ResInvolution: An Involution-ResNet Fused Global Spatial Relation Leveraging Model for Histopathological Image Analysis under Federated Learning Environment

by

Shakib Mahmud Dipto
22166030

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

Department of Computer Science and Engineering
Brac University
May 2024

Declaration

It is hereby declared that

1. The thesis submitted is my original work while completing my degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Shakib Mahmud Dipto
22166030

Approval

The thesis/project titled "ResInvolution: An Involution-ResNet Fused Global Spatial Relation Leveraging Model for Histopathological Image Analysis under Federated Learning Environment" submitted by

1. Shakib Mahmud Dipto (22166030)

Of Spring, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.Sc. in Computer Science and Engineering on May 25, 2024.

Examining Committee:

Supervisor:
(Member)

Md. Ashraful Alam, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Internal Examiner:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

External Examiner:
(Member)

Ashis Talukder, PhD
Associate Professor
Department of Management Information Systems
Dhaka University

Program Coordinator:
(Member)

Md Sadek Ferdous, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Accessing image data in the domain of medical image analysis is challenging owing to concerns regarding privacy. Federated Learning is the approach used to get rid of this challenge. With millions of learning parameters, Residual Network (ResNet) is one of the most advanced architectures for classifying medical images. Because of its resource-hungry nature, using this ResNet architecture in the Federated learning framework has an impact on the entire system. This research introduces a novel architecture called Residual Involution (ResInvolution), specifically developed for analyzing histopathological images within a federated learning environment. The architecture utilizes a cutting-edge model, the Involution-ResNet Fused Global Spatial Relation Leveraging model, to enhance the analysis process. This model is impressively lightweight, boasting less than 190,000 parameters. Its efficiency and ease of deployment make it ideal for medical image analysis tasks. By incorporating involution operations into the ResNet framework, it becomes possible to adjust the spatial weighting of features dynamically. The proposed model enables a comprehensive analysis of intricate structures that exceed the capabilities of traditional convolutional networks. This model has been deployed within a federated learning environment, where privacy is prioritized. Also utilize decentralized data sources, thereby eliminating the necessity of centralizing sensitive medical images. This approach ensures strict adherence to medical data privacy regulations while simultaneously leveraging collective insights from multiple institutions. The model has undergone rigorous testing on three distinct datasets: GasHisSDB, NTC-CRC-HE-100K, AND LC25000. In Federated Learning scenarios, the model achieves accuracies of 91%, 95%, and 99% on these datasets, respectively. However, in the context of federated learning, the accuracies exhibited are 91%, 93%, and 97%, respectively. The model's effectiveness is evaluated through various performance metrics, including the confusion Matrix, Accuracy, Precision, Recall, F1-Score, Receiver operating Characteristic (ROC) curve, and Area under the ROC Curve (AUC) Score. The results highlight the model's ability to adapt to various challenges, such as limited data and irregular data distribution, commonly encountered in federated learning environments. ResInvolution sets a revolutionary benchmark in medical image analysis, enhancing the ability to interpret intricate medical images and paving the way for future advancements in scalable, privacy-preserving deep learning technologies.

Keywords: ResInvolution; ResNet; Involution; Fusion; Federated Learning; Histopathological images; Involution Neural Network, INN, CNN

Acknowledgement

Firstly, I would like to express my gratitude to the Great Allah for allowing me to complete my thesis without any significant disruptions successfully.

Secondly, I would like to express my gratitude to my advisor, Dr. Md. Ashraful Alam and Dr. Md. Golam Rabiul Alam, for their invaluable support and guidance throughout our project. He was always there for me whenever I needed assistance.

Lastly, I would like to express my heartfelt gratitude to my parents for their unwavering love and support throughout my educational journey. Because of their constant encouragement, I am now on the brink of graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AUC Area under the ROC Curve

CNN Convolutional Neural Network

$CNNS$ Convolutional Neural Networks

$DenseNET$ Densely Connected Convolutional Networks

DRN Deep Residual Network

FL Federated Learning

FN False Negative

FP False Positive

INN Involution Neural Network

LR Logistic Regression

ML Machine Learning

NAR Network-Auto-Reduction

$R-CNN$ Region-Based Convolutional Neural Network

$ResInvolution$ Residual Involution

$ResNet$ Residual Network

ROC Receiver operating Characteristic

SE Squeeze and Excitation

SVM Support Vector Machine

TN True Negative

TP True Positive

VGG Visual Geometry Group

Chapter 1

Introduction

A crucial aspect in the field of medical diagnostics is Histopathological image analysis, it plays a vital role in various disease detection and characterization. Histopathological images can be collected through microscopic medical examination of tissue samples. These images provide very valuable insights into the morphological characteristics of tissues and cellular structure. In the medical field, it's very crucial and important for histopathological images to be detected and examined accurately and efficiently. This detection cannot be overstated as it directly affects the patients and their future medical procedures. Early detection can be a beneficial factor for effective and precise treatment planning. As it can be a great help for improving results. However, the process of these histopathological image analyses can be mostly time-consuming and challenging as it subjects to subjectify and inter-observer variability [1]. The continuous expansion of medical data repositories, in disease manifestation, demands the formulation of sophisticated and precise methodologies implementation towards histopathological image analysis. Methodologies have revolutionized in this field in the past days and still are updating with ongoing time including disease detection and histopathological image classification [2]. These advanced computational models have developed remarkable and revolutionary performances in extracting patterns and features from histopathological images which enables accurate disease classification and diagnosis [3]. Deep learning application in histopathological image classification offers numerous advantages. These models can be trained on large datasets of annotated images. It enables them to learn the complex relationship between image features and disease characteristics. After training these models can be applied to new unseen images for automated classification and disease detection, It potentially reduces the burden of pathologists and improves diagnosing the disease with accuracy. Furthermore, deep learning models have the potential to identify subtle patterns and associations that may not be readily apparent to human experts, leading to new insights and biomarkers in disease mechanisms [4]. The implementation of deep learning methods in histopathological image analysis is possible but filled with many challenges. One of the most significant challenges is computational complexity. Also collecting the resources for the data to train is another complicated issue. These architectures which are designed for medical image analysis mostly involve a large number of parameters and require extra time to employ [5]. Also, because of privacy concerns in clinics or hospitals, there can be resource-constrained environments which is one of the major challenges. Then, another critical concern in the field of histopathological image analysis is data se-

curity as histopathological images are highly sensitive and subject to strict privacy regulations in most of hospitals and clinics [6]. So, most of hospitals and clinics institutions are often hesitant to share patient data publicly, which becomes an issue to the development and evaluation of deep learning models on diverse and representative datasets. To nullify these challenges, we propose a lightweight, deep-learning architecture for the efficient classification of histopathological images which is involution-based. Involution is a recently introduced operation that aims to capture spatial information. It works more effectively while reducing the computational costs [7]. By implementing the involution layers into our deep learning model, we aim to achieve better performance than the traditional CNN-based architectures. It can also be implemented while requiring fewer computational requirements.

In order to detect damaged tissue and other defects in the synthesized histopathology images, we need to retain the anatomical structures of the organs for a higher-quality medical image translation. The anomalies may be located at a specific position in the images but these positions of anomalies may determine different types of diseases. Compared to enhancement in medical images, many tumors exhibit spatially heterogeneous patterns. These patterns reflect different spatial properties based on various factors. This problem is stated in various medical imaging studies [8]–[10]. Therefore, regardless of an object's location on an image, filters should be able to identify its features. Additionally, by leveraging federated learning techniques privacy concerns surrounding medical data are addressed [11]. Federated learning is a distributed machine learning approach that can enable collaborative model training without the need to centralize sensitive data. By incorporating federated learning, our proposed novel method can be trained on decentralized datasets from multiple preserving data privacy which can be benefiting from the diversity of data sources.

1.1 Problem Statement

Conventional generative models that extract features based on convolution are limited because convolutions are limited by spatial-agnostic and channel-specific constraints. While the spatial-compact and spatial-agnostic characteristics make sense for improving efficiency and understanding translation equivalency, they deprive convolution kernels of their capacity to adjust to a variety of visual patterns concerning various spatial positions. Furthermore, locality limits the convolution's receptive area, making it difficult to capture distant spatial interactions in a single frame. Other than that, the lack of consideration of the mean of affine transformation in normalization can result in the loss of spatial information. Here, the anomaly is in an odd location, taking the mean may cause a misidentification or image generation. For medical image processing professionals such as radiologists, physicians, and specialists, noisy images provide a significant difficulty. We can characterize noise as an undesired and random signal that has the potential to eliminate vital information (e.g. disease lesion, tumor area, cell anomaly, etc.) that could be used in the future identification of a disease, such as a tumor in a patient. For convolutions, this vital information may get lost due to its locality limits mentioned in the previous paragraph. Vision transformers can come in handy in this case but immense computational power requirement and its heavy feature extraction process may hamper the cause. Therefore a feature extractor with fewer weight requirements and normaliza-

tion level modification will fulfill the goals. To propose such a solution, involution can be an important component to integrate into the model due to its local and global receptive fields and lightweight nature. However, setting up a proper configuration for the involution layer can be challenging as the number of layers, kernels, and placement in the model needs to be determined.

1.2 Research Contribution

In this study, we present a novel contribution to the field of histopathological image classification through the development of an Involution-Based ResNet Architecture, named ResInvolution. This architecture leverages the power of involution layers to enhance feature extraction and improve classification accuracy. The ResInvolution model is designed to be lightweight and efficient, with a significantly reduced number of trainable parameters compared to traditional deep learning models. The primary innovation lies in the integration of involution layers within the ResNet framework, which allows for more effective and precise identification of relevant features in histopathological images. This is particularly beneficial in medical image analysis, where accurate classification is crucial for diagnosis and treatment planning. Furthermore, we have extended the application of the ResInvolution model by employing it within a Federated Learning environment. Federated Learning is a decentralized approach that enables the training of machine learning models across multiple institutions without the need to share sensitive data. By implementing ResInvolution in this setting, we have demonstrated its potential to maintain high classification performance while preserving data privacy and security. This is especially important in the medical field, where data privacy concerns often limit the availability of large, centralized datasets. To sum up, the key contributions of this research are:

- We have introduced the ResInvolution architecture, which optimizes feature extraction and classification in histopathological images.
- We designed the model to take advantage of combined local and dynamic global spatial attention, making the model achieve similar performance to that of larger models with many more parameters.
- Our proposed model is highly efficient and perfectly suited for training and evaluating in low-end computing systems with limited resources, because of its lightweight architecture.
- We have successfully integrated this lightweight architecture into a federated learning framework, ensuring high performance while maintaining data privacy.
- This study has a significant potential for improving medical image analysis and paving the way for further research and application in privacy-preserving machine learning in healthcare.

1.3 Thesis Orientation

The thesis is structured into five chapters. The following chapters will cover related works, methodology, result analysis, and conclusion. Chapter 2 provides a comprehensive analysis of the existing literature, highlighting its strengths, limitations, and areas that require further research. Chapter 3 provides a comprehensive overview of the research methodology, design, and integration of involution layers into the ResNet framework. It also discusses the datasets used, model specification, and experimental setup in detail. Chapter 4 provides a thorough analysis of the performance of various versions of the proposed model in both traditional and federated learning environments. Chapter 5 provides a comprehensive overview of the key findings, highlighting the significant advancements made in creating a highly efficient and accurate model for classifying histopathological images in federated learning. This chapter also acknowledges the limitations of the study and proposes potential avenues for further research.

Chapter 2

Related Work

For conducting this research, the we have gone through relevant studies that integrate novel operations such as involution with residual networks, custom deep learning models for histopathological image analysis also for cancer detection. In addition, privacy-preserving techniques like federated learning for collaborative models in the field of medical imaging and healthcare are also explored.

A novel method called Network-Auto-Reduction (NAR) is proposed for simplifying Convolutional Neural Networks (CNNS) that reduce their computational costs during interference while maintaining model accuracy [12]. Although the depth, width, and input resolution dimensions are maintained in the NAR in a precise way with less than 4-time computations, the vanishing gradient problem is still an issue. Primarily, NAR has shown promising results in classifying Tumors in histopathological images. A similar approach for identifying histopathological images is undertaken by the authors in [13] to assist medical personnel in making diagnostic decisions for lymphoma. Data augmentation played a vital role in training their network named R-CNN. Numerous optimizers such as ADAM and SGD were integrated to find the best results. The optimal model achieved an accuracy of 83.3% and a recall of 71.8% using the ADAM optimizer. The faster R-CNN showed potential in identifying histopathological images, that can serve as desired tool for the pathologists. The only issue, with the change of optimizer, the model failed to perform significantly well. Another approach for automatic histopathological image analysis for cancer was proposed in [14]. A novel CNN architecture with a Squeeze and Excitation (SE) is employed based on a hybrid squeezing method. Color-based spatial squeezing and channel-wise pooling adaptively boost meaningful features and suppress less important ones. The utilized dataset are Camelyon16 and BreaKHis datasets, the hybrid squeezing method outperformed regular CNN available in the Keras. The computational load was also decreased by 35% in the case of trainable parameters. The model is yet to be evaluated with other available State-of-the-art models available in the concerned sector.

Skin disease detection through dermatoscopic images with the aid of a novel CNN method was introduced in [15]. To ignore the hassle of pigmented skin lessor and manual detection of skin diseases, the authors proposed a hybrid CNN-DenseNET architecture to enhance diagnostic accuracy and efficiency. Furthermore, to make the model more proficient in hybridization with DenseNet architecture and feature

extraction using ResNET. The approach showed an impressive accuracy of 95.7% on the HAM10000 and 91.07% on the PAD-UFES-20 dataset. The paper compared favorably with recent works, demonstrating the potential of the proposed model to automate the diagnosis of various skin diseases effectively. From the point of view of Machine Learning(ML), automatic classification of breast cancer from histopathological images using pre-trained CNNs combined with traditional classifiers such as Support Vector Machine(SVM), Logistic Regression (LR) are proposed in [16].

The authors propose a two-phase model, wherein the first phase, features are extracted from the images using pre-trained CNNs like VGG16, VGG19, Xception, and ResNet50, and in the second phase, these extracted features are used by the SVM and LR classifiers for binary classification (benign or malignant) and multi-class classification (based on magnification factors of 40x, 100x, 200x, and 400x). The performance is evaluated using various metrics like accuracy, precision, recall, and F1-score. The experimental results demonstrate that the ResNet50 network combined with LR achieves the maximum accuracy of 93.27% for binary classification, and the CNN+LR model outperforms CNN+SVM for classifying different magnification factors, with the highest accuracy of 86.73% achieved for the 100x magnification factor. Another work [17] proposed deep residual learning, a framework for training deep CNN. Primarily, a good number of training error had been observed that led to overfitting. To overcome the situation authors focused on residual learning, reformulating the layers to fit a residual mapping instead of the original unreferenced mapping. This is implemented through shortcut connections that skip one or more layers. Comprehensive experiments on ImageNet and CIFAR-10 datasets show that these residual networks are easier to optimize, gain accuracy from increased depth, and outperform previous deep networks.

