancer at 11.6% [2]. Nevertheless, it is clear that lung cancer is the leading cause of death among all cancer-related deaths, with a mortality rate of 18%. There is no other cancer that even comes close to having this enormous number of fatalities.
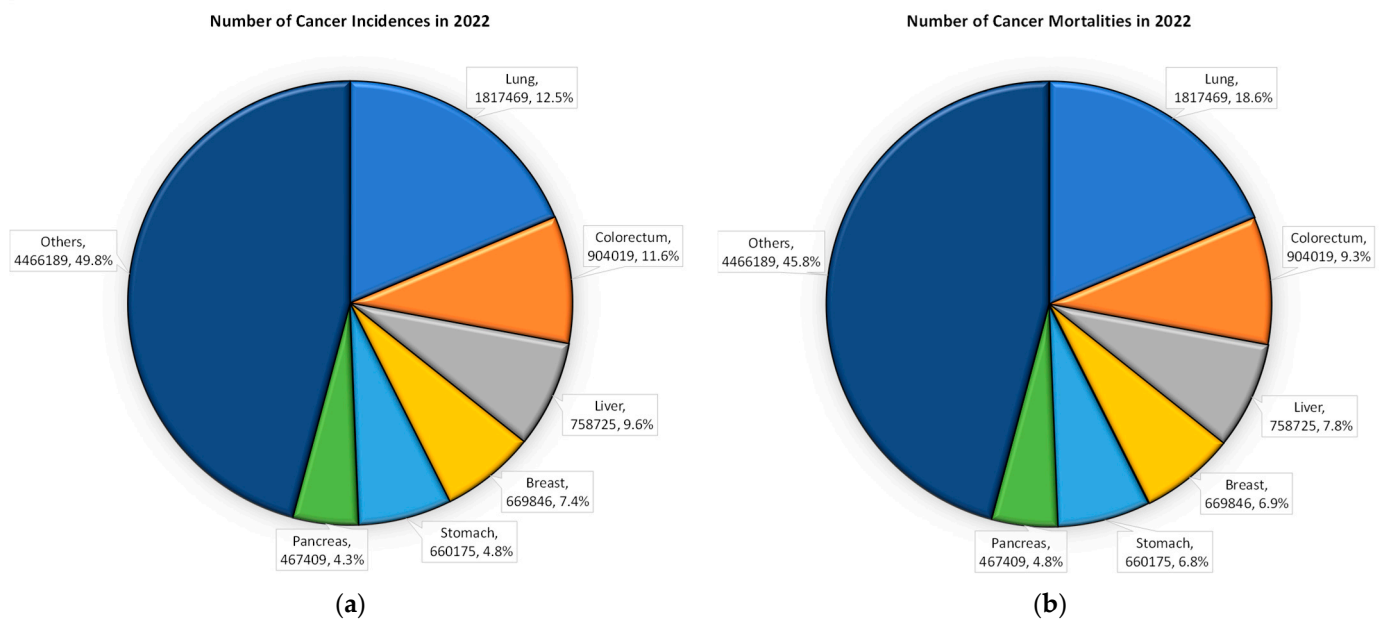
Figure 1. Total number and proportion of cancer (**a**) incidences and (**b**) mortalities in 2022 (data taken from [1]).

On the basis of the analysis, it can be concluded that colorectal cancer ranks as the second leading cause of mortality and the third highest in terms of incidence globally. In 2022, it ranked second in terms of new cases, accounting for 9.6% of all cases, just after breast cancer. Additionally, it constituted 9.3% of the total number of deaths, ranking second only to lung cancer [1]. The global impact of this phenomenon resulted in approximately 1.93 million new cases and 935,000 deaths, representing approximately 10% of all new cancer cases and fatalities worldwide. With these prevalence rates rising sharply annually, it is projected that by 2040, there will be 3.2 million new instances of colorectal cancer (CRC), posing a serious threat to global public health [3]. Hence, it is crucial to employ expeditious and efficient protocols for diagnostic decision making to formulate a tailored treatment strategy that maximizes patient survival rates in every individual case.

Multiple factors contribute to the development of cancer. These include exposure to physical carcinogens, such as radiation and ultraviolet rays, and certain behaviors, such as having a high body mass index and using alcohol and tobacco, in addition to specific biological and genetic factors [4]. While symptoms may vary among individuals and even different types of cancers, none of these signs are exclusive to cancer, and not all patients will encounter them. In light of this, cancer detection may be difficult in the absence of specialized diagnostic tools such as ultrasound, positron emission tomography (PET), computed tomography (CT), MRI, or biopsy. Prompt identification is crucial for the detection of both lung and colon cancers. In the field of clinical medicine, symptoms of these particular types of cancers typically manifest during the advanced stages of the disease. Physicians face difficulties in the early diagnosis of lung cancer through exclusive reliance on visual assessment of CT images. The application of computer-aided diagnosis (CAD) in the examination of histopathological images continues to be a prominent area of emphasis in the field of cancer detection.

Privacy is a significant concern for healthcare facilities. As a result, regulations such as HIPAA [5] and GDPRs [6] pose significant challenges for institutions with respect to disclosing patient data, including anonymous information. To overcome these difficulties, the adoption of a decentralized method for machine learning called federated learning has been suggested. Currently, several deep learning techniques have shown excellent accuracy when predicting the lung cell class. Nevertheless, none of these entities have adopted a thorough strategy to address data privacy regulations.

It is challenging for clinicians to diagnose cancer at an early stage by relying merely on a visual assessment of CT images, as evidenced by the fact that symptoms of cancer often manifest in the advanced stages of the disease. By pooling information, healthcare providers can better pinpoint areas for improvement and perhaps diagnose lung cancer at an earlier, more treatable stage. By sharing data, organizations can increase efficiency and maximize resources, leading to greater productivity and cost savings. Medical data sharing is a powerful tool that can drive significant advancements in the field of medicine. Hospitals in the healthcare industry or individual devices are examples of the many parties or devices used across federated learning methodologies in healthcare.

Figure 2 illustrates a federated learning model where multiple healthcare providers collaborate to train a shared machine learning model without sharing patient data directly. Each healthcare provider trains its local model using its own data, and only the model updates are sent to a central server. The central server aggregates these updates to create a global model, which is then shared back with all the healthcare providers, who then continue to refine the global model by performing further training with their local data, and this iterative process continues until the model is fully trained. By collaborating and sharing data, organizations can unlock the potential for groundbreaking advancements in our current understanding of diseases and treatments.

This study aims to design and implement a distributed federated learning (FL) architecture that allows for adaptive thresholding and the customization of local neural networks (NNs) in the context of medical image classification, with a specific focus on lung and colon cancer diagnoses. The goal is to harness the power of federated learning to facilitate collaboration among healthcare organizations while maintaining a decentralized approach. Through this decentralized approach, each institution can train a local neural network on its own data while contributing to a shared global model that benefits from the collective knowledge of all participants. By doing so, individual institutions can contribute to a shared global model without directly sharing sensitive patient data, ensuring compliance with stringent healthcare privacy and security regulations. This study not only seeks to increase the accuracy of cancer diagnoses but also aims to create a scalable and secure solution that can be applied to a wide range of medical imaging tasks. To summarize, this study makes the following significant contributions:

(1) A robust federated learning architecture was designed and developed, particularly for medical image classification, which is demonstrated through lung and colon cancer classifications. The framework seamlessly consolidated data from many healthcare organizations while upholding data privacy and security regulations.

(2) The federated learning workflow was streamlined for smooth global model updates after each communication round, with local model weights adjusted to align with the global model. A comprehensive evaluation process was also applied, assessing each client's model performance after every training epoch, enhancing transparency and identifying performance variations or underperforming clients.

(3) Explainable AI techniques were integrated to provide visual and quantitative insights into the model's decision-making process and provide further interpretability.

(4) The performance of the proposed federated learning (FL) model is evaluated against well-known transfer learning (TL) models and other current state-of-the-art (SOTA) approaches.

The main novelty of this work lies in comparing the performance of federated and centralized machine learning models for the LC-25000 dataset, which is specifically applied to lung and colon cancer classifications. This research is the first to demonstrate the efficiency of a unique combination of federated learning in this context, highlighting the collaborative nature of the framework, introducing technical innovations, and emphasizing the potential impact on clinical practices.

The subsequent sections of the article are structured as follows: Section 2 provides a concise overview of prior research that is similar to this study. Section 3 presents a detailed representation of the dataset used in this study and the methodology proposed for the

federated learning model. The experimental conditions under which the proposed method is compared are presented in Section 4. Section 5 provides a comprehensive analysis and evaluation of the experiment's performance, highlighting its findings. Section 6 concludes the proposed research and provides recommendations for further enhancement. Finally, the article concludes by providing a concise overview of the results.
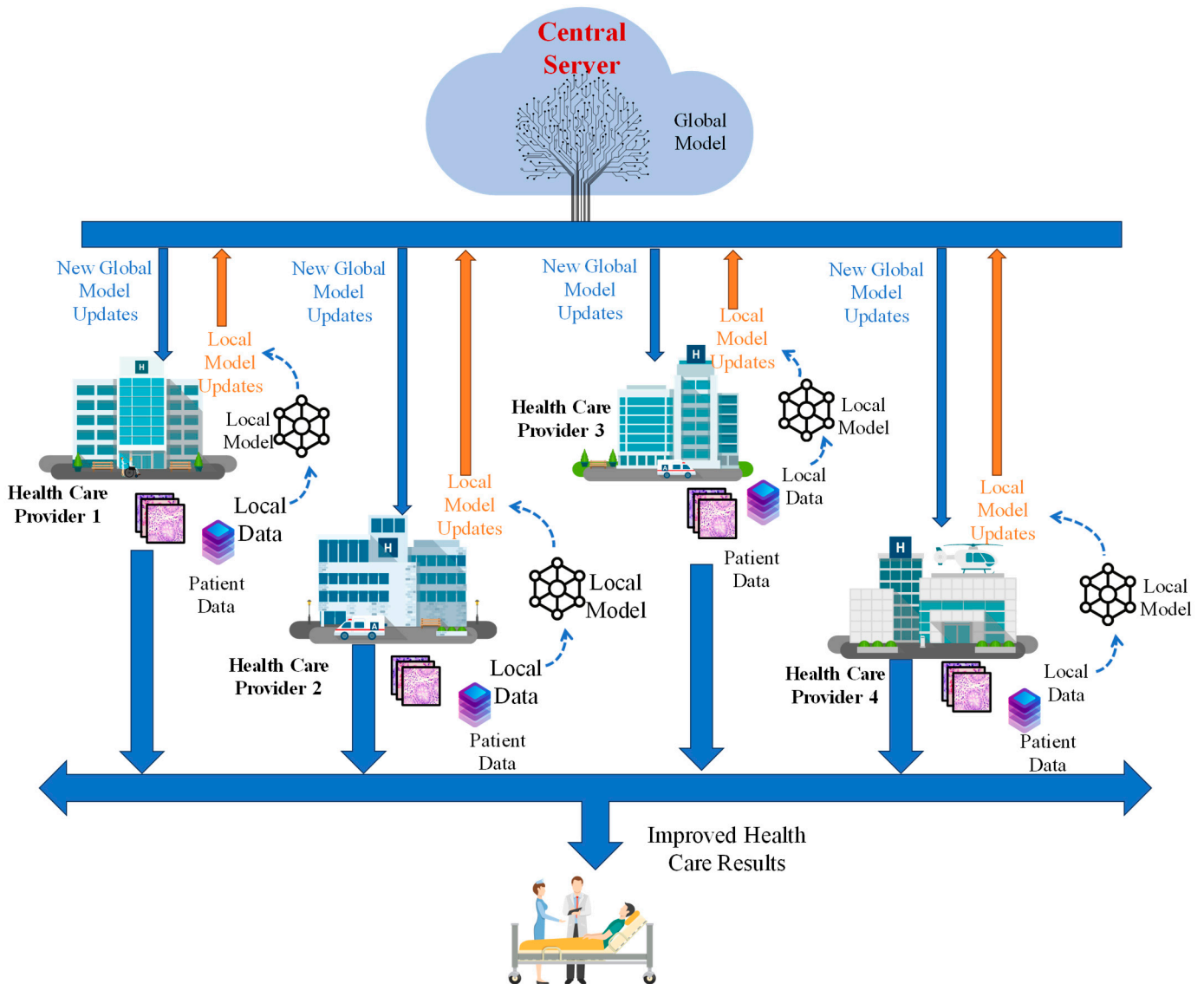
Local Model updates: Client-side weight or parameter adjustments based on local data. Global Model Updates: Server-side aggregation of local client updates to refine the local model. Patient Data: Test Data, Personal Data, Vital Signs

Figure 2. Federated learning conceptual framework applicable in healthcare sector.

