CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD

Melanoma Disease Classification using Federated Learning

by

Muhammad Irfan Ashraf

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Faculty of Engineering
Department of Electrical Engineering

2023

I would like to dedicate this work to my family, their true support and motivation has made it possible to successfully complete this research.

CERTIFICATE OF APPROVAL

Melanoma Disease Classification using Federated Learning

by

Muhammad Irfan Ashraf

(MEE193016)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Tehseen Zia	COMSATS, Islamabad
(b)	Internal Examiner	Dr. Noor Muhammad Khan	CUST, Islamabad
(c)	Supervisor	Dr. Imtiaz Ahmad Taj	CUST, Islamabad

Dr. Imtiaz Ahmad Taj
Thesis Supervisor
Sept, 2023

Dr. Noor Muhammad Khan
Head
Dept. of Electrical Engineering
Sept, 2023

Dr. Imtiaz Ahmad Taj
Dean
Faculty of Engineering
Sept, 2023

Author's Declaration

I, **Muhammad Irfan Ashraf** hereby state that my MS thesis titled " **Melanoma Disease Classification using Federated Learning** " is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Muhammad Irfan Ashraf)

Registration No: MEE193016

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Melanoma Disease Classification using Federated Learning**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Muhammad Irfan Ashraf)
Registration No: MEE193016

Acknowledgement

I would like to express my deepest appreciation to all those who provided me with the possibility to complete this research. I would like to give special gratitude, to my research thesis supervisor Dr. Imtiaz Ahmad Taj whose contribution in suggestions and encouragement, helped me to complete this research.

I owe a tremendous debt of gratitude to my family, as their unwavering support made this work possible. I would like to extend a heartfelt appreciation to my wife Dr. Benish, who has been with me throughout this entire journey. Her assistance and guidance has been invaluable. Her constant encouragement and belief in my abilities kept me going, urging me to see this work through to completion.

Further, I want to express my special thanks to my beloved daughter, Aizal Irfan. She is my inspiration and motivation to complete my thesis and keep progress in life.

I am truly fortunate to have such an incredible support system in my family, and I am profoundly grateful for their contributions to this achievement.

(Muhammad Irfan Ashraf)

Abstract

The early identification of melanoma, which is a form of skin cancer, is crucial as it significantly improves the chances of survival. Over the years, statistical machine learning techniques, such as deep learning, are being used more and more to automate the detection process. These techniques acquire large datasets for training. However, health institutions are inhibited when it comes to disclosing patients' data because of privacy concerns and confidentiality of the individuals involved. In order to overcome this limitation, federated learning offers a solution that allows health institutions to foster collaboration that trains a global model without the requirement of sharing their raw data.

This work presents an analysis of ResNet50 model using both conventional deep learning and federated learning framework. In recent literature, many federated learning techniques have been used to improve the accuracy of melanoma disease. However, its performance on unbalanced datasets has not been thoroughly examined. This work addresses this gap by evaluating the effectiveness of federated learning on both balanced and unbalanced datasets. To address the challenges posed by unbalanced datasets, a new strategy 'Imbalance Weighted Average Strategy (IWAS)' on the server side is proposed, that uses an imbalance factor to adjust the weight of each client's contribution based on the imbalance ratio in their dataset.

The results are evaluated on the publically available ISIC-2020 dataset. The results illustrate that the federated learning approach performs better on average 2% than the conventional deep learning model. After splitting the datsets into both balanced and unbalanced, the results indicate that federated learning outperforms on the balanced dataset because of the federated averaging algorithm on the server side. In the end, experiments are conducted on the proposed IWAS to evaluate the performance on an unbalanced dataset. The result shows the effectiveness of the proposed approach, as it significantly enhances the average performance of the federated learning model by about 4%.

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial intelligence
ARL-CNN	Residual Learning convolutional neural network
AUC	Area Under the Curve
BCI	Brain-Computer Interfaces
CAD	Computer-aided diagnosis
CNN	Convolutional Neural Network
cPDS	cluster Primal-Dual Splitting
DL	Deep Learning
DIA	Dermoscopy image analysis
FL	Federated Learning
FTL	federated transfer learning
IOT	Internet of Things
IRRCNN	Inception Recurrent Residual Convolutional Neural Network
ISIC	International Skin Imaging Collaboration
MICCAI	Medical Image Computing and Computer-Assisted Intervention
ML	Machine Learning
non-IID	non-independent and identically distributed
R2U-Net	Recurrent Residual U-Net
SVM	Support vector machine
TFF	TensorFlow Federated

Chapter 1

Introduction

1.1 Background

In modern times, cancer is one of the primary factors contributing to mortality, responsible for a significant number of deaths each year. In 2018, there were 9.5 million cancer-related deaths reported and approx 17 million new cancer cases arrived according to the American Cancer Society [1]. This translates to a death rate of one in every 6 people [2].

There are several factors that contribute to skin cancer development. Major factors include a weak immune system and exposure to direct ultraviolet radiation emitted by the sun and genetics. Regardless of skin color, skin cancer can affect anyone but those certain populations who have fair skin or a background of skin cancer within the family and a past experience of sunburns are at higher risk of skin cancer [3]. In particular, skin cancer remains a significant health concern worldwide, with melanoma being the deadliest form that emerges when melanocytes become hyperactive. It is caused by the unusual multiplication of pigment-producing cells, leading to the formation of malignant tumors that give color to the skin (as shown in figure 1.1).

The ABCDE acronym [4], [5] is a useful approach to detect the signs of melanoma. The acronym stands for asymmetry, irregular borders, multiple colors, diameter

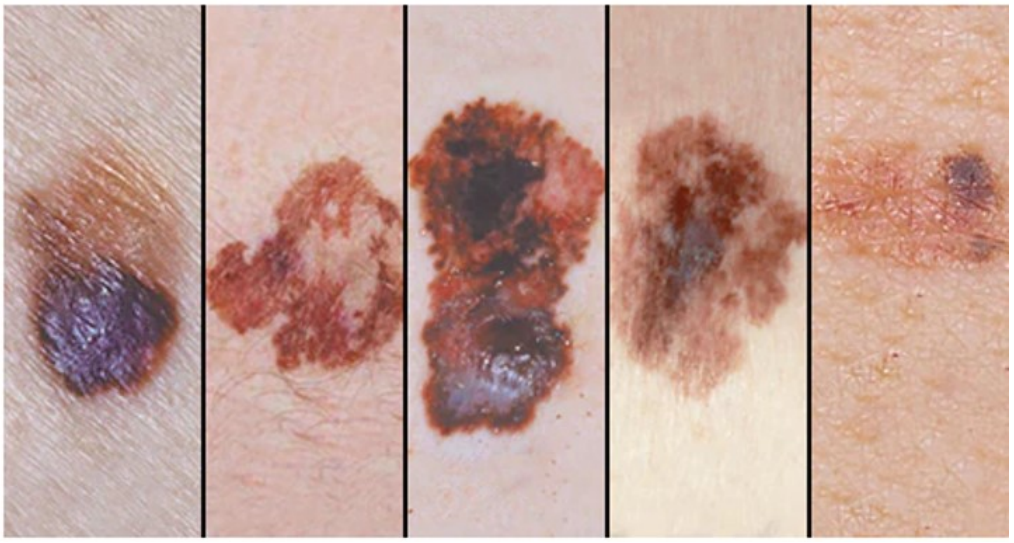


FIGURE 1.1: Melonama Images.
Picture Credits: https://www.mayoclinic.org/

larger than 6mm, and enlarging lesion, all are potential signs of melanoma. If any of these symptoms are present, it is important to seek a thorough evaluation by a medical professional to determine if further investigation or treatment is necessary. Regular skin self-examinations can also help in determining any abnormalities in skin or moles growth.

For successful treatment early detection of melanoma is important. If detected early the survival rates are considerably high which is estimated to be around 99%. However, the survival rate significantly drops to about 20% if detected late [6].

Medical imaging is a useful tool in the treatment and diagnosis of many diseases and continuously improving the health sector with a visual representation of the exterior and interior of the body. Imaging techniques including CT scans, PET scans, MRI, X-rays, and dermatoscopy, use various forms of radiation, radio frequencies, and sound waves to capture images of different parts of the body.

Dermatoscopy [7], a non-invasive imaging technique that uses polarized light to capture images of the skin has shown significant results in improving diagnostic accuracy for skin cancer. According to research, the dermatoscopy has increased diagnostic accuracy by 10-27% when compared to human experts relying on their

knowledge and naked eye examination alone [8].

Advancements in computer science have given rise to artificial intelligence (AI) and its potential to bring a revolutionary change in the diagnostic process. AI algorithms can analyze endless amounts of medical imaging data in real time, potentially outperforming human experts in detecting and diagnosing diseases, including skin cancer.

By automating the diagnostic process, AI possesses the capability to enhance accuracy, reduce diagnostic errors, and ultimately improve patient outcomes. As AI continues to grow, it is probable to play an increasingly critical role in diagnosis and medical imaging.

A group of researchers from Germany, France, and the United States developed an AI-based system to help detect and classify harmless and dangerous skin lesions [9]. Over 100,000 images of skin lesions are fed into Convolutional Neural Network (CNN), to help diagnose skin lesions. The system was able to correctly identify the lesions with 95% accuracy. In comparison, human experts were only able to do this with 86.6% accuracy.

1.2 Melanoma Detection using Conventional Learning Approaches

Significant improvements have been made in utilizing conventional machine learning (ML) and deep learning (DL) methods to help in the diagnosis of melanoma. Since melanoma is a malignant form of skin disease, therefore early detection and treatment are very important.

The task of ML/DL models is to ascertain whether a skin lesion is melanoma or not. DL algorithms have proven to be highly effective in the detection of melanoma. Skin lesions are in many forms, each with its own associated changes. Despite this diversity, these lesions can broadly be classified into two main types. Understanding these types is crucial for accurate diagnosis and treatment decisions.

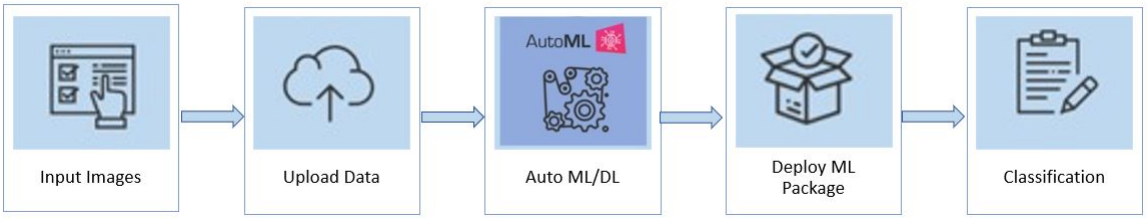


FIGURE 1.2: Conventional ML/DL Model

Skin lesions are broadly classified as:

1. Benign - non-cancerous
2. Malignant - cancerous

The overview of DL framework for melonama detection is illustrated in figure 1.2. Initially, a set of images is fed into the system as input, which is then uploaded to the server for processing. Next, the system is trained using ML/DL models, resulting in an ML package that is capable of classifying the problem. The trained package can then be deployed for use in the diagnosis of melanoma, providing a reliable and efficient solution.

A huge number of data is required for training to build an efficient machine-learning model. However, traditional machine learning methods require data to be collected centrally in one location, that can lead to risks such as data leakage. Also, collecting sensitive data such as healthcare data violates user privacy [10].

The conventional method of collecting datasets poses significant complications such as sensitive data confidentiality and legal restrictions related to ethics, privacy, and data protection. For example, rare disease information that is available only at a particular hospital cannot be shared with others due to legal restrictions [11]. Moreover, buying a dataset from a hospital for research can be expensive, and it may not be enough for a good study. Therefore, multiple datasets are required from many hospitals, which would be even more expensive. Due to these challenges, traditional ML/ DL models are hard to use especially in healthcare.

Federated Learning (FL) [12] is a new way of training ML and DL models where the data is stored in different places instead of one central location. FL is different

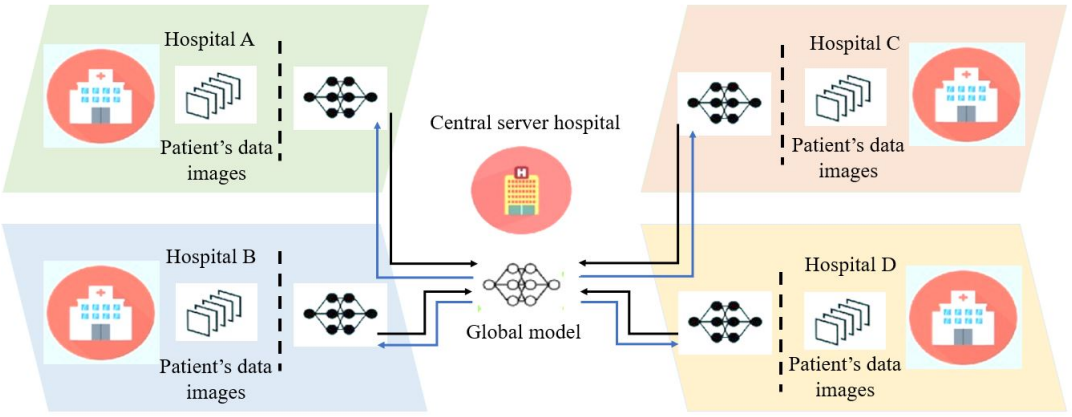


FIGURE 1.3: FL Model

from traditional way of training models, where all the data is gathered in one place. Therefore, FL method provides better data privacy compared to traditional methods.

1.3 Federated Learning

A different ML approach called federated learning performs training on a large collection of distributed data and its core idea revolves around "data privacy". This is different from traditional methods where data is sent to a central location. With FL, the data stays where it is and the model is trained on each device separately. This helps keep sensitive data safe because data remains in the device. After training, only the model is shared, not the data as shown in figure 1.3. A central server combines each model from the user's devices to create a global model that everyone can use.

FL has totally evolved DL and hospitals and other organizations would now be able to create a state of the art ML/DL models without exposing the raw data. Further FL has been categorized into three groups [13]:

1. Distributed learning
2. Cross-silo
3. Cross-device

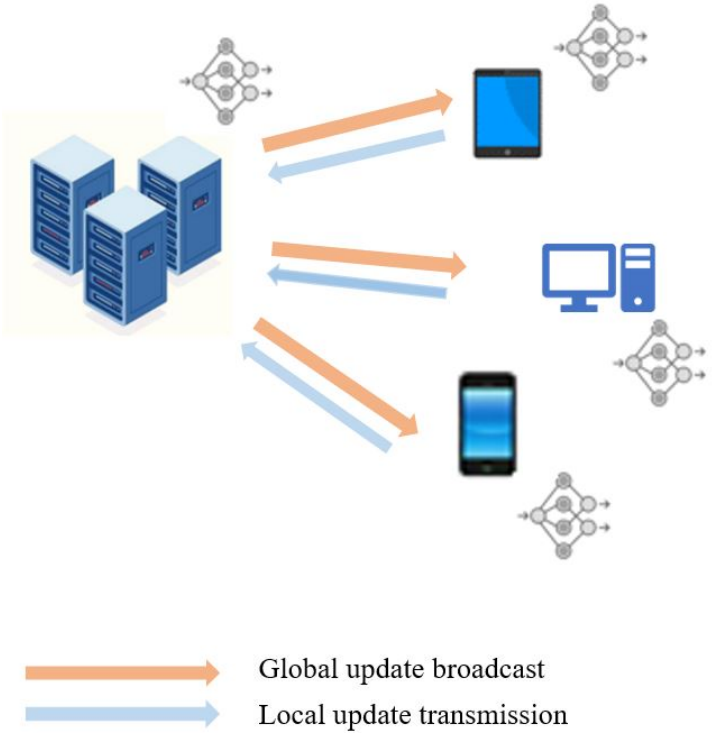


FIGURE 1.4: Distributed Learning Framework

1.3.1 Distributed Learning

In distributed learning, model training on a large dataset is a common scenario where the clients are computing nodes located in a single cluster or large data center. The data is centrally stored and can be balanced and shuffled over clients to ensure fairness in the training process.