Involution networks such as Deep Residual Network(DRN) and a baseline Deep Residual Network (DRN) for hyperspectral image classification [18]. In comparison with these two models, DRN consistently outperformed DRN in overall accuracy with fewer parameters in the case of HyRANK dataset. Overall, DRIN reflects 3.23 times fewer trainable parameters than DRN. The lightweight involution operation, with each kernel corresponding to an individual feature channel, enables DRN's superior performance over convolutional counterparts. Another important architecture is the I-CNet, a custom model that combines both involution and convolution for image classification tasks. By integrating the transfer learning model ResNet, I-Cnet optimizes feature extraction and achieves higher classification accuracy than another competitive algorithm in Chinese MNIST, Medical MNIST, and Kaokore datasets. Another novel [7] involution model for visual representation learning in deep neural networks was used for spatial-agnostic and channel-specific convolution. A model named 'Rednet" was proposed replacing the 3X3 convolutions at bottleneck positions. The proposed involution and custom RedNet model show superiority on instance segmentation compared to conventional convolutional counterparts while being model efficient in classification tasks. Another study [19] investigates privacy-preserving machine learning for healthcare applications especially in the automated classification of skin diseases using federated learning (FL). The authors first discuss the motivations and challenges of maintaining privacy in machine learning models, such as ethical concerns, legal regulations like GDPR,

and the trade-off between privacy and utility. They then review existing privacy-preserving techniques like differential privacy, secure multi-party computation, and federated learning. The paper also covers previous work on skin lesion classification using deep-learning models. The authors propose an FL-based approach where hospitals train local models on their data and share only the model updates with a central aggregator, thus preserving patient privacy. In [20] provides a comprehensive review of federated learning (FL) applications in medical image analysis with deep neural networks (DNNs) for disease diagnosis. For brain disorders, it worked significantly well where X-ray and CT scans are provided to the model for training purposes. This paper recognized the security concerns associated with FL. The possible solutions for poisoning attacks and model inversion techniques were also discussed. FL is also explored in [21] where systematic surveys of the current state-of-the-art model are used for medical analysis purposes. It extracted valuable information from 22 articles related to that field. No novel models are proposed in the paper rather than models such as CNN, Encoders, and other models are discussed thoroughly along with their application in the concerned sector. A study [22] presents an up-to-date overview of federated learning techniques for medical image analysis. It discusses the challenges of conventional centralized and distributed machine learning approaches, such as data scarcity, privacy concerns, and computational costs. The authors highlight federated learning as a promising involution that enables collaborative training of deep learning models across multiple institutions without sharing sensitive patient data. The paper covers various federated learning architectures, including client-server and peer-to-peer models, and the integration of transfer learning techniques. Additionally, it explores the potential of combining federated learning with blockchain technology for enhanced data security and traceability. Another study [23] proposes a new federated learning approach called FedDropoutAvg for training generalizable deep learning models on distributed medical imaging data without sharing the raw data. It builds upon existing federated learning methods like FedAvg and FedProxthat perform weighted averaging of local model parameters. FedDropoutAvg introduces randomness in client selection and federated averaging, aiming to reduce overfitting to individual sites' data.

The literature review emphasizes the importance of developing efficient lightweight deep-learning models specially optimized for analyzing histopathological images. These models are crucial for the early detection and accurate classification of different types of cancer. Such models are essential in initial cancer screenings, where rapid and accurate diagnosis is crucial. Nevertheless, the review highlights a notable obstacle in the limited accessibility of public datasets in this domain, affecting the progress and enhancement of robust deep learning models. In order to address this challenge, distributed learning emerges as a promising solution. This approach addresses the challenges of limited data availability and allows for the use of distributed data, which improves privacy and data security. In order for distributed learning systems to be truly effective, it is crucial to utilize models that strike the perfect balance between being lightweight and delivering high performance on both small and large datasets. This dual requirement ensures that the models are versatile and adaptable to varying data volumes and characteristics, making them well-suited for widespread deployment in medical image analysis.

Chapter 3

Methodology

We have gathered several datasets of histopathological images, ensuring that each set has been processed individually to preserve their distinct characteristics. The preprocessing stages for each dataset involved resizing and scaling the images to conform to specified dimensions and intensities. Afterward, the dataset was divided into two parts: 75% for training and 25% for testing. We further subdivided the training segment of each dataset into two parts: 20% for model validation and 80% for training. For every dataset individually, we implemented a ResNet-Involution combined model referred to as ResInvolution. Every model was trained and validated using specific training and validation datasets.

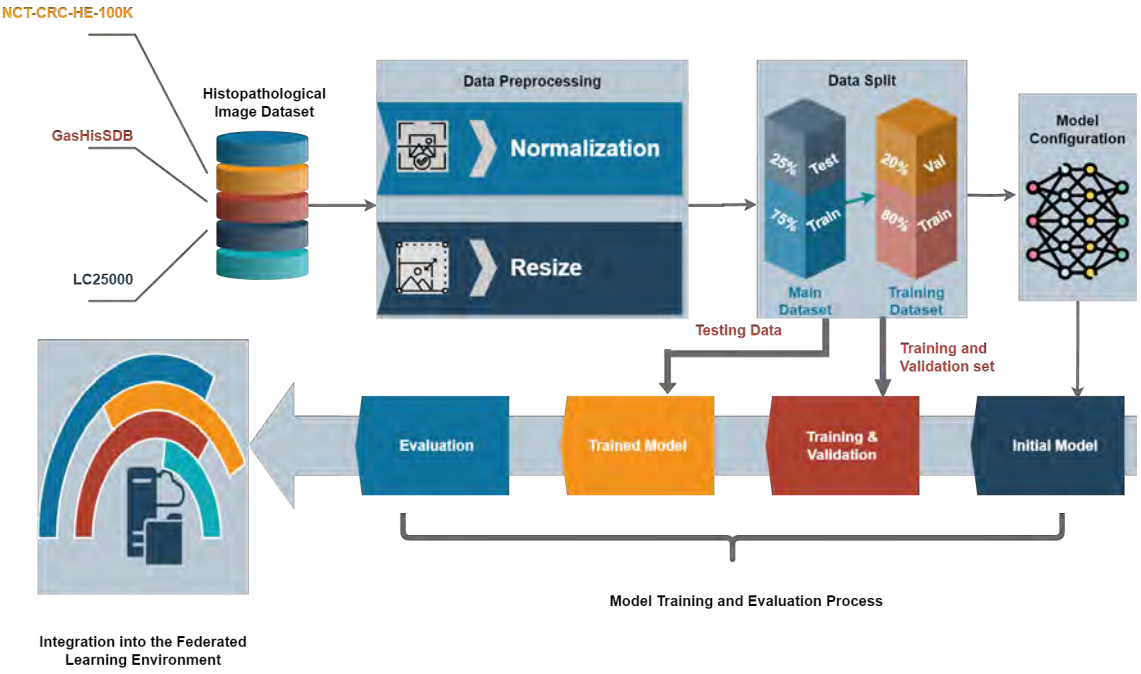


Figure 3.1: Overview of the Proposed System

Subsequently, we assessed the efficacy of the ResInvolution model by employing the testing set from each dataset separately. The evaluation was conducted on five versions of the ResInvolution architecture, each differing in the number of involution blocks included, ranging from 0 to 4. The most effective architecture was

subsequently chosen and implemented in a federated learning environment. In this environment, the training set, which consisted of 75% of the data from each dataset, was evenly divided across 10 clients. Each client divides their assigned data into 90% for training their local models and 10% for validation. After training and validating the local models, the aggregate trained weights of the models were computed. Subsequently, the combined weight was transmitted to the global model in each communication cycle. This training was conducted over a span of one hundred communication sessions. After the completion of the training compilation, we evaluated the effectiveness of the global model by utilizing the designated 25% testing subset from each dataset. In the last phase, we conducted a thorough evaluation of the model's performance in the federated learning framework. Figure 3.1 Illustrate the overview of our research methodology.

3.1 Description of the Datasets

In this study, we have acquired three datasets of histopathological images for analyzing the performance of our proposed methodology. The datasets are GasHisSDB [24], NCT-CRC-HE-100K[25], and LC25000 [26].

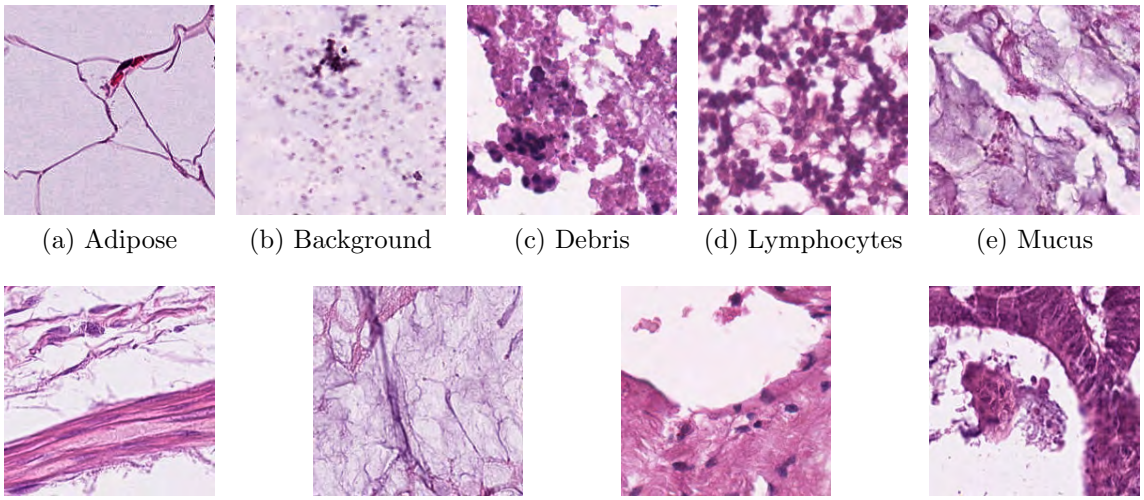
3.1.1 NCT-CRC-HE-100K dataset

The NCT-CRC-HE-100K dataset consists of 100,000 non-overlapping image patches extracted from 86 H&E stained human cancer tissue slides and normal tissue samples. These slides are from the National Center for Tumor Diseases (NCT) biobank and the University Medical Center Mannheim (UMM) pathology archive. The dataset contains images from colorectal cancer (CRC) primary tumors, CRC liver metastases, and non-tumorous regions from gastrectomy specimens to increase variability in the normal tissue classes. The images are 224x224 pixels at 0.5 microns per pixel resolution. Color normalization was applied using Macenko's method. The tissue types included in the representation are Adipose (ADI), Background (BACK), Debris (DEB), Lymphocytes (LYM), Mucus (MUC), Smooth muscle (MUS), Normal colon mucosa (NORM), Cancer-associated stroma (STR), and Colorectal adenocarcinoma epithelium (TUM). Below are the descriptions of the classes.

- **Adipose (ADI):** This class represents fat cells or adipocytes found in connective tissue. Adipose tissue provides insulation and serves as an energy storage reserve.
- **Background (BACK):** This class covers any areas outside of biological tissue, such as the slide background or empty space between tissue sections.
- **Debris (DEB)**Debris refers to cellular fragments, necrotic tissue, or other biological material that is not easily identifiable as a specific tissue type
- **Lymphocytes (LYM):** Lymphocytes are a type of white blood cell crucial for the immune system's response against infections and diseases like cancer
- **Mucus (MUC):** Mucus is a viscous, lubricating substance produced by specialized cells lining the colon and other organs. It protects and lubricates the epithelial surfaces.

- **Smooth Muscle (MUS):** This class covers smooth muscle cells, which form layers of hollow organs like the colon and facilitate contractions for functions like peristalsis.
- **Normal Colon Mucosa (NORM):** This represents healthy epithelial cells lining the inner surface of the colon, exhibiting normal glandular architecture
- **Cancer-Associated Stroma (STR):** Stroma refers to the connective tissue surrounding and supporting the tumor cells, often exhibiting an altered composition in cancer.
- **Colorectal adenocarcinoma Epithelium (TUM):** This class captures the malignant epithelial cells of colorectal adenocarcinoma, the most common form of colon cancer.

An illustrative instance of this dataset is presented in Figure 3.2.



(a) Adipose (b) Background (c) Debris (d) Lymphocytes (e) Mucus

(f) Smooth Muscle (g) Normal Colon Mucosa (h) Cancer Associated Stroma (i) Colorectal adenocarcinoma Epithelium

Figure 3.2: Sample Data from the NCT-CRC-HE-100K Dataset

3.1.2 GasHisSDB

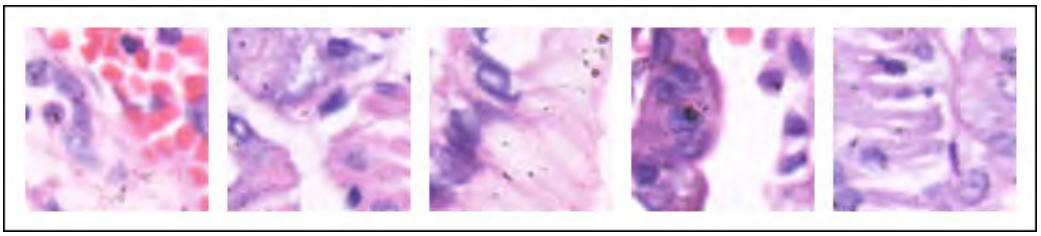
The Gastric Histopathology Sub-size Image Database (GasHisSDB) is a new publicly available dataset for computer-aided diagnosis of gastric cancer. This dataset was developed by researchers from Northeastern University and Liaoning Cancer Hospital and Institute. The original images were obtained from the Pathology Department and were stained using the Hematoxylin and Eosin (H&E) technique. The original whole-slide images were cropped into sub-size images, and the abnormal regions were selected as regions of interest. The sub-size images were further processed by randomly rotating and shuffling the order to reduce the correlation between images from the same original slide. Three types of visual features are provided for each image: color histogram, Local Binary Patterns (LBP), and Gray-Level Co-occurrence Matrix (GLCM) features. Each image is annotated as either normal (no

cancerous cells) or abnormal (containing cancerous cells).

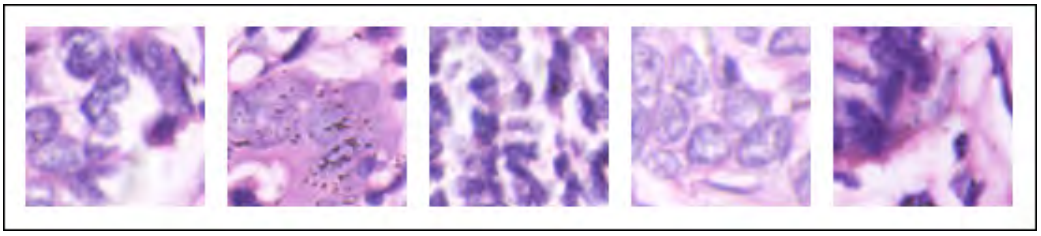
The dataset contains a total of 245,196 sub-size tissue case images. The images are provided in three sub-sizes: 160x160 pixels, 120x120 pixels, and 80x80 pixels. The 160x160 sub-size set contains 20,160 normal and 13,124 abnormal images, the 120x120 sub-size set contains 40,460 normal and 24,801 abnormal images and the 80x80 sub-size set contains 59,151 abnormal and 87,500 normal images. We have used the 80x80 sub-sized set for our study. Each sub-sized sets contains two two classes: normal and abnormal. Below are the descriptions of the classes.

- **Normal:** This class contains histopathological images showing the absence of cancerous cells, representing normal or benign gastric tissue sections.
- **Abnormal:** This class contains histopathological images showing the presence of cancerous cells, representing gastric cancer tissue sections.