2. Literature Survey

2.1. Lung and Colon Cancer Diagnoses

Lung and colon cancers are prevalent and highly fatal diseases that have a significant global impact, affecting many individuals annually. Despite their organ-specific origins, there are several shared features and important distinctions between the two in terms of prognosis, diagnostic criteria, and therapeutic options. Numerous research groups have achieved substantial advancements in the detection of lung and colon cancers in recent years. These developments include the application of deep learning methods on the basis of histopathological image analysis. The works are organized into three categories for classifying lung cancer (adenocarcinoma, squamous cell carcinoma, and benign), two

categories for classifying colon cancer subtypes (adenocarcinoma and benign), and five categories for classifying both lung and colon cancer categories.

In a previous study [7], a deep learning-based supervised learning technique was developed to classify lung and colon cancer tissues into five distinct categories. The approach implemented utilized two methods for feature extraction: 2D Fourier features and 2D wavelet features. The final accuracy of the work was 96.33%. Another study [8] utilized feature extraction from histopathology images and various machine learning classifiers, such as random forest (RF) and XGBoost, to classify lung and colon cancers. The study achieved impressive accuracies of 99%. A CNN pretrained diagnostic network was specifically designed for the detection of lung and colon cancers [9]. The network demonstrated a high level of accuracy in diagnosing colon and lung cancers, achieving accuracies of 96% and 97%, respectively. Convolutional neural networks (CNNs) using the VGG16 model and contrast-limited adaptive histogram equalization (CLAHE) were used by other researchers [10] to classify 25,000 histopathology images. Transformers have advanced medical image analysis but struggle with feature capture, information loss, and segmentation accuracy; CASTformer addresses these issues with multi-scale representations, a class-aware transformer module, and adversarial training [11]. Furthermore, incremental transfer learning (ITL) offers an efficient solution for multi-site medical image segmentation by sequentially training a model across datasets, mitigating catastrophic forgetting, and achieving superior generalization to new domains with minimal computational resources and domain-specific expertise [12].

One study [13] discussed the use of histogram equalization as a preprocessing step, followed by the application of pretrained AlexNet, to improve the classification of lung and colon cancers. Toğaçar et al. [14] utilized a pretrained DarkNet-19 in conjunction with support vector machine classifiers to attain a 99.69% accuracy rate in their study. Using DenseNet-121 and RF classifiers, Kumar et al. [15] achieved a 98.6% accuracy rate in their classification. Another study utilized feature extraction and ensemble learning techniques, along with the incorporation of high-performance filtering, to attain an impressive accuracy rate of 99.3% when using LC25000 data [16]. The use of artificial neural networks (ANNs) with merged features from the VGG-19 and GoogLeNet models was covered in [17]. The ANN achieved an accuracy of 99.64% when the fusion features of VGG-19 and the handcrafted features were combined. In a separate study, the authors employed a convolutional neural network (CNN) with a SoftMax classifier, which they named AdenoCanNet [18]. The accuracy of the entire LC25000 dataset was 99.00%.

In addition to the previously discussed methods, recent studies have made significant advancements in learning-based methods for medical image segmentation. Contrastive learning and distillation techniques have shown promise in addressing the challenges of limited labeled data and segmentation accuracy in medical image analysis, with methods like contrastive voxel-wise representation learning (CVRL) [19] and SimCVD [20] advancing state-of-the-art voxel-wise representation learning by capturing 3D spatial context, leveraging bi-level contrastive optimization, and utilizing simple dropout-based augmentation to achieve competitive performance even with less labeled data. Additionally, ACTION (Anatomical-aware ConTrastive dIstillatiON) [21] tackles multi-class label imbalance by using soft negative labeling and anatomical contrast, improving segmentation accuracy and outperforming state-of-the-art semi-supervised methods on benchmark datasets. Finally, ARCO enhances semi-supervised medical image segmentation by introducing stratified group theory and variance-reduction techniques, addressing tail-class misclassification and model collapse, and demonstrating superior performance across eight benchmarks [22].

When the disadvantages of the current models used in lung and colon cancer classifications are analyzed, several challenges and limitations can be identified:

(1) Data Privacy Concerns: Many existing models require centralized data collection, where medical images from different institutions are pooled together in a single repository. This raises serious privacy concerns, especially in healthcare, where patient

data are highly sensitive. Centralized models can be susceptible to data breaches and may not comply with regulations such as HIPAA or GDPR.

(2) Limited Generalization: Centralized models are often trained on data from a limited number of sources or geographic locations, which can result in poor generalizability to other patient populations. This lack of diversity in the training data can lead to biases and reduced effectiveness when applied to new datasets, limiting the model's ability to handle variations in medical imaging from different institutions or regions.

(3) Computational Requirements: Modern models for cancer classification, such as deep convolutional neural networks (CNNs), demand significant computational resources. This can be a barrier for smaller institutions with limited access to high-performance computing infrastructure. Moreover, training large-scale models can be time-consuming and energy-intensive.

(4) Imbalance in Class Distribution: Medical datasets, including lung and colon cancer imaging datasets, often suffer from class imbalance, where the number of images of cancerous tissues is much lower than that of non-cancerous ones. This imbalance can bias the model, making it more likely to misclassify cancer cases, which is especially problematic in clinical settings where false negatives can be life-threatening. Work reported by You et al. [23] introduced adaptive anatomical contrast with a dynamic contrastive loss, which better handles class imbalances in long-tail distributions.

(5) Difficulty in Handling Heterogeneous Data: Medical imaging data can be highly heterogeneous due to differences in imaging equipment, protocols, and settings across institutions. Current models may struggle to handle this heterogeneity, leading to reduced performance when applied to data from sources other than the training data.

2.2. Federated Learning Applications

The integration of massive amounts of data can benefit machine learning models, as stated previously. Access to data in the medical field is highly limited because of the strict considerations of user privacy and data security. In this context, decentralized collaborative machine learning algorithms that protect privacy are appropriate for creating intelligent medical diagnostic systems. The notion of federated learning, which was initially introduced by Google in 2016 [24,25], has since been expanded to encompass scenarios involving knowledge integration and collaborative learning between organizations.

A client server-based method called federated averaging (FedAvg) was used for breast density classification in [26]. This method incorporates local stochastic gradient descent (SGD) on each client with a server that performs model averaging. In their publication [27], the authors proposed a federated learning approach utilizing pretrained deep learning models for the purpose of COVID-19 detection. The clients aimed to collaborate to achieve a global model without the need to share individual samples from the dataset. Another federated learning framework [28] for lung cancer classification utilizing histopathological images demonstrated 99.867% accuracy, while imposing significant limitations on data sharing between institutions. Zhang et al. [29] introduced a dynamic fusion-based approach for COVID-19 detection. An image fusion technique was employed to diagnose patients with COVID-19 on the basis of their medical data. The evaluation parameters yielded favorable outcomes. However, the lack of consideration for patient data privacy was a significant oversight in the proposed medical image analysis.

In the healthcare industry 5.0 domain, researchers have proposed that the Google net deep machine learning model is utilized for precise disease prediction in the smart healthcare industry 5.0 [30]. The proposed methodology for secure IoMT-based transfer learning achieved a 98.8% accuracy rate, surpassing previous state-of-the-art methodologies used in cancer disease prediction within the smart healthcare industry 5.0 on the LC25000 dataset. In a parallel investigation of Society 5.0 [31], researchers presented data as a service (DaaS) along with a suggested framework that uses the blockchain network to provide safe and decentralized transmission and distribution of data and machine learning systems on the cloud. The main contributions and shortcomings of previous federated learning research can be found in Table 1.

Table 1. Summary of the main contributions and shortcomings of previous FL research.

Previous Study	Main Contribution of the Research	Limitations of the Work
Zhang et al. [29]	Dynamic fusion-based approach for CT scan image analysis to diagnose COVID-19.	Concerns regarding the appropriateness of controls for patient data privacy and authenticity.
Roth et al. [26]	In a real-world collaborative setting, the author employed FL to develop medical imaging classification models.	The proposed model is overly simplistic and requires additional simulations.
Khan et al. [30]	Proposed a secure IoMT-based transfer learning methodology	Focusing more on the application of IoMT devices, intended for industry 5.0 application. There is chance of data corruption through IoMT devices.
Florescu et al. [27]	A federated learning (FL) system was implemented for COVID-19 detection using CT images, with clients deployed locally on a single machine.	Doubts about the suitability of safeguards for maintaining the confidentiality and integrity of patient data.
Peyvandi et al. [31]	Proposed blockchain-based DCIaaS framework enhances data and computational intelligence quality, equality, and privacy for machine learning, demonstrating improved accuracy in biomedical image classification and hazardous litter management.	Potential complexity and computational overhead introduced by using blockchain technology, which could affect the efficiency and scalability of the system.

The major gaps in the literature concerning lung and colon cancer classifications that inspired the current study framework are briefly summarized below.

- There is a noticeable absence of sufficient measures to guarantee the privacy and security of patient data.
- There are instances where the computational cost becomes considerably higher owing to the substantial increase in the data scale, making it challenging to maintain efficiency and performance.

3. Data and Methodology

This section provides insights into the dataset and various approaches to implementing federated learning. Figure 3 illustrates the overall study path. Initially, the preprocessed dataset was divided into training, testing, and validation sets. These datasets are distributed across multiple healthcare institutions, each training their local neural networks independently using their own data, without sharing any sensitive patient information. Each institution trains its local neural network independently, ensuring that patient data remain private. The locally trained models are then aggregated in a central server to update a global model, which is shared back with the institutions for further refinement. Explainable AI techniques are applied to enhance model interpretability by visualizing the features that drive predictions. The process culminates in performance evaluation, ensuring accuracy and transparency in predictions, while fostering collaboration across healthcare organizations.

The data that are subsequently distributed among clients via an independent and identically distributed (IID) approach. Local models are first developed through training on the data, and then the clients send the model parameters to the server. After training the local models, the results are aggregated in a secure, centralized server to update a global model, which represents the combined knowledge of all institutions. This process involves training data on individual client devices and subsequently merging the local models on a central server. The workflow for federated learning via Inception-V3 is illustrated in Figure 4.

3.1. Dataset, Preprocessing, and Splitting

The dataset LC25000, published in 2020 by A. Borkowski and colleagues [32], was utilized in this study. The collection contains images of lung and colon tissues, which are categorized into five distinct classes. There are three distinct types of lung tissue images: adenocarcinoma, squamous cell carcinoma, and benign. Some sample images of the classes can be seen in Figure 5. The production of this content was made possible through the

provision of resources and utilization of facilities at James A. Haley Veterans' Hospital. It is collected from patients through keen observation by physiologists.
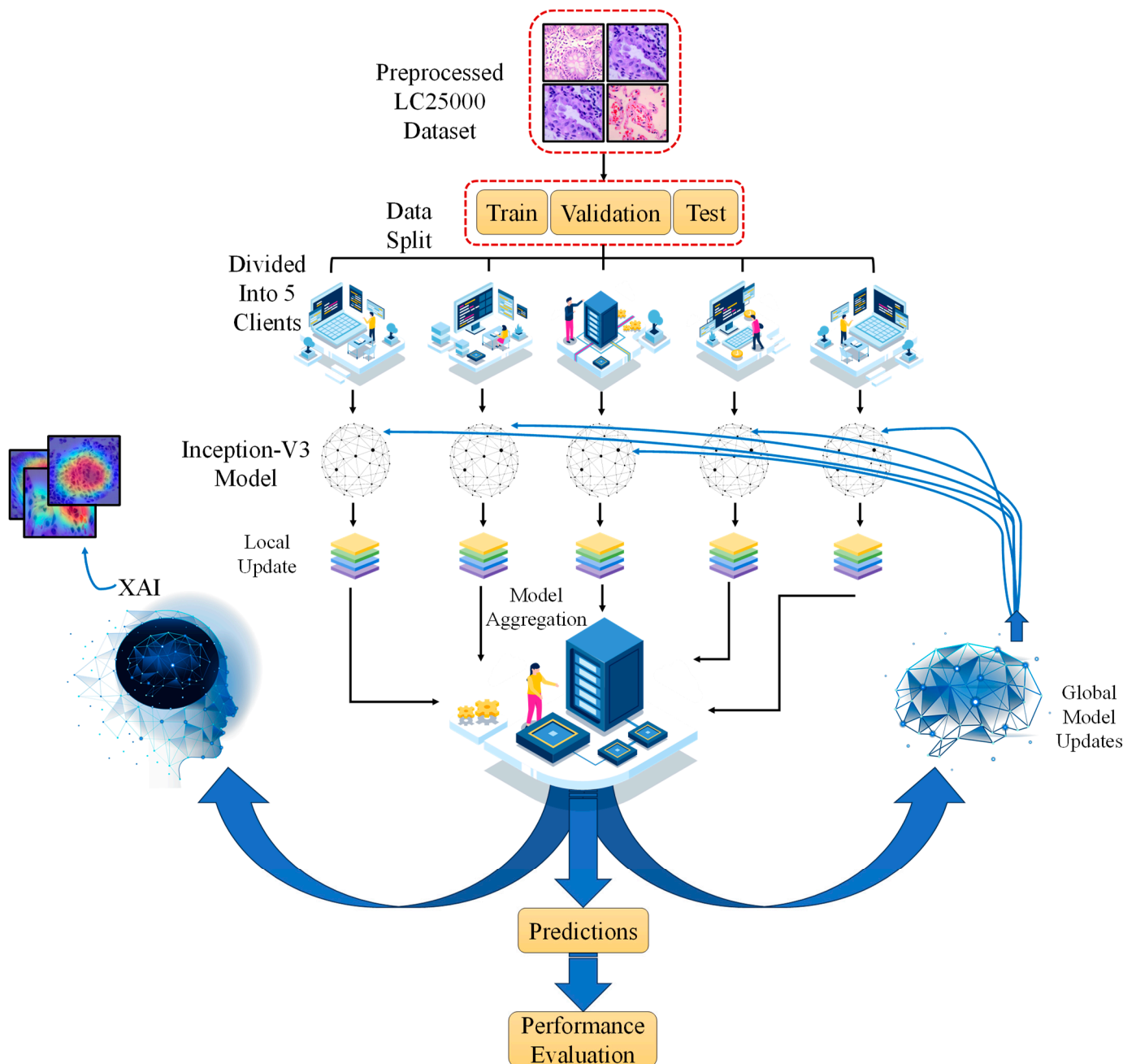
Figure 3. Diagram illustrating the entire path of the federated learning study, including steps from data collection to model evaluation and explainable AI integration.