The basic framework is shown in figure 1.4. Any client has the ability to access and read any portion of the dataset as the data is centrally orchestrated. Centrally orchestrated means that there is a central server that coordinates and manages the training process. The number of clients commonly ranges from 1 to 1000. The approach assumes that the computation is the bottleneck in the data center and fast networks are generally required.

In this scenario, the training process is simple, which means that each client has the potential to participate in every round of the computation, updating the weights throughout each iteration. Additionally, the data can be flexibly partitioned and redistributed among all the clients to optimize the training and evaluation process.

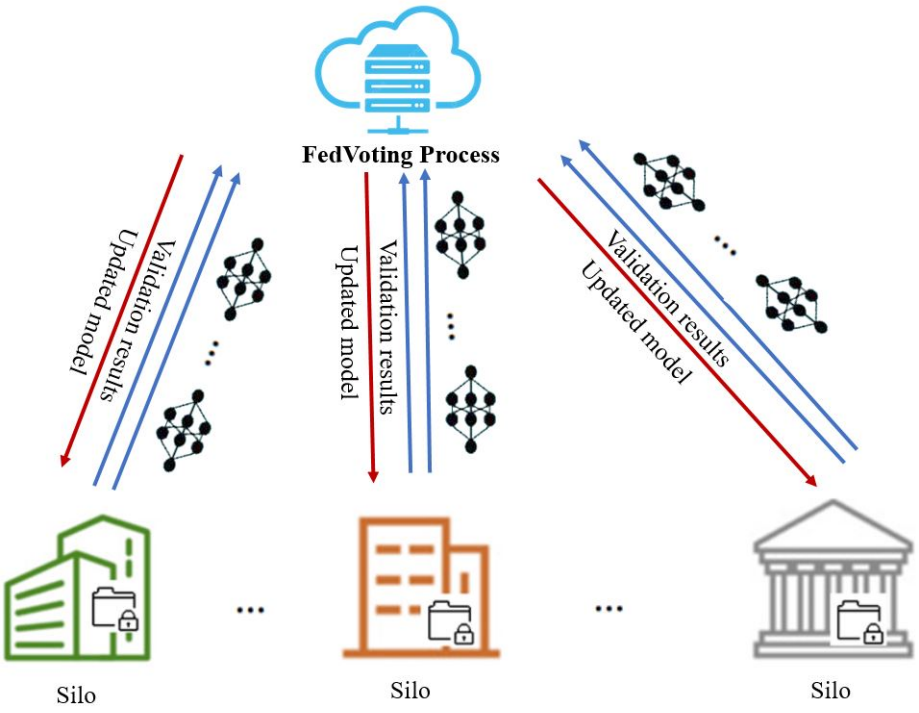


FIGURE 1.5: Cross-silo Framework

Overall, this approach to training a model on a large dataset is a common and efficient method that takes advantage of the capabilities of modern data centers.

1.3.2 Cross-silo

Cross-silo [14] means that the clients are organizations like hospitals or banks as depicted in figure 1.5. They have a lot of data to train on and are usually reliable. Only a few clients work together in each round of training to help in creating a global model and these clients are persistent.

Each client in the FL system has a unique identity or name that grants the ability to access a specific entity or object. Furthermore, each client is stateful, meaning it may participate in every computation round, maintaining state from round to round.

Finally, the partition remains constant and can be either horizontally or vertically divided. This means that the data is divided based on the client's characteristics, and the model tries to learn from this data.

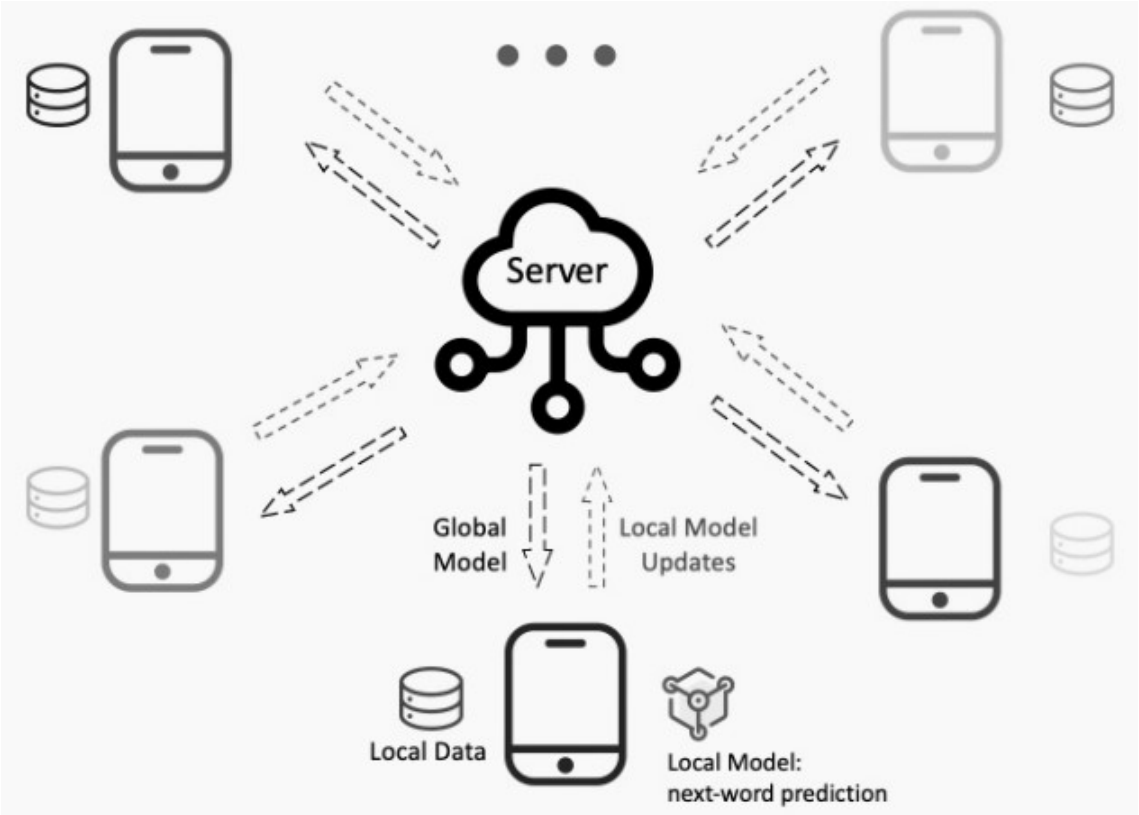


FIGURE 1.6: Cross-device Framework

1.3.3 Cross-device

Cross-device means that the clients are mobile devices like phones or tablets as shown in figure 1.6.

They don't have a lot of data to train on and may not always be reliable due to poor network connections, as the computations are performed across devices that are connected through slower connections, such as Wi-Fi or cellular networks. Hence, the system should be designed in a way that minimizes communication overhead.

It has many clients who only participate in one round of training and they are transient and clients cannot be indexed directly, which means that is unable to target specific clients for computations using client identifiers.

Therefore, the system needs to use other techniques, such as random sampling or clustering, to select a subset of clients for each computation round.

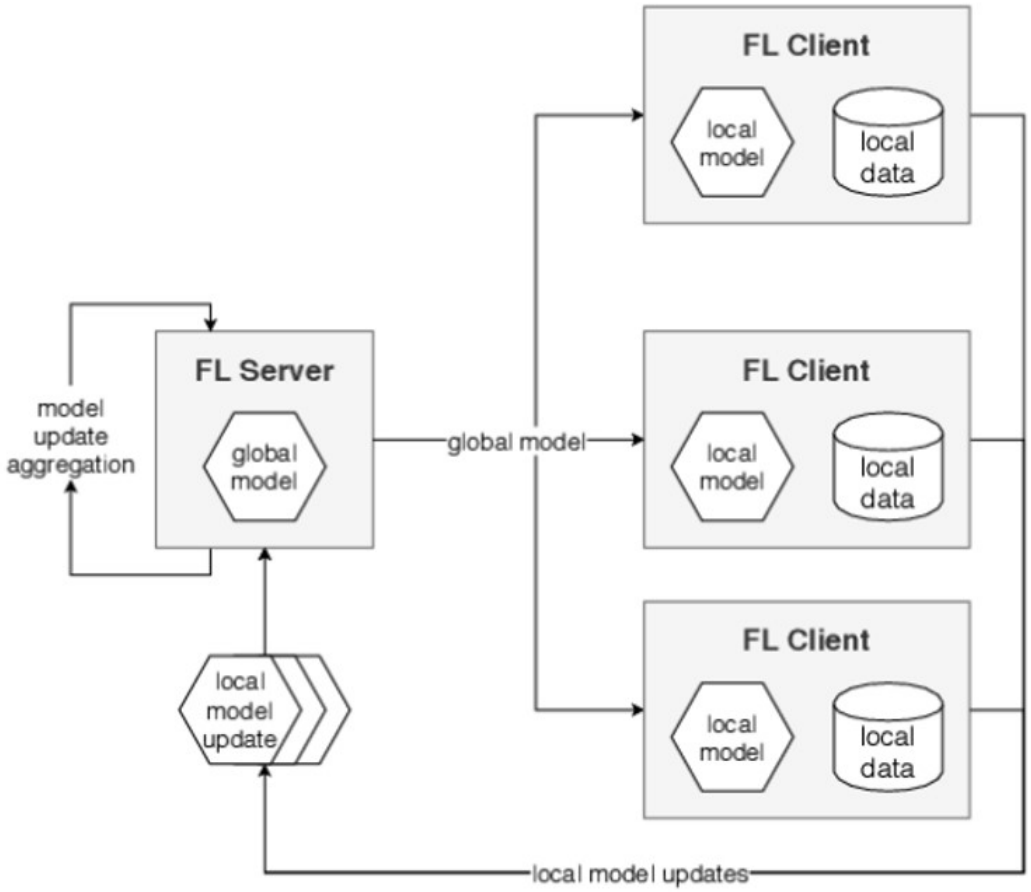


FIGURE 1.7: Flow Diagram of FL

1.4 Federated Learning Working

The training in FL involves a series of rounds where the model is improved over time. The communication rounds are depicted in Figure 1.7. The process involves the server starting with a base model round follows some steps:

- First the server binds a number of clients as per certain criteria.
- Selected clients receive the base model from the server.
- Then models are trained on the local data of each client by running multiple times (i.e. number of local epochs).
- The updated weights are sent back to the server by the clients.
- Server aggregates client weights based on any algorithm normally.

- The aggregated model is to be shared as a base model for the next round.

Despite the advantages, FL has several challenges as well, such as the trade-off between accuracy and privacy, system heterogeneity, communication bottleneck, and poisoning. A detailed explanation of challenges is mentioned in chapter 3.

1.5 Thesis Motivation

Early detection of melanoma, a type of skin cancer is essential to enhance the likelihood of survival rates. To train an ML/DL model for melanoma classification requires bringing all the data to one central server, which can be a challenge especially when dealing with medical data because of privacy concerns.

Sharing sensitive data can lead to serious privacy issues and may be refused by hospitals and healthcare providers. Also, exchanging data with pharmaceutical companies is not always possible.

Moreover, because a lot of data is generated on smartphones and other wearable devices, transferring all this data to a central location can be time-consuming and expensive.

To overcome these challenges, FL can be used. It trains the models on decentralized data, so sensitive data doesn't need to be shared. This minimizes the risk of personal data leakage. Under this framework, all devices can collaborate to train a high-quality model without breaching each other's private information.

Moreover, FL is capable to provide a scalable solution without using much bandwidth as well as it doesn't require the transmission of raw data, which leads to lower communication costs.

Although FL has many benefits, there are still some limitations such as data heterogeneity and accuracy issues when dealing with an unbalanced dataset. The main focus of this thesis is to classify melanoma skin disease and improve accuracy on unbalanced data sets, which will enhance user privacy.

1.6 Thesis Objective

- Perform a comparative analysis between conventional and federated learning models to classify/detect melanoma disease.
- Propose an improved federated learning framework for melanoma disease detection to improve accuracy on unbalanced data-set, thus enhancing user privacy.

1.7 Thesis Organization

The thesis is structured in the following way:

Chapter 1: Introduction

The first chapter signifies the importance of melanoma disease detection and discusses the traditional ML/DL methods that are used for the classification of the disease. The chapter also covers the importance of federated learning, how it works, and discusses various frameworks, and challenges. Finally, the chapter concludes with the motivation and objectives of the research.

Chapter 2: Literature Review and Problem Formulation

This chapter presents a detailed literature survey on previous work of melanoma disease detection using DL and FL methods. Moreover, their limitations and challenges are discussed. Through this analysis, critical gaps in the research are identified. Additionally, based on the findings from the literature survey, the chapter concludes the problem statement.

Chapter 3: Federated Learning

This chapter provides a comprehensive overview of federated learning, the steps involved in the training process and the challenges related to FL. Moreover, the need of FL is explored and highlighted especially in medical domain. Additionally, different FL frameworks are examined and explored the utilization of FLOWER framework based on its distinct advantages.

Chapter 4: Methodology

In this chapter, our methodology is presented. The dataset and the client/server connection in FL are discussed. Also, the performance of the proposed algorithm with ResNet50 model is explored on two types of datasets: balanced and imbalanced distribution of melanoma images. Finally, the proposed algorithm is explained for aggregation of model updates on the server.

Chapter 5: Results and Evaluation

The results and evaluation of DL and FL models are presented in this chapter. The results obtained from both balanced and unbalanced distribution are discussed. The results showed that it outperformed on the balanced dataset but there is a significant increase in performance when the proposed imbalanced weighted-averaging strategy is applied.

Chapter 6: Conclusion and Future Work

This chapter has conclusions based on comparative analysis and evaluation of results presented in the previous section. Additionally, potential future research directions; that can be pursued based on the findings of this study; are discussed. By considering these aspects, this chapter aims to provide a comprehensive conclusion to the research and set the stage for future advancements in the field.

Chapter 2

Literature Review

This chapter explains how classical training helps in the detection of melanoma and then discusses ML/DL techniques. It also talks about how FL is becoming a better approach compared to traditional machine learning. At the end of the chapter, the research gap and a problem statement are identified based on the analysis of the literature. The research methodology and thesis contributions are also presented.

2.1 Introduction

Melanoma is a type of skin cancer that is becoming more common, and many researchers are working on ways to detect and analyze it using computer software. In the past, mainly research focused on early detection to determine if a melanoma is cancerous or not, and many studies used ML and DL techniques to classify melanoma. However, there are still concerns about privacy when using these methods, so more recent research has been focused to protect personal information by using FL.

FL uses the power of multiple devices to collectively train a global model without the need to share raw data. This decentralized approach significantly reduces the risk of personal data leakage.

The literature review reviewed various Learning based techniques that have been used by researchers to detect melanoma. This chapter highlights the contributions and limitations of their work.

2.2 Related Work

This section of the thesis discusses how researchers have worked on detecting melanoma using traditional methods in the past. The summary of the contributions made in this field is demonstrated in figure 2.1. Then, how ML/DL methods are been used to improve the accuracy of detection. Finally, it discusses how FL is becoming a suitable way to detect melanoma.

The section also addresses the contributions and limitations of previous research work. Researchers have done a lot of work to improve the accuracy of melanoma detection, but there is still some room for improvements to the current methods. Researchers can identify these areas to improve the accuracy of melanoma detection.

2.2.1 Traditional Techniques

Dermoscopy [7] is a method to inspect pigment skin lesions used by doctors with the help of a magnifying instrument (as shown in figure 2.2). The main purpose is to decide if the lesion should be monitored or can be left untouched or have to remove it. The technique helped overcome the gaps between the early stage of a lesion and a microscopic examination.