This binary dataset is intended to serve as a benchmark for developing and evaluating computer-aided diagnostic techniques for gastric cancer, particularly using weakly supervised learning approaches. The sub-size image format and the availability of multiple features make the dataset suitable for testing the performance of various machine learning and deep learning classification methods. Some samples from the dataset have been provided in figure 3.3



(a) Normal



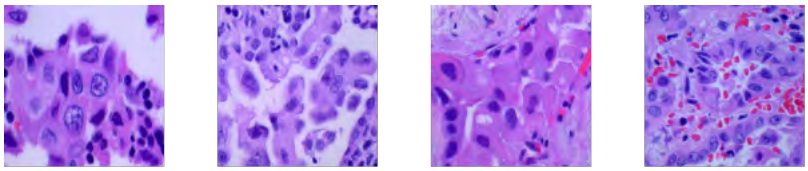
(b) Abnormal

Figure 3.3: Sample Data from the GasHisSDB Dataset

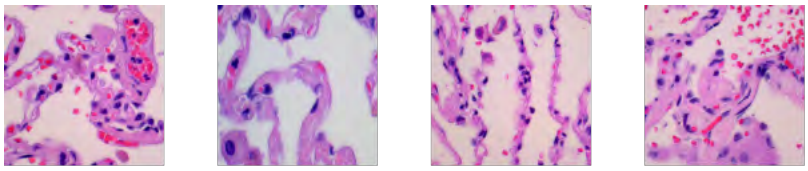
3.1.3 LC25000

The LC25000 dataset comprises a substantial compilation of 25,000 histopathology images depicting lung and colon tissue samples. It was meticulously selected for the purpose of categorizing various forms of lung and colon malignancies based on microscopic tissue images. The dataset is a great resource for the development and evaluation of machine learning models, for the automated diagnosis of cancer using histopathology pictures. This dataset consists of photos from 5 distinct

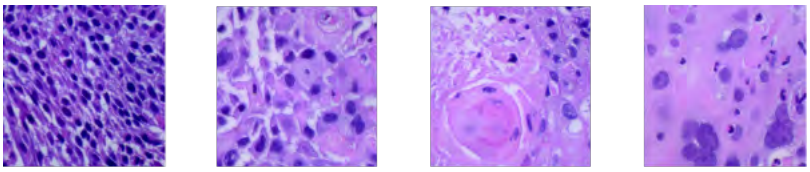
categories: lung adenocarcinoma, lung squamous cell carcinoma, benign lung tissue, colon adenocarcinoma, and benign colon tissue. Every class consists of 5,000 photos, guaranteeing an equitable distribution throughout the classes.



(a) Lung Adenocarcinoma



(b) Benign Lung Tissue



(c) Lung Squamous Cell Carcinoma

Figure 3.4: Sample Data of the used classes from LC25000 Dataset

A detailed description of the classes is mentioned below.

- **Lung Adenocarcinoma:** These images represent a form of non-small cell lung cancer originating from glandular cells lining the lung. Lung Squamous Cell Carcinoma: These images represent a type of non-small cell lung cancer originating in squamous cells that line the airways in the lungs.
- **Lung Squamous Cell Carcinoma:** This is a kind of non-small cell lung cancer that originates from the squamous cells, which are flat cells lining the airways in the lungs.
- **Benign Lung Tissue:** These images are captured from samples of tissue taken from the lung; they are not cancerous and do not show the presence of any form of malignancy.
- **Colon Adenocarcinoma:** These images are captured from a specific type of colorectal cancer that develops from glandular cells of the colon or the rectum.
- **Benign Colon Tissue:** These images are of colon tissue; they are not cancerous and they depict a normal and healthy slice of tissue.

The LC25000 dataset consists of images of both lung and colon tissues with both cancerous and non-cancerous samples. This dataset has been created for classifying and diagnosing the particular cancers properly, based on the histopathological images. For the experiment, we consider the three classes of lung cancer. Sample images of the employed classes can be seen in Figure 3.4.

3.2 Model Specification

In this research, we have utilized ResNet50, Involution, and Federated Learning framework to analyze the histopathological images.

3.2.1 ResNet50

Resnet50 is a 50-layer convolutional neural network with 48 convolutional, one max-pool and one average pool layer [17]. Its a powerful image classification model that can be trained in large dataset and can obtaining state-of-the-art result. One of the core innovations of this model is the use of residual connections which allow the network to learn a set of residual functions that help the model to map the input to the desired output. It also addressed the vanishing gradient problem in the case of very deep layers of neural networks [27].

The input image is initially passed through a convolutional layer with a kernel size of 7x7, which is followed by batch normalization, ReLU activation, and max pooling layers (Stage 1). The subsequent stages which are Stage 2 to Stage composed of identity blocks and convolutional blocks. Identity blocks are often used when the input and output dimensions remain the same. Convolutional blocks are employed when the dimensions need to be changed. Those blocks contain convolutional layers, batch normalization, and then ReLU activation functions, with residual connections that allow the input to skip some layers and be added back to the output. After all the stages, the feature maps are subjected to an average pooling layer. The resulting output is fed into a fully connected layer for image classification. It allows the gradients to propagate more easily through the network during training and enables the model to learn more effectively. It leads to improved performance, especially for deeper-layer architectures.

3.2.2 Involution Neural Network

The Involution Neural Network is the recently introduced architecture that represents a novel approach to efficiently capture spatial data while reducing computational overhead compared to other CNN models [7]. The key part of this network lies in the involution operation, which functions as an alternative to the traditional convolution process. It involves inverting some of the design principles which are behind the commonly used convolution operation. While the traditional CNN operation applies the same fixed kernel (a square matrix), involution instead incorporates using distinct kernels for each spatial location, but it shares these kernels across the channels.

The figure 3.5 portrays the core mechanism behind involution neural networks. An innovative neural architecture designed to capture spatial data efficiently while minimizing computational overhead compared to conventional convolutional approaches. The process commences by partitioning the input tensor into multiple subgroups, each undergoing a series of linear transformations influenced by normalization and activation layers. Subsequently, a spatial kernel is derived by aggregating information across the input's spatial dimensions.

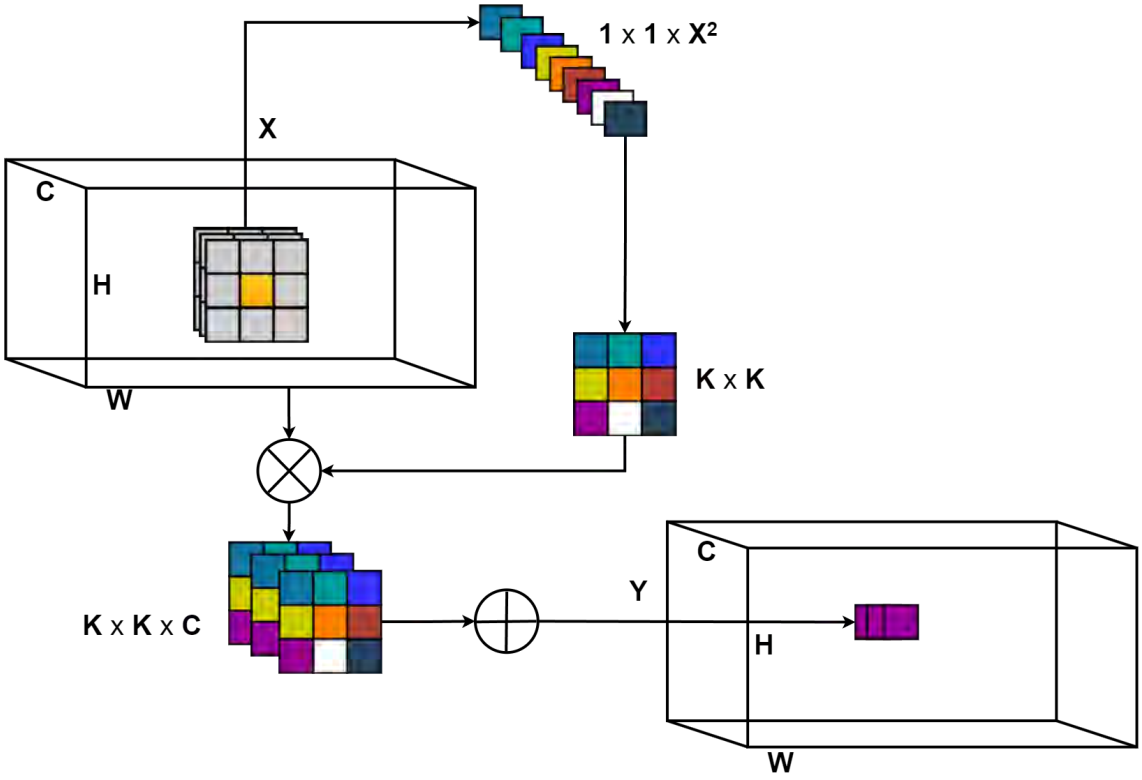


Figure 3.5: Illustration of kernel production of involution

This specially crafted kernel is then reshaped and applied to the input, yielding an output tensor that encapsulates the spatial relationships present in the original data. Through this strategic manipulation of spatial information, involution networks aim to deliver performance on par with or surpassing traditional convolutional counterparts, while demanding fewer computational resources, thereby rendering them well-suited for resource-constrained scenarios. By leveraging the involution neural operation, we can aim to achieve comparable performance or superior to traditional CNNs while requiring fewer computational resources, rendering them particularly well-suited for resource-constrained environments.

3.2.3 Federated Learning

Federated Learning is a decentralized and privacy-preserving approach to machine learning, which enables training models on data distributed across multiple devices or organizations without exposing or sharing the raw data. In the context of histopathological image detection, federated learning can be particularly beneficial due to the sensitive nature of medical data and the need to maintain patient privacy. Figure 3.6 represents the architecture of a federated learning framework.

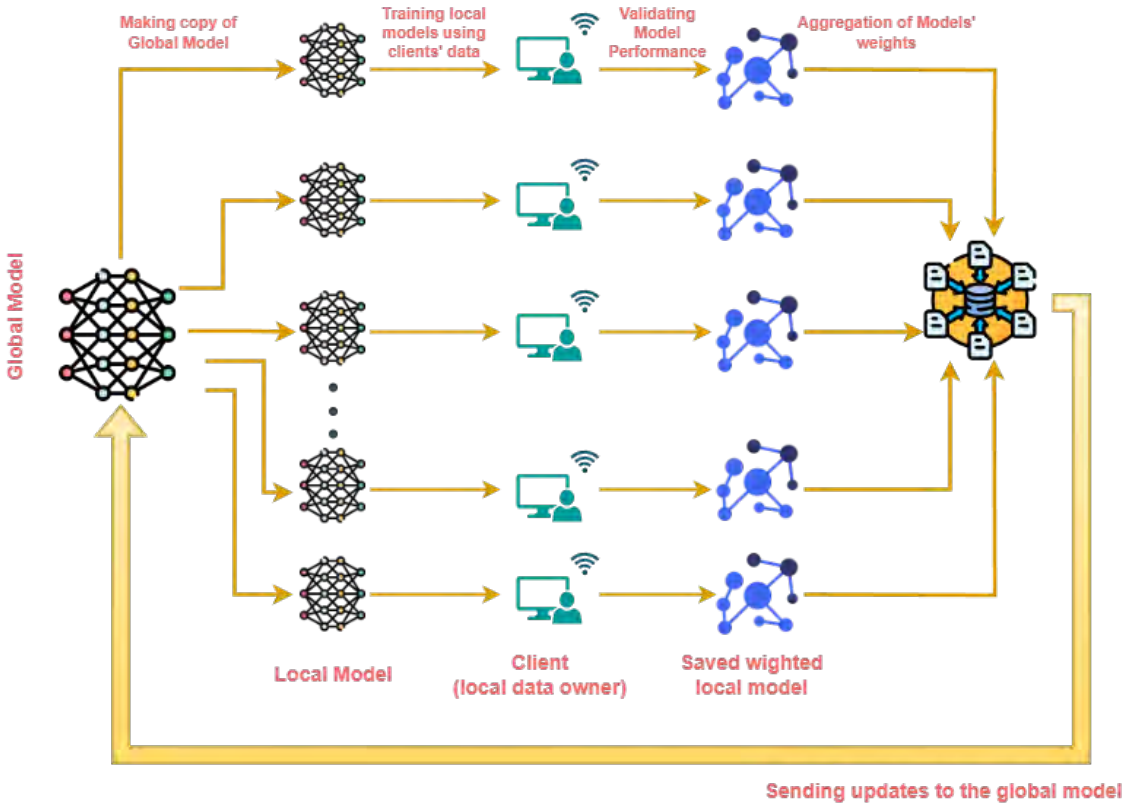


Figure 3.6: Architecture of the Federated Learning Enviroment

Instead of centralizing all the data in a single location, the data remains distributed across multiple sources, such as hospitals, medical research centers, or individual devices. Each participating entity holds a portion of the overall dataset [28]. A central server initializes a global model with initial parameters or weights. This global model is then distributed to the participating entities [29]. Each participating entity trains the global model locally on its data. This local training process involves updating the model's weights based on the entity's respective image data. During this process, the raw data never leaves the local device or organization [30]. After the local training is completed, each participating entity sends model parameters or updates in the weights to the central server. The parameters do not contain any raw data, which is another guarantee of the privacy of data [31]. As a next step, the central server aggregates the received weight updates from each participating entity, usually by averaging or applying more sophisticated methods of aggregation. Hence, the updated global model incorporates the learning from all the data sources in a distributed manner. The new global model is then sent back to all the participating entities for the next round of local training. The process of steps 3 to 6 is repeated for many more iterations until the global model converges or reaches the desired performance level. Following are some of the benefits of federated learning with respect to the detection of histopathological images:

- **Privacy Preservation:** The raw histopathological images and data of the patient remain at the original sites. This does not create any need to share and centralize data. This can potentially provide security to the privacy of patients and be able to comply with data protection.

- **Data Parallelism:** With the use of distributed data resources, federated learning may potentially aggregate larger and more diverse data from different sources to improve model performance. Besides, it may improve generalization.
- **Collaborative Learning:** Many institutions or organizations can train a shared model without directly giving out their sensitive data, which may motivate research collaboration and knowledge sharing.
- **Scalability:** Scalability of federated learning to a large number of entities. Each entity needs to maintain its data during local training. Sometimes, customized models can be learned by federated learning for individual data sources or devices, and can sometimes benefit from the accuracy of histopathological image detection for some populations or settings.

3.3 Proposed ResInvolution Architecture

The proposed ResInvolution architecture consists of three primary parts, the projection block, the residual block, and the fully connected classifier block. That model starts with a projection block with two parallel streams, one stream consisting of involution layers while the other stream consists of a convolution layer being directly added to the stream with the involution layers, serving as a skip connection. Convolution layers have weaknesses due to having a limited local receptive field, preventing them from getting access to global special patterns. Thus, tasks that require access to global special information, and convolution layers may not perform as effectively. Meanwhile, involution layers, due to having dynamic kernels, can extract both local and global special information when necessary. Therefore, the projection block includes involution layers with an attempt to extract specialized features with its dynamic kernel, and in case it fails, the parallel convolution stream delivers the regular convolution-extracted features and concatenates with the overall feature volume, providing a safe involution-convolution combined feature extraction space. In short, the combination of convolutional layers and involutional operations is designed to enhance the model's capability to adapt its processing to different local and global spatial contexts, potentially improving performance over traditional architectures that rely solely on convolutional layers. In addition, due to the parallel structural nature of the involution and convolution layers, the backpropagation should be less affected by the depth of the involution block.

The features of the projection block are then passed through regular residual blocks, which consist of parallel skip connections as the projection block. The residual block is composed of three identity blocks, each containing convolution, batch normalization, and ReLU activation. Meanwhile, a skip connection is attached from the previous block's activation to the given block's output feature, preventing the model from facing the vanishing gradient issue. There are two residual blocks in total. The expectation is to process the concatenated local and global special features extracted from the projection block and find additional features from it.