The LC25000 dataset consisted primarily of 1250 pathology slide images of lung and colon tissues. Borkowski et al. [32] used an Augmentor library to apply preprocessing techniques to the images and increased the size of our dataset to a total of 25,000 images. This was achieved through the implementation of various augmentations, including left and right rotations with a maximum angle of 25 degrees and a probability of 1.0. Additionally, horizontal and vertical flips were applied with a probability of 0.5. Consequently, the dataset was expanded to a total of 25,000 images, which were further categorized into five distinct categories. Each category contained 5000 images. The images were resized to dimensions of 768×768 prior to the application of augmentation techniques. To guarantee

privacy and unrestricted usage, these images underwent validation and adhered to the regulations set forth by the Health Insurance Portability and Accountability Act (HIPAA). Table 2 displays the designated names and IDs assigned to each class of images within the dataset and an overview of the dataset split. To reduce computational complexity, we downsized the images in our training and test directories from our pre-existing dataset, which had 100×100 pixels. The utilization of the training and test directories is justified by the fact that the test directory's images are utilized to test the global model, whereas the training directory's images are disseminated to end devices/clients for local data training. The dataset containing 25,000 lung and colon cancer images is organized into training, testing, and validation sets at an 80:10:10 ratio.
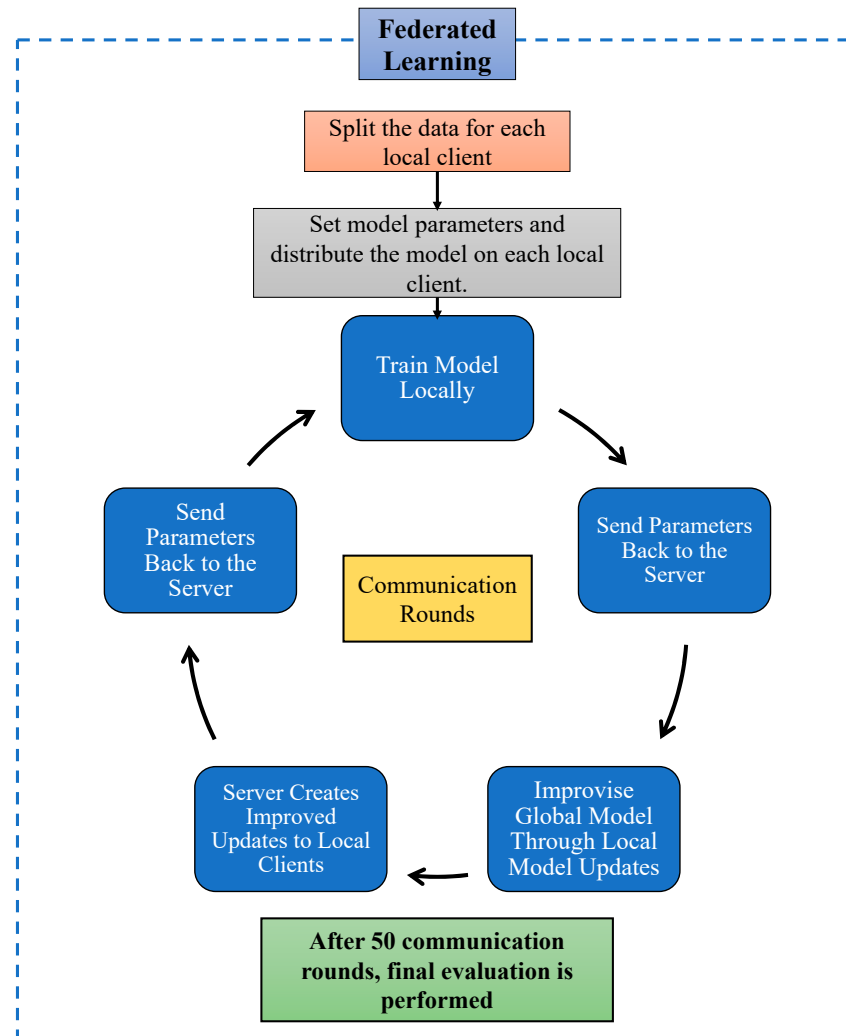
Figure 4. Federated learning with Inception-V3 methodology.

Table 2. Description of the employed dataset.

Image Type	Folder Title	Total Images	Training Set	Testing Set	Validation Set
Lung Adenocarcinoma	lung_aca	5000	4000	500	500
Lung Benign	lung_bnt	5000	4000	500	500
Lung Squamous Cell Carcinoma	lung_scc	5000	4000	500	500
Colon Adenocarcinoma	colon_aca	5000	4000	500	500
Colon Benign	colon_bnt	5000	4000	500	500
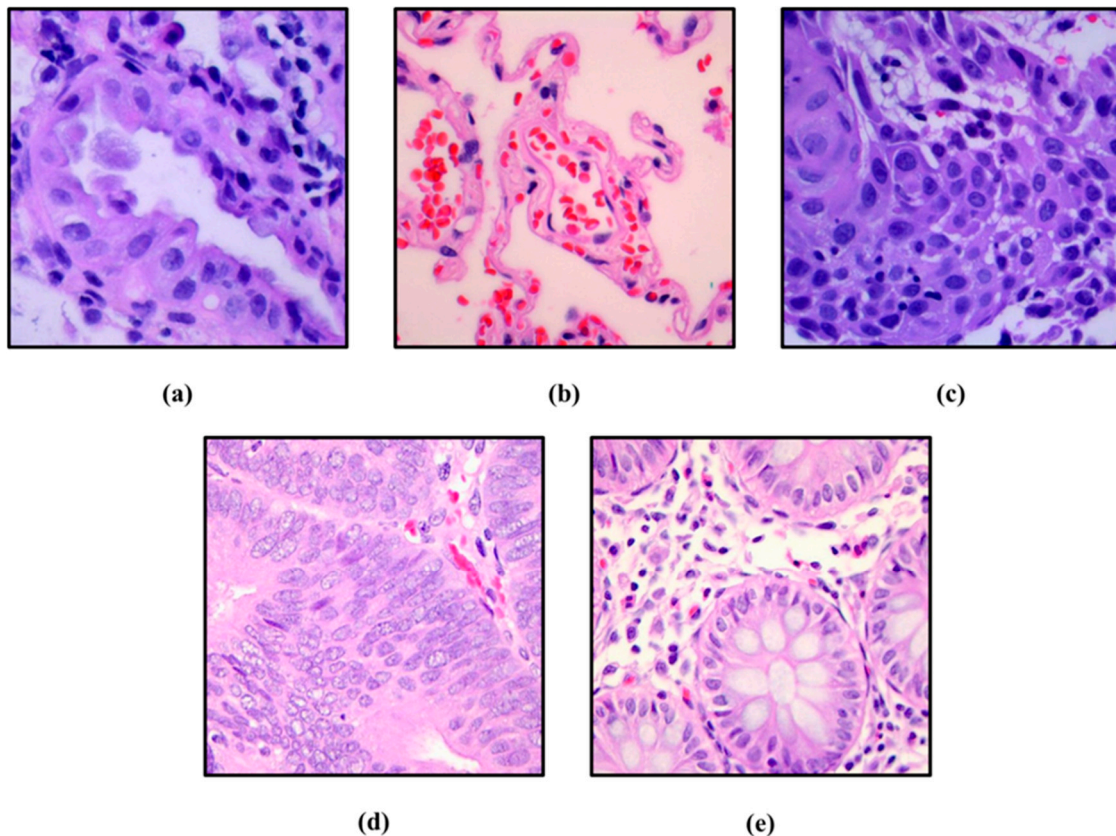
Figure 5. Histopathological images from LC25000 dataset where (**a**) lung adenocarcinoma; (**b**) lung benign; (**c**) lung squamous cell carcinoma; (**d**) colon adenocarcinoma; and (**e**) colon benign.

3.2. Description of the Classes

3.2.1. Lung Adenocarcinoma

Lung adenocarcinoma represents the prevailing form of primary lung cancer observed inside the United States. This particular condition is classified within the category of non-small cell lung cancer (NSCLC) and is closely linked to a history of tobacco smoking. Although there has been a decrease in incidence and mortality rates, cancer continues to be the primary cause of death related to this disease in the United States. Adenocarcinoma of the lung typically arises from the mucosal glands and accounts for approximately 40% of the total cases of lung cancer [33].

3.2.2. Lung Benign

The lung and bronchus encompass a diverse collection of benign tumors, which typically manifest as single, peripheral lung nodules or, less frequently, as endobronchial lesions that result in obstructive symptoms. These tumors commonly occur without any noticeable symptoms. Surgical removal of all endobronchial lesions is recommended to ease symptoms and prevent potential damage to distal lung tissue.

3.2.3. Lung Squamous Cell Carcinoma

Squamous cell carcinoma (SCC) of the lung, alternatively referred to as squamous cell lung cancer, represents a subtype of non-small cell lung cancer (NSCLC). Squamous cell lung cancers frequently manifest in the central region of the lung or the primary airway, specifically the left or right bronchus. Tobacco smoke is well recognized as the primary causal agent responsible for cellular change. The prevalence of smoking-related lung cancer is estimated to be approximately 80% in males and 90% in females [34].

3.2.4. Colon Adenocarcinoma

Colon adenocarcinoma is a type of colorectal cancer that originates from glandular cells that line the inner surface of the colon. It generally arises from pre-existing adenomatous polyps, which can gradually become cancerous. Colon adenocarcinoma is characterized by alterations in bowel patterns, bleeding from the rectum, and discomfort in the abdomen. Screening techniques, such as colonoscopies, are vital for identifying and preventing cancer at an early stage by enabling the elimination of precancerous polyps before they develop into cancerous growths.

3.2.5. Colon Benign

Benign conditions in the colon are characterized by the presence of non-malignant growths or anomalies that do not pose any risk to one's health. Examples include non-malignant polyps, diverticulosis, and inflammatory bowel illnesses such as ulcerative colitis. Although some disorders may not be malignant, they can nonetheless induce symptoms and may necessitate medical attention and treatment.

3.3. Federated Learning (FL)

The introduction of federated learning technology facilitates the training of a model by incorporating a central server while simultaneously maintaining decentralized training data on distributed clients. The objective is to leverage FL technology to ensure the confidentiality of the user data and facilitate data expansion. In a concurrent manner, this approach enables participants to collectively train a global model without the need to share their individual private data, as illustrated in Figure 6. This diagram illustrates a federated learning architecture where multiple decentralized data sources (represented by servers) locally train machine learning models on their own data. The locally trained models are then encrypted and sent to a central server, which aggregates them into a global model without accessing the raw data. This decentralized approach preserves data privacy while improving the global model through collaborative learning. The regional data need to undergo preprocessing for each contributor. This involves making modifications, digitizing, and standardizing the data to transform it into a standardized format while ensuring privacy. The distribution of images from our dataset among various clients is achieved by dividing the total size of the image by the number of clients. The dataset is divided uniformly, resulting in the generation of independent and identically distributed (IID) data.