Dermoscopic diagnosis involves two steps: to determine if the lesion is melanocytic or not and if it is, then classify it as benign or malignant. Various algorithms and techniques have been developed to assist in this process. It can also diagnose non-melanocytic pigmented lesions such as haemangioma, pigmented basal cell carcinoma, lichen planus-like keratosis, and seborrhoeic keratoses.

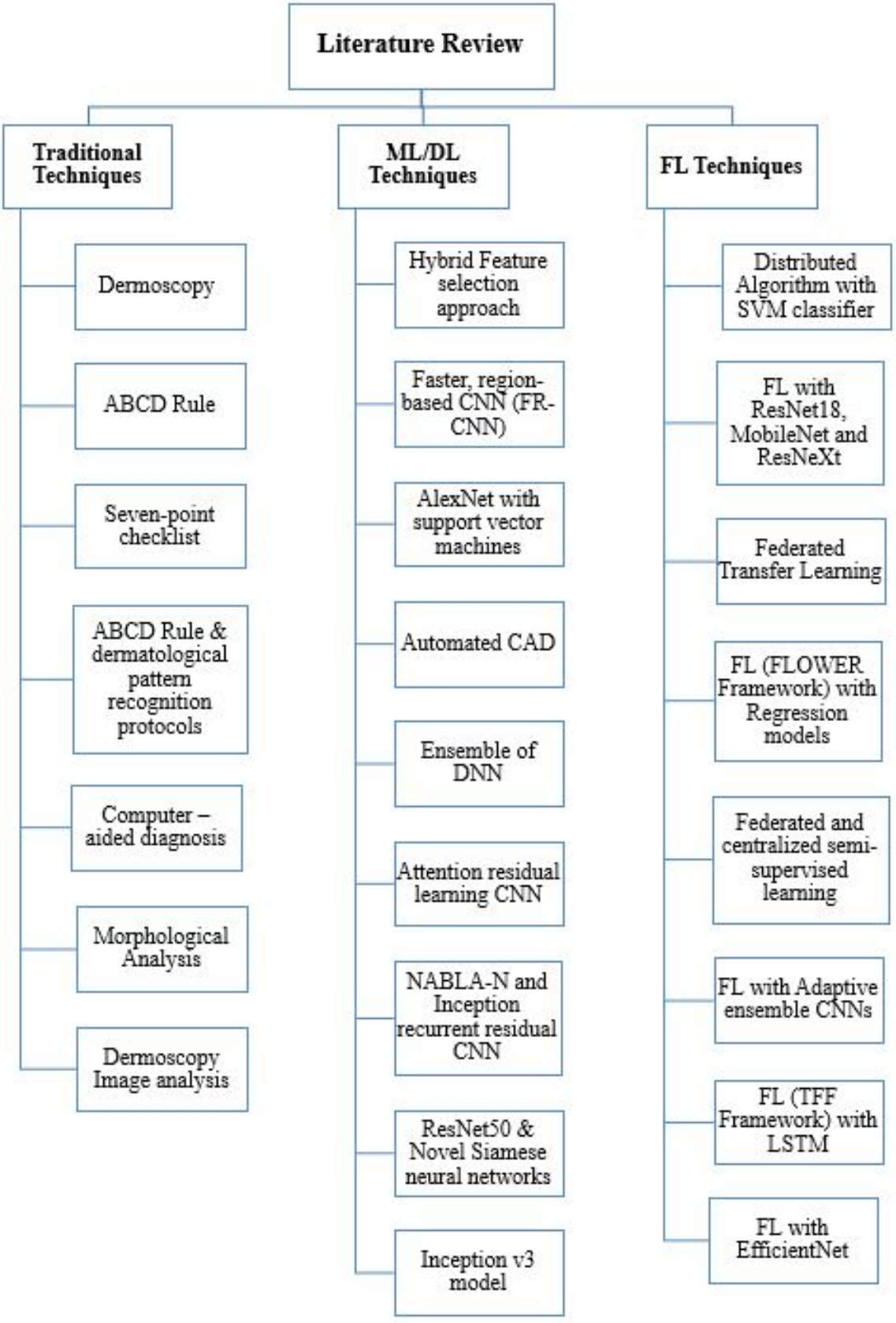


FIGURE 2.1: Melanoma Disease Detection Techniques

FIGURE 2.2: Dermoscopy
Picture Credits: https://nextstepsinderm.com/

The ABCD rule [5] is created to help detection of melanoma. This study used 172 pigmented skin lesions to check if the ABCD rule could accurately predict which ones are melanoma. Tumors that scored above 5.45 are mainly categorized as malignant, whereas those with lower scores are categorized as benign. The ABCD rule concluded a high positive predictive value and negative predictive value, establishing its reliability in the detection of melanoma. Moreover, this rule can be easily learned and promptly applied to provide an objective and consistent diagnosis.

A study is conducted to determine the diagnostic performance of various methods for identifying melanoma, including a revised seven-point checklist [15]. In the current clinical setting, the revised checklist had demonstrated improved effectiveness in detecting melanomas and atypical melanocytic naevi. The study involved eight dermatologists who assessed images of naevi, excised melanomas and monitored naevi using pattern analysis along with both the standard and revised seven-point

checklists. The findings showed that the revised checklist recommended more excisions than the standard checklist. However, it also showed greater sensitivity in identifying melanomas and atypical naevi.

In another research, [16] used a computer program that helped doctors to diagnose melanoma disease, a type of skin cancer. A hybrid technique is used consisting of ABCD Rule and dermatological Pattern Recognition protocols. This program has various algorithms for measuring the color, size, asymmetry, border, and diameter of melanoma plus it also includes three built-in algorithms for detecting specific patterns in the images. Results are evaluated by medical professionals after extensive testing on large carps of datasets. They successfully demonstrated that this approach is accurately predicted certain patterns of melanoma and a valuable tool to diagnose melanoma by medical experts.

Chang et al. [17] demonstrated an automated tool Computer-aided diagnosis (CADx) to helped doctors to classify melanomas. This tool used standard dermatological protocols and algorithms to evaluate the color, size, symmetry, border, and diameter of a melanoma. Furthermore, it has three algorithms that are capable of identifying specific patterns within melanoma, which could help to calculate certain characteristics in melanoma images. they tested their system on a database of 160 images and are observed to be more accurate according to medical professionals. The algorithms for pattern recognition showed a high accuracy rate of over 85% and also indicated the tool's reliability and its potential to contribute to accurate melanoma diagnoses. Figure 2.3 displays a basic view of a CAD (Computer-Aided Diagnosis) system that takes images as a prepossessing first step then the image augmentation stage and feature extraction to be followed. The CAD system is specifically designed to help in making diagnoses of melanoma for medical experts.

The number of people dying from melanoma skin cancer is rapidly increasing every year, so there is a need for better and optimal ways to detect it. One way to do this is with a computer program that can look at pictures of skin lesions and find the ones that might be melanoma. *Olugbara et al.* [18] discussed a new way to do

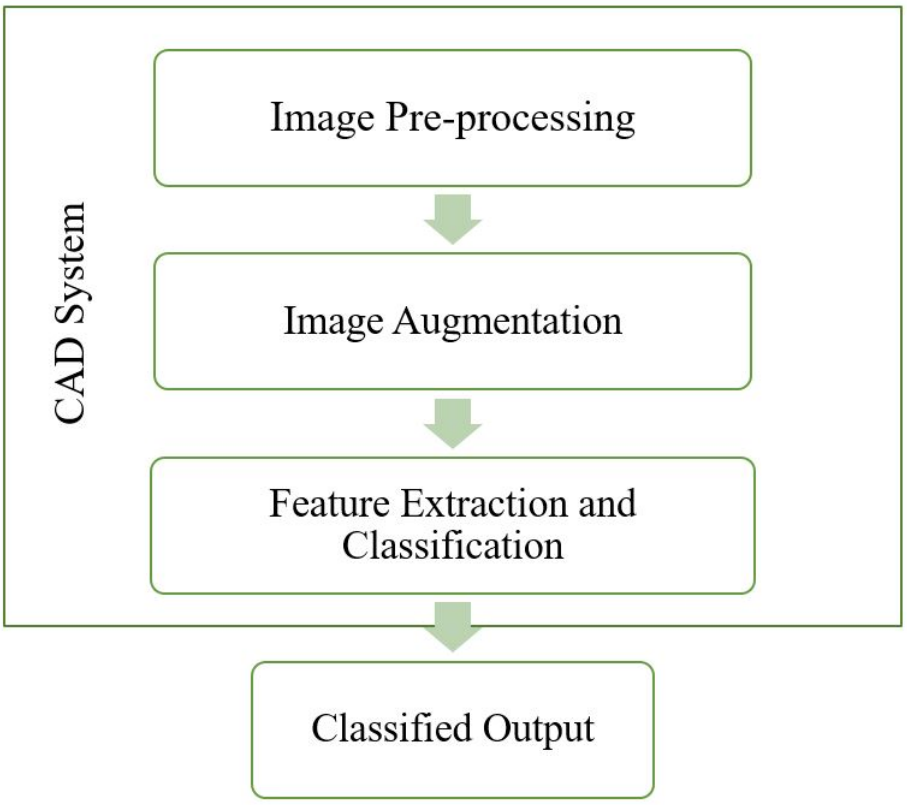


FIGURE 2.3: CAD System Overview

this using a special algorithm that analyzes the color of the lesion and sees how different it is from the rest of the skin. The algorithm is tested on large publicly available datasets of skin lesions and the results are very satisfactory. The new algorithm worked better than other algorithms that are currently available for finding melanoma.

Dermoscopy image analysis (DIA) [19] is an area of research where researchers analyzed skin lesions pictures to help identify melanoma. There are many different methods being developed in this field, but it can be hard to keep track of all of them. The main aspect of this thesis is feature extraction. This is where researchers aimed to identify distinctive characteristics of the images that could assist in diagnosis. The authors of this thesis thoroughly examined the various types of features that have been used so far, discussed their advantages and limitations, and proposed areas for future research. By doing this, their intention is to help the researchers understand what has already been done and to uncover potential areas for new discoveries.

2.2.2 Machine Learning/Deep Learning Techniques

The recent improvements and advancements in ML/DL have empowered the development of new techniques that utilize ML and DL-based algorithms for diagnosing a diverse range of medical conditions, including melanoma. These techniques are commonly applied in the field of computer vision and related fields.

Several studies [20], [21] have proposed the use of ML and DL-based detectors for the early and accurate diagnosis of various medical conditions, demonstrating the potential of these techniques to significantly improve patient outcomes.

Due to rapid advancements in ML technology, the potential of artificial intelligence has been increasing, leading to active research on its applications in dermatology. Recent studies have explained the effectiveness of convolutional neural networks (CNNs) in classifying melanoma images, with accuracies comparable to those of dermatologists.

Until now, there have been no reports on how well a CNN performs in a clinical image classification task using only clinical images of pigmented skin lesions when compared to dermatologists.

In this study [22], the researchers collected a total of 5846 clinical images of pigmented skin lesions from a diverse group of 3551 patients, including both malignant and benign tumors. They randomly selected 666 patients to create a test dataset and a training dataset with the remaining images, where they gave bounding-box annotations to 4732 images. A faster region-based CNN (FRCNN) was trained using the training dataset and tested its performance on the test dataset. The study included a group of ten board-certified dermatologists and ten dermatologic trainees who underwent the same tests and the diagnostic accuracy of the dermatologists is compared with that of the FRCNN.

The accuracy of FRCNN is found to be better than that of the dermatologists for both six-class and two-class classification tasks. The researchers wanted to make this system available to the public so that people can use it to improve the diagnosis of skin cancer.

In recent years, deep learning-based techniques have been increasingly used to improve the accuracy of classifying similar features. One such approach, presented in a study referenced as [23], is a hybrid model that combined shallow learning-based pre-trained models, like AlexNet, with deep learning-based techniques such as support vector machines.

Skin diseases are more prevalent than other types of diseases, and they can be caused by various factors such as fungal infections, bacteria, allergies, or viruses. Laser and Photonics-based medical technologies have greatly improved the accuracy and speed of diagnosing skin diseases. However, the cost of such a diagnosis is still very high, making it unaffordable for many people. Thus, image processing techniques can be used to develop automated screening systems for dermatology at an early stage. The extraction of features played a crucial role in accurately classifying skin diseases. Computer vision techniques have been used in a variety of methods for detecting skin diseases. The proposed method worked by taking color images as input, resizing the images, and extracting features using a pre-trained convolutional neural network. The extracted features are then classified using a multiclass SVM algorithm. The system has been successfully tested to detect three different types of skin diseases with an accuracy rate of 100%.

Skin diseases have a tendency to appear in various forms, lack of qualified dermatologists, and the need for timely and accurate diagnosis create the need for automated CAD. *Bajwa et al.* [24] aimed to expand the previous research in CAD by exploring how to classify a large number of skin diseases using deep learning. Their goal is to enhance the accuracy of disease classification and make use of disease taxonomy to improve the system's overall performance. The authors employed state-of-the-art Deep Neural Networks (DNN) trained on the two largest skin image datasets, DermNet and ISIC Archive which are publicly available. This classification improved the performance of these models. The results showed significant advancements in skin disease classification using Deep Learning. Especially on DermNet, they achieved remarkable performance with an accuracy of 80% and an Area Under the Curve (AUC) of 98% in accurately classifying 23 diseases. They also extended this achievement by successfully classifying all 622

unique sub-classes within the dataset, achieving an accuracy of 67% and an AUC of 98%. In context of ISIC Archive, they achieved an average accuracy of 93% and an AUC of 99% in classifying all 7 diseases. The study concluded that DL has immense potential in accurately classifying a wide range of skin diseases, nearing the level of human accuracy. This showed that it has a significant impact on practical real-time skin disease diagnosis by supporting medical experts in large-scale screening through the analysis of dermoscopic images.

Milton et al. [25] demonstrated the findings of the International Skin Imaging Collaboration's (ISIC) largest skin image analysis challenge in the world. ISIC is a global partnership that has developed the world's most comprehensive public repository of melanoma images. The challenge is conducted in 2018 during the Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference in Granada, Spain, and covered over 12,500 images spread over three different tasks. During the challenge, out of 900 people who registered for data download, 115 participated in the task of lesion segmentation, 25 individuals took part in the lesion attribute detection task, and in the disease classification task, 159 participants are involved.

The researchers established new protocols to evaluate the performance of segmentation algorithms and to test their ability to generalize. The results demonstrated that even the top segmentation algorithms failed on over 10% of the images on average. Furthermore, algorithms with similar performance on test data may have different abilities to generalize, highlighting the importance of generalization in machine learning applications. These findings are critical for regulatory agencies overseeing the increasing number of machine learning tools in the healthcare domain and set a higher benchmark for future public challenges in the healthcare field.

Zhang et al. [26] talked about using an automated system to classify skin lesions in medical images to improve diagnosis and reduce melanoma deaths. They mentioned that while deep learning models have made breakthroughs in image classification, skin lesion classification remained challenging because there is not

enough data and the images can be very similar to one another. To overcome these challenges, the authors proposed a new model called the attention residual learning convolutional neural network (ARL-CNN) that used a new attention learning mechanism to focus on the crucial aspects of the image. They tested their model on an ISIC2017 dataset of skin lesion images and found that it outperformed other models in accurately classifying the lesions.

Alom et al. [27] presented a novel DL architecture called the NABLA-N network that incorporated advanced feature fusion techniques employed in decoding units to enhance the accuracy and efficiency of dermoscopic image segmentation tasks. The proposed model aimed to improve the representation of low to high-level feature maps for better semantic segmentation. In terms of qualitatively and quantitatively, N-NET showed better results with similar network parameters compared to other models. Furthermore, skin cancer classification is performed using the Inception Recurrent Residual Convolutional Neural Network (IRRCNN) model. They compared the proposed N-Net and IRRCNN models on benchmark datasets from the ISIC-2018 for the segmentation and classification of skin cancer. The results obtained from the experiments highlighted the effectiveness of segmentation tasks over the Recurrent Residual U-Net (R2U-Net). additionally, the classification model demonstrated a testing accuracy of 87% for skin cancer classification on the ISIC-2018 dataset.