Finally, at the tail end of the feature extractor, a very lightweight classifier of just the output layer integrated with softmax activation is attached. The number of neurons in the output layer corresponds to the number of classes. The primary feature

Algorithm 1 Proposed ResInvolution Architecture

Input: C_b: Dictionary of batched data,
C_r: Number of epochs,
I_s: Input Shape,
C: Number of classes

Output: Predicted Disease Class

function ResInvolution($inputImages$)
$image \leftarrow$ preprocess($inputImages$)
$x \leftarrow$ conv_batchnorm_relu($image, filters = 64, kernel = 7, strides = 2$)
$x \leftarrow$ MaxPoolingLayer($x, pool = 3, strides = 2$)
$x \leftarrow$ ResNetBlock($x, filters = 64, reps = 3, strides = 1$)
$x \leftarrow$ GlobalAvgPool2D()(x)
$x \leftarrow$ DenseLayer($x, units = C, activation =' relu'$)
$model \leftarrow$ Model($inputs = input, outputs = x$)
return $model$
end function
function ResNetBlock($x, filters, reps, strides$)
$x \leftarrow$ projection_block($x, filters, strides$)
for all $j = 0$ to $reps - 1$ **do**
$x \leftarrow$ identity_block($x, filters$)
end for
return x
function PROJECTION_BLOCK($tensor, involution, filters, strides$)
for all $j = 0$ to $involution$ **do**
if $j = 1$ **then**
$x \leftarrow$ Involution($tensor, channel = 3, group = 1, kernel = 3, stride = 1, reduction = 2$)
else
$x \leftarrow$ Involution($x, channel = 3, group = 1, kernel = 3, stride = 1, reduction = 2$)
end if
end for
$x \leftarrow$ Conv2D($x, filters = 4 * filters, kernel_size = 1, strides = 1$)
$x \leftarrow$ BatchNormalization()(x)
$y \leftarrow$ Conv2D($filters = 4 * filters, kernel_size = 1, strides = 1$)
$y \leftarrow$ BatchNormalization()(y)
$x \leftarrow$ Add()(y, x)
$x \leftarrow$ ReLU()(x)
return x
end function
end function

extractor is expected to extract very effective distinguishable features combining the properties of involution and convolution and thus, the classifier itself was rather kept minimalistic. With a small-sized classifier, reduced amount of intermediate layers, and low-cost involution layers, the model is extremely lightweight while being effective. Consequently, it should be suitable for a federated learning environment where the reduction of bandwidth cost is necessary. Algorithm 1 shows step by step workflow of the ResInvolution Architecture.
We have incorporated four versions of ResInvolution Architecture in this study. Each version of the architecture has a different number of Involutions in the Projection Blocks.

In Figure 3.7, we can see that the lightweight customized ResNet consisted of only convolution layers. No involution layer exists in the projection blocks of the model.

In Figure 3.8, One involution layer on the left stream of the first projection has been added to the architecture. All the other layers and functions remain unchanged.

Additional involution can be observed in Figure 3.9. This variant of the ResInvolution architecture consisted of two involution layers in the first projection block. It also has the same number of convolution layers and other functions as the previous variant.

Three Involution layers have been added to the first projection block of the lightweight custom ResNet model to develop the three involution-based ResInvolution architecture. Figure 3.10 represents the architecture of the three involution variants of ResInvolution architecture.

The four involution based ResInvolution architecture has four involution layers in the left stream of the first projection block illustrated in 3.9. This variant of the ResInvolution architecture consisted of nine convolution layers with ReLU activation and following one batch normalization.

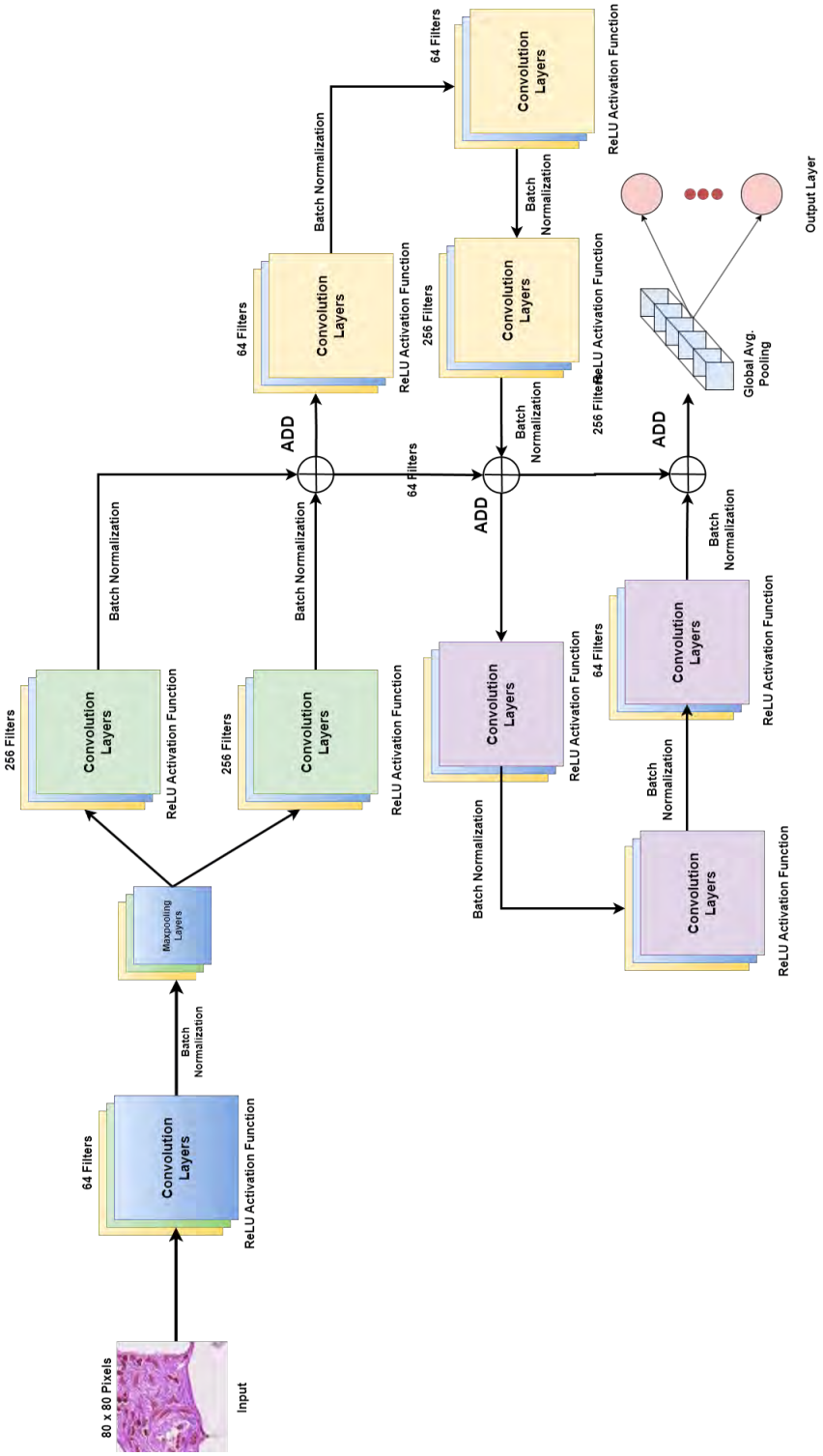


Figure 3.7: A Lightweight Blueprint of ResNet Architecture with No Involution Layer