After receiving the model parameters from the clients, the server will summarize the information on the basis of the structure of the central server. It updates the parameters of the existing model and stores it for the subsequent round of training parameter upload and collection from the participants before it is redistributed. The initial FL iterative procedure is subsequently followed by the iterative process of our comprehensive model. In this process, a CNN was integrated by tailoring for cancer tissue data samples and modifying the model to enable continuous iterations. In the context of dispersed machine learning, the prevailing approach for aggregating models involves ensuring that all participants possess an equal number of training samples. In the context of federated learning, it is common for participating members to have varying quantities of data at their disposal. The aggregation of local models is achieved by applying a weighting scheme based on the quantity of training samples available for each model. Consequently, the models that possess a greater quantity of samples are given preference over those with a limited number. The aforementioned method has a straightforward nature, yet it has demonstrated a notable level of prevalence and effectiveness in scenarios involving federated learning. The global model parameter can be defined as

$$\omega_{t+1} \leftarrow \sum_{k=1}^{k} \frac{n_k}{n} \omega_{t+1}^k \qquad (1)$$

where

k = participants;

n_k = samples of participants k;

n = samples of all participants;

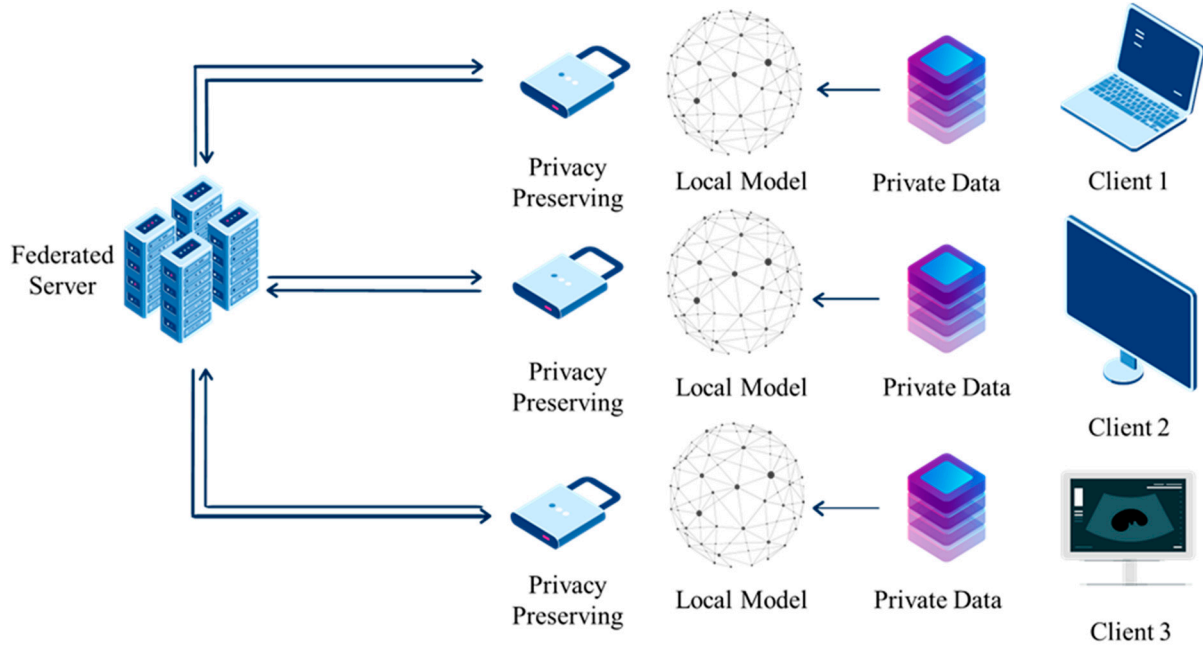w_{t+1} = local parameter of k participants.

Figure 6. Workflow of federated learning within a federated server to a local client.

A federated round is the name given to each iteration of this process, and it consists of concurrent training, update aggregation, and parameter distribution. The following are the primary control parameters that are utilized in the process of computing FL:

C = customers or contributors who took part in an update cycle;

E = number of local epochs that each contributor has been responsible for;

B = smallest batch size that can be utilized for a local update.

β_1 and β_2 are considered hyperparameters.

During the regional model optimization step, the clients execute a specific number of local epochs. The Adam optimizer is utilized for the first- and second-order moments to overcome local minima [35]:

$$\omega_{i,t} \leftarrow \omega_{i,t} - \eta \frac{\sqrt{1 - \beta_2^n}}{1 - \beta_1^n} \times \frac{m_{i,n}}{\sqrt{v_{i,n}} + \sigma} \tag{2}$$

3.4. Inception-V3 Model

Inception-V3 is an advanced deep convolutional neural network (CNN) architecture that has been meticulously crafted by Google as an integral component of the Inception project [36]. The algorithm has been specifically developed to cater to a range of computer vision applications, with a particular focus on image classification and object recognition by enabling both efficiency and accuracy. The exceptional performance and efficiency of Inception-V3 are attributed to its utilization of a distinctive inception module, which integrates multiple filter sizes into a single layer, enabling the capture of features at various scales.

The primary technological advancement of Inception-V3 lies within its inception modules, which encompass various building blocks, such as convolutions, average pooling, max pooling, concatenations, dropouts, and fully connected layers. Figure 7 represents the architecture of Inception-V3. The architecture is composed of multiple modules, including

convolutional layers, pooling layers, and inception modules that combine various filter sizes (e.g., 1×1, 3×3, and 5×5) to capture different types of image features. The model extensively utilizes batch normalization, which is applied to activation inputs. The computation of loss involves the utilization of Softmax. The Inception-V3 model also includes the integration of global average pooling, a technique that replaces the conventional fully connected layers located at the final stage of the network. By reducing overfitting and parameterizing the model size, the efficiency of the system is enhanced and its adaptability to new tasks is improved.
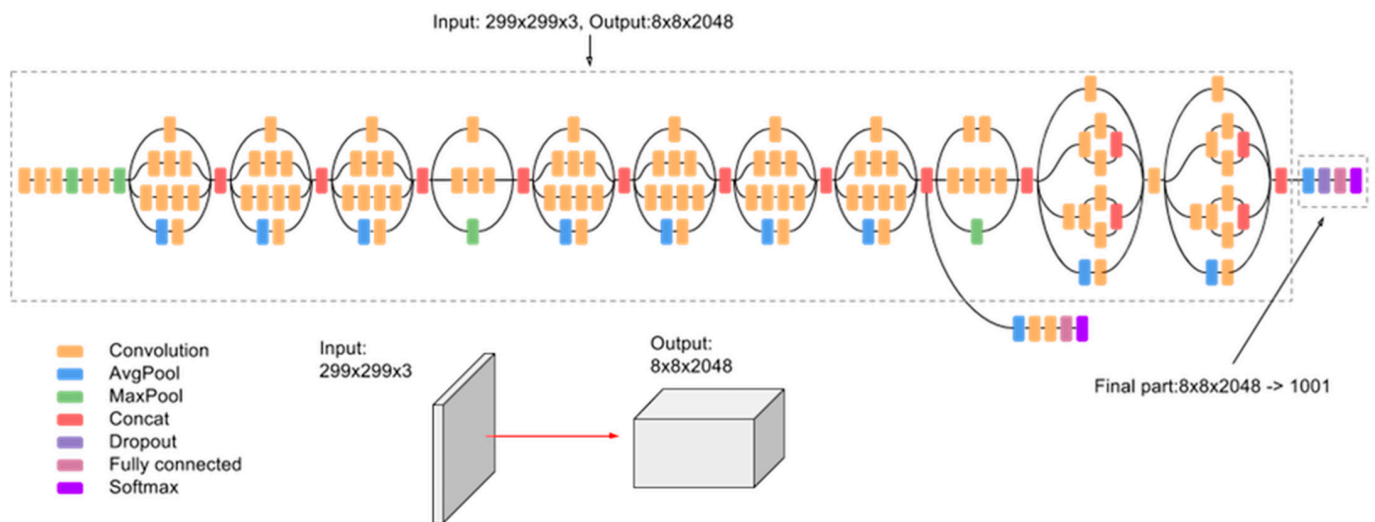

Figure 7. Inception-V3 architecture [36].

3.5. Proposed Workflow for FL

This section provides an understanding of various implementations of federated learning. This encompasses the procedure of training data on individual client devices and subsequently aggregating the local models on a central server.

3.5.1. Local Device (Client) Creation

Benign conditions in the colon are characterized by the presence of non-malignant growths or anomalies that do not pose any risk to one's health. Examples include non-malignant polyps, and diverticulosis.

Data distribution among end devices is achieved through a technique called data sharding. This process involves the distribution of smaller datasets, referred to as logical shards or chunks, from a larger dataset. In the context of a real-world application, it is important to note that each client may possess datasets of varying sizes and variations. However, for developing the federated learning (FL) prototype, the shards among the clients in the model were evenly distributed. The size of each shard was determined via the following formula:

$$\text{Shard Size} = \frac{Total\ No.\ of\ Images}{No.\ of\ Clients} \tag{3}$$

The data shards subsequently underwent processing and were allocated to the clients via the batching procedure. Upon conclusion of the process, each client obtained their respective local datasets and prepared them for training purposes.

3.5.2. Integration of Inception-V3 and Its Configuration

As previously stated, the local datasets underwent training on local devices. The Inception-V3 architectural model was implemented on each client to facilitate the training of their local data. Additionally, it was utilized in the global model for testing purposes. Additionally, three additional convolutional neural network (CNN) layers were incorporated

into the implementation. These layers included the dense layer, flatten layer, and dropout layer. The ImageNet dataset was utilized as the default weight in the model, similar to the implementation observed in the CNN. The Adam optimizer with a learning rate of 0.00001 was used to optimize the accuracy of the Inception-V3 model. Additionally, the categorical cross-entropy method was implemented to calculate the loss function.

3.5.3. Communication Rounds

The communication round was executed a total of 50 times to achieve a specific level of accuracy for our global model. The communication cycle involved the clients, which hold the datasets previously generated and acquire their weights on the basis of the global model's weight. Each of the five clients then proceeded to train their respective local data and generate the accuracy of their individual local models. The clients subsequently proceeded to transmit their trained models to the central server, which is commonly referred to as the global model, for the purpose of aggregation. This aggregation process involves performing an averaging operation on the federated learning (FL) model. After the initial aggregation, the global model generated a new weight. In the subsequent round, the clients adjusted their weights on the basis of the updated global weight. The weight scaling factor was incorporated into the model to facilitate this particular operation.

3.6. Experimental Setup and Hyperparameter Settings

The experiments are performed on Google Colab, and the hyperparameters of the experiments are shown in Table 3. The proposed federated learning model was run on an NVIDIA A100 40 GB GPU and 85 GB RAM provided by Google Colab services. The images for the original dataset had dimensions of 100×100 and were trained in this environment. The hyperparameter settings were set to achieve the maximum computational capacity for the best possible performance on the whole dataset. The hyperparameters were optimized through trial and error, as shown in Table 4.

Table 3. Description of the system configuration of Google Colab.

System	Specification
Processor	Intel Xeon CPU
CPU	~2.30 GHz
RAM	85 GB
GPU	NVIDIA A100
GPU RAM	40 GB
Hard Disk	80 GB

Table 4. Description of system hyperparameters.

Hyperparameter	Value
Optimizer	Adam
Loss	Categorical Crossentropy
Batch Size	16
Image Size	100×100
No. of Epochs	50
No. of Clients	5

3.7. Evaluation Metrics

Several evaluation criteria were used to evaluate the performance of our proposed model. They are the accuracy, precision, recall, F1 score, specificity, and confusion matrix. The metric commonly used to evaluate the performance of machine learning models is accuracy. The metric reflects the frequency at which the model accurately predicts the positive class. Precision is the proportion of accurate forecasts. It is the ratio of the number of accurate positive forecasts to the entire number of positive predictions. Recall indicates

the proportion of accurate predictions relative to the ground truth. The F1 score takes the harmonic mean of both precision and recall to create a single metric. The respective formulas are presented here.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1\_score = \frac{2 \times (Precision \times Recall)}{Precisin + Recall} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

where *TP* = True Positive, *TN* = True Negative, *FP* = False Positive, and *FN* = False Negative.

4. Experimental Results

The suggested technique is evaluated via histopathology images obtained from the LC25000 dataset [32]. This section presents a concise overview of the outcomes obtained by the proposed federated learning model when provided with Inception-V3 for the categorization of histological images from the lung and colon cancer datasets. The results were categorized into three sections: lung, colon, and combined lung–colon outcomes for the model.

Every experiment involved the training of five distinct transfer learning models, namely, Inception-V3, VGG16, ResNet-50, ResNeXt50, and Xception. Initially, transfer learning models were utilized to train the initial lung cancer images to evaluate the models' performance on the dataset. A comparative analysis of the outcomes of the models was conducted to determine which model was more suitable for the federated learning approach.

4.1. Lung Cancer

Table 5 and Figure 8 present a concise overview of the performance of the federated learning model in comparison to various base models when it is applied to the lung cancer images. The Inception-V3, VGG16 [37], ResNet50 [38], ResNeXt50 [39], and Xception [40] models achieved average classification accuracies of 99.16%, 98.33%, 99.20%, 99.20%, and 99.27%, respectively. The Inception-V3 and Xception models exhibited the best accuracy, with a precision and recall of 99.27%. Conversely, the VGG16 model had the lowest performance. Inception-V3 displayed a commendable performance, approaching the outcomes achieved by Xception. Upon integrating the federated learning model with Inception-V3, it became clear that it significantly outperformed all the other findings. The model attained 99.87% accuracy, with a precision and recall of 1.0.