Zafar et al. [28] reviewed the different techniques that are used to analyze skin lesions, containing preprocessing, feature extraction, segmentation, selection, and classification. Even though results of these techniques are far better, there are still challenges in analyzing complex and uncommon features of skin lesions. Their prime research goal is to analyze all the existing techniques that are being used for the detection of skin cancer and to enhance the previous techniques to help future research.

Mijwil et al. [30] used DL to analyze over 24,000 images of skin cancer. Different models are tested with different settings to see which one worked the best at identifying whether a skin cancer is malignant or benign. High-quality images are

used and the InceptionV3 model worked the best upon testing with an accuracy of about 87%.

The conventional method used for detecting changes in melanoma screening is called short-term monitoring. In this method, experts observe changes in skin lesions over a span of the first two to three months. Although, this method focused on melanoma type and also depends on clinicians' experience. In this research [29], a new approach utilizing deep learning is introduced to automatically detect short-term lesion changes. This method consisted of a deep neural network to analyze the similarity between two dermoscopy images captured for the purpose of skin cancer diagnosis within a brief time frame. The network is trained to detect whether the lesion has shown some changes or not. The proposed method used a basic structure to extract features from the images of the lesions and incorporates a segmentation loss term to replicate the decision-making process of clinicians. The research team used 1,000 pairs of lesion images dataset to evaluate the proposed method, and the results suggested that the model performed well for screening melanoma disease.

2.2.3 Federated Learning Techniques

Data privacy is very essential, especially for medical data where it's important to keep patient information private. But large amounts of datasets are needed for research and training ML/DL models that can help with personalized treatments. FL has been suggested as a solution, as it relies on sharing the models instead of personal data. This means that person's data stays where it is and never leaves the user's device or the organization, in which it is collected. FL is an emerging research area and can be applied in many domains, including confidential healthcare datasets. A systematic literature review [12] has been conducted to explore the concept of FL and its usefulness for confidential healthcare data.

Huang et al. [14] highlighted the digital health future with FL. They emphasized how FL could be the solution for medical in the future because of privacy concerns

and data being stored separately. They also discussed the challenges and things that need to be considered when using FL in healthcare.

Another main feature of FL in (Internet of Things) IOT is discussed in this research paper [31]. The more we have IOT devices, the more data we have, some of which might contain private information. The traditional way of learning and processing this data in a centralized way is becoming too costly and raises privacy concerns. FL has come up as a promising alternative. In FL, clients collaborate in training models without having to share data in centeralized server, which reduces communication and storage costs and protects user privacy. However, implementing FL in IoT networks have some challenges, and this paper also discussed those challenges and some recent approaches to addressing them.

FL is categorized into two main types. One is cross-device FL where the machines are mobile or edge devices and there can be millions of them, and cross-silo FL where the machines are organizations or sites and there are usually fewer of them. This paper [14] is focused on cross-silo FL, and discussed its applications and challenges They reviewed the existing solutions and suggested future research directions.

The healthcare industry has a huge amount of data that is spread across different points and owned by different organizations or entities. In the past, researchers have focused their attention on centralized algorithms where all the data is placed in one central location, This paper [33] objective is to develop a way to predict hospitalizations for cardiac events using a distributed algorithm that allows multiple data holders to collaborate without sharing raw data. They used a method called soft-margin l1-regularized sparse Support Vector Machine (sSVM) classifier and developed an algorithm called iterative cluster Primal-Dual Splitting (cPDS) to address the issue through a decentralized approach. The cPDS algorithm converged faster than centralized methods and achieved similar prediction accuracy. They are able to identify significant features uncovered by the algorithm that have predictive value for future hospitalizations, which can help inform prevention efforts.

Another application of FL can help identify COVID-19 through the use of DL and computer vision with Chest X-ray images. However, patient data is a bit concerning for the hospital in order to collect Chest X-ray images which is a challenge. They proposed FL is a better solution to address this issue. They used different machine learning models (MobileNet, ResNet18, ResNeXt) to train the models and training results are compared with and without the FL technique. The results showed that ResNet18 is the best model for training, while ResNeXt is the best model for identifying COVID-19-labeled images. MoblieNet had the least number of parameters. Overall, the study [34] suggested that ResNeXt and ResNet18 are better models for identifying COVID-19.

The field of Brain-Computer Interfaces (BCI) uses DL methods to classify brain readings, but it has been hard to get enough data to train the models. This is because electroencephalographic (EEG) signals are private and can't be shared easily.

Ng et al. [36] proposed a new way to protect privacy and still use deep learning to classify EEG recordings. They called it federated transfer learning (FTL) and it used a technique called domain adaptation to extract useful information from multiple EEG datasets without sharing the actual data. They tested FTL on a dataset for motor imagery classification and found that it performed better than other DL techniques, even when it had fewer data to work with. FTL achieved higher accuracy without actually sharing any private data.

Rahman, et al. [37] used the FL technique to make a machine learning model that predicts if someone with COVID-19 will die within 7 days. FL is a way to use data from different places without sharing the actual data. The researchers collected data from five hospitals in the same health system. They made two different kinds of models: one in which model trained on data from each hospital and the other in which model trained on data from all hospitals. The results demonstrated that the FL models outperformed just as well as the local models. They aimed that FL can help us make better DL models for COVID-19 without breaking patient privacy.

The COVID-19 pandemic has produced an urgency for better management and accurate diagnosis of the disease. The current guideline recommends using a test called RT-PCR. Chest CT scans are another tool for detecting COVID-19 patterns in the lungs. To help doctors analyze CT scans and diagnose COVID-19, ML and DL models have been developed. However, these models used large data that comes from various hospitals and can cause data heterogeneity problems plus also when it uploads to a central server, it leads to compromises of hospital's private data. Then they found a technique called "federated and semi-supervised learning" [38]. This technique helps the utilization of data from different hospitals without compromising hospital's private data plus it also reduced the need for medical experts to manually annotate the data, which can be time and effort-consuming. They tested this technique on three different countries China, Italy, and Japan to detect COVID-19 using CT scans. The results showed the significant potential to help doctors diagnose and manage COVID-19 patients more accurately.

Another FL approach is used in this study [37] to predict how long a patient will stay in the hospital. This study helped the hospital's management so they can arrange their resources and provide good treatment. ML/DL can help with this, but it's hard to get enough data because hospitals keep patient information private. This study proposed a solution where hospitals can train their own models on their own data, and then send just the important information to a central server. The server then combined all the important information to make a prediction about how long a patient will stay. They tested this approach using data from ten different hospitals and found that the combined model is better than each hospital's model alone. The combined model worked better when many hospitals gave their data.

Dou, et al. [39] worked on detecting lung abnormalities in COVID-19 patients. Their prime goal is to create a model that can accurately classify the disease using chest CT scans. The authors used an FL approach in their system for training and testing the model that involved the data used from multiple hospitals in different countries without sharing the data between them plus ensuring data privacy and security. The system is tested on patients from seven different centers, and it

is found to be effective at detecting the disease. The authors also performed case studies on longitudinal scans to estimate the lesion burden for hospitalized COVID-19 patients. FL can be a useful tool to develop models during pandemics across institutions and countries without sharing large amounts of sensitive data.

The domain of medical imaging has been greatly improved by the use of AI in assisting radiologists with patient diagnosis. However, building accurate DL models using small datasets can be difficult for individual medical sites.

Radiologists have to do a lot of work to prepare the datasets because medical images are not labeled properly for AI training. This process can be unsuitable with the growing number of medical images captured annually. To solve this problem, this research proposed [36] using a new learning framework called FL. This framework allows individual sites to work together to train a global model without directly sharing datasets, which protects patient privacy. FL can also improve the accuracy of the model by putting together the results from multiple sites. This approach can help in solving the issue of inadequate supervision when training DL models with small datasets. However, there are still some challenges left that need to be addressed in future while adopting FL.

Skin diseases are rapidly increasing worldwide and are considered major chronic diseases. in particular, Skin tumors can be life-threatening if not diagnosed early. Doctors use various tools, including manual and computer vision-based tools, to diagnose skin tumors, but they can sometimes misinterpret the disease and take longer to analyze it. However, using cutting-edge technologies such as DP with a federated machine learning approach, doctors can diagnose the level of severity of skin conditions more accurately.

Hashmani, et al. [40] proposed an FL-based adaptive ensemble as the primary classifier to help dermatologists diagnose skin diseases better. The proposed architecture comprises a local edge and a global server point that collaborate to diagnose diseases and continuously enhance diagnostic accuracy over time. Experiments are conducted using dermoscopy images from the ISIC-2019 dataset to assess the accuracy and flexibility of the model's classification. This study has

the potential to the development of a new hardware device to help dermatologists diagnose skin tumors using federated machine learning.

Agbley, et al., [41] focused on the detection of melanoma skin disease with FL technique. The researchers used two types of data: pictures of skin lesions and clinical information about patients. They compared the FL model performance with a Centralized ML/DL model that uses all the data in one place. The FL model performed equally in comparison with centralized models and even did better in some cases. The FL model also showed better results in identifying melanoma disease. This means FL can be a better solution for creating better ML models while still protecting patient privacy.

2.3 Research Gaps

Studies have shown great potential in exploring of FL techniques with different DL models to achieve better results. However, there is still potential for enhancing the accuracy of detecting melanoma.

Most existing models have been developed and tested on uniformly distributed datasets, which limits their generalizability on unbalanced datasets. There is a need to develop models that can perform well on the unbalanced distribution of skin disease datasets.

2.4 Problem Statement

- The existing conventional methods for classifying melanoma usually based on ML/DL fall short of addressing privacy concerns effectively.
- Although FL models for melanoma detection have been studied with balanced datasets, there hasn't been a significant exploration into how these models perform with unbalanced datasets.

2.5 Research Methodology

This research work is categorized into two phases. The first phase is to implement the state-of-the-art DL model and compare the performance with FL model in the context of melanoma detection. In the second phase, an improved FL approach is proposed that aimed to enhance the accuracy of FL models.

To achieve this, a dataset of skin lesion images from the International Skin Imaging Collaboration (ISIC-2020) [42] is collected. After pre-processing the dataset to ensure consistency and to be suitable for analysis, the dataset is divided into training, validation, and testing sets in a ratio of 60:20:20, respectively. It helps to evaluate how well the model is performing on new and unseen data.

For the conventional DL model, the ResNet50 model is employed. The model is used for training and testing purposes, which takes melanoma images and iteratively updates its weights and biases to improve its accuracy in making predictions.

For FL model, a federated learning framework is used that allows multiple institutions to collaboratively train a global model without sharing their data. The FL framework is implemented using FLOWER architecture that is open source and customizable. Training and testing are performed on the same dataset. After testing the performance of the FL approach on unbalanced datasets, it is found that the results are not satisfactory. As a result, an improved FL approach is proposed to address this issue using a fine server aggregation algorithm.

The proposed approach is then evaluated using the same testing set and its performance is compared with the ResNet50 model.

2.6 Research Contribution

With the rapid advancement in technology and the need for efficient healthcare services, Smart Healthcare Systems have emerged to meet the needs of the new

era. One of the critical aspects of healthcare is the detection of skin diseases, particularly melanoma, which is important for timely diagnostics and treatment.

The main research contributions are as follows:

- To perform a comparative analysis between existing DL model and the FL model for melanoma detection over the publicly available ISIC-2020 dataset [42]
- To improve the accuracy of detection for unbalanced datasets.
- To purpose an improved aggregation technique on the server side for an unbalanced dataset.

Chapter 3

Federated Learning

3.1 Overview

In 2017, Google [43] introduced a new approach to machine learning called Federated Learning.

In this approach, multiple clients work together to solve an ML problem, with a central server. Each client has their own set of data that they keep locally and do not share with others. Instead, updates are made to the ML model on each client's local data, and only these updates are shared with the central server for aggregation and analysis. This approach ensures that privacy is maintained, as raw data doesn't leave the client's device and only the necessary updates are shared with others.

FL aims to minimize loss functions across multiple clients [44], [43]. In other words, the goal is to reduce the amount of error or inaccuracy in the ML/DL models used by each client.

The local loss functions are averaged across all clients, with each client's contribution weighted based on the number of samples they have. The weights are then updated to improve the overall model. The process continues until the best level of accuracy is achieved.

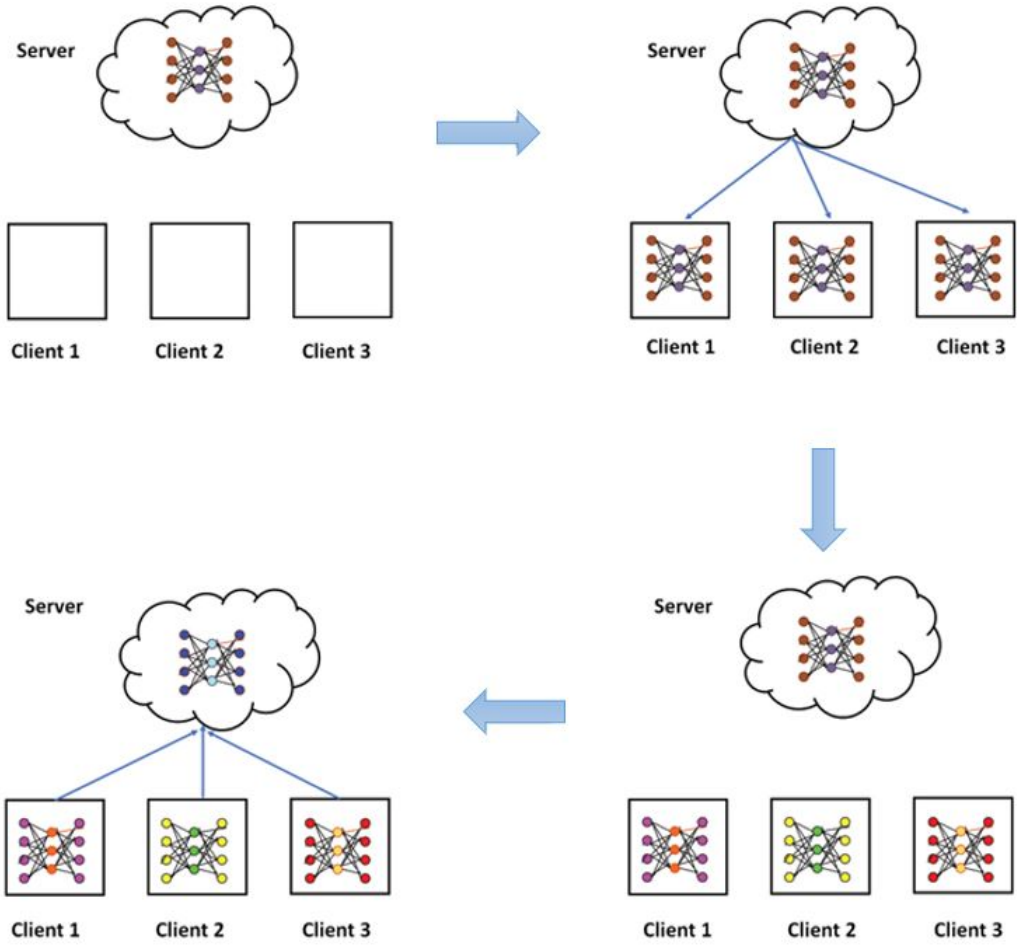


FIGURE 3.1: Training Process in FL

3.2 Training Process

The process of training in federated learning usually involves the following steps [12]:

1. Client Selection: It involves selecting clients based on specific eligibility criteria. For instance, in a cross-device setting, only idle, charging, and devices connected to a free wireless network may be selected for training, as seen in Google's Gboard for Android devices [45]. Similarly, in cross-silo settings, it is important for the clients to be more reliable to maximize the availability of clients for training.