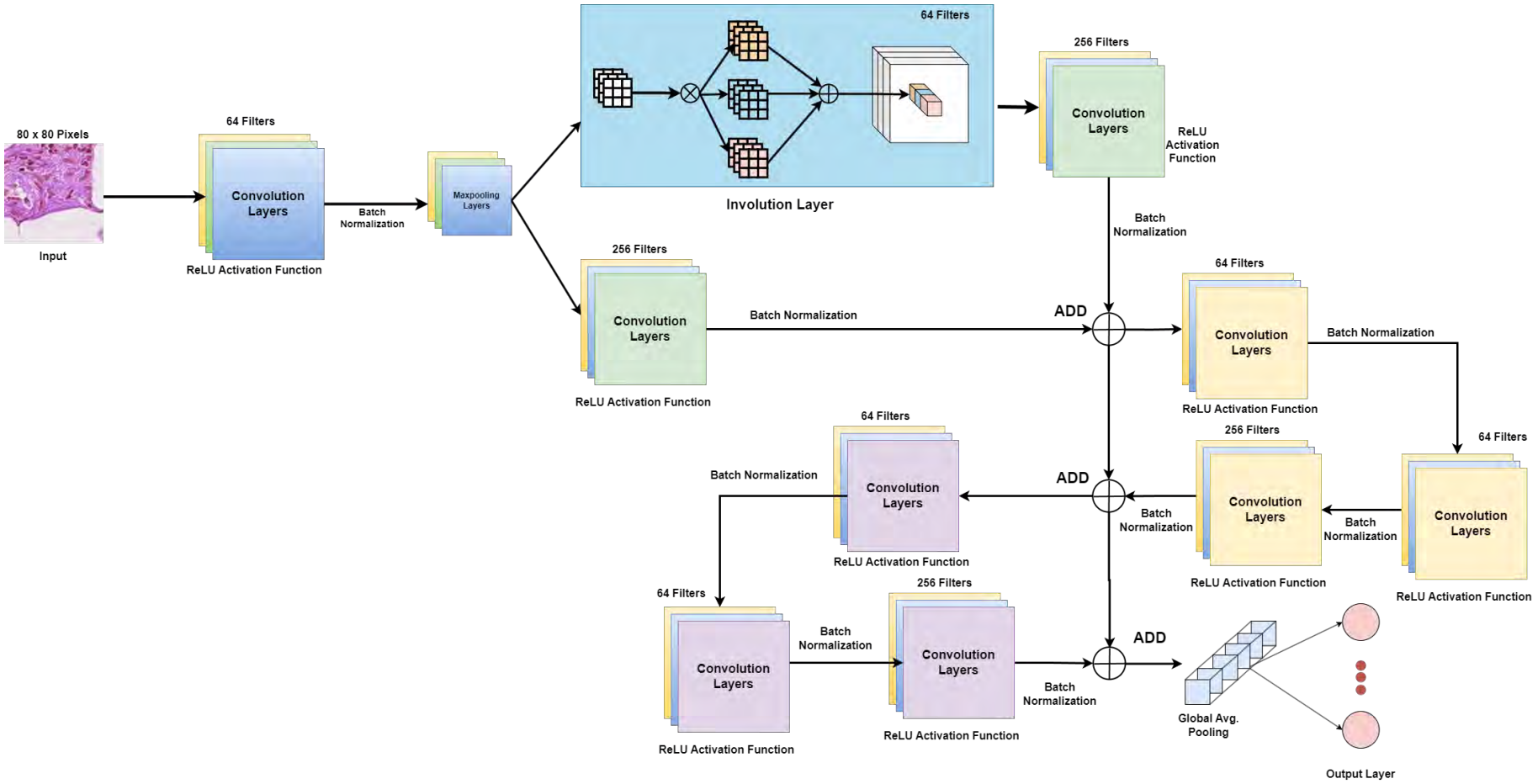


Figure 3.8: ResInvolution Architecture with One Involution Block

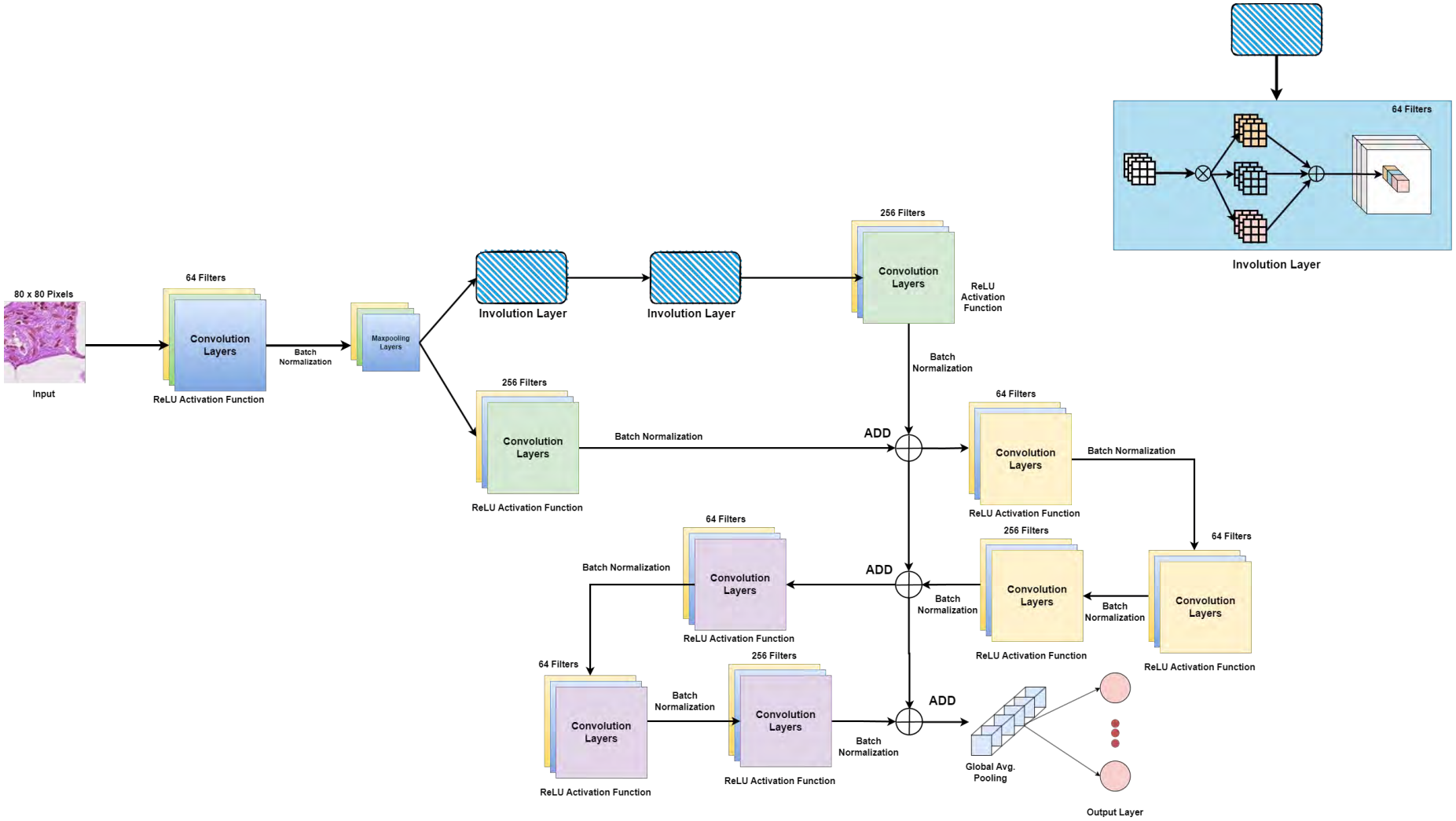


Figure 3.9: ResInvolution Architecture with Two Involution Block

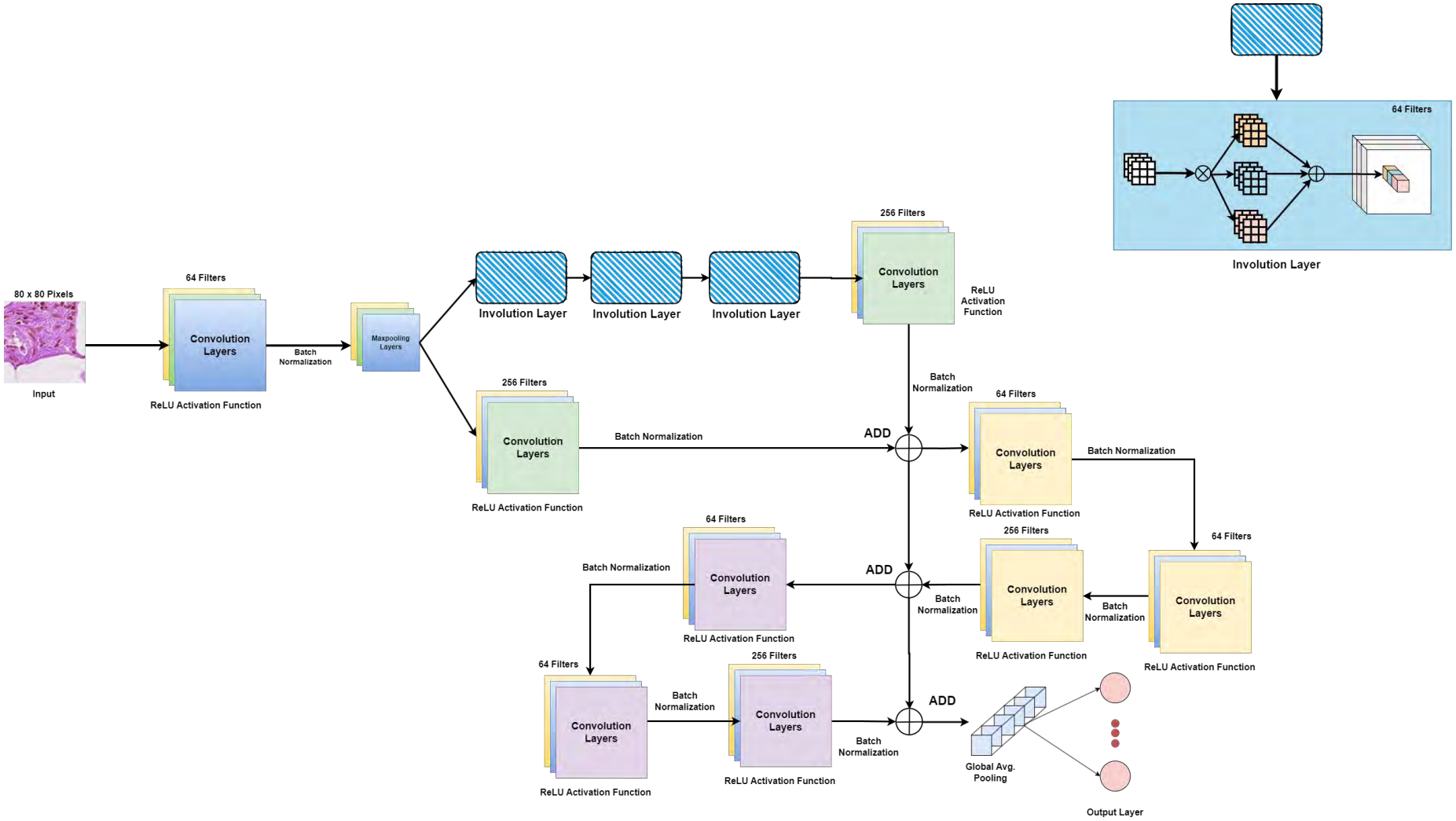


Figure 3.10: ResInvolution Architecture with Three Involution Block

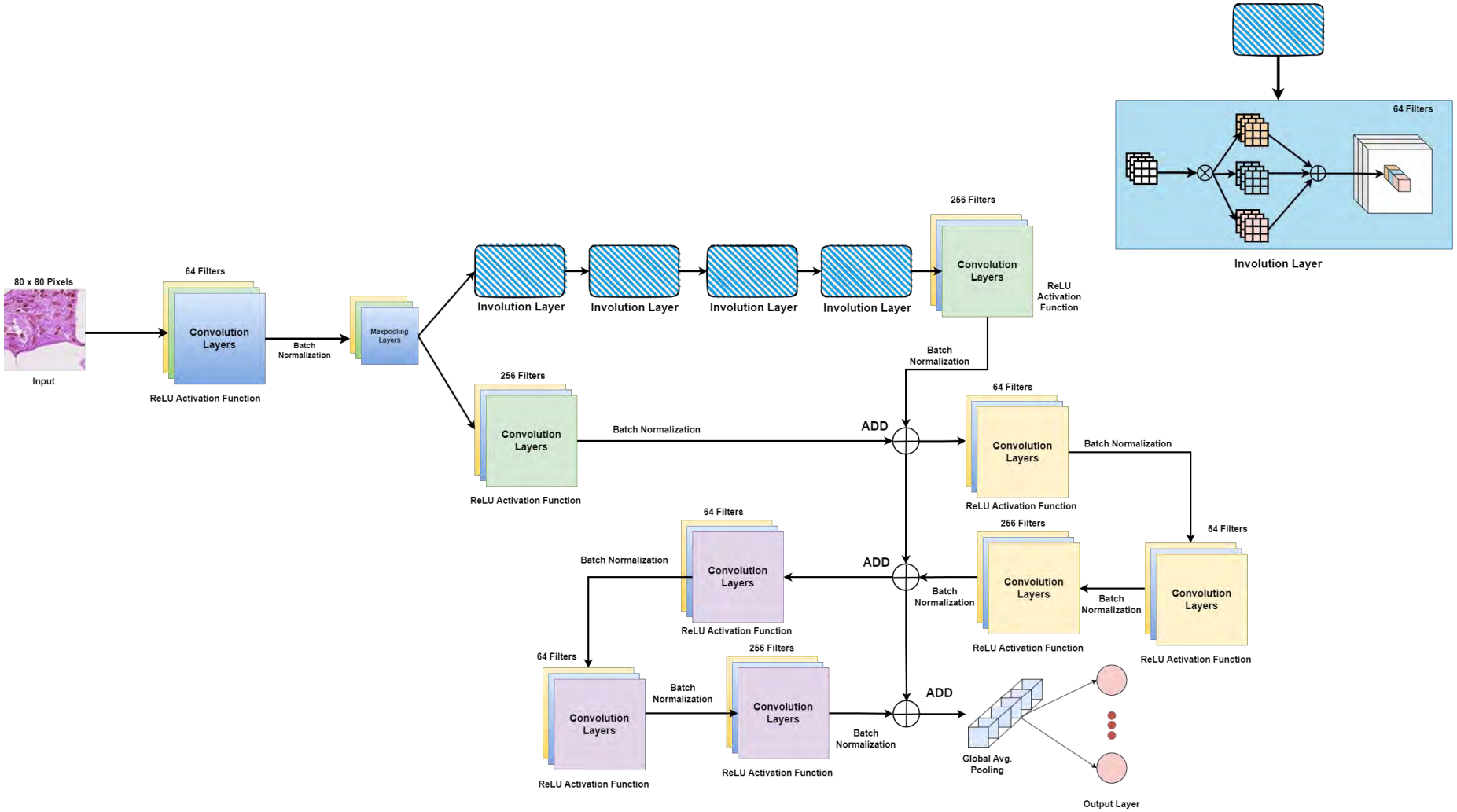


Figure 3.11: ResInvolution Architecture with Four Involution Block

Chapter 4

Result Analysis

In this chapter, we are going to discuss the performance analysis of the ResInvolution architecture in traditional learning and a Federated Learning environment for three different datasets.

4.1 Performance Evaluation Metrics

To evaluate the performance of the model, confusion matrix, accuracy, precision, recall, F1 score, AUC score, ROC curve have been taken into account. These are essential evaluation metrics used to assess the performance of deep learning models, particularly in classification tasks.

- **Accuracy:** This measures the share of correctly predicted instances in comparison to the whole set of instances. That gives a general measure of correctness of the model's predictions [32].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

- **Precision:** It is the ratio of correctly predicted positive instances to the total predicted positive instances. It measures the model's ability to avoid false positives [33].

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

- **Recall:** It is the ratio of correctly predicted positive instances to the total actual positive instances. It measures the model's ability to identify all positive instances [33].

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

- **F1-score:** It is the harmonic mean of precision and recall, providing a balanced measure of both metrics [34].

$$f1 - score = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.4}$$

- **Confusion Matrix:** A tabular representation summarizing the performance of the classification model in terms of counts of right and wrong predictions made by the model for each class [35]. This confusion matrix summarizes the number of correct and incorrect predictions in a very detailed way. The entries in the matrix are the number of correctly and incorrectly classified instances: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).
- **ROC Curve:** It is a plot that illustrates the diagnostic ability of a binary classifier by plotting the true positive rate (recall) against the false positive rate (1 - specificity) at various threshold settings [36].
- **AUC Score:** It is the area under the ROC curve, which summarizes the ROC curve into a single value. It measures the degree of separability between classes, with a value of 1 indicating perfect separability and 0.5 indicating no separability [36].

4.2 Hardware Configuration

This research was conducted using a local workstation equipped with an Intel Core i5 10400 processor, which has 6 cores and 12 threads, running at a frequency of 2.9 GHz. The machine also has 32GB of Ram and an Nvidia Geoforce RTX 4070 Ti Super GPU with 16GB of memory.

4.3 Performance Analysis of ResInvolution Architecture in Traditional Learning Environemnt

In this section we are going to analyze the performance of the variants of our proposed ResInvolution architecture in traditional learning environement. The test performance along with some comparative study on NCT-CRC-HE-100K, GasHisSDB and LC25000 datasets has been analyzed.

4.3.1 Performance Analysis on NTC-CRC-HE-100K Dataset

We have tested all instances of the NTC-CRC-HE-100K dataset at its highest validation accuracy peak after fifty epochs to analyze the performance of the ResInvolution Architecture. In order to assess the performance, we have considered a comprehensive classification report, a confusion matrix, and a Receiver Operating Characteristic (ROC) curve. These possess a crucial role in evaluating the effectiveness of a classification model established to differentiate between different classes.

4.3.1.1 Base Architecture with No Involution Layers

From the confusion matrix in Figure 4.1 we can see that The base architecture with no involution layers demonstrates impressive performance, especially for classes such as MUC and TUM, where the true positives are high, indicating successful detection of these instances. The true negatives are consistently strong across different categories, indicating a high level of specificity. Nevertheless, some instances of incorrect

positive and negative classifications, particularly in the BACK and DEB categories, indicate areas for potential improvement in the model.

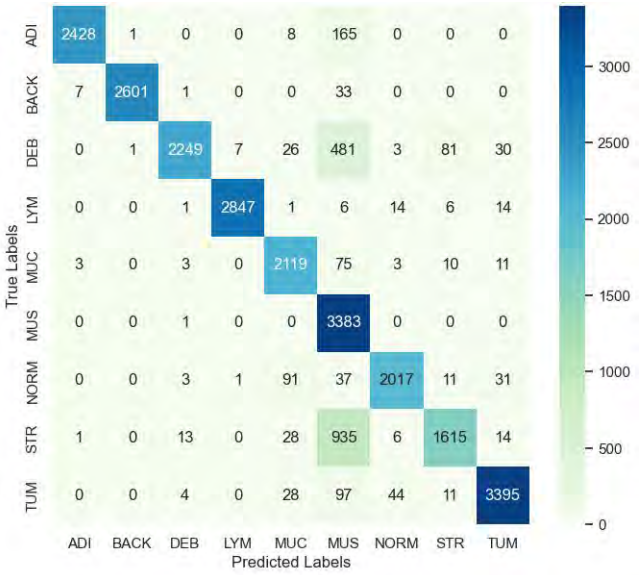


Figure 4.1: Confusion Matrix Generated based on the prediction of Base Architecture with No Involution Layers

The classification report of the base architecture with no involution layers in Figure 4.2 exhibits strong performance with high precision, recall, and F1-scores, especially in categories such as ADI, BACK, LYM, and TUM. Although there are areas that could be enhanced, one particular category that stands out is MUS. The precision in this category is noticeably lower, suggesting that there is room for improvement in terms of feature recognition capabilities.

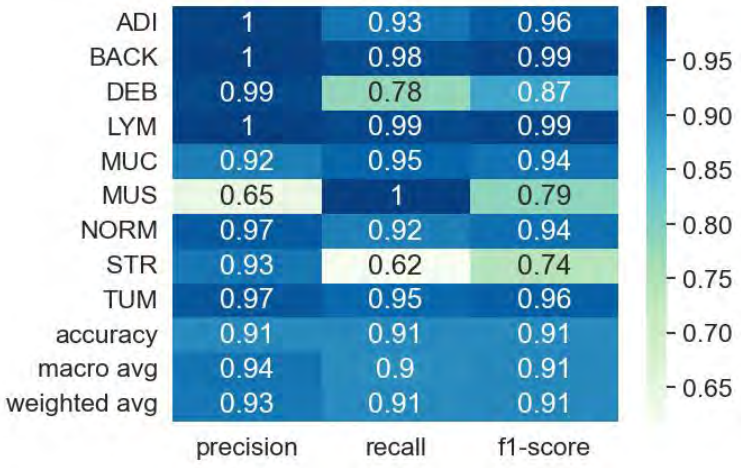


Figure 4.2: Classification Report of the Base Architecture with No Involution Layers

The base architecture showcases impressive performance, especially in categories such as BACK and LYM, with AUC scores reaching an exceptional 0.99. Nevertheless, certain categories, such as STR, have an AUC of 0.81, suggesting room for

improvement in the base model's ability to distinguish between classes. Figure 4.3 displays the ROC curves and the corresponding AUC scores.

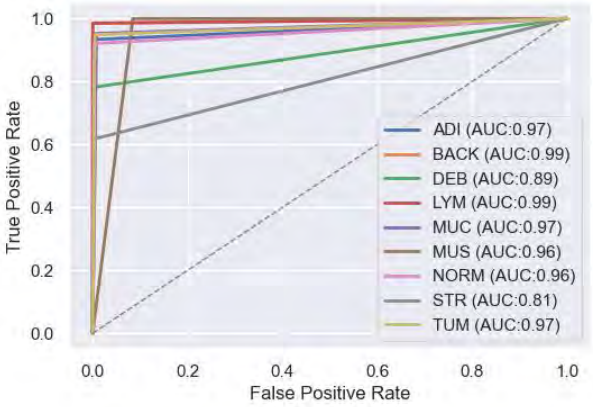


Figure 4.3: ROC Curve and AUC Score of the Base Architecture with No Involution Layers

4.3.1.2 ResInvolution Architecture with One Involution Layers

Based on the confusion matrix illustrated in Figure 4.4, it is evident that the one involution variant demonstrates an enhancement in reducing false negatives in classes like LYM and MUC.

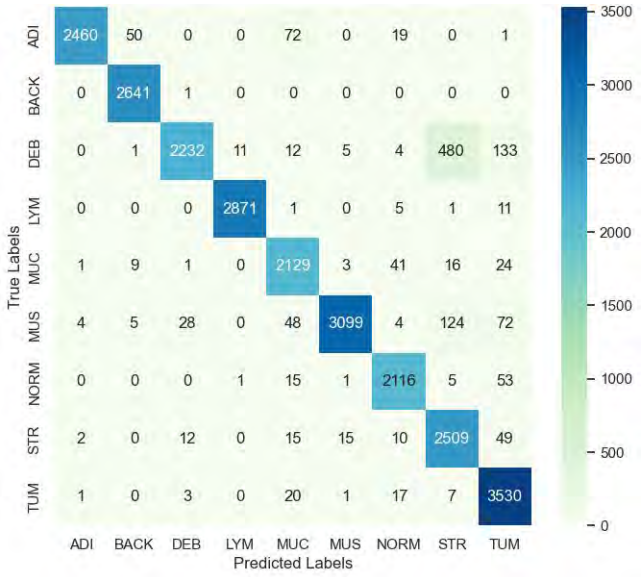


Figure 4.4: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with One Involution Layer

This improvement in sensitivity surpasses the previous variant. Additionally, there is a noticeable decrease in incorrect positive predictions in multiple categories, indicating that the incorporation of an involution block contributes to improving the model's ability to extract meaningful features and enhance its overall classification accuracy.

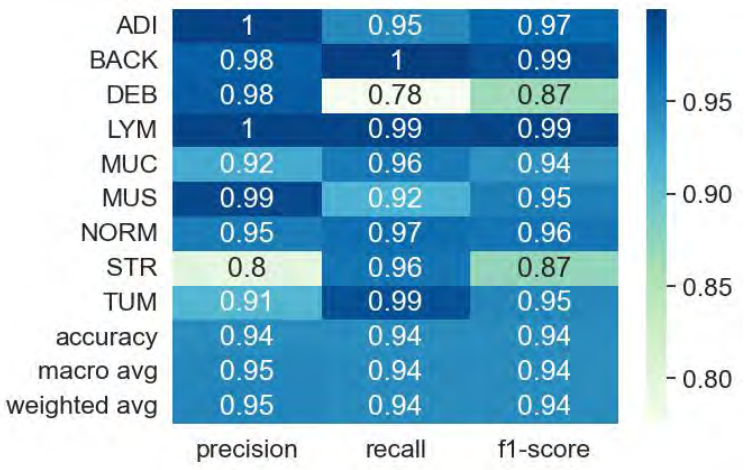


Figure 4.5: Classification Report of the ResInvolution Architecture with One Involution Layer

The classification report of the One Involution variant mentioned in Figure 4.5 demonstrates notable advancements compared to the No Involution variant, specifically in the DEB and MUS classes. Both precision and recall have experienced substantial improvements in these areas. This improvement indicates that incorporating an involution block enhances the model's ability to effectively capture and handle complicated features, resulting in a higher level of accuracy in classification across all categories. This version demonstrates a more balanced distribution of high performance across all categories, suggesting the positive impact of the architectural modification.