Table 5. Performance analysis of the lung cancer images compared to the base models.

Classification Model	Precision	Recall	Accuracy
"Federated Learning with Inception-V3"	1.0	1.0	99.87%
Inception-V3	0.9916	0.9916	99.16%
VGG16	0.9833	0.9833	98.33%
ResNet-50	0.9926	0.992	99.20%
ResNeXt50	0.992	0.9927	99.20%
Xception	0.9927	0.9927	99.27%

LUNG CANCER PERFORMANCE ANALYSIS

	Federated Learning with Inception-V3	Inception-v3	VGG16	ResNet-50	ResNeXt50	Xception
Precision	1	0.9916	0.9833	0.9926	0.992	0.9927
Recall	1	0.9916	0.9833	0.992	0.9927	0.9927
Accuracy	99.87%	99.16%	98.33%	99.20%	99.20%	99.27%

Figure 8. Performance comparisons of federated learning with Inception-V3 with other centralized transfer learning models, i.e., Inception-V3, VGG-16, ResNet-50, ResNeXt-50, and Xception, on lung cancer images.

Figure 9 displays the confusion matrix generated by the proposed technique for lung cancer. The data in Table 6 clearly indicate that the proposed model achieved 100% accuracy in detecting squamous cell carcinoma (lung_scc) and benign lung (lung_bnt) images. Additionally, the model demonstrated 99.60% accuracy in correctly identifying lung adenocarcinoma (lung_aca) images.

Figure 9. Confusion matrix for federated learning with Inception-V3 for lung cancer classification.

Table 6. Class-wise performance analysis of the lung cancer images using federated learning with Inception-V3.

Type of Class	Precision	Recall	F1 Score	Specificity	Accuracy
lung_aca	99.60%	100.00%	99.80%	99.80%	99.60%
lung_bnt	100.00%	100.00%	100.00%	100.00%	100.00%
lung_scc	100.00%	100.00%	100.00%	100.00%	100.00%
Macro Average	99.87%	100.00%	99.93%	99.93%	99.87%

4.2. Colon Cancer

Table 7 and Figure 10 provide a comprehensive overview of the performance of the federated learning model in comparison to other base models when analyzing colon cancer images. The Inception-V3, VGG16, ResNet50, ResNeXt50, and Xception models achieved average classification accuracies of 100%, 99.60%, 99.70%, 100%, and 100%, respectively. The Xception and ResNeXt models achieved the highest accuracy, precision, and recall rates of 100%, 1.0, and 1.0, respectively, whereas the VGG16 model achieved the lowest performance. Inception-V3 demonstrated strong performance, comparable to the results achieved by Xception. After applying our federated learning model with Inception-V3, it became clear that it significantly outperformed all the other methods. The achieved results include an accuracy rate of 100%, precision score of 1.0, and recall score of 1.0.

COLON CANCER PERFORMANCE ANALYSIS

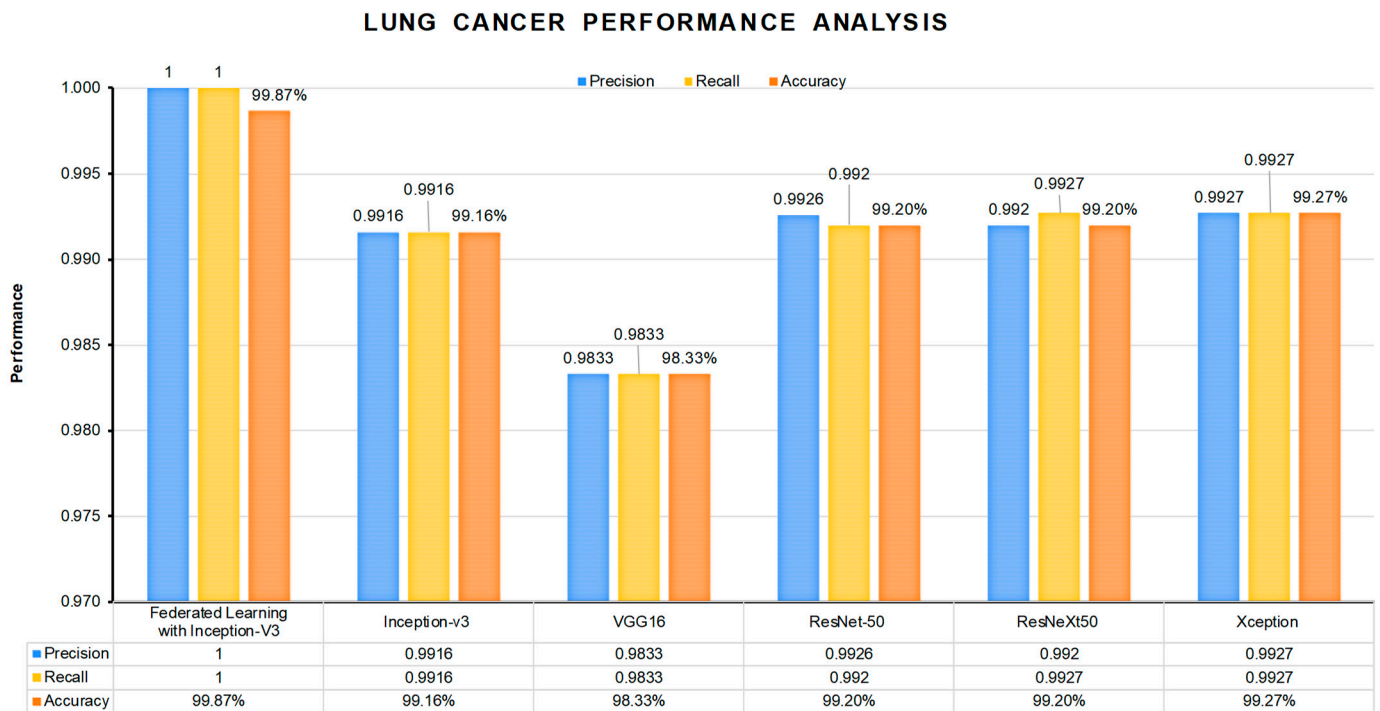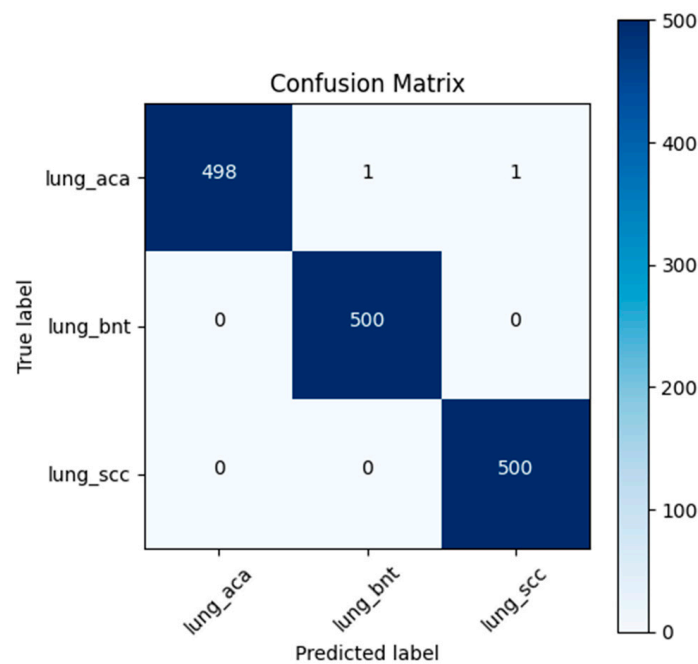|	Federated Learning with Inception-V3	Inception-v3	VGG16	ResNet-50	ResNeXt50	Xception
Precision	1	0.996	0.994	0.997	1	1
Recall	1	0.996	0.994	0.997	1	1
Accuracy	100.00%	99.60%	99.40%	99.70%	100.00%	100.00%
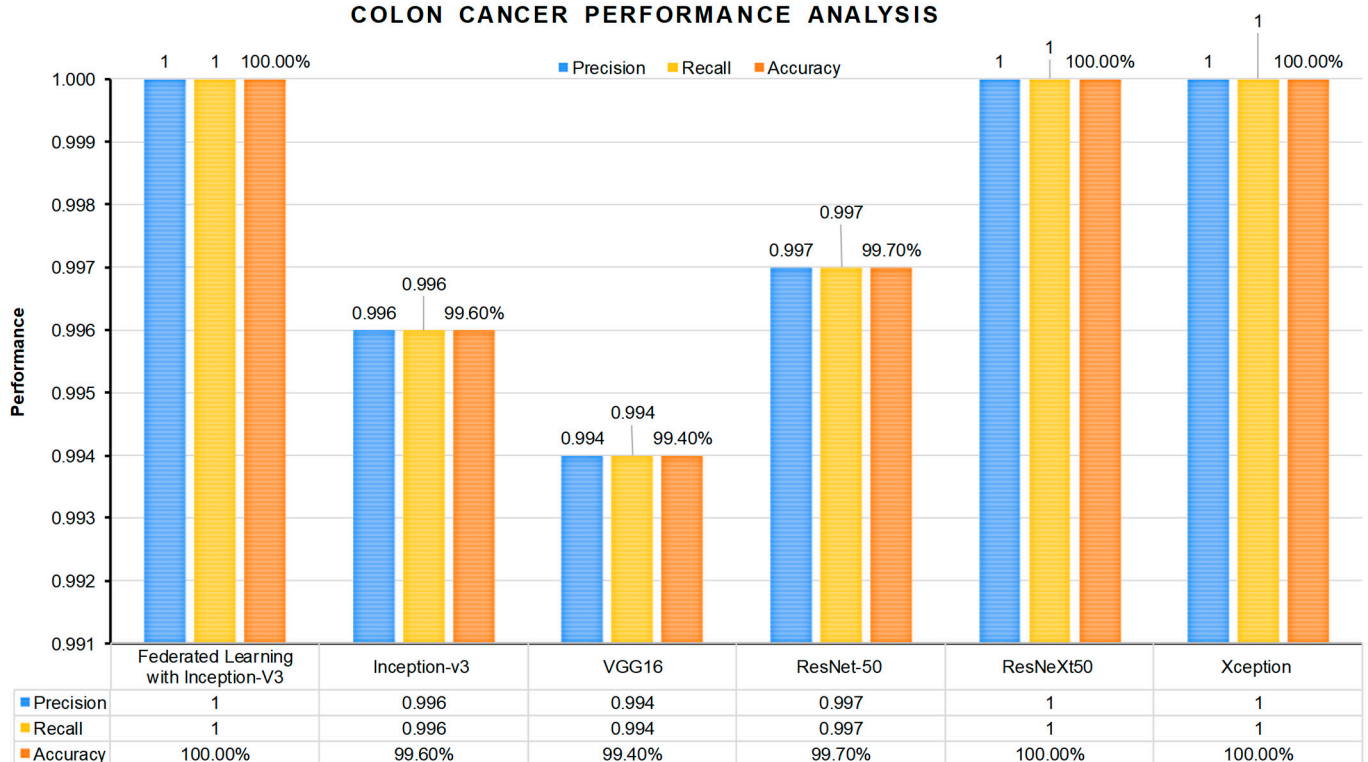
Figure 10. Performance comparisons of federated learning with Inception-V3 with other centralized transfer learning models, i.e., Inception-V3, VGG-16, ResNet-50, ResNeXt-50, and Xception, on colon cancer images.

Figure 11 displays the confusion matrix of the proposed method for colon cancer. On the basis of the data presented in Table 8, the proposed model clearly exhibited a remarkable ability to accurately detect colon adenocarcinoma (colon_aca) and colon benign (colon_bnt) images, achieving a 100% accuracy rate.

Table 7. Performance analysis of colon cancer images compared to the base models.