2. Broadcast: Current weights of the model are shared with the selected clients.

3. Client Computation: The selected clients use the training program, such as Stochastic Gradient Descent (SGD), to update their local models.

4. Aggregation: The server performs aggregation after receiving updated weights from each client.

5. Model Update: After the aggregation step, the server shares the updated model weights locally for the next round of the training process.

Overall, federated learning aims to train a model without directly accessing the user's data from multiple clients. The training process is also depicted in figure 3.1. This approach is useful in cases where privacy concerns or data decentralization make traditional centralized training methods impractical or impossible.

3.3 Federated Learning for Medical Images

FL is becoming increasingly important in the field of medical images. There are several reasons that make FL important in the medical domain [31]. Some key advantages are as follows (also shown in figure 3.2):

1. Privacy: Each organization involved in the system would only learn the information that is necessary for it to perform its intended role. During the FL process, the raw data never leaves the individual devices, and only the model updates are sent to the central server. This approach helps to minimize the risk of medical data leakage, which is a critical concern in the medical industry where patient privacy and data security are of utmost importance.

2. Performance: The limitations of hospital data can sometimes prevent them from gathering sufficient data to build a high-quality ML/DL model. However, using the FL framework, multiple hospitals can collectively train a high-quality model that benefits from the data collected by other hospitals without compromising their privacy.

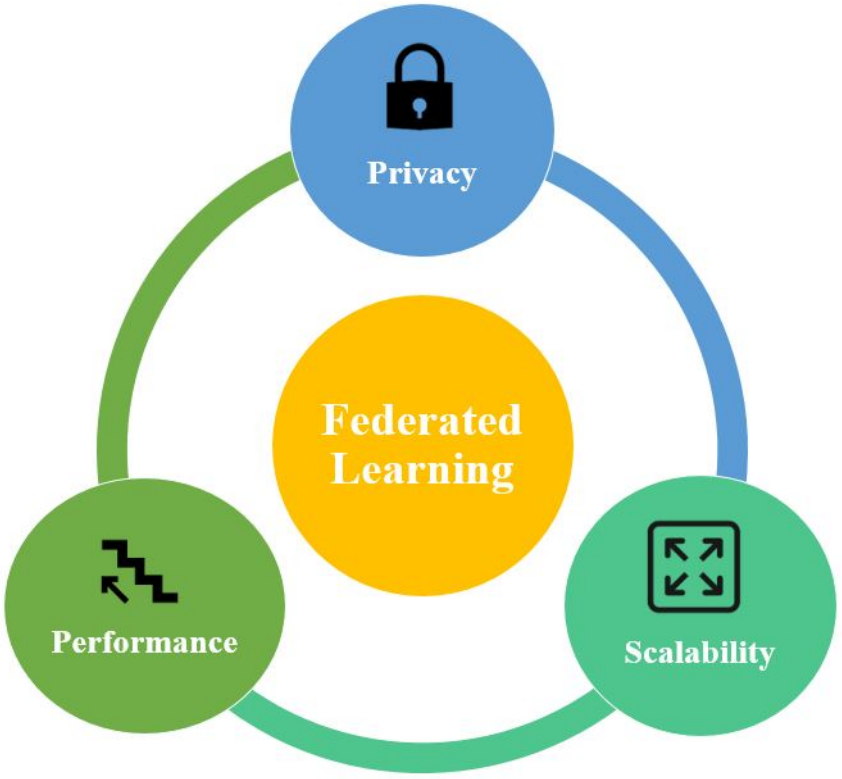


FIGURE 3.2: Advantages of FL in Medical Domain

The FL framework also allows for periodic updates to the local model, ensuring that the edge node can improve its model performance continuously. This approach is highly effective in enhancing the model's performance, which may not be achievable by individual hospitals on their own.

3. Scalability: The distributed nature of FL enables the utilization of computation resources located at multiple hospitals in a parallel manner. With the increase in the hardware capability of edge devices, the data size of each entity becomes large. Centralizing all data to the server can waste computing resources at the edge or create high traffic on wireless communication networks, which can be an obstacle to network scalability. With the nature of distributed learning FL enhances the scalability by adding more hospitals or organizations to the framework network without increasing the extra burden on the central server. Additionally, FL eliminates the amount of raw data that needs to be sent, making it more scalable and cost-effective for communication in low-bandwidth IoT networks.

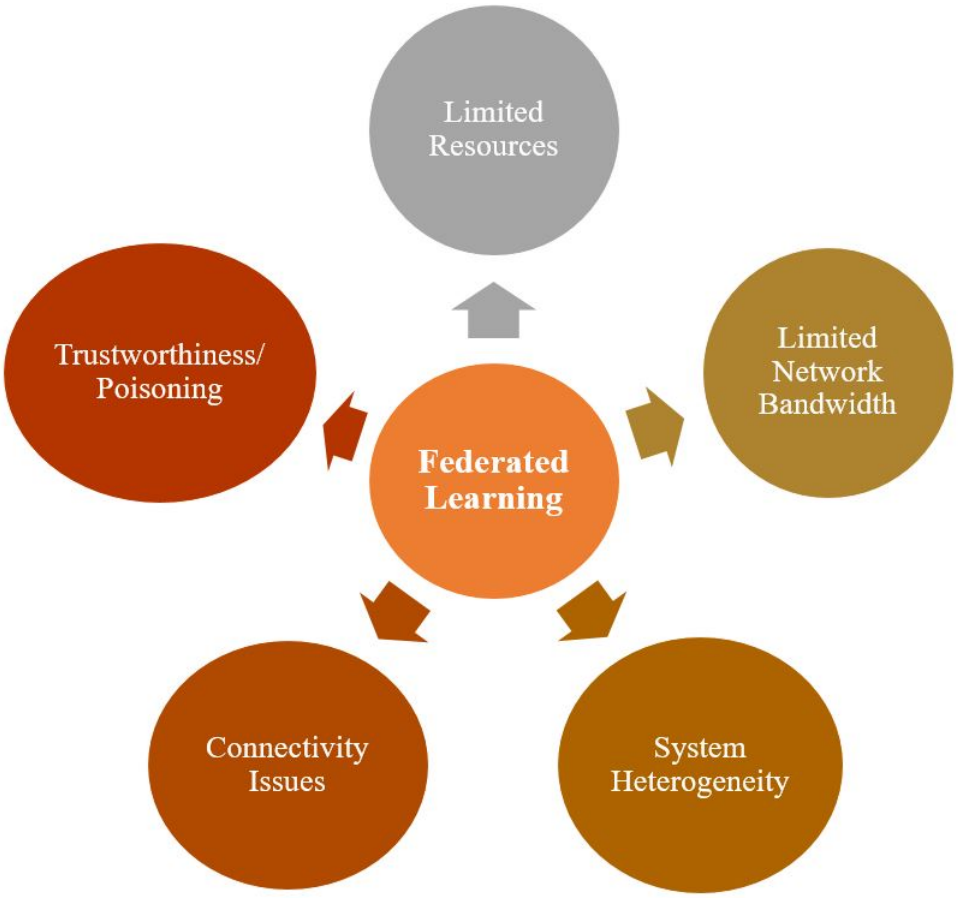


FIGURE 3.3: Challenges in FL

3.4 Challenges in Federated Learning

In order to fully utilize the benefits of FL in the medical domain, there are several challenges that need to be addressed [46]. These challenges can be considered major obstacles in enabling FL to its fullest potential.

1. Limited Resources: The deployment of deep neural networks, on each node, is challenging due to limited resources like computation power and memory. It is difficult to process deep neural network models on resource-constrained devices for storing model weights, parameters, and intermediate results.

   To address this challenge, Researchers are working on two ways to make it more efficient. Like creating models that are easier for computers to handle, and also compressing models to make them smaller and faster. This will help to save energy and computing power.

2. Limited Network Bandwidth: Communication bottleneck is a challenge in an FL-based environment due to the limited bandwidth of wireless networks. This leads to inefficient communication between the server and clients.

   To reduce the bandwidth demand during FL, various methods have been explored, including gradient compression, sampling-based framework, and client-edge-cloud hierarchical aggregation framework. These techniques can reduce bandwidth costs

3. Connectivity Issues: The intermittent connectivity of devices in large-scale systems is a serious obstacle to the efficient management and scheduling of clients in the FL framework.

   Most FL studies are based on synchronous updates, but in real-world settings, the local training speed and availability of devices vary, making it hard to update. Asynchronous aggregation schemes are being used to reduce the risk of connectivity issues.

4. System Heterogeneity: In designing an FL framework on different hospital devices due to the diversity in their hardware, software, and data collection patterns.

   Different devices have different resources and use different DL frameworks, resulting in various model formats that need to be aggregated. These diversities make it difficult to design an FL framework that can adapt to any new data and distribute computational workload over various devices. However, the heterogeneity beforehand and making adjustments accordingly can help in overcoming this problem.

5. Trustworthiness/Poisoning: The security challenges posed by client fake data and the potential for distributed denial of service attacks.

   The detection of malicious or broken devices and the implementation of a lightweight security protocol using FL are important to avoid poisoning the training process and negatively affecting the global model. However, using secure aggregation protocols protects personal data while minimizing the computational burden on every device.

3.5 Federated Learning Frameworks

Numerous open-source FL frameworks are available. In this thesis, thorough investigation and exploration are performed and different FL frameworks are examined based on their features, in order to identify their strengths and weaknesses.

1. **TensorFlow Federated (TFF):** TFF [47] is an open-source framework that allows for the implementation of FL systems. It expands the popular ML framework TensorFlow to allow training models on distributed data sources while maintaining data privacy. TFF is appropriate for collaborative learning situations where numerous clients desire to contribute data and maintain the privacy of their data as it supports the distributed computation of model training and evaluation. For building of FL system, this framework provides a high-level API's that helps researchers and developers to focus on the design of the algorithm, rather than implementation details.

   Limitations: Despite its benefits, TFF has also several limitations like communication overhead. It involves additional latency in communication with a central server and client devices. So, performance can be degraded when large number of clients to handle and where network connectivity is poor.

   Another main limitation of TFF is that supports training models with homogeneous architecture i.e. recurrent neural networks and feed-forward neural networks. However, more complex model architectures like multi-layers neural networks with custom layers, require additional effort to integrate and adapt with TFF.

   In conclusion, TFF is a useful platform for the implementation of FL systems. It offers a high-level API, support to target mobile devices and IoT devices, and privacy-preserving technology. Additional study is still ongoing to overcome the framework's shortcomings and standardize the FL field.

2. **FedML:** FedML [48] is another framework that is being used in FL systems. It provides three computing paradigms: on-device training, especially

for edge devices, single-machine simulation, and distributed computing. The model training module mainly uses PyTorch but users can implement their own model and manage it according to their needs. FedML library includes two main components: FedML-API which provides high-level APIs that contain for testing and evaluation and FedML-core for low-level APIs that support various communication protocols, including socket-based communication and HTTP-based communication that makes it easy to integrate with different platforms and devices.

Limitations: Despite its advantages, It also has some limitations like FedML offers a comprehensive set of tools for developing and evaluating FL algorithms but it may not cover all corner cases or user requirements, which limits its functionality. For new users to get started with this framework is difficult due to the limited documentation and online resources available as compared to other established frameworks like TensorFlow Federated, PySyft and FLOWER.

3. **PySyft:** Pysyft [49] is a Python-based open-source framework for secure and privacy-preserving DL. It uses privacy-enhancing mechanisms to ensure privacy and separates private data from model training. The following are the core features of PySyft like capacity to send and receive data securely, run remote functions, and the ability to create a virtual worker to simulate a remote machine. Furthermore, it offers a number of high-level interfaces for basic ML operations like data loading, model training, and evaluation. Popular frameworks such as Pytorch and Tensorflow can be integrated with PySyft framework and can be used in conjunction with these frameworks to enable secure collaborative ML.

   Limitations: PySyft is a powerful framework that enables private and secure ML in federated learning but it also has some limitations in model learning as compared to other FL frameworks because of its focus on security and privacy. Another main limitation of PySyft is fewer resources available for finding solutions and troubleshooting issues despite having a growing community.

In a nutshell, PySyft is a flexible and powerful FL framework, but its complexity and performance may not make it the best choice for all use cases.

4. **FLOWER:** A FRIENDLY FEDERATED LEARNING FRAMEWORK [50] is an open-sourced library that can work with different types of devices and can be used for research and development purpose under the license of Apache 2.0.

   The basic aim of FLOWER is to provide an easy-to-use and secure federated learning experience especially for developers and data scientists.

   It is a lightweight and scalable Python-based framework that streamline implementation of FL algorithms and also includes a number of built-in features like the ability to adjust the learning rate, batch size, and other hyperparameters that make it easy to customize the training process. Additionally, FLOWER supports various types of ML and DL algorithms plus also helpful in reinforcement learning.

   Basically, It has been designed to work in a decentralized way, where each client share only model instead of raw data and can decide when to participate and when to drop out.

   Furthermore, It also provides mechanisms for every clients to communicate securely with the central server, which is crucial for ensuring the privacy and security of the data. In terms of capability, Flower facilitates in order to execute FL experiments on a large scale and in system heterogeneous scenarios. With only a pair of high-end GPUs it can perform FL experiments up to 15M client size. Researchers can migrate experiments into real-life scenarios to examine other parts of the design.

3.5.1 Flower Architecture

Flower provides high-level and easy-to-use API for building federated learning systems. It has numerious set of rules and libraries for training and evaluating ML/DL models in a distributed and decentralized manner. At its

core, Flower allows multiple clients to train ML/DL models using data that is distributed across multiple devices without having to share the raw data. This means to train the model locally on each device or in organization, and only share the model updates with the central server.

The protocol called Remote Procedure Call (RPC) that enables a program to request services from other programs on a network without having to know the underlying netwrok configuration. gRPC is a high-performance, open-source RPC framework that can operate in various environment. Flower uses the gRPC framework to enable the server to execute different steps of the training process.

The Flower framework consists of fundamental elements that work together to facilitate FL. These components include:

- Client: In the Flower framework, clients are responsible for training and evaluating ML/DL models on local data. A client can be a device or machine that participates in the FL process. Each client execute its own instance of the DL model and sends the updated model to the server periodically. It extends on the NumpyClient class object which has 3 methods inside it:
  (a) getParameters: It downloads the current model from the server.
  (b) fit: It trains the model on a device and sends the updated weights back to the server once completed.
  (c) evaluate: This method is called by the server when testing is required.
- Server: In the federated learning system server is responsible for coordinating and managing the training process across all of the clients. Clients sends the updates to server then server aggregates them and sends new model updates back to the clients. It includes the averaging algorithm and the overall training strategy. In order to participate in the training process, clients usually establish a secure gRPC connection with the server and indicate their readiness. Minimum of two clients

are required to join the process then server initiates procedures over gRPC.

- Strategy: The strategy is implemented on the server side and indicates a set of rules and algorithms that determine how the model updates will be aggregated on the server. There are several different strategies available in the Flower framework i.e. Fed avg, weighted averaging, each with its own strengths and weaknesses.
- Backend: The backend is the framework used to implement the ML/DL model. The Flower framework can be integrated with several different backends like PyTorch, TensorFlow, and Keras.
- Communication: The communication layer in the Flower framework is implemented using HTTP or WebSockets, depending on the requirements of the application. The layer is responsible for transmitting data between the clients and the server.

3.5.2 Advantages of Flower Framework

FLOWER is a popular open-source federated learning framework that provides several advantages over other federated learning frameworks:

(a) Ease of use: Flower has a user-friendly interface and simplifies much of the complexity involved in FL. It allows easy integration with popular ML libraries such as TensorFlow, PyTorch and keras, making it a great choice for developers who are new to federated learning.

(b) Scalability: It is highly scalable. Computing resources may be dynamically allocated and released depending upon the no. of clients with its built-in elastic resource management. This indicates that FLOWER can train models effectively with minimal wastage of resources and can add or remove clients on the fly.