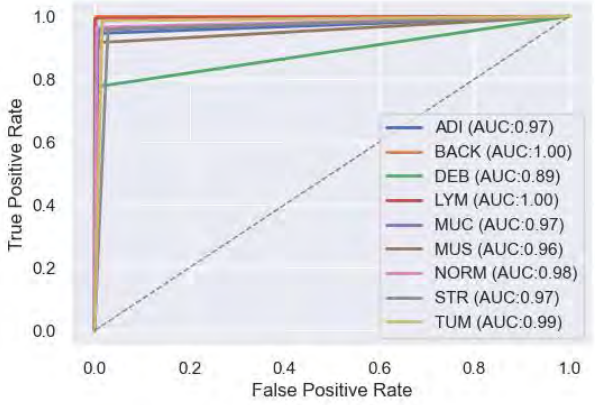


Figure 4.6: ROC Curve and AUC Score of the ResInvolution Architecture with One Involution Layer

Figure 4.6 demonstrates that the inclusion of a single involution block results in significant enhancements across multiple categories. The AUC scores for BACK, DEB, and TUM are extremely high, indicating a substantial improvement in the model's classification accuracy for these conditions. By incorporating an involution block, the model's sensitivity and specificity are significantly improved, showcasing its efficacy in enhancing feature extraction and boosting the overall classification accuracy.

4.3.1.3 ResInvolution Architecture with Two Involution Layers

By analyzing the confusion matrix in Figure 4.7, it becomes evident that there is a significant improvement in both true positives and true negatives when advancing to two involution layers. This improvement is particularly notable in challenging categories such as STR and NORM. This version greatly reduces the occurrence of incorrect positive and negative results in various categories, highlighting the importance of incorporating deeper involutional integration to achieve more precise and accurate diagnostic outcomes.

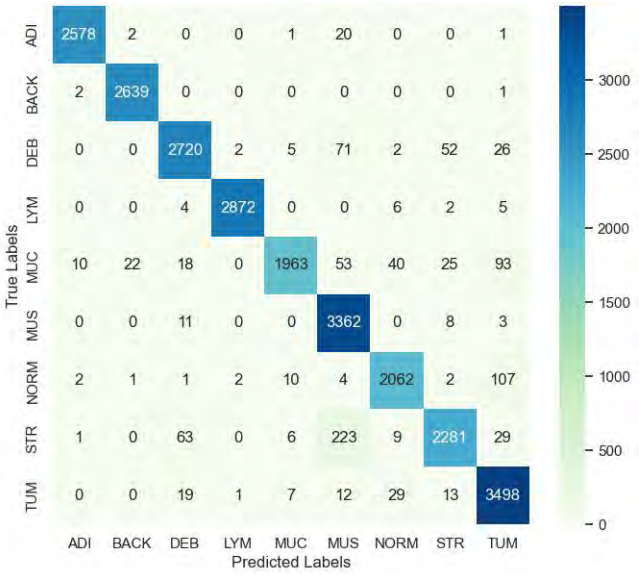


Figure 4.7: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Two Involution Layers

Based on the classification report in Figure 4.8, it is evident that the Two Involution layers significantly enhance the model's performance. especially in difficult categories like MUC and NORM.

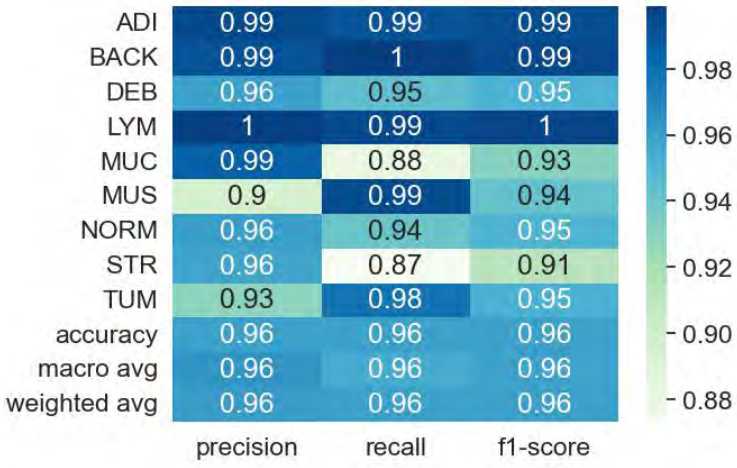


Figure 4.8: Classification Report of the ResInvolution Architecture with Two Involution Layers

The addition of the second involution block not only improves the overall accuracy but also enhances performance across the board, as shown by the increased macro and weighted average scores to 0.96. It appears that there is an improvement in the accuracy and dependability of the model as more involution layers are added.

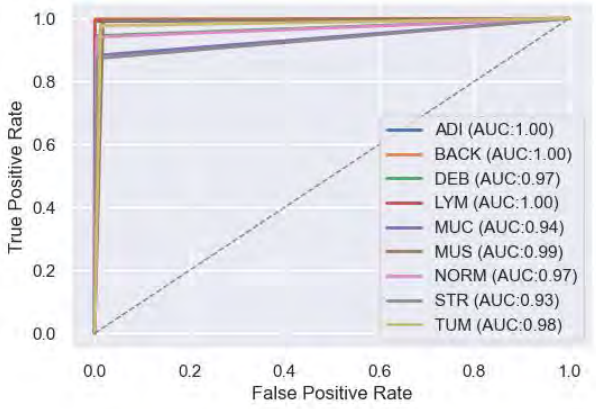


Figure 4.9: ROC Curve and AUC Score of the ResInvolution Architecture with Two Involution Layers

The model's performance is further enhanced with the incorporation of a second involution block, as evidenced by the ROC curves and AUC scores presented in Figure 4.9. Perfect scores (1.00) are achieved in challenging classes such as LYM and MUC, while there is a significant improvement in STR, which rises to 0.93. These findings indicate that the inclusion of two involution layers enhances the model's ability to effectively handle the intricacies of histopathological image features. As a result, the model exhibits improved classification performance, demonstrating greater accuracy and robustness across a wider spectrum of conditions.

4.3.1.4 ResInvolution Architecture with Three Involution Layers

The confusion matrix in Figure 4.10 demonstrates that the Three Involution variant of the ResInvolution architecture delivers the most significant enhancements, particularly in terms of true positives for challenging categories like STR and NORM. This highlights the substantial diagnostic capabilities that can be achieved by incorporating additional involution layers. The significant decrease in both incorrect positive and negative results across different categories highlights the model's strength and dependability for clinical applications.

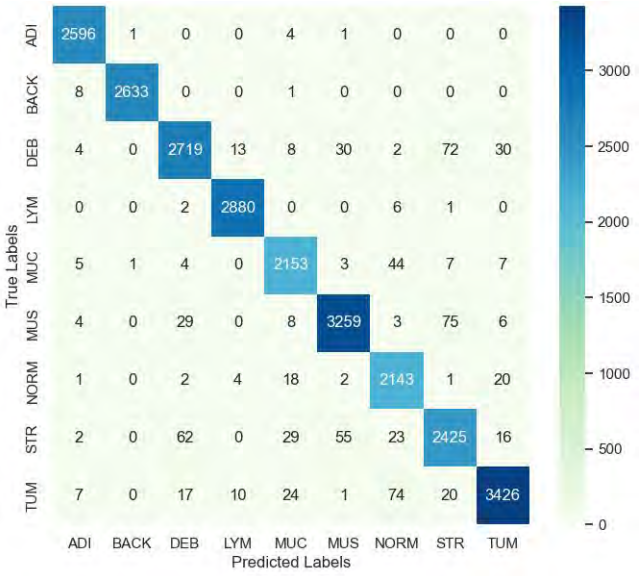


Figure 4.10: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Three Involution Layers

The effectiveness of the ResInvolution architecture appears to be optimized, as indicated by the classification report for the Three Involution variant in Figure 4.11. Almost all diagnostic classes now show F1-scores of 0.96 or higher, with significant improvements in precision and recall in categories such as STR and MUS that were previously lacking. This model configuration seems to strike a good balance, maximizing the effectiveness of the architecture without making it unnecessarily complex.

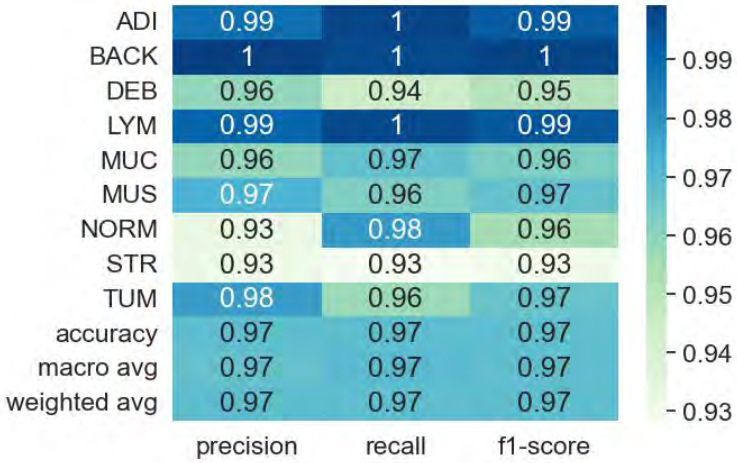


Figure 4.11: Classification Report of the ResInvolution Architecture with Three Involution Layers

As illustrated in Figure 4.12, the inclusion of a third involution block marginally improves the area under the curve (AUC) values, thereby establishing a consistent level of excellence in almost all domains. Specifically, STR and NORM attain AUCs of 0.96 and 0.99, respectively. This variant appears to strike a perfect balance, maximizing diagnostic accuracy while keeping the model simple and straightforward.

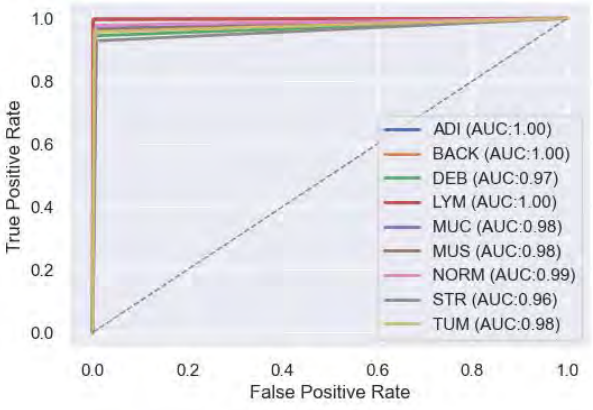


Figure 4.12: ROC Curve and AUC Score of the ResInvolution Architecture with Three Involution Layers

4.3.1.5 ResInvolution Architecture with Four Involution Layers

The confusion matrix shown in Figure 4.13 demonstrates that incorporating a fourth involution block leads to slight enhancements in true positives and true negatives for certain categories such as ADI and BACK. However, it also brings about some increases in false positives, especially in categories such as DEB, which may indicate potential diminishing returns with an excessive number of involution layers.

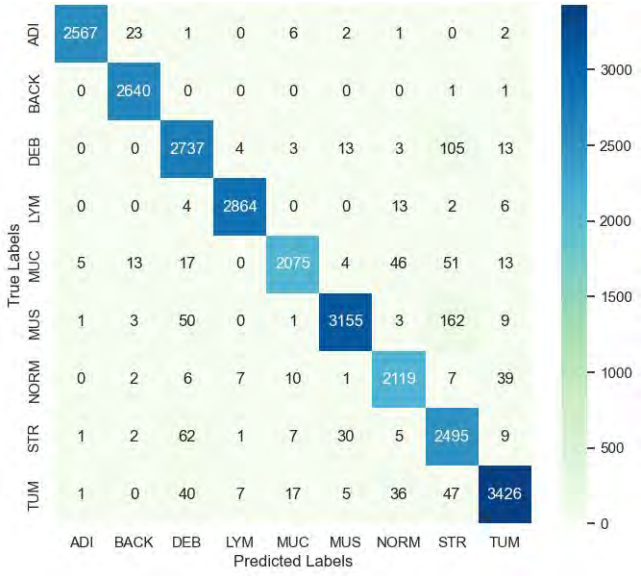


Figure 4.13: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Four Involution Layers

The overall performance, although impressive, suggests that there could be a maximum threshold for the number of involution layers that can be incorporated before the model's practical effectiveness is compromised by increased complexity.

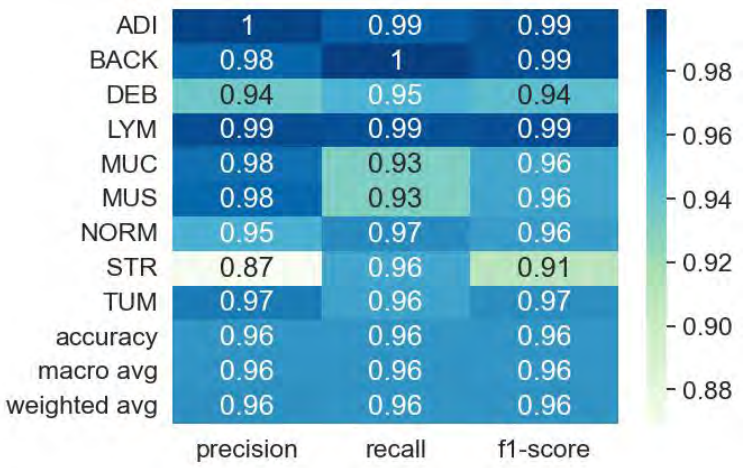


Figure 4.14: Classification Report of the ResInvolution Architecture with Four Involution Layers

Figure 4.14 shows the classification report of the Four Involution variant. It reveals that including a fourth block does not lead to substantial improvements. In fact, in certain cases, there are slight decreases in metrics such as precision in MUS and STR. This could suggest the limited benefits of increased complexity. Although the performance across categories remains consistently high, there are minor fluctuations that indicate the potential effectiveness of using three involution layers. This configuration seems to strike the optimal balance between complexity and performance efficacy.

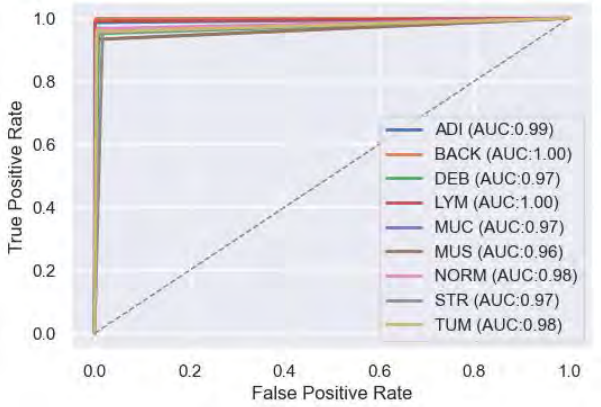


Figure 4.15: ROC Curve and AUC Score of the ResInvolution Architecture with Four Involution Layers

Based on the AUC scores and ROC curves presented in Figure 4.15, it appears that adding a fourth involution block does not result in a significant improvement in the AUC scores compared to the configuration with three involution layers. In fact, in certain cases like STR, the AUC slightly decreases to 0.97. Based on the data, it appears that including additional involution layers may not necessarily improve performance. In fact, it could potentially result in diminishing returns, where the model becomes too complex without any significant benefits.

4.3.2 Performance Analysis on GasHisSDB Dataset

For analyzing the performance of ResInvolution Architecture on the GasHisSDB dataset we have tested all the instances at it's highest validation accuracy peak after fifty epochs. To analyse the performance a comprehensive classification report, a confusion matrix, and a Receiver Operating Characteristic (ROC) curve have neem taken into account. These are instrumental in assessing the performance of a classification model designed to distinguish between 'Normal' and 'Abnormal' classes for Gastric Cancer Classification.