Classification Model	Precision	Recall	Accuracy
"Federated Learning with Inception-V3"	1.0	1.0	100.00%
Inception-V3	0.996	0.996	99.60%
VGG16	0.994	0.994	99.40%
ResNet-50	0.997	0.997	99.70%
ResNeXt50	1.0	1.0	100.00%
Xception	1.0	1.0	100.00%
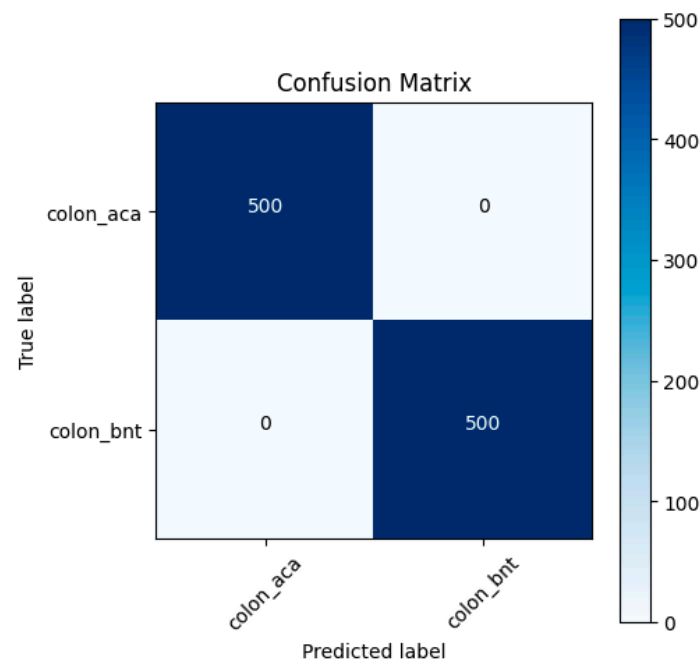
Figure 11. Confusion matrix for federated learning with Inception-V3 for colon cancer classification.

Table 8. Class-wise performance analysis of the colon cancer images using federated learning with Inception-V3.

Type of Class	Precision	Recall	F1 Score	Specificity	Accuracy
colon_aca	100%	100%	100%	100%	100%
colon_bnt	100%	100%	100%	100%	100%
Macro Average	100%	100%	100%	100%	100%
Micro Average	100%	100%	100%	100%	100%

4.3. Lung and Colon Cancers

Table 9 and Figure 12 provide a comprehensive overview of the performance of the federated learning model in comparison to other base models on lung and colon cancer images. The Inception-V3, VGG16, ResNet50, ResNeXt50, and Xception models achieved average classification accuracies of 98.96%, 98.36%, 98.88%, 98.88%, and 99.10%, respectively. The Xception model demonstrated the highest accuracy, precision, and recall, all at 99.10%. Conversely, the VGG16 model exhibited the lowest performance. Inception-V3 demonstrated a strong performance, comparable to the results achieved by Xception. After implementing the federated learning model with Inception-V3, it became clear that it significantly outperformed all the other methods. The achieved accuracy, precision, and recall were all 99.72%.

Table 9. Performance analysis of lung and colon cancer images compared to the base models.

Classification Model	Precision	Recall	Accuracy
"Federated Learning with Inception-V3"	99.72%	99.72%	99.72%
Inception-V3	98.96%	98.96%	98.96%
VGG16	98.36%	98.36%	98.36%
ResNet-50	98.96%	98.84%	98.88%
ResNeXt50	98.88%	98.88%	98.88%
Xception	99.10%	99.10%	99.10%

Figure 12. Performance comparisons of federated learning with Inception-V3 with other centralized transfer learning models, i.e., Inception-V3, VGG-16, ResNet-50, ResNeXt-50, and Xception, on combined lung and colon cancer images.

Figure 13 displays the confusion matrix of the proposed method for lung and colon cancers. On the basis of the data presented in Table 10, the proposed model exhibited a high level of accuracy in detecting benign lung (lung_bnt) images, with a 100% accuracy rate. Similarly, the model demonstrated 99.72% accuracy in correctly identifying lung adenocarcinoma (lung_aca) and squamous cell carcinoma (lung_ssc) images. For the colon cancer images, 100% accuracy was achieved for both classes.

Table 10. Class-wise performance analysis of the lung and colon cancer images using federated learning with Inception-V3.

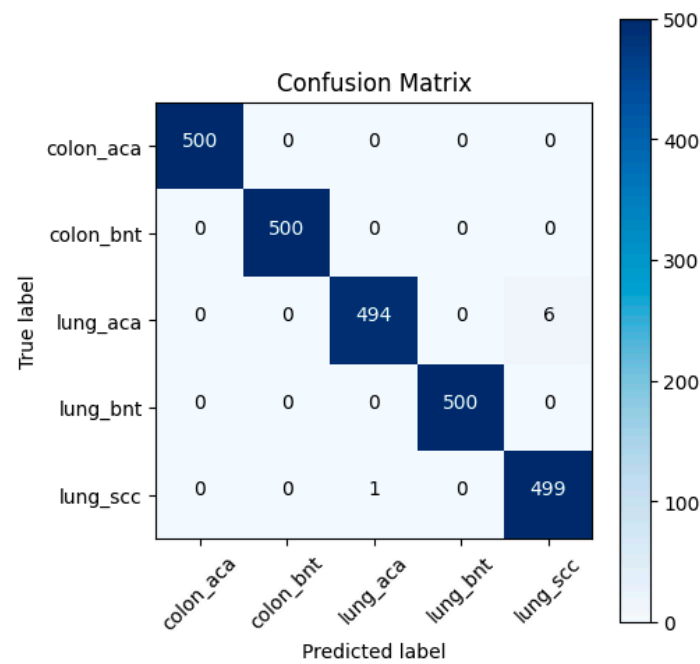| Type of Class	Precision	Recall	F1 Score	Specificity	Accuracy
colon_aca	100%	100%	1.000	100%	100%
colon_bnt	100%	100%	1.000	100%	100%
lung_aca	98.80%	99.80%	0.999	99.70%	99.72%
lung_bnt	100%	100%	1.000	100%	100%
lung_scc	99.80%	98.812%	0.993	99.95%	99.72%
Macro Average	99.72%	99.72%	0.9984	99.93%	99.88%
Micro Average	99.72%	99.72%	99.72%	-	99.72%

Figure 13. Confusion matrix for federated learning with Inception-V3 for combined lung and colon cancer classifications.

4.4. Client-Wise Results

Figure 14 illustrates the iterative process of updating and optimizing the global model following each communication event and how the clients' individual accuracies vary from the global accuracy. After a total of 50 communication rounds, the accuracy reached 99.867% for lung cancer classification, 100% for colon cancer classification, and 99.72% for lung and colon cancer classifications. In the local context, each communication round encompassed comprehensive training sessions for all clients. To facilitate the simulation, a single epoch was conducted for the purpose of local communication rounds. Following the completion of the simulation, accuracy, loss, and categorical accuracy metrics for all of the esteemed clients were successfully acquired. Despite not being immediately evident, the findings indicate a positive upward trend (thicker red lines). This implied that there was a possibility of an increase in the accuracy of the client, although it cannot be guaranteed.

It was evident that, in accordance with the initial predictions, the clients demonstrated increased accuracy with each successive communication round. In the client-wise accuracy measure, there was a steady increase in accuracy, indicating that performance improved with each communication round.

4.5. Explainable AI (XAI)

XAI methods [41], such as Grad-CAM, heatmaps, and saliency maps, were implemented to explore their usefulness in the classification of lung and colon cancers. Grad-CAM [42] was utilized in deep learning to visualize the key regions of an input image that had the greatest impact on the output of a convolutional neural network (CNN) model. The heatmap demonstrated the correlation between the image's features and the model's prediction, highlighting the significant areas that played a role in determining the final prediction. As saturation increases, the weight assigned to those pixels also increases. Figure 15 presents a clear visualization of the significant features (pixels) in the image that the model deemed crucial for its prediction. The saliency map was generated by calculating the gradients of the predicted class score with respect to the input image pixels. It is a method similar to the heatmap. The algorithm identifies the pixels in the input image that have the greatest impact on the predicted class score.
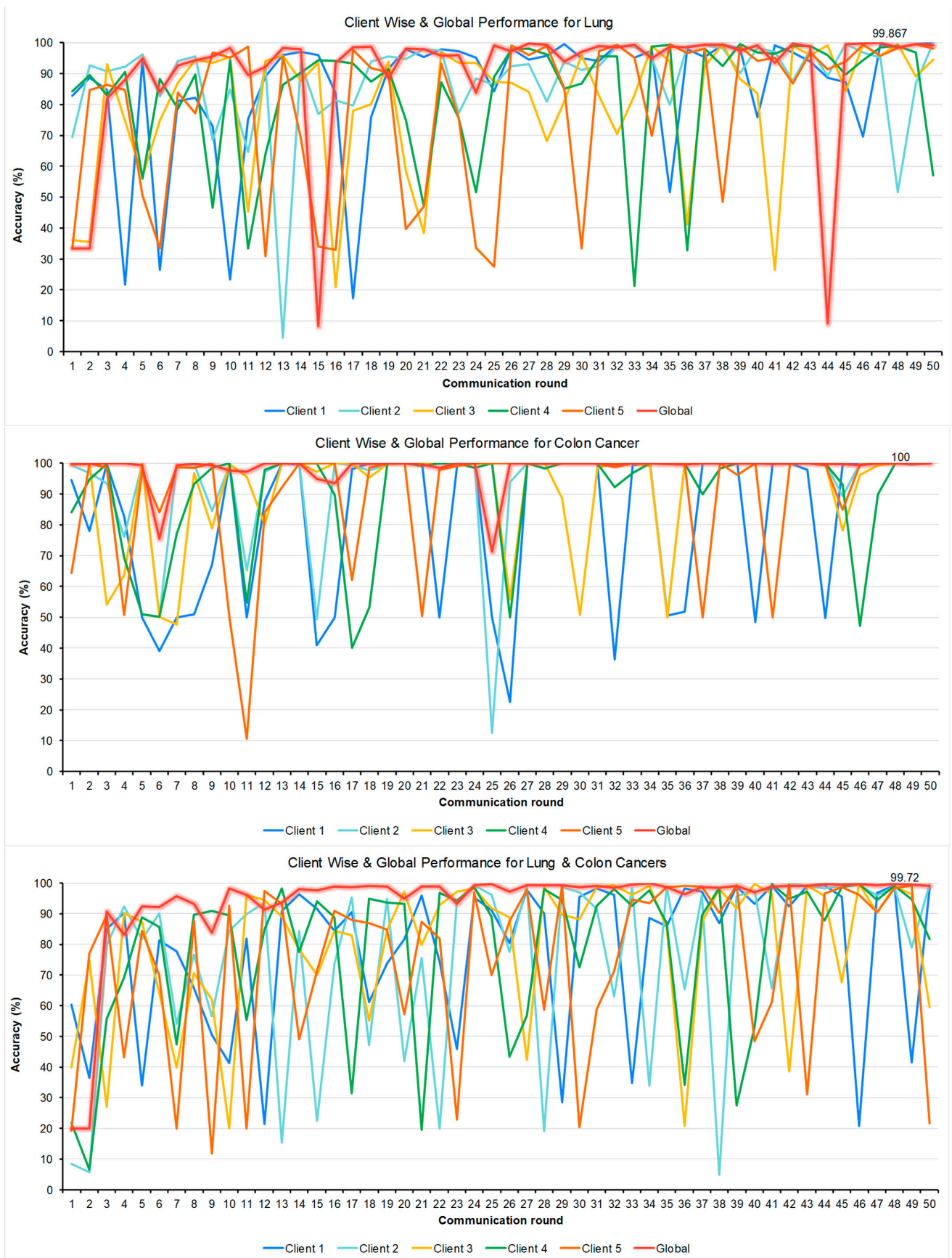
Figure 14. Global accuracy vs. local accuracy for the combination of the lung, colon and lung–colon methods.
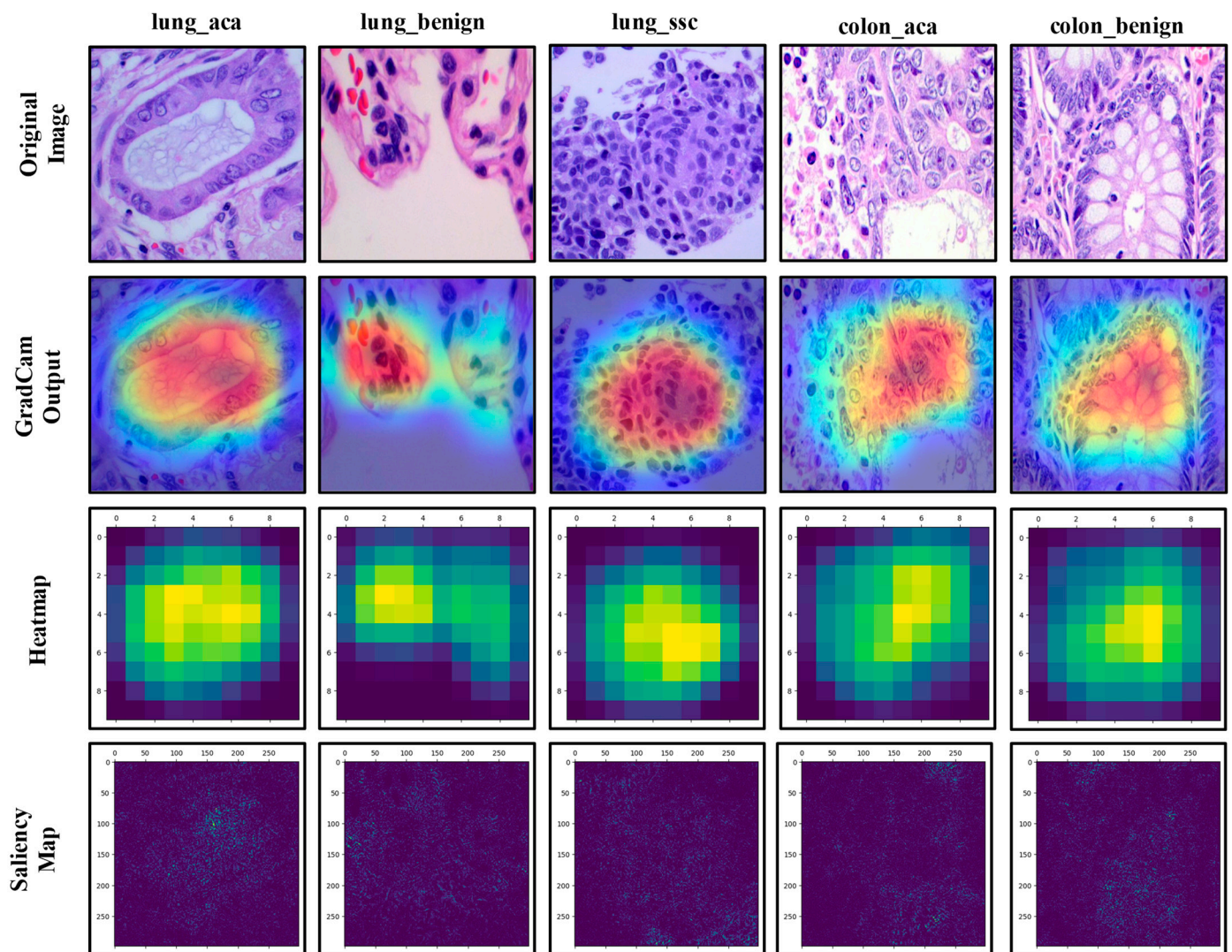
Figure 15. Attention visualization of images employing GradCAM, heatmaps, and saliency maps.