(c) Flexibility: Flower is flexible in its nature and can be used to train a wide range of ML models. It supports deep neural networks and

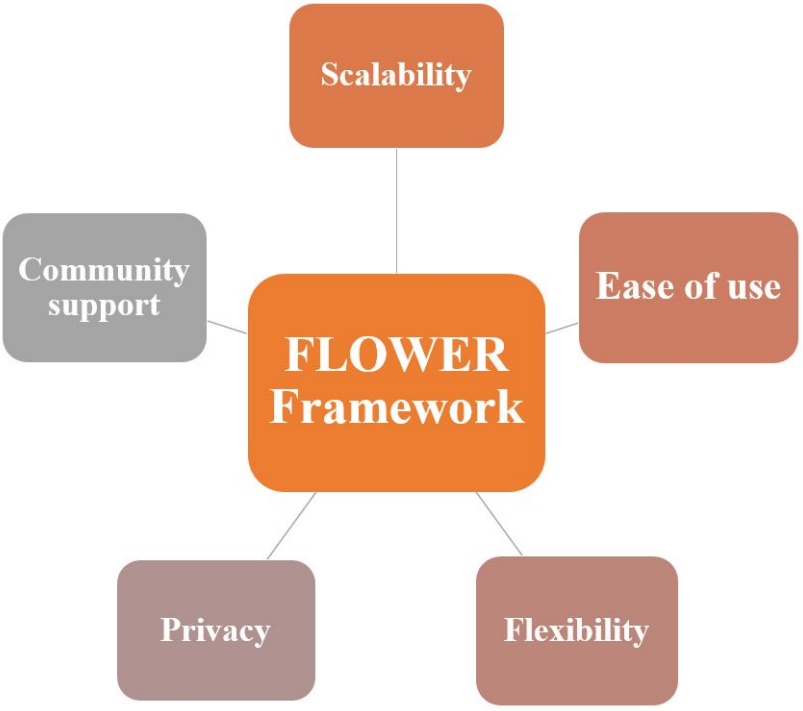


FIGURE 3.4: Advantages of Flower Framework

different learning algorithms that can be customized to meet specific requirements.

(d) Privacy: Flower ensures that sensitive user data is protected throughout the FL process. It is designed with privacy in mind and provides several privacy-preserving features such as differential privacy, secure aggregation, and federated identity management.

(e) Community support: It has a growing community of developers who actively contribute to the flower framework and provide support to users. This means that users can take advantage from ongoing improvements and updates, along with access to a vast array of resources and knowledge.

Overall, Flower is a powerful and flexible FL framework that provides many advantages over other frameworks. Its ease of use, scalability, flexibility, privacy-preserving features, and strong community support make it an excellent choice for researchers and developers which are looking to implement or explore FL in their projects.

3.6 Summary

In this chapter, the latest approach to machine learning called federated learning is presented. Then it signifies the step-by-step process of federated learning and how models are trained collaboratively on decentralized devices using distributed data. Further, the challenges that arise in federated learning are explored and the specific need for FL in the medical domain is discussed. Finally, different FL frameworks are examined that have been developed by researchers and industry experts and explored the selection of FLOWER framework is based on its unique advantages.

Chapter 4

Methodology

FL includes multiple clients have their own models and a central server has a global model. Initially, the global model's parameters are used to start the clients' models. Then, each client does training epochs on their dataset. After the local training, the client models are sent to the server, which aggregates them using a method called federated averaging and updates the global model. The updated model's parameters are sent back to each client for further training. This cycle continues until a desired result is achieved on the evaluation metrics or model stops improving.

The overview of the proposed FL setup is shown in figure 4.1.

4.1 Dataset

The 2020 International Symposium on Biomedical Imaging Challenge used a collection of annotated dermoscopy images of melanoma provided by the International Skin Imaging Collaboration (ISIC) [42]. The images are publicly available in the ISIC archive and stored in JPG format with varying sizes as RGB images. The dataset includes a total of 33,126 training images, of which 32,542 are Benign and 584 are malignant images. Figure 4.2 displays some sample images from the dataset.

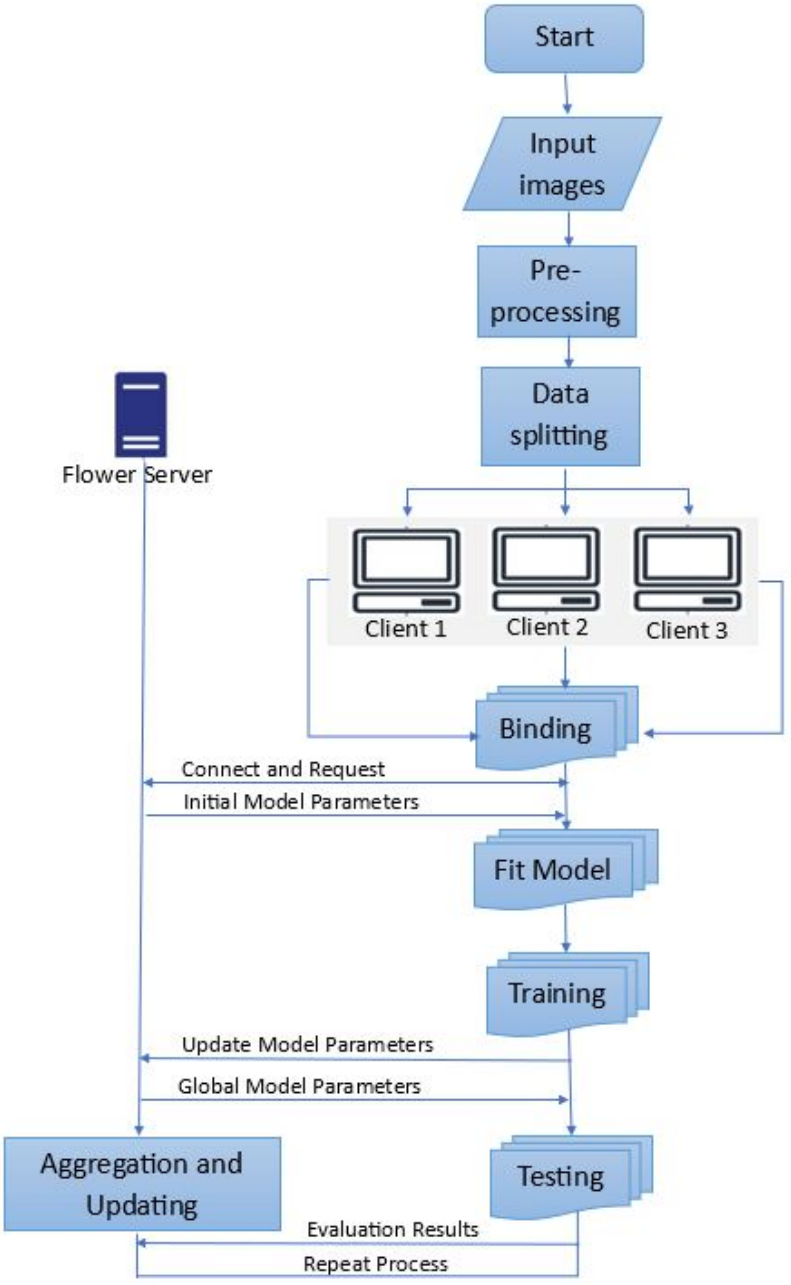


FIGURE 4.1: Proposed FL Framework

4.2 Pre-processing

To utilize DL models for analyzing melanoma skin lesions in dermoscopy images, the proposed approach used preprocessed labeled melanoma dermoscopy images. The raw input images go through pre-processing technique before being fed into the deep learning model. To simplify the problem and optimize the model's performance, all the images were resized to a fixed dimension of 224x224x3 using

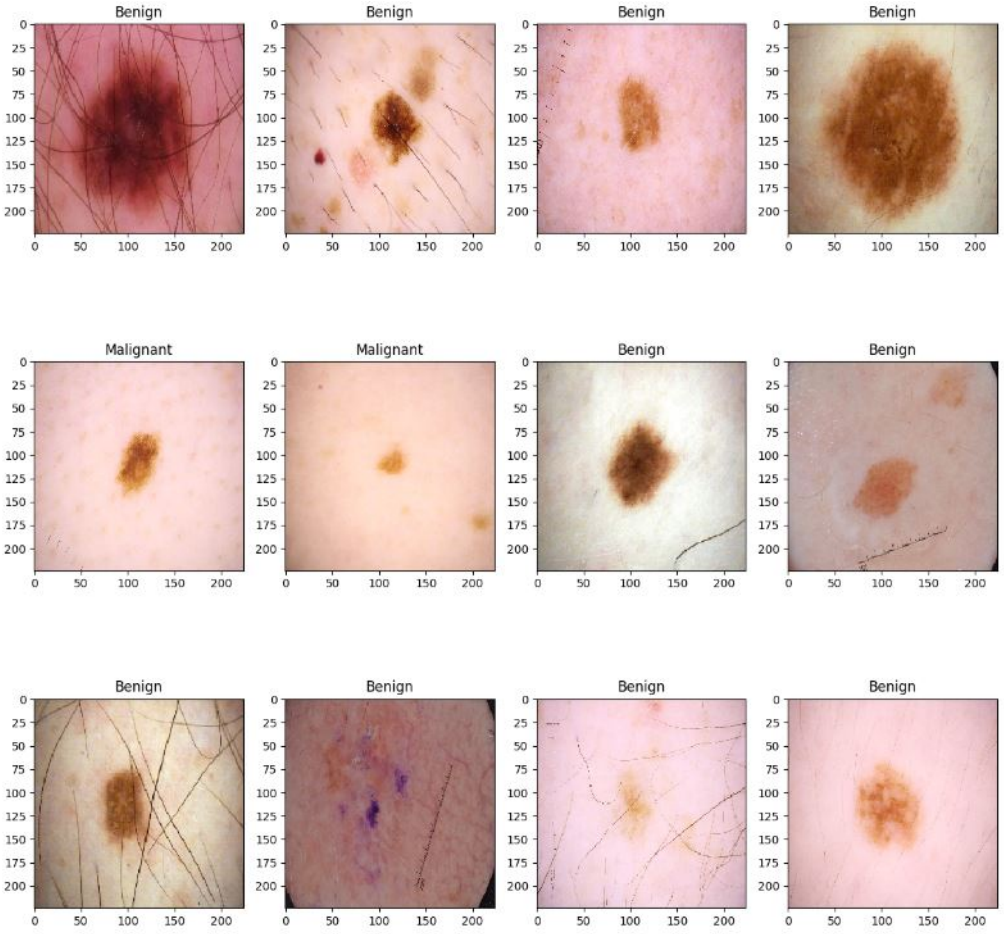


FIGURE 4.2: ISIC Dataset Sample Images

OpenCV. This resizing approach reduces the number of parameters, resulting in lower computational requirements and a faster training process as well as ensures that the images are uniform and suitable for the analysis of DL models. It also helps to handle large datsets because less amount of memory is required to store the images.

The available melanoma dermoscopy image data have been partitioned into 60% training, 20% validation, and 20% testing sets for evaluation of the performance of our model. However, since the test labels are not publicly available on the ISIC website so, only the training data has been split into three sets for three different clients. Each client's private data was used for their respective training and validation sets, while the test data is used to evaluate the overall performance of the global model. The discussed approach is helpful to ensure that the model is tested on completely new and unseen data, simulating a real-world scenario.

TABLE 4.1: Balanced Distribution of Datasets into Three Clients

Images	**Client 1**	**Client 2**	**Client 3**
Total Images	795	794	795
Benign	600	600	600
Malignant	195	194	195

As there are constraints related to time and memory, only a subset of the images are used for training and evaluation purposes. There are a total of 33,126 training images, out of which 2284 images are selected for this purpose (1700 Benign and 584 Malignant). The images are selected keeping in mind that it represents the overall dataset and can provide insights into the model's performance. After selecting the appropriate subset from the dataset, the training and evaluation process can be completed more effectively i.e. saving time and optimizing the utilization of memory without sacrificing the accuracy of the results.

In order to evaluate the performance of our FL model, initially a balanced data distribution approach is adopted by splitting the dataset as shown in Table 4.1.

With the help of this technique, it is ensured that each client in the proposed FL system receives an equal number of benign and malignant images, which helps in avoiding any potential bias in the training process.

With the adaptation of balanced data distribution approach, the aim of fairness and transparency in the FL system by providing each client with an equal opportunity to contribute to the model is achieved.

4.3 Client/Server Connection

In the FLOWER framework [50], a communication channel is used to establish a connection between the server and the client. The client sends a request to the server that contains its IP address and port number to initiate the connection. The server adds the client when it receives the request and sends a response back to the client that confirms the registration. Once registered, the client can contribute to the FL process by downloading the model from the server, training the model

locally, and uploading the updated model back to the server for aggregation.This process is repeated for multiple rounds until the best accuracy is achieved. The communication between the server and the client is usually achieved through a RESTful API or gRPC [50], that allows for efficient data transfer and message passing.

4.4 ResNet50 Model

ResNet50 is a pre-trained CNN that is used for the classification of images. In the proposed approach, ResNet50 model is used as the DL model to classify benign and malignant cases. Due to its deeper architecture than other DL models, ResNet50 model has the tendency to learn more complex features and representations from the input data. Also, it uses residual connections, which allow the model to learn by skipping over unnecessary layers. That helps to prevent the vanishing gradient problem and allows the model to train effectively. Also, It has high accuracy on a wide range of tasks, especially in image classification and object detection.

ResNet50 model takes input images and compiles the model using the Adam optimizer and binary cross-entropy loss, and returns the compiled model. The function then adds some additional layers to the model, including a global average pooling layer, a dropout layer, a batch normalization layer, and a fully connected layer with softmax activation.

The figure 4.3 shows the modified ResNet50 model, which takes an image with size (224x224x3) as input and passes it through the pre-trained layers. The output from the pre-trained layers is then passed through a global average pooling layer to reduce the spatial dimensions of the output coming from the CNN layers to a one-dimensional array. This is achieved by computing the average of each feature map in the output over its entire spatial dimensions. Then a dropout layer is added to prevent overfitting. It helps to learn more robust features that are useful for making predictions on unseen data. The outcome of the dropout layer is passed through a batch normalization layer to improve the stability and speed up the

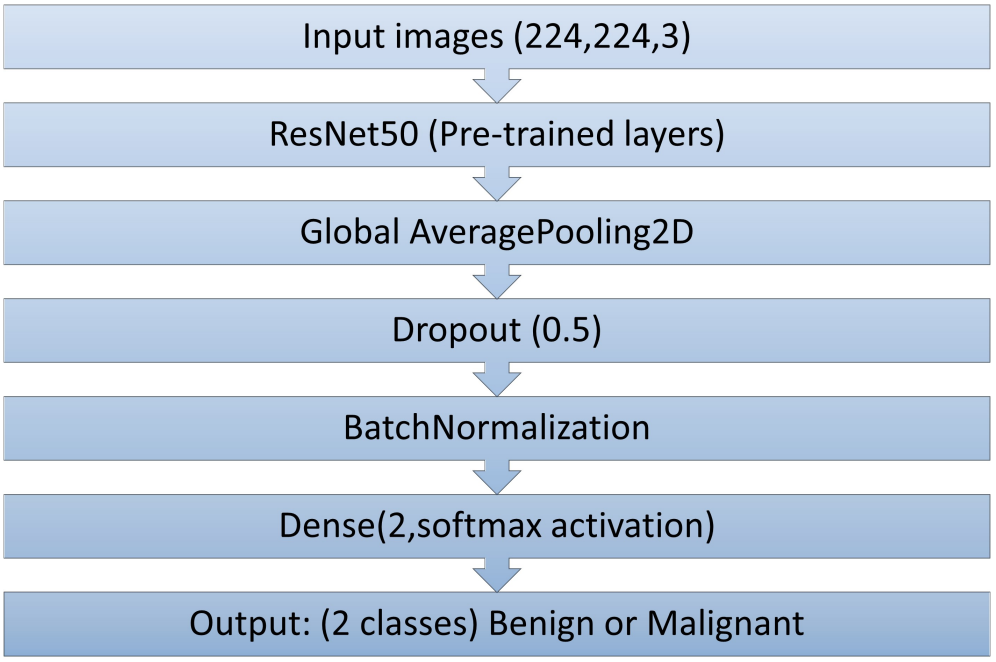


FIGURE 4.3: Flow Diagram of ResNet50 Model

training of the model. It normalizes each layer of the input and acts as a form of regularization, that can help to prevent overfitting of the model. At last, a fully connected layer is achieved that consists of 2 units and a softmax activation function. It acts as a binary classifier that maps the learned features to the output classes (e.g. benign or malignant).