4.3.2.1 Base Architecture with No Involution Layers

The confusion matrix in Figure 4.16 reflects the performance of the base architecture without any involution layers. The variant of the ResInvolution architecture demonstrated a success rate of 19,919 true negatives (correctly identified normal cases) compared to 12,401 accurate positive predictions (correctly identified abnormal cases). In contrast, 1,956 false positives (normal cases erroneously identified as abnormal) and 2,387 false negatives (abnormal cases erroneously identified as normal) occurred. Subsequent comparisons are conducted with this variant serving as the foundation.

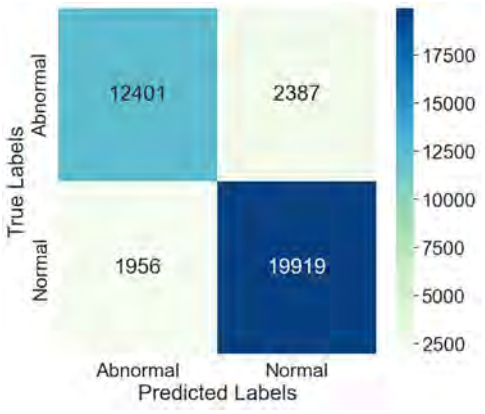


Figure 4.16: Confusion Matrix Generated based on the prediction of Base Architecture with No Involution Layers

Based on Figure 4.17, the base architecture shows moderate performance metrics, achieving an overall accuracy of 0.8815 without any involution layers. The precision and recall scores for the abnormal category are 0.8638 and 0.8386, respectively. These results indicate that accurately identifying abnormal cases poses certain challenges. This configuration establishes a foundation for future advancements, highlighting the promising possibilities that can arise from incorporating involution layers.

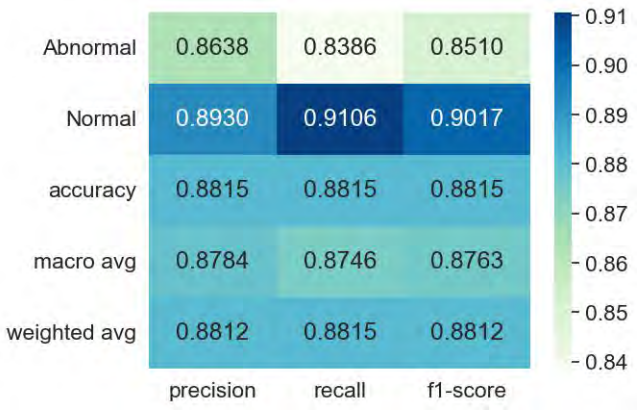


Figure 4.17: Classification Report of the Base Architecture with No Involution Layers

The ROC curve represented in Figure 4.21 plots the true positive rate against the false positive rate at various threshold settings. The AUC (Area Under Curve) is 0.87, indicating a good ability of the model to differentiate between the 'Normal' and 'Abnormal' classes.

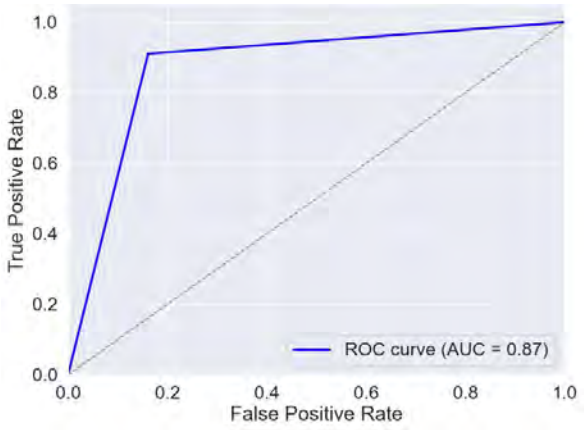


Figure 4.18: ROC Curve and AUC Score of the Base Architecture with No Involution Layers

4.3.2.2 ResInvolution Architecture with One Involution Layers

Figure 4.19 shows that significant performance enhancement was observed when one involution layer was incorporated, as opposed to the variant without the involution layer. The number of true positives has increased to 13,755, while the true negatives have also shown improvement, reaching 18,332. This suggests a more accurate identification of both abnormal and normal cases. The number of false negatives has decreased to 1,033, while there has been a slight increase in false positives to 3,543 compared to the baseline. This indicates a heightened ability to detect abnormalities, although there may be a slight increase in incorrectly classifying normal cases as abnormal. While comparing it to the previous variant, the One Involution variant demonstrates better accuracy in predicting true positive cases

and a notable decrease in false negatives. This highlights how incorporating one involution layer can enhance the detection of abnormal cases.

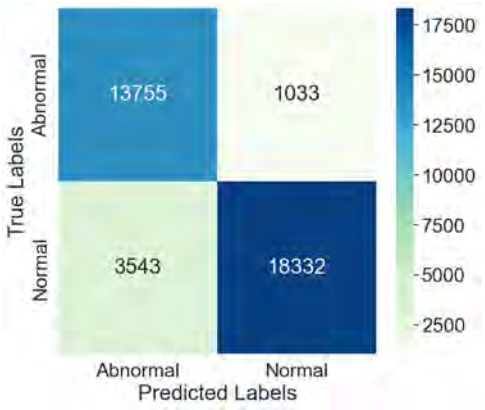


Figure 4.19: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with One Involution Layer

By incorporating an involution layer, the classification performance in Figure 4.20 demonstrates a noticeable enhancement compared to the variant without an involution layer. Specifically, the recall for the abnormal class experiences an increase, reaching 0.9301. Despite a slight decrease in precision for abnormal cases to 0.7952, the incorporation of involution layers enhances the overall balance between precision and recall, highlighting their effectiveness in accurately detecting true abnormal conditions.

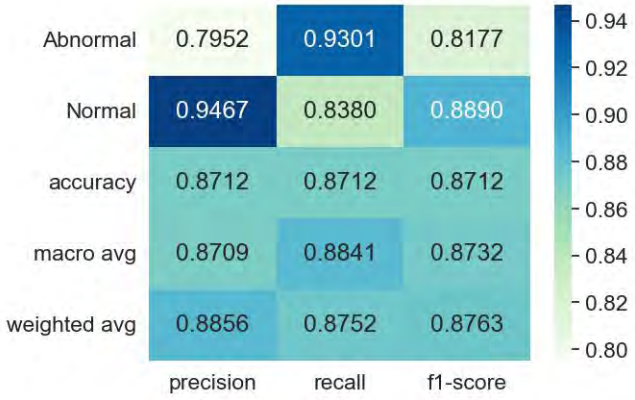


Figure 4.20: Classification Report of the ResInvolution Architecture with One Involution Layer

The ROC curve represented in Figure 4.21 maintains a higher trajectory slightly longer indicating it provides a better balance between detecting true positives while keeping false positives low across a broader range of decision thresholds. The AUC (Area Under Curve) is 0.88, indicating a good ability of the model to differentiate between the 'Normal' and 'Abnormal' classes. This ROC curve suggests a slight improvement in model performance then the No Involution variant of the base architecture, as evidenced by the higher AUC score and the more favorable curve shape.

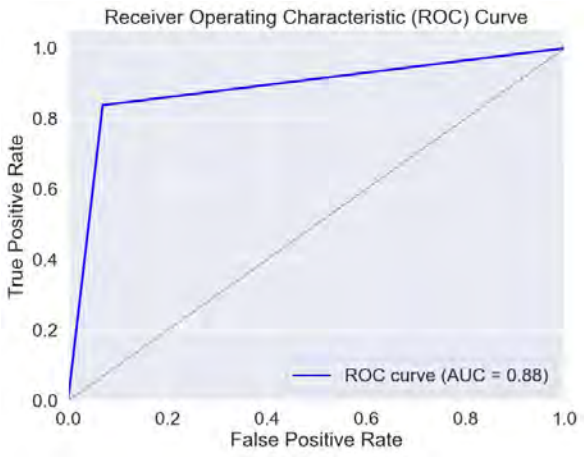


Figure 4.21: ROC Curve and AUC Score of the ResInvolution Architecture with One Involution Layer

4.3.2.3 ResInvolution Architecture with Two Involution Layers

Enhanced predictive capabilities were achieved by incorporating an additional involution layer. From Figure 4.22 it is observed that the number of true negatives exhibited a significant improvement, reaching 21,349 and a slight true positives got decreased to 10,591. The number of false positives significantly reduced to 526, while the number of false negatives increased to 4,197. This configuration showcases a shift towards placing greater emphasis on accurately identifying normal cases. This matrix demonstrates a notable enhancement in accurately identifying negative cases when compared to previous variants. It also shows a substantial reduction in false positives, resulting in an improved precision of the model in distinguishing normal cases.

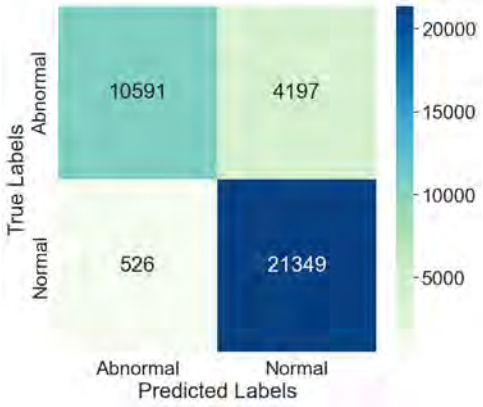


Figure 4.22: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Two Involution Layers

The capacity of the model to reliably categorize circumstances is further improved with the addition of two involution layers, as can be seen in Figure 4.23. The precision for abnormal cases shows a significant improvement, reaching an impressive 0.9527. This suggests a reduction in false positives. Additionally, the recall for

normal cases also achieves a high value of 0.9760. This configuration outperforms the One Involution and No Involution variants, showcasing substantial improvements in accurately identifying true conditions and minimizing incorrect labels.

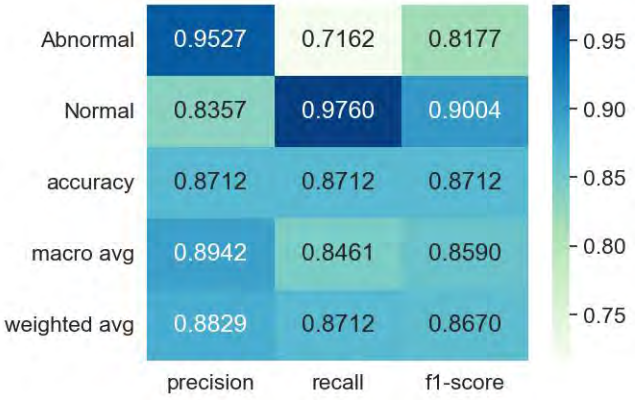


Figure 4.23: Classification Report of the ResInvolution Architecture with Two Involution Layers

From Figure 4.24, The ROC curve and the AUC score of 0.87 which is similar to the the No involution variant's performance. The ROC curve and the AUC score of Two Involution Variants of ResInvolution architecture suggest that the classifier performs well, particularly at distinguishing between the Normal and Abnoraml classes efficiently at various thresholds. However, as always with classification tasks, especially in sensitive fields such as healthcare, it's crucial to consider the specific needs and costs associated with false positives and false negatives to optimally set the decision thresholds. The analysis indicates that while the model is robust, careful consideration and potentially further tuning might be needed depending on the specific application to maximize its effectiveness and minimize potential risks or costs.

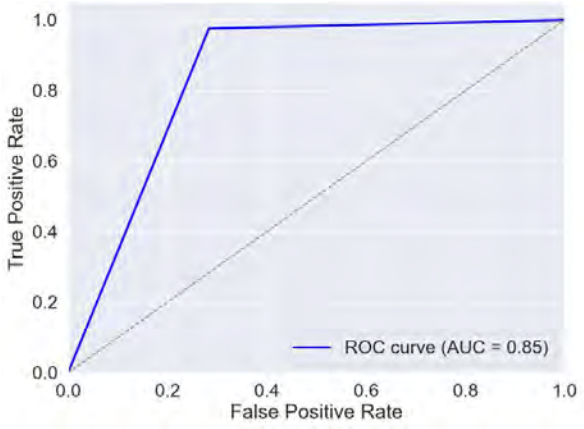


Figure 4.24: ROC Curve and AUC Score of the ResInvolution Architecture with Two Involution Layers

4.3.2.4 ResInvolution Architecture with Three Involution Layers

The addition of another layer of involution demonstrates even more progress. The confusion matrix presented in Figure 4.25 shows that there were a total of 12,963 true positives and 20,608 true negatives. It's worth noting that there were decreases in both false negatives, which went down to 1,825, and false positives, which decreased to 1,267. This indicates a well-rounded enhancement in the ability to accurately identify both abnormal and normal cases. This matrix demonstrates a significant overall enhancement compared to previous all variants, with improved detection of both abnormal and normal conditions, while also achieving lower rates of false positives and false negatives.

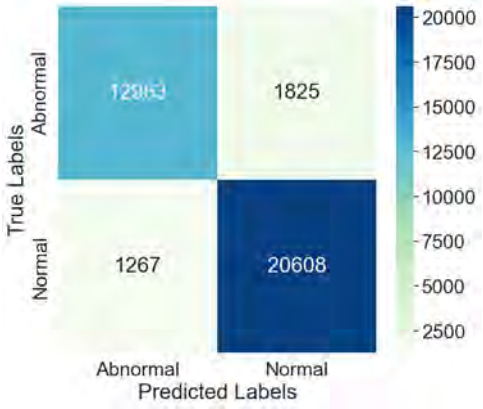


Figure 4.25: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Three Involution Layers

The optimal performance of the ResInvolution Architecture is achieved by the Three Involution variant, as per the classification report shown in Figure 4.26.

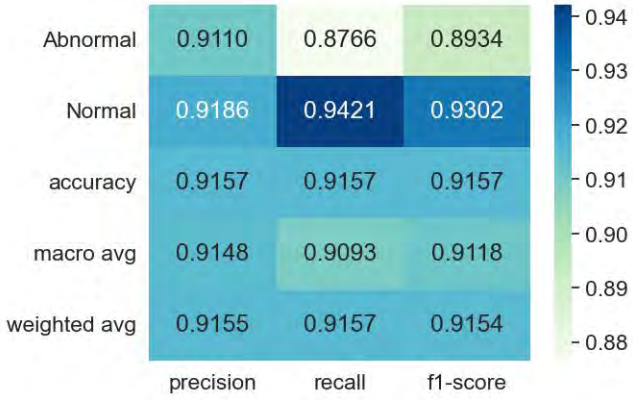


Figure 4.26: Classification Report of the ResInvolution Architecture with Three Involution Layers

This is accomplished by increasing the precision for abnormal cases to 0.9110 and the recall for normal cases to 0.9421. The overall accuracy of 0.9157 demonstrates a significant improvement compared to previous variants. This highlights the effectiveness of incorporating additional involution layers in achieving a strong balance

between detecting abnormalities and maintaining high accuracy across classifications. This configuration outperforms previous versions, showcasing substantial improvements in accurately identifying genuine conditions and minimizing incorrect classifications.

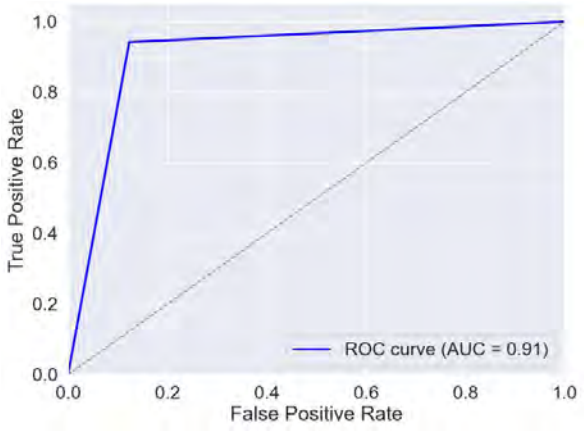


Figure 4.27: ROC Curve and AUC Score of the ResInvolution Architecture with Three Involution Layers

The ResInvolution Architecture with three Involution Layers exhibits a commendable performance in discriminating between class labels as evidenced in Figure 4.27 by the ROC curve and the AUC score of 0.91. This architecture seems particularly well-suited for scenarios where high accuracy is crucial, and the ability to fine-tune the balance between sensitivity and specificity is beneficial. However, careful calibration of the decision threshold is essential to optimize performance according to specific operational requirements, especially in critical applications like medical diagnosis or real-time anomaly detection.