Grad-CAM determines the gradients of the target class in relation to the last convolutional layer of the model. Grad-CAM uses gradients to identify the influential regions of an image that contribute to the classification decision. The heatmap illustrates the significance of each pixel in contributing to the model's classification decision. The color red represents regions that have greater significance for the classification, whereas cooler colors such as blue indicate regions that are less important. Saliency maps are generated by calculating the gradient of the predicted class score of the model with respect to the pixels of the input image. The purpose of this technique is to identify the regions in the input image that have the greatest impact on the model's output. The images were chosen randomly from the dataset and may behave differently from one image to the other.

By examining the GRAD-CAM output and heatmaps, it was observed that the classifier distinguishes lung adenocarcinoma by focusing on cell near-white regions, which are indicative of mucosa or connective tissue. For lung squamous cell carcinoma, the classification is based on the fish-scale appearance of the cells under the microscope. In healthy lung tissue (lung_benign), the classifier identifies red blood cells as a key distinguishing feature. For colon adenocarcinoma, the classifier relies on areas with irregular glandular structures, atypical cell shapes, and disrupted tissue architecture, highlighting the presence of desmoplastic stroma and inflammatory cell infiltrates. In contrast, benign colon tissue is characterized by regular, well-organized glandular structures and uniform cell shapes.

5. Discussion

5.1. Comparative Analysis

To demonstrate the effectiveness of the proposed method, a comparative analysis was conducted with previous studies that utilized the LC25000 dataset. There are very few studies that are exclusively concerned with lung and colon malignancies, and the majority of them employed CNN-based centralized deep learning techniques. The methodologies discussed here do not ensure data privacy, as they necessitate access to all training images and their corresponding classification labels. The proposed methodology not only outperforms the most advanced methods for detecting lung and colon cancers separately and shows comparable performance for detecting lung and colon cancers combinedly but also maintains the privacy element of patient data (Table 11).

Table 11. Comparative analysis of previous studies on LC25000.

Previous Studies	Year	Approaches	Performance		
			Colon	Lung	Lung and Colon
Mangal et al. [9]	2020	Deep learning approach using CNN	Accuracy: 96.00%	Accuracy: 97.89%	-
Tasnim et al. [43]	2021	CNN with max pooling	Accuracy: 99.67%	-	-
Talukder et al. [16]	2021	Deep feature extraction and ensemble learning	Accuracy: 100%	Accuracy: 99.05%	Accuracy: 99.30%
Shandilya et al. [44]	2021	Pretrained CNN	-	Accuracy: 98.67%	-
Hadiyoso et al. [10]	2022	VGG-19 architecture and CLAHE framework	Accuracy: 98.96%	-	-
Karim et al. [45]	2022	Extreme learning machine (ELM)-based DL	-	Accuracy: 98.07%	-
Raju et al. [46]	2022	Extreme learning machine (ELM)-based DL	Accuracy: 98.97% Precision: 98.87% F1 Score: 98.84%	-	-
Chehade et al. [8]	2022	XGBoost	Accuracy: 99.00% Precision: 98.6% F1 Score: 98.8%	Accuracy: 99.53% Precision: 99.33% F1 Score: 99.33%	Accuracy: 99%
Ren et al. [47]	2022	Deep convolutional GAN (LCGAN)	-	Accuracy: 99.84% Precision: 99.84% F1 Score: 99.84%	-
Mehmood et al. [13]	2022	Transfer learning with class selective image processing	-	-	Accuracy: 98.4%
Khan et al. [30]	2023	Transfer learning with a secure IoMT-based approach	-	-	Accuracy: 98.80%
Toğaçar et al. [14]	2022	DarkNet-19 model and SVM classifier	-	-	Accuracy: 99.69%
Attallah et al. [48]	2022	CNN features with transformation methods	-	-	Accuracy: 99.6%
Masud et al. [7]	2022	Deep learning (DL) and digital image processing (DIP) techniques	-	-	Accuracy: 96.33% Precision: 96.39% F1 Score: 96.38%
Al-Jabbar et al. [17]	2023	Fusion of GoogleNet and VGG-19	-	-	Accuracy: 99.64% Precision: **100%**
Ananthakrishnan et al. [18]	2023	CNN with an SVM classifier	Accuracy: 99.8%	Accuracy: 98.77%	Accuracy: **100%**
Proposed Model	2024	Federated learning with Inception-V3	Accuracy: **100%** Precision: **100%** F1 Score: **100%**	Accuracy: **99.87%** Precision: **99.87%** F1 Score: **99.87%**	Accuracy: 99.72% Precision: 99.72% F1 Score: **99.72%**

Note: Bold numerical values indicate best results.

5.2. Strengths of the Proposed Model

The initial approach involves the use of the Inception-V3 model to process local models on individual client devices, leveraging the data specific to each device. However, it is important to note that individual devices retain ownership of their respective local data. The second process involves aggregating local models to generate a global model with enhanced accuracy. The global server receives only the parameters and weights. Both centralized training and decentralized training are conducted with identical hyperparameter settings. In the centralized case, the models are trained directly on the overall training data with 50 epochs. During the FL experiment, the training processes of five clients are stuck with their local data, each with 50 communication rounds. This component is an integral part of the federated learning system and is designed specifically to address and mitigate privacy concerns.

Both lung cancer and colon cancers are prevalent types of cancers with high mortality rates. Identifying a condition at an early stage significantly improves the chances of survival. Physicians may face challenges in precisely identifying lung cancer with CT images. Sharing patient data for research purposes among health organizations is a growing concern because of the numerous restrictions in controlling patient data. Adopting suitable data-sharing techniques can increase the likelihood of detecting cancer through CT scans.

In this context, we introduce the proposed federated learning methodology, which can effectively guarantee adherence to regulations while addressing patient data and precisely categorizing cancer cells with great accuracy.

The federated averaging technique aggregates model updates from multiple clients to compute a global model. This aggregation process combines knowledge from diverse sources while mitigating biases and noise present in individual client datasets. As a result, the global model benefits from the collective intelligence of all participating clients, leading to enhanced performance in cancer classification tasks. The framework used here hosts five clients to demonstrate the federated learning process, but it is possible to scale efficiently to a large number of devices. Each device performs local computations, and only model updates are aggregated centrally, reducing the burden on central servers.

The utilization of federated learning techniques for lung and colon cancer classifications represents a significant advancement with important clinical implications. Through the utilization of decentralized data sources from various healthcare institutions, federated learning models present unique possibilities to improve diagnostic accuracy and customize treatment strategies. By engaging in collaborative data sharing while prioritizing patient privacy, these models have the ability to identify complex patterns within different types of cancers. This allows for personalized interventions and well-informed clinical decision making. Real-time decision support systems, powered by federated learning algorithms, enable healthcare providers to gain timely insights, facilitating proactive management and enhancing patient outcomes. In addition, the iterative process of federated learning allows for the ongoing improvement of models, ensuring their ability to adapt to changing clinical practices and contributing to advancements in precision oncology and population health management. Federated learning is a revolutionary method in cancer care that promotes cooperation, creativity, and advancement in the pursuit of better patient care and results.

5.3. Challenges in Federated Learning for Medical Diagnostics

Federated learning presents unique challenges in medical diagnostics, particularly when applied to lung and colon cancer classifications. One of the foremost difficulties is data heterogeneity across healthcare institutions. Medical images can vary significantly on the basis of equipment used, imaging protocols, and patient demographics. This variation can cause discrepancies in local model performance, making it difficult to achieve consistent convergence of the global model. Ensuring that the federated model can be generalized effectively across diverse imaging environments while maintaining high diagnostic accuracy remains a critical challenge.

Another challenge is the increased communication overhead inherent in federated learning. Unlike traditional centralized models, where all data are aggregated in one place, federated learning requires frequent communication between institutions to share model updates without sharing raw data. This decentralized nature can lead to slower global model updates, especially when dealing with large datasets typical of medical imaging. Additionally, network connectivity issues or differences in computational resources across institutions can further delay the training process and affect the overall performance of the model.

An additional challenge in federated learning for medical diagnostics is the reluctance of healthcare institutions to share data, even in a decentralized framework. Despite the promise of privacy-preserving methods, many clinics are hesitant to participate because of concerns about data ownership, security, and compliance with stringent regulations such as the HIPAA or GDPR. Institutions often worry about losing control over their data or exposing themselves to potential breaches or misuse, which makes collaboration difficult. Overcoming this challenge requires building trust among participants through robust legal agreements, transparent protocols, and guarantees that ensure data sovereignty while still allowing for effective collaboration in training the global model.

5.4. Limitations and Future Work

One major limitation of the model is that it cannot handle heterogeneous data as IID data were chosen for this study. However, user data can vary significantly from institution to institution. This means that the data distribution varies significantly across different devices, leading to challenges in model convergence and performance consistency. The non-independent and identically distributed nature of data across clients can hinder model convergence and lead to suboptimal global model performance. This needs to be implemented to test the capability of the proposed model.

This study implemented federated learning to classify lung and colon cancer images on IID data. The entire LC25000 dataset was divided among clients. However, in implementing federated learning across multiple organizations, there is an enormous challenge due to variations in data distribution and quality across participating institutions. Addressing data heterogeneity is essential for improving the robustness of the federated learning model. The size of the federated learning dataset might have been constrained by the number of participating institutions. Expanding the dataset or exploring methods for effective knowledge transfer with a limited sample size could enhance model generalizability.

Despite efforts to ensure privacy through federated learning, concerns regarding the security of patient data transmission persist. Individuals can employ various techniques to reconstruct the original data from these updates, such as model inversion or gradient leakage. Further exploration of advanced privacy-preserving techniques, such as secure aggregation or differential privacy, is warranted. With that in mind, diverse and heterogeneous data from various sources, such as multi-modal imaging (CT and PET), genetic data, and clinical records, could be incorporated to increase generalizability. Advanced privacy-preserving techniques, including homomorphic encryption or federated learning with differential privacy, are researched and implemented to strengthen the security and confidentiality of patient data during the federated learning process. The feasibility of real-time implementation of federated learning models for cancer classification is particularly valuable for a timely diagnosis and treatment planning in clinical settings.

6. Conclusions

The proposed federated learning approach with the Inception-V3 model to classify lung and colon histopathological images yielded a significant outcome in accurately distinguishing between three subtypes of lung cancer and two subtypes of colon cancer from histopathological images. The model demonstrated a remarkable accuracy of 99.720%, as well as a recall, a precision, and an F1 score of 99.720%. The detection of colon cancer achieved 100% accuracy in classifying both classes. The accuracy of lung cancer classification for the three classes was 99.867%. This proves that the proposed federated learning methodology can effectively guarantee adherence to regulations while dealing with patient data and precisely categorizing cancer cells with great accuracy.