4.5 Server Aggregation Algo

For Balanced data distribution, federated averaging is used for aggregating the model updates from three clients. The FedAvg algo works as follows:

1. The server first initializes the global model with default parameters, which are initially the same for all three clients.

2. Each client downloads the current global model parameters from the server and performs local training on their respective datasets.

3. Each client then computes the new model parameters based on training and sends them back to the central server.

4. The central server aggregates all the model updates by taking their average and updating the global model accordingly.

5. The updated global model sends back to the clients for the next round of training.

The process is repeated for multiple rounds until the desired accuracy is achieved.

The Federated Averaging algorithm is represented in eq 4.1:

$$w_{i+1} = \frac{\sum_{i=1}^{C} n_c w_{i,c}}{\sum_{c=1}^{C} n_C} \tag{4.1}$$

where:

- $w_{i,c}$ is the local model parameter of client C at round i.
- n_C is the number of data samples in the client $c's$ dataset.
- C is the total number of clients.
- w_{i+1} is the global model parameter at round $i + 1$, which is obtained by averaging the local models from each client.

The weights are proportional to the size of each client's dataset. With this approach, it can ensure that the global model accurately reflects the data distribution across all clients.

4.6 Unbalanced Dataset

Once the training and testing of the Flower framework on a balanced dataset are completed, it is important to evaluate its performance on an unbalanced dataset. In order to accomplish this, the dataset is partitioned into an unbalanced distribution and divided into three clients. Table 4.2 displays the distribution of data amongst the clients.

TABLE 4.2: Unbalanced Distribution of Datasets into Three Clients

Images	**Client 1**	**Client 2**	**Client 3**
Total Images	795	794	795
Benign	745	600	455
Malignant	50	194	340

Afterward, additional experiments were performed to evaluate the model's performance when encountering unbalanced datasets.

As shown in table 4.2, melanoma images of client 2 and client 3 are significantly higher as compared to client 1. By using this unbalanced dataset, the aim is to test the robustness of the Flower framework and assess its ability to handle data imbalances in real-world scenarios.

4.7 Proposed Server Aggregation Algo (IWAS)

In the proposed aggregation algo, an "imbalanceFactor" of each client is calculated, and the weights are multiplied by these factors. The default value for "imbalanceFactor" is set to 1.0, meaning that it will not affect the weights. The proposed "imbalanceFactor" is calculated as:

$$IF = S * C/Ex_i \tag{4.2}$$

where:

- IF represents the imbalance factor
- S is no. of minimum melanoma examples
- C is no. of classes
- Ex_i is the total no of examples for each client

The imbalance factor is used to adjust the weight of each client's contribution based on the imbalance ratio in their dataset.

The function first calculates the total number of examples used during the training of each client and then sums up the number of examples of all clients.

$$Ex_T = Ex_1 + Ex_2 + ... + Ex_N \tag{4.3}$$

Suppose for each client i (where $i = 1, 2, ..., N$), the local model has L layers. The weight of each client at L is represented as:

$$w_i = (w_{i,1}, w_{i,2}, ..., w_{i,L}) \tag{4.4}$$

The number of training examples and imbalance factor of each client are used to train the local model.

Afterward, it creates a list of updated weights, where each weight is multiplied by the related number of examples and the client's imbalance factor. The updated weights are calculated as:

$$\begin{aligned} W_1 &= [IF_1 * Ex_1 * w_{1,1}, IF_1 * Ex_1 * w_{1,2}, ..., IF_1 * Ex_1 * w_{1,L}] \\ W_2 &= [IF_2 * Ex_2 * w_{2,1}, IF_2 * Ex_2 * w_{2,2}, ..., IF_2 * Ex_2 * w_{2,L}] \\ &\dots \\ W_N &= [IF_N * Ex_N * w_{N,1}, IF_N * Ex_N * w_{N,2}, ..., IF_N * Ex_N * w_{N,L}] \end{aligned} \tag{4.5}$$

Finally, Aggregated weights are calculated by taking the element-wise sum of the updated weights list and dividing it by the total number of examples Exi used by all clients. Aggregated weights are defined as:

$$W` = (W_1 + W_2 + \dots + W_N)/Ex_T \tag{4.6}$$

These updated weights will be shared with each client for the next round of the training process. This will continue until the desired level of accuracy is attained.

Chapter 5

Results and Evaluation

This chapter presents the evaluation results and comparison between the conventional DL model and the FL model. Further, this chapter also includes model training and testing evaluation. The results are evaluated on the basis of training loss, training accuracy, validation loss, validation accuracy, and testing accuracy for each model. Finally, this chapter provides a comparative analysis of the improved FL framework.

5.1 Conventional ResNET50 Model

ISIC-2020 dataset [42] has been utilized in the conducted experiments. Out of a total of 33,126 training images, 2284 images are selected from which 1700 are Benign and 584 are Malignant. Firstly the dataset is divided into 60% training, 20% validation, and 20% testing sets. The data distribution of the training dataset is shown in Figure 5.1.

In our experiment, a ResNet50 model is trained on the dataset with 50 epochs, a batch size of 32, and learning rate is 1e-4. To optimize the performance of the model, various combinations of learning rates and batch sizes are tried to determine whether the mentioned values produced the best results. The training accuracy achieved is 0.8857, with a validation loss of 0.4763 and a validation accuracy of

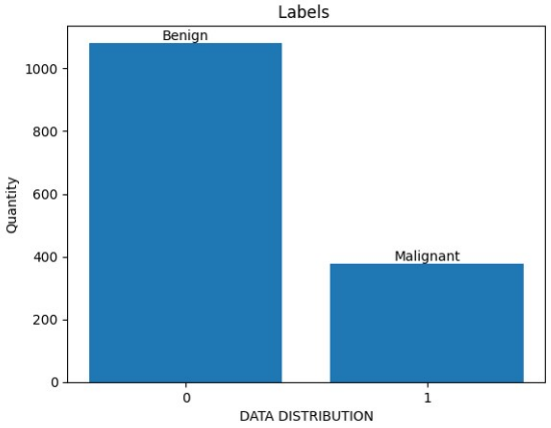


FIGURE 5.1: Data Distribution Used in Conventional DL Model

TABLE 5.1: ResNet50 Model Results on Conventional DL

Loss	0.2882
Accuracy	0.8857
Val loss	0.8306
Test Accuracy	0.8030

0.8306. Upon evaluating the model on test data, an accuracy of 0.80 is obtained. Overall, the results in table 5.1 demonstrate the effectiveness of the proposed approach in training a ResNet50 model on the given data distribution.

5.2 ResNET50 Model – Federated Learning

In the proposed FL setup, there are three clients, each having its own local dataset. Each client is trained on its local dataset for 20 epochs, with a batch size of 32, which is the same as in conventional DL.

After each round, the server aggregated the parameters of all clients. The process is repeated for 10 rounds. Adam optimizer is used, with an initial learning rate of 1e-4.

To measure the performance of the model, Binary Cross-Entropy loss is used with softmax as the criterion. The experiment is carried out using the FLOWER framework and NVIDIA GeForce GTX 1050 Ti computing resources with a 4GB RAM size.

TABLE 5.2: ResNET50 Model Results on FL - Balanced Datset

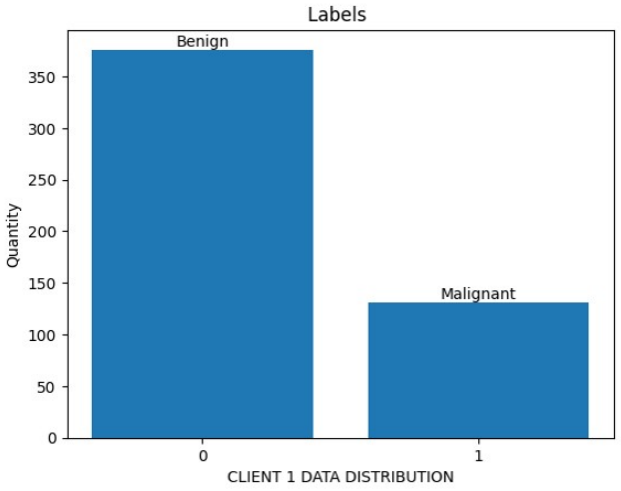
Results	**Client 1**	**Client 2**	**Client 3**
Loss	0.1530	0.1422	0.1795
Accuracy	0.9496	0.9528	0.9272
Val loss	0.5752	0.9572	0.8215
Val accuracy	0.8281	0.7480	0.8125
FL test accuracy	0.8301	0.8427	0.7924

5.2.1 Balanced Dataset

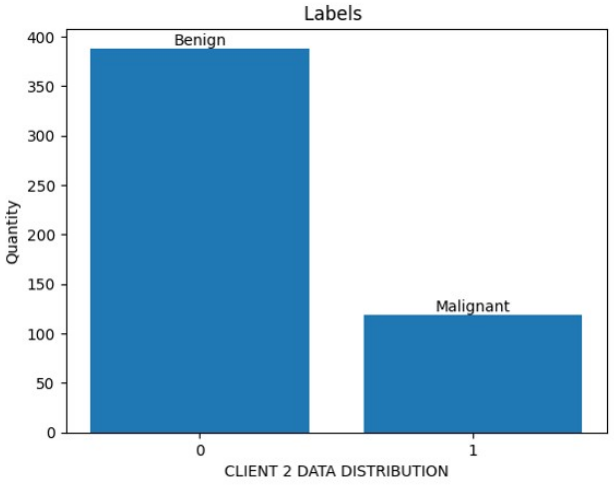
The same dataset is used as in the above conventional DL model. Then it was split into three clients for the purpose of the experiment. The distribution of benign and melanoma images is nearly equal among the three clients to facilitate comparison with the results obtained using a conventional DL model. Table 4.1 shows the number of benign and melanoma images for each client. The data distribution of the training dataset is shown in fig 5.2.

Each client's data is trained on the same ResNet50 model, and the results are compared. The training accuracy achieved by client 1 is 0.9496, with a validation loss of 0.5752 and a validation accuracy of 0.8281. Similarly, client 2 achieved a training accuracy of 0.9528, with a validation loss of 0.9572 and a validation accuracy of 0.8125. Whereas, client 3 achieved a training accuracy of 0.9272, with a validation loss of 0.8215 and a validation accuracy of 0.8125. The trained model is then evaluated on test data, resulting in accuracies of 0.8301, 0.8427, and 0.7924 for clients 1, 2, and 3, respectively. The results shown in table 5.2 indicate that the FL framework is not only improving the training results but also enhancing user privacy.

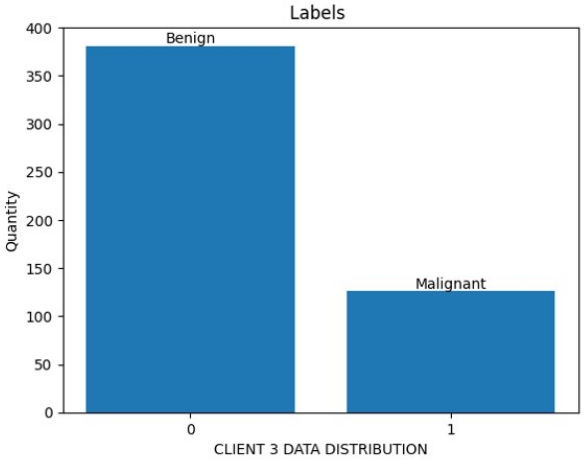
In comparison to the conventional DL results in table 5.1, the application of FL has shown improved performance. The conventional ML approach resulted in 88% on the training data and 80% accuracy on the testing data, while the FL model achieved a training accuracy for all three clients are 94%, 95%, and 92% and testing accuracy are 83%, 84%, and 79% respectively. This indicates a significant improvement in overall accuracy when using FL compared to traditional ML techniques.



(a)



(b)



(c)

FIGURE 5.2: Data Distribution in Balanced Dataset

TABLE 5.3: ResNET50 Model Results on FL - Unbalanced Datset

Results	**Client 1**	**Client 2**	**Client 3**
Loss	0.1130	0.1866	0.2253
Accuracy	0.9685	0.9272	0.8917
Val loss	0.2542	0.8634	0.9667
Val accuracy	0.9219	0.8110	0.7891
FL test accuracy	0.9685	0.7861	0.6666

Overall, the results demonstrate the potential of FL as a technique to improve the performance of DL models in a distributed and privacy-preserving setting.

5.2.2 Unbalanced Dataset

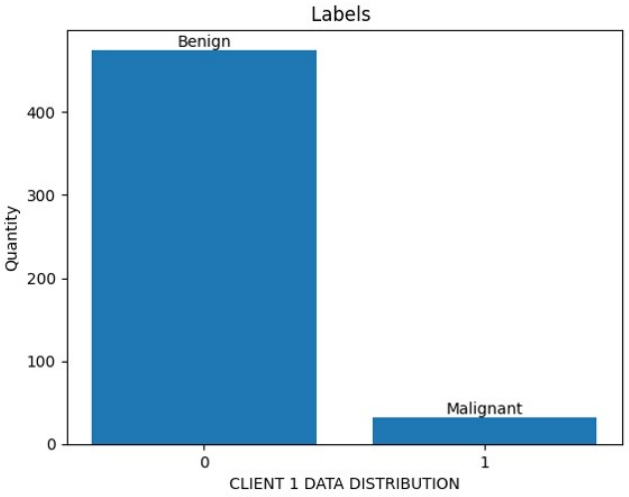
The same dataset is split into three clients with an unbalanced distribution with an equal number of total images.

For Client 1, there is an increase in benign images but limited malignant images. Client 2 has almost the same distribution as in table 4.1. But Client 3 has a higher number of malignant images and less benign images.

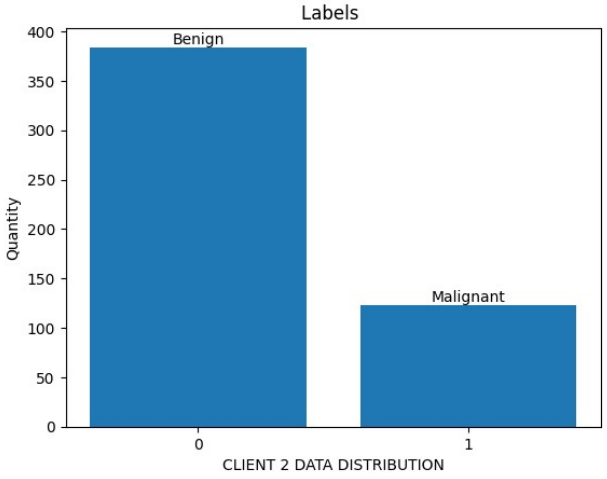
This data distribution will help in the testing FL approach in real-world scenarios where the data is usually nonuniform. Table 4.2 shows the number of benign and melanoma images for each client. The data distribution of the training dataset is shown in fig 5.3.

After applying the same ResNet50 model to each client's dataset. Client 1 achieved a training accuracy was 0.9685, with a validation loss of 0.2542 and a validation accuracy of 0.9219. Client 2 achieved a training accuracy was 0.9272, with a validation loss of 0.8634 and a validation accuracy of 0.80110, And client 3 achieved a training accuracy was 0.8917, with a validation loss of 0.9667 and a validation accuracy of 0.7891. Upon evaluating the model on our test data, we obtained an accuracy of 0.9685, 0.7861, and 0.6666 respectively.

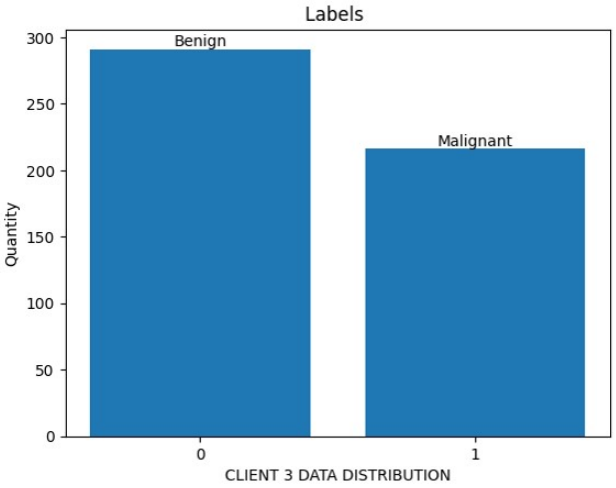
At the server side of the FL process, the weights of the client models are averaged to obtain the global model which propagates back to each client. However, if any



(a)



(b)



(c)

FIGURE 5.3: Data Distribution in Unbalanced Dataset

client has a significantly more unbalanced data distribution, where one class has a higher proportion of samples compared to the other class, it can have an impact on the overall results.

In this test, it can be seen in table 5.3 that client 1 has more benign images and only a few melanoma images. This imbalance in data distribution can affect the performance of the global model. This effect can be observed in the testing accuracy of client 2 and client 3, where their accuracy is significantly dropped to 78% and 66% respectively. This highlights the challenges encountered in real-world situations where data distribution is usually unbalanced or uneven.

In FL, it is important to keep data distribution variations in consideration among the participating clients. In real each hospital contributor may have varying amounts of melanoma images but their contribution is valuable regardless of the distribution of data. By addressing the issue of unbalanced data and designing appropriate strategies, the FL framework can accommodate such scenarios and ensure fair and effective collaboration among every participant.

5.3 ResNET50 Model - IWAS

After implementing the imbalanced weighted-averaging strategy on the server side where imbalance factor of each client has been added. This factor represents the difference in the number of melanoma images in each client's dataset. By considering this factor, the aim is to address the issue of unbalanced data distribution in FL.

In order to calculate the imbalance factor for each client, the distribution of samples in their respective datasets is analyzed by calculating the number of melanoma samples in each client. This imbalance factor is used to adjust the weight of each client's contribution based on the imbalance ratio in their dataset.

After obtaining the imbalance factor for each client, it is added into the server strategy for weighted averaging. This ensures that the contribution of each client

TABLE 5.4: Proposed IWAS Strategy Results

Results	**Client 1**	**Client 2**	**Client 3**
Loss	0.1340	0.1800	0.2507
Accuracy	0.9803	0.9390	0.9134
Val loss	0.2660	0.9152	0.8217
Val accuracy	0.9219	0.7638	0.8125
FL test accuracy	0.8653	0.8456	0.8298

to the global model is adjusted according to the imbalance factor in their dataset. Clients with a higher imbalance factor would have a greater influence in updating the global model, helping to mitigate the bias caused by class imbalance.

The same ResNet50 model is applied to each client's unbalanced dataset. Client 1 achieved a training accuracy was 0.9803, with a validation loss of 0.2660 and a validation accuracy of 0.9219. Client 2 achieved a training accuracy was 0.9390, with a validation loss of 0.9152 and a validation accuracy of 0.7638, And client 3 achieved a training accuracy was 0.9134, with a validation loss of 0.8217 and a validation accuracy of 0.8125. Upon evaluating the model on our test data, we obtained an accuracy of 0.8653, 0.8456, and 0.8298 respectively.

In comparison to the results obtained in table 5.3 using the unbalanced dataset, the proposed imbalanced weighted-averaging strategy has shown improvements in overall results as shown in table 5.4. Specifically, clients 2 and 3 have shown significant improvements in their testing accuracy due to their higher imbalance factors. In fact, the test accuracy of client 3 has now reached a level comparable to that of clients 1 and 2, which was not the case previously due to bias caused by the imbalanced dataset.

5.4 Summary

The chapter presents the results and evaluation of deep learning and the FL model. The FL framework is applied and evaluated on both balanced and unbalanced datasets to determine its effectiveness. It is observed that the FL model performs

well on the balanced dataset, however, there is a significant improvement in performance when the proposed imbalanced weighted-averaging strategy is applied.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

The early detection of melanoma, a type of skin cancer, is essential for increasing the chances of survival rates. In recent years, ML/DL approaches have been developed to automate the detection process. However, concerns regarding the privacy of medical data have made it really challenging to acquire large datasets for these approaches. To address this limitation, federated learning is a solution where data owners can collaborate on training a global model without sharing their raw data.

The aim of this thesis is to investigate the effectiveness of ML/DL and FL model in classifying benign and malignant skin cancer. Specifically, we compared the performance of the ResNet50 DL model using both conventional and federated learning approach. The results highlight improvement in the FL model over the conventional DL model.

Furthermore, performance of FL framework is evaluated on both balanced and unbalanced datasets to check its efficiency. The findings show that the FL framework performed well when the dataset is balanced but its performance deteriorated on unbalanced datasets. To overcome the impact of class imbalance, an imbalance weighted-averaging strategy is proposed which is implemented on the server

side. This strategy aims to improve the accuracy of the FL model on unbalanced datasets.

The experimental results demonstrate the efficacy of the proposed imbalanced weighted-averaging strategy, as it significantly improves the performance of FL model. This approach can be used as a benchmark for future studies concerning FL and imbalanced datasets.

6.2 Future Work

Overall, this thesis highlights the effectiveness of FL in handling imbalanced datasets and the importance of developing strategies on server side to cater class imbalance issues. This work can be extended by fine-tuning the imbalance factors and exploring other techniques to enhance the accuracy of all clients and achieve even better overall performance. Further, pre-processing techniques can be explored in order to cater variations in datasets. Also, more deep-learning models can be used in the future.

Furthermore, the work can be extended to explore more developing techniques that allow each client to customize and personalize the global model according to their unique needs and preferences. This would enhance the federated learning approach more effective and useful.

Bibliography

[1] "Global Cancer Facts and Figures," American Cancer Society. [**Online**]. Available: https://www.cancer.org/research/cancer-facts-statistics/global.html, Accessed on: September 2023.

[2] "Skin Cancer Facts and Statistics," Skin Cancer Foundation. [**Online**]. Available: https://www.skincancer.org/skin-cancer-information/skin-cancerfacts, Accessed on: September 2023.

[3] R. T. Narendhirakannan and M. A. C. Hannah, "Oxidative Stress and Skin Cancer: An Overview," *Indian Journal of Clinical Biochemistry*, vol. 28, pp. 110-115, 2013.

[4] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2019," *CA:A Cancer Journal for Clinicians*, vol. 69, pp. 7-34, 2019.

[5] F. Nachbar, W. Stolz, T. Merkle *et al.*, "The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551-559, 1994.

[6] S. Amutha and R. M. Devi, "Early detection of malignant melanoma using random forest algorithm," *International Journal for Trends in Engineering Technology*, vol. 13, no. 1, 2016.

[7] K. A. Crotty and S. W. Menzies, "Dermoscopy and its role in diagnosing melanocytic lesions: a guide for pathologists," *Pathology*, vol. 36, no. 5, pp. 470-477, 2004.

[8] H. Kittler, H. Pehamberger, K. Wolff, and M. Binders, "Diagnostic accuracy of dermoscopy," *The Lancelet Oncology*, vol. 3, no. 3, pp. 159-165, 2002.

[9] "["Man against machine: AI is better than dermatologists at diagnosing skin cancer," European Society for Medical Oncology [**Online**]]. Available: www.sciencedaily.com/releases/2018/05/180528190839.htm, Accessed on: September 2023.

[10] Y. Li and L. Shen, "Skin Lesion Analysis Towards Melanoma Detection Using Deep Learning Network," *Sensors*, vol. 18, no. 2, 2018.

[11] B. Harangi", "Skin lesion detection based on an ensemble of deep convolutional neural network," *Journal of BioMedical Informatics*, vol. 86, pp.25-32, 2018.

[12] B. Pfitzner, N. Steckhan, and B. Arnrich, "Federated Learning in a Medical Context: A Systematic Literature Review," *ACM Transaction on Internet Technology*, vol. 21, no. 2, 2021.

[13] P. Kairouz, H. B. McMahan, B. Avent, *et al.*, "Advances and Open Problems in Federated Learning," *arXiv*, 2021.

[14] C. Huang, J. Huang, and X. Liu,, "Cross-Silo Federated Learning: Challenges and Opportunities ," *arXiv*, 2022.

[15] G. Argenziano, C. Catricalà, M. Ardigo, *et al.*, "Seven-point checklist of dermoscopy revisited," *British Journal of Dermatology*, vol. 164, no. 4, pp. 785-790, 2011.

[16] A. G. Isasi, B. G. Zapirain, and A. M. Zorrilla, "Melanomas non-invasive diagnosis application based on the ABCD rule and pattern recognition image processing algorithms," *Comput Biol Med.*, vol. 41, no. 9, pp. 742-755, 2011.

[17] W.-Y. Chang, A. Huang, C.-Y. Yang, *et al.*, "Computer-Aided Diagnosis of Skin Lesions Using Conventional Digital Photography: A Reliability and Feasibility Study," *PLoS One*, vol. 8, no. 11, pp. e76212, 2013.

[18] O. O. Olugbara, T. B. Taiwo, and D. Heukelman, "Segmentation of Melanoma Skin Lesion Using Perceptual Color Difference Saliency with Morphological Analysis," *Mathematical Problems in Engineering*, vol. 2018, pp. 1-19, 2018.

[19] C. Barata, M. E. Celebi, J. S. Marques, "A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer," *IEEE Journal of Biomedical and Health Informatics* , vol. 23, no. 3, pp. 1096 - 1109, 2019.

[20] A. U. Haq, J. P. Li, M. H. Memon, *et al.*, "A novel integrated diagnosis method for breast cancer detection," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 2, pp. 2383 – 2398, 2020.

[21] A. U. Haq, J. P. Li, M. H. Memon, *et al.*, "Recognition of the parkinson's disease using a hybrid feature selection approach," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 1, pp. 1319-1339, 2020.

[22] S. Jinnai, N. Yamazaki, Y. Hirano, *et al.*, "The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning," *Biomolecules*, vol. 10, no. 8, pp. 1123, 2020.

[23] N. S. A. ALEnezi, "A Method Of Skin Disease Detection Using Image Processing And Machine Learning," *Procedia Computer Science*, vol. 163, pp. 85-92, 2019.

[24] M. N. Bajwa, K. Muta, M. I. Malik, *et al.*, "Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks," *Applied Sciences*, vol. 10, no. 7, pp. 2488, 2020.

[25] M. A. A. Milton, "Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge," *Computer Vision and Pattern Recognition*, 2019.

[26] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention Residual Learning for Skin Lesion Classification," *IEEE Trans Med Imaging.*, vol. 38, no. 9, pp. 2092-2103, 2019.

[27] M. Z. Alom, T. Aspiras, T. M. Taha, *et al.*, "Skin Cancer Segmentation and Classification with NABLA-N and Inception Recurrent Residual Convolutional Networks ," *CoRR*, 2019.

[28] M. Zafar, M. I. Sharif, M. I. Sharif, *et al.*, "Skin Lesion Analysis and Cancer Detection Based on Machine/Deep Learning Techniques: A Comprehensive Survey," *Life*, vol. 13, no. 1, pp. 46, 2023.

[29] B. Zhang, Z. Wang, J. Gao, *et al.*, "Short-Term Lesion Change Detection for Melanoma Screening With Novel Siamese Neural Network," *IEEE Transactions on medical imaging*, vol. 40, no. 3, pp. 840-851, 2021.

[30] M. M. Mijwil, "Skin cancer disease images classification using deep learning solutions," *Multimedia Tools and Applications*, vol. 80, pp. 26255–26271, 2021.

[31] T. Zhang, L. Gao, C. He, *et al.*, "Federated Learning for Internet of Things: Applications, Challenges, and Opportunities," *CoRR*, 2021.

[32] T. S. Brisimi R. Chen, T. Mela, *et al.*, "Federated learning of predictive models from federated Electronic Health Records," *International Journal of Medical Informatics*, vol. 112, pp. 59-67, 2018.

[33] T. S. Brisimi R. Chen, T. Mela, *et al.*, "Federated learning of predictive models from federated Electronic Health Records," *International Journal of Medical Informatics*, vol. 112, pp. 59-67, 2018.

[34] B. Liu, B. Yan, Y. Zhou, *et al.*, "Experiments of Federated Learning for COVID-19 Chest X-ray Images," *arXiv*, 2020.

[35] C. Ju, D. Gao, R. Mane, *et al.*, "Federated Transfer Learning for EEG Signal Classification," *arXiv*, 2020.

[36] D. Ng, X. Lan, M. M.-S. Yao, *et al.*, "Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets," *Quant Imaging Med Surg.* , vol. 11, no. 2, pp. 852-857, 2021.

[37] M. M. Rahman, D. Kundu, S. A. Suha, *et al.*, "Hospital patients' length of stay prediction: A federated learning approach," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 7874-7884, 2022.

[38] D. Yang, Z. Xu, W. Li, *et al.*, "Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan," *Medical Image Analysis*, vol. 70, pp. 101992, 2021.

[39] Q. Dou, T. Y. So, M. Jiang, *et al.*, "Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study," *npj Digit. Med.*, vol. 4, no. 60, 2021.

[40] M. A. Hashmani, S. M. Jameel, S. S. H. Rizvi, *et al.*, "An Adaptive Federated Machine Learning-Based Intelligent System for Skin Disease Detection: A Step toward an Intelligent Dermoscopy Device," *Applied Sciences*, vol. 11, no. 5, pp. 2145, 2021.

[41] B. L. Y. Agbley, J. Li, A. U. Haq, *et al.*, "Multimodal Melanoma Detection with Federated Learning," *18th International Computer Conference on Wavelet Active Media Technology and Information Processing*, 2022.

[42] "International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset," *International Skin Imaging Collaboration*, https://doi.org/10.34970/2020-ds01, Accessed on: September 2023.

[43] H. B. McMahan, E. Moore, D. Ramage, *et al.*, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *ArXiv*, 2023.

[44] J. Konečný, H. B. McMahan, D. Ramage, *et al.*, "Federated Optimization: Distributed Machine Learning for On-Device Intelligence," *ArXiv*, 2016.

[45] A. Hard, K. Rao, R. Mathews, *et al.*, "Federated Learning for Mobile Keyboard Prediction," *ArXiv*, 2018.

[46] N. Rieke, J. Hancox, W. Li, *et al.*, "The future of digital health with federated learning," *npj Digit. Med.*, vol. 3, no. 119, 2020.

[47] "TensorFlow Federated: Machine Learning on Decentralized Data," **[Online]**. Available: https://www.tensorflow.org/federated, Accessed on: September 2023.

[48] C. He, S. Li, J. So, *et al.*, "FedML: A Research Library and Benchmark for Federated Machine Learning," *ArXiv*, 2020.

[49] A. Ziller, A. Trask, A. Lopardo, *et al.*, "PySyft: A Library for Easy Federated Learning," *Federated Learning Systems. Studies in Computational Intelligence*, vol. 965, 2021.

[50] D. J. Beutel, T. Topal, A. Mathur, *et al.*, "Flower: A Friendly Federated Learning Research Framework," *ArXiv*, 2022.