Author Contributions: All authors had an equal contribution in preparing and finalizing the manuscript. Conceptualization: M.M.H., M.R.I., M.F.A., M.A. and J.H.; methodology, M.M.H., M.R.I., M.F.A., M.A. and J.H.; validation: M.M.H., M.R.I., M.F.A., M.A. and J.H.; formal analysis: M.M.H., M.R.I., M.F.A., M.A. and J.H.; investigation: M.M.H., M.R.I., M.F.A., M.A. and J.H.; data curation: M.M.H., M.R.I. and M.F.A.; writing—original draft preparation: M.M.H., M.R.I. and M.F.A.; writing—review and editing: M.M.H., M.R.I., M.F.A., M.A. and J.H.; supervision: M.R.I., M.A. and J.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following table alphabetically lists all the acronyms along with their definitions.

Acronym	Stands for
ANN	Artificial Neural Network
CAD	Computer-aided Diagnosis
CNN	Convolutional Neural Network
CT	Computed Tomography
colon_aca	Colon Adenocarcinoma
colon_bnt	Colon Benign
DaaS	Data as a Service
DL	Deep Learning
DT	Decision Tree
ELM	Extreme Learning Machine
FL	Federated Learning
FedAvg	Federated Averaging
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
IID	Independent and Identically Distributed
IoMT	The Internet of Medical Things
LCC	Large Cell Carcinoma
lung_aca	Lung Adenocarcinoma
lung_bnt	Lung Benign
lung_scc	Lung Squamous Cell Carcinoma
MRI	Magnetic Resonance Imaging
NSCLC	Non-small Cell Lung Cancer
RF	Random Forest
SCC	Squamous Cell Carcinoma
SGD	Stochastic Gradient Descent
TL	Transfer Learning
WHO	World Health Organization
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **2024**, *74*, 229–263. [CrossRef] [PubMed]
2. Cancer Today. Available online: https://gco.iarc.fr/today/online-analysis-pie?v=2020&mode=cancer&mode_population=continents&population=900&populations=900&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group[]=0&ages_group[]=17&nb_items=7&group_cancer=1&include_nmsc=1&include_nmsc_other=1&half_pie=0&donut=0 (accessed on 13 January 2024).
3. Xi, Y.; Xu, P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl. Oncol.* **2021**, *14*, 101174. [CrossRef] [PubMed]
4. Cancer. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 13 January 2024).
5. N.A. and R.A. Office of the Federal Register. Public Law 104-191-Health Insurance Portability and Accountability Act of 1996. govinfo.gov, August 1996. Available online: https://www.govinfo.gov/app/details/PLAW-104publ191 (accessed on 13 January 2024).
6. I (Legislative Acts) Regulations Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)'. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 10 January 2024).
7. Masud, M.; Sikder, N.; Nahid, A.-A.; Bairagi, A.K.; Al Zain, M.A. A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework. *Sensors* **2021**, *21*, 748. [CrossRef]
8. Chehade, A.H.; Abdallah, N.; Marion, J.-M.; Oueidat, M.; Chauvet, P. Lung and colon cancer classification using medical imaging: A feature engineering approach. *Phys. Eng. Sci. Med.* **2022**, *45*, 729–746. [CrossRef]
9. Mangal Engineerbabu, S.; Chaurasia Engineerbabu, A.; Khajanchi, A. Convolution Neural Networks for Diagnosing Colon and Lung Cancer Histopathological Images. September 2020. Available online: https://arxiv.org/abs/2009.03878v1 (accessed on 1 February 2024).

10. Hadiyoso, S.; Aulia, S.; Irawati, I.D. Diagnosis of lung and colon cancer based on clinical pathology images using convolutional neural network and CLAHE framework. *Int. J. Appl. Sci. Eng.* **2023**, *20*, 1–7. [CrossRef]

11. You, C.; Zhao, R.; Liu, F.; Dong, S.; Chinchali, S.; Topcu, U.; Staib, L.; Duncan, J. Class-aware adversarial transformers for medical imagesegmentation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 29582–29596. [PubMed]

12. You, C.; Xiang, J.; Su, K.; Zhang, X.; Dong, S.; Onofrey, J.; Staib, L.; Duncan, J.S. *Incremental Learning Meets Transfer Learning: Application to Multi-Site Prostate MRI Segmentation*; Springer: Cham, Switzerland, 2022; Volume 13573, pp. 3–16. [CrossRef]

13. Mehmood, S.; Ghazal, T.M.; Khan, M.A.; Zubair, M.; Naseem, M.T.; Faiz, T.; Ahmad, M. Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning with Class Selective Image Processing. *IEEE Access* **2022**, *10*, 25657–25668. [CrossRef]

14. Toğaçar, M. Disease type detection in lung and colon cancer images using the complement approach of inefficient sets. *Comput. Biol. Med.* **2021**, *137*, 104827. [CrossRef]

15. Kumar, N.; Sharma, M.; Singh, V.P.; Madan, C.; Mehandia, S. An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. *Biomed. Signal Process. Control* **2022**, *75*, 103596. [CrossRef]

16. Talukder, A.; Islam, M.; Uddin, A.; Akhter, A.; Hasan, K.F.; Moni, M.A. Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Syst. Appl.* **2022**, *205*, 117695. [CrossRef]

17. Al-Jabbar, M.; Alshahrani, M.; Senan, E.M.; Ahmed, I.A. Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features. *Bioengineering* **2023**, *10*, 383. [CrossRef]

18. Ananthakrishnan, B.; Shaik, A.; Chakrabarti, S.; Shukla, V.; Paul, D.; Kavitha, M.S. Smart Diagnosis of Adenocarcinoma Using Convolution Neural Networks and Support Vector Machines. *Sustainability* **2023**, *15*, 1399. [CrossRef]

19. You, C.; Zhao, R.; Staib, L.H.; Duncan, J.S. *Momentum Contrastive Voxel-Wise Representation Learning for Semi-Supervised Volumetric Medical Image Segmentation*; Springer: Cham, Switzerland, 2022; Volume 13434, pp. 639–652. [CrossRef]

20. You, C.; Zhou, Y.; Zhao, R.; Staib, L.; Duncan, J.S. SimCVD: Simple Contrastive Voxel-Wise Representation Distillation for Semi-Supervised Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2022**, *41*, 2228–2237. [CrossRef] [PubMed]

21. You, C.; Dai, W.; Min, Y.; Staib, L.; Duncan, J.S. *Bootstrapping Semi-Supervised Medical Image Segmentation with Anatomical-Aware Contrastive Distillation*; Springer: Cham, Switzerland, 2022; Volume 13939, pp. 641–653. [CrossRef]

22. You, C.; Dai, W.; Min, Y.; Liu, F.; Clifton, D.A.; Zhou, S.K.; Staib, L.; Duncan, J.S. Rethinking Semi-Supervised Medical Image Segmentation: A Variance-Reduction Perspective. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 9984–10021.

23. You, C.; Dai, W.; Min, Y.; Staib, L.; Sekhon, J.; Duncan, J.S. *ACTION++: Improving Semi-Supervised Medical Image Segmentation with Adaptive Anatomical Contrast*; Springer: Cham, Switzerland, 2023; Volume 14223, pp. 194–205. [CrossRef]

24. Konečn, J.K.; Brendan, H.; Google, M.; Google, D.R.; Richtárik, P. Federated Optimization: Distributed Machine Learning for On-Device Intelligence. October 2016. Available online: https://arxiv.org/abs/1610.02527v1 (accessed on 1 February 2024).

25. Konečn, J.; McMahan, H.B.; Yu, F.X.; Suresh, A.T.; Google, D.B.; Richtárik, P. Federated Learning: Strategies for Improving Communication Efficiency. October 2016. Available online: https://arxiv.org/abs/1610.05492v2 (accessed on 1 February 2024).

26. Roth, H.R.; Chang, K.; Singh, P.; Neumark, N.; Li, W.; Gupta, V.; Gupta, S.; Qu, L.; Ihsani, A.; Bizzo, B.C.; et al. *Federated Learning for Breast Density Classification: A Real-World Implementation*; Springer: Cham, Switzerland, 2020; Volume 12444, pp. 181–191. [CrossRef]

27. Florescu, L.M.; Streba, C.T.; Şerbănescu, M.-S.; Mămuleanu, M.; Florescu, D.N.; Teică, R.V.; Nica, R.E.; Gheonea, I.A.; Florescu, L.M.; Streba, C.T.; et al. Federated Learning Approach with Pre-Trained Deep Learning Models for COVID-19 Detection from Unsegmented CT images. *Life* **2022**, *12*, 958. [CrossRef]

28. Hossain, M.; Ahamed, F.; Islam, R.; Imam, R. Privacy Preserving Federated Learning for Lung Cancer Classification. In Proceedings of the 2023 26th International Conference on Computer and Information Technology, ICCIT 2023, Cox's Bazar, Bangladesh, 13–15 December 2023. [CrossRef]

29. Zhang, W.; Zhou, T.; Lu, Q.; Wang, X.; Zhu, C.; Sun, H.; Wang, Z.; Lo, S.K.; Wang, F.-Y. Dynamic-Fusion-Based Federated Learning for COVID-19 Detection. *IEEE Internet Things J.* **2021**, *8*, 15884–15891. [CrossRef] [PubMed]

30. Khan, T.A.; Fatima, A.; Shahzad, T.; Rahman, A.U.; Alissa, K.; Ghazal, T.M.; Al-Sakhnini, M.M.; Abbas, S.; Khan, M.A.; Ahmed, A. Secure IoMT for Disease Prediction Empowered with Transfer Learning in Healthcare 5.0, the Concept and Case Study. *IEEE Access* **2023**, *11*, 39418–39430. [CrossRef]

31. Peyvandi, A.; Majidi, B.; Peyvandi, S.; Patra, J.C. Privacy-preserving federated learning for scalable and high data quality computational-intelligence-as-a-service in Society 5.0. *Multimed. Tools Appl.* **2022**, *81*, 25029–25050. [CrossRef] [PubMed]

32. Borkowski, A.A.; Bui, M.M.; Thomas, L.B.; Wilson, C.P.; DeLand, L.A.; Mastorides, S.M. Lung and Colon Cancer Histopathological Image Dataset (LC25000). December 2019. Available online: https://arxiv.org/abs/1912.12142v1 (accessed on 1 February 2024).

33. Bhimji, S.S.; Wallen, J.M. Lung Adenocarcinoma. StatPearls, June 2023. Available online: https://www.ncbi.nlm.nih.gov/books/NBK519578/ (accessed on 1 February 2024).

34. Walser, T.; Cui, X.; Yanagawa, J.; Lee, J.M.; Heinrich, E.; Lee, G.; Sharma, S.; Dubinett, S.M. Smoking and Lung Cancer: The Role of Inflammation. *Proc. Am. Thorac. Soc.* **2008**, *5*, 811–815. [CrossRef]

35. Ma, Z.; Zhang, M.; Liu, J.; Yang, A.; Li, H.; Wang, J.; Hua, D.; Li, M. An Assisted Diagnosis Model for Cancer Patients Based on Federated Learning. *Front. Oncol.* **2022**, *12*, 860532. [CrossRef]

36. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]

37. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015. Available online: http://www.robots.ox.ac.uk/ (accessed on 17 February 2024).

38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. Available online: http://image-net.org/challenges/LSVRC/2015/ (accessed on 17 February 2024).

39. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995. [CrossRef]

40. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [CrossRef]

41. Ribeiro, M.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, San Diego, CA, USA, 12–17 June 2016; pp. 97–101. [CrossRef]

42. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int. J. Comput. Vis.* **2016**, *128*, 336–359. [CrossRef]

43. Tasnim, Z.; Chakraborty, S.; Shamrat, F.M.J.M.; Chowdhury, A.N.; Alam Nuha, H.; Karim, A.; Zahir, S.B.; Billah, M. Deep Learning Predictive Model for Colon Cancer Patient using CNN-based Classification. *Int. J. Adv. Comput. Sci. Appl.* **2021**, 12. [CrossRef]

44. Shandilya, S.; Nayak, S.R. *Analysis of Lung Cancer by Using Deep Neural Network*; Springer: Singapore, 2022; Volume 814, pp. 427–436. [CrossRef]

45. Karim, D.Z.; Bushra, T.A. Detecting Lung Cancer from Histopathological Images using Convolution Neural Network. In Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Auckland, New Zealand, 7–10 December 2021; pp. 626–631. [CrossRef]

46. Raju, M.S.N.; Rao, B.S. Lung and colon cancer classification using hybrid principle component analysis network-extreme learning machine. *Concurr. Comput. Pr. Exp.* **2022**, *35*, e7361. [CrossRef]

47. Ren, Z.; Zhang, Y.; Wang, S. A Hybrid Framework for Lung Cancer Classification. *Electronics* **2022**, *11*, 1614. [CrossRef]

48. Attallah, O.; Aslan, M.F.; Sabanci, K. A Framework for Lung and Colon Cancer Diagnosis via Lightweight Deep Learning Models and Transformation Methods. *Diagnostics* **2022**, *12*, 2926. [CrossRef] [PubMed]