4.3.2.5 ResInvolution Architecture with Four Involution Layers

The confusion matrix in Figure 4.28 demonstrates a high percentage of both 'Normal' and 'Abnormal' cases.

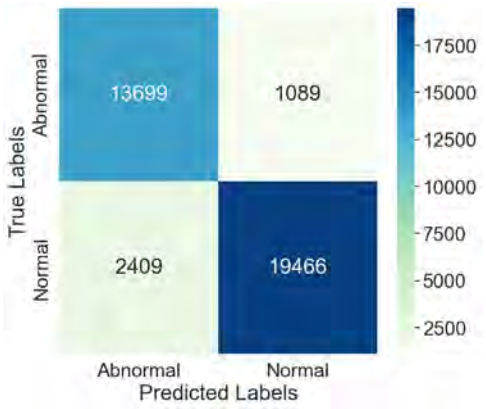


Figure 4.28: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Four Involution Layers

The model successfully identifies a significant number of true positives (19,466) and true negatives (13,699), showcasing its effective discrimination between the two categories. Although there are some false negatives (2,409) and false positives (1,089), the overall results indicate that the model is both precise in its predictions and reliable in various settings. However, the Three Involution variant of the ResInvolution model has an improved ability to identify 'Normal' conditions than this variant.

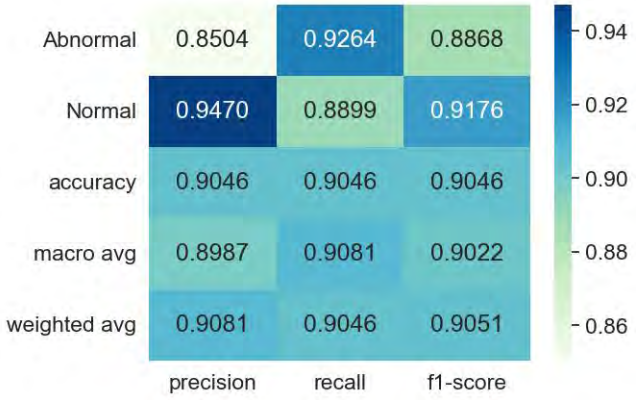


Figure 4.29: Classification Report of the ResInvolution Architecture with Four Involution Layers

The Four Involution variant, as shown in Figure 4.29, achieves a remarkable overall accuracy of 0.9046, demonstrating excellent precision and recall metrics across all categories. It demonstrates an impressive precision of 0.8504 for abnormal cases and an outstanding recall of 0.9264, showcasing its ability to accurately detect abnormal conditions. In typical scenarios, the precision stands at 0.9470, accompanied by a recall of 0.8899. This version provides exceptional performance overall, but it falls short of the three-layer model, which achieves a better trade-off between precision and recall, resulting in superior performance metrics. The Three involution variant continues to demonstrate superior performance in achieving the highest scores across precision, recall, and F1-score, suggesting a more reliable capability to accurately classify both abnormal and normal conditions.

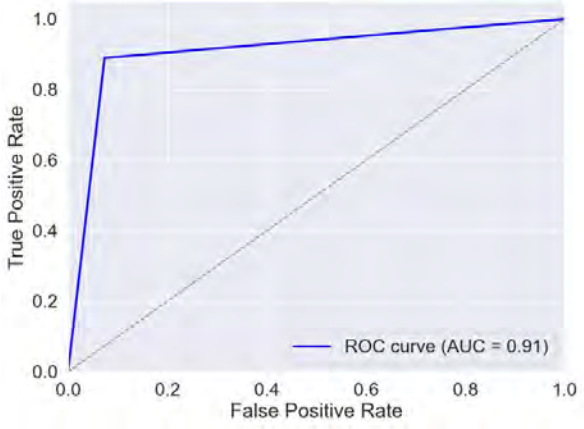


Figure 4.30: ROC Curve and AUC Score of the ResInvolution Architecture with Four Involution Layers

The ROC curve and its corresponding AUC score of 0.91 from Figure 4.30 illustrate that the ResInvolution Architecture with Four Involution layers is effective in handling classification tasks, balancing between detecting true positives and minimizing false positives efficiently. The ROC curve and the AUC score is similar to the Three Involution variant of ResInvolution architecture. This performance is ideal for applications where precision and recall are critically important, such as medical imaging or any other high-stakes field where accurate detection significantly impacts outcomes.

4.3.3 Performance Analysis on LC25000 Dataset

We have conducted extensive testing on all instances for the lung_aca, lung_n, and lung_scc classes of the LC25000 dataset at its highest validation accuracy peak after fifty epochs to thoroughly analyze the performance of the ResInvolution Architecture. To evaluate the performance, we have analyzed a thorough classification report, a confusion matrix, and a Receiver Operating Characteristic (ROC) curve. These play a vital role in assessing the effectiveness of a classification model used to distinguish between various classes.

4.3.3.1 Base Architecture with No Involution Layers

The confusion matrix for the base architecture without involution layers demonstrates in Figure 4.31 a remarkable level of accuracy in classifying different types of lung cancer. The number of true positives is very high for Lung_ACA (1199), Lung_N (1250), and Lung_SCC (1211), with very few false negatives, which suggests a high level of sensitivity. Nevertheless, there are instances of false positives (FP), specifically with Lung_ACA being incorrectly identified as Lung_SCC (50) and vice versa (39), indicating the need for enhanced specificity.

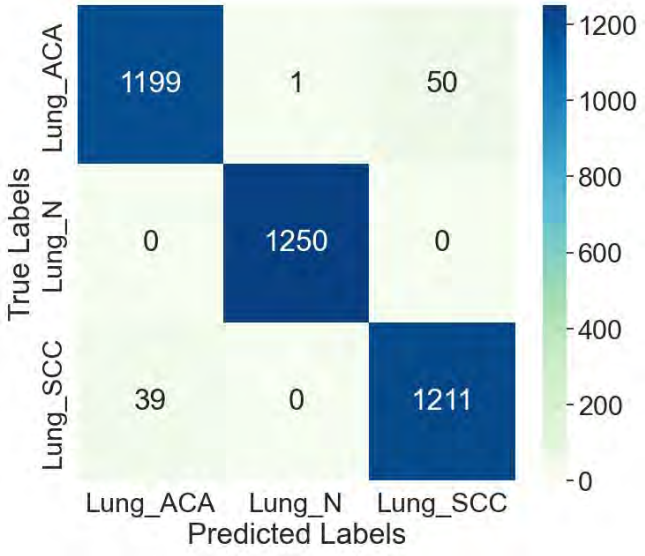


Figure 4.31: Confusion Matrix Generated based on the prediction of Base Architecture with No Involution Layers

Figure 4.32 illustrates that the classification report for the base architecture without

involution layers demonstrates robust performance metrics for all three lung cancer classes: Lung_ACA, Lung_N, and Lung_SCC. There is a consistently high precision, recall, and F1 score; in particular, Lung_N has a flawless recall score. Overall, the performance is outstanding, with both weighted and macro averages across measures of 0.9760.

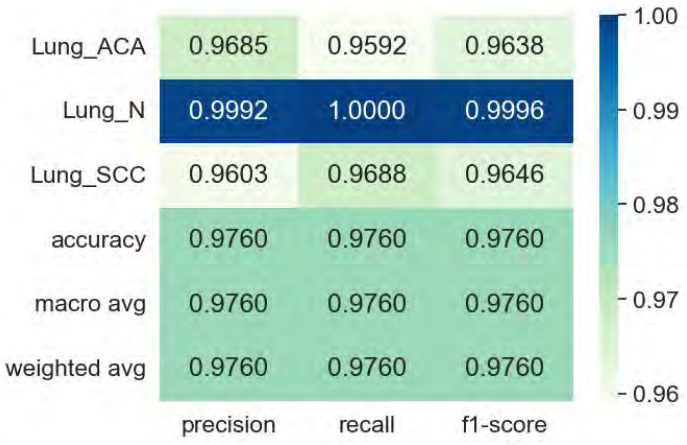


Figure 4.32: Classification Report of the Base Architecture with No Involution Layers

The ROC curve depicted in figure 4.33 for the base architecture without involution layers demonstrates outstanding performance across the classes. It achieves AUC scores of 0.97 for both Lung_ACA and Lung_SCC, and a flawless score of 1.00 for Lung_N. This demonstrates a solid foundation, suggesting that even without involution layers, the model excels at differentiating between the various cancer classes, particularly for Lung_N.

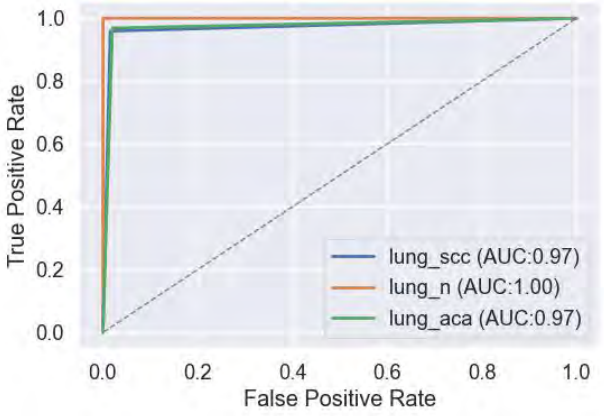


Figure 4.33: ROC Curve and AUC Score of the Base Architecture with No Involution Layers

4.3.3.2 ResInvolution Architecture with One Involution Layers

In comparison to the base architecture, the addition of one involution layer results in a slight improvement in classification. The confusion matrix in Figure 4.34 displays

a notable increase in the number of true positives for Lung_ACA. Additionally, there is a decrease in the misclassification of Lung_ACA as Lung_SCC, as well as a decrease in the misclassification of Lung_SCC as Lung_ACA. When we add involution layers, the matrix shows a modest enhancement in sensitivity and specificity across the classes. This indicates that the model's capability to differentiate between similar classes has improved with the inclusion of involution layers.

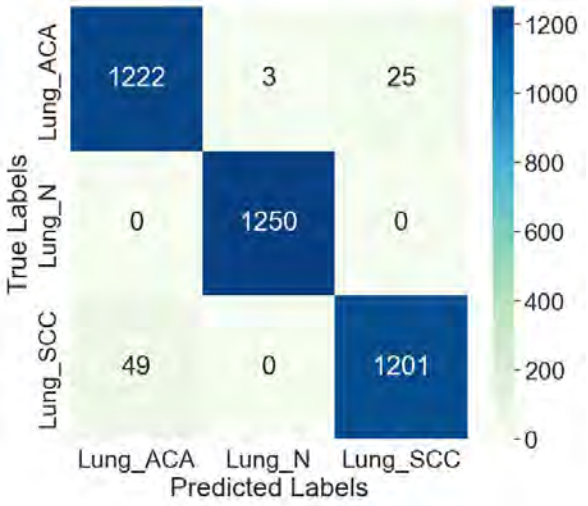


Figure 4.34: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with One Involution Layer

Incorporating a single involution layer results in a slight enhancement in precision and F1-scores for Lung_ACA and Lung_SCC, as indicated in the classification report shown in Figure 4.35. This highlights the layer's contribution to improving the model's precision and overall performance. The recall performance of Lung_N remains flawless, showcasing unwavering accuracy in classifying this particular class. The macro and weighted averages show a slight improvement of 0.9795, indicating that the inclusion of an involution layer has a positive impact on the model's ability to discriminate.

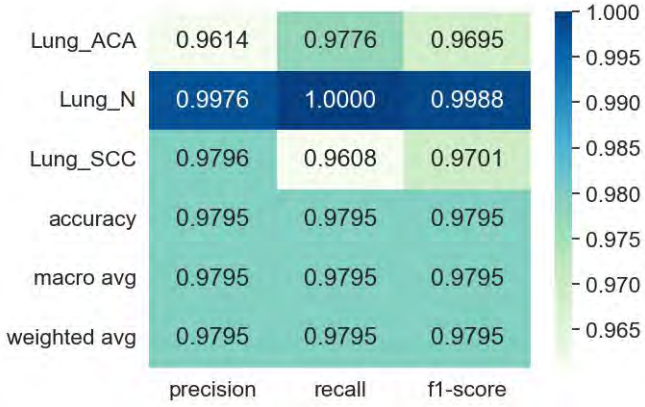


Figure 4.35: Classification Report of the ResInvolution Architecture with One Involution Layer

In Figure 4.36, the addition of one involution layer leads to a slight enhancement in the AUC scores for Lung_ACA and Lung_SCC, with both increasing to 0.98, when compared to Figure 4.33. It appears that incorporating an involution layer improves the model's capacity to accurately distinguish between different types of cancer compared to the configuration without the layer. This leads to a slight reduction in false positive rates.

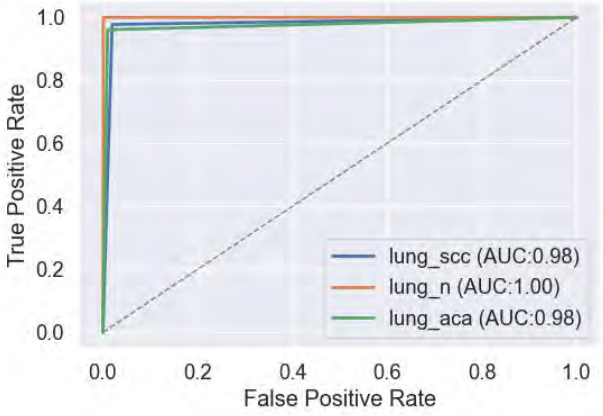


Figure 4.36: ROC Curve and AUC Score of the ResInvolution Architecture with One Involution Layer

4.3.3.3 ResInvolution Architecture with Two Involution Layers

Based on the confusion matrix in Figure 4.37, it is evident that incorporating a second involution layer further enhances the model's performance in comparison to the previous two variants.

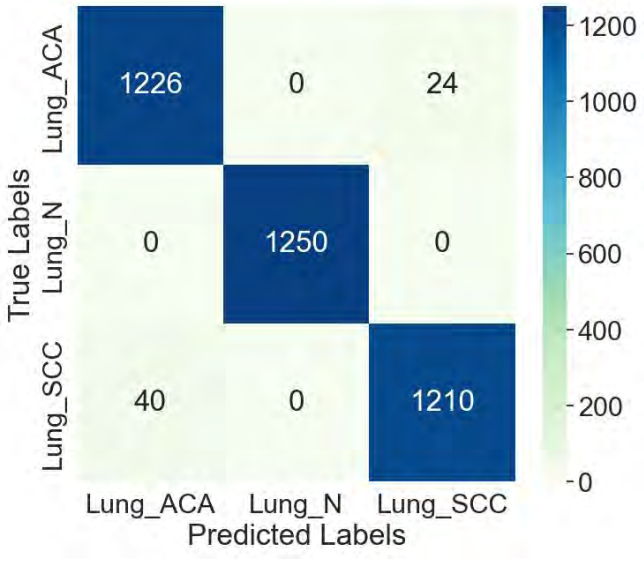


Figure 4.37: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Two Involution Layers

The number of true positives for Lung_ACA has increased to 1226, while the number of false positives for Lung_SCC misclassified as Lung_ACA has decreased to 24.

This matrix demonstrates a noticeable pattern where the inclusion of more involution layers leads to a decrease in classification errors, specifically in distinguishing between Lung_ACA and Lung_SCC. This indicates a gradual enhancement in the accuracy of the model. Including a second involution layer significantly improves the model's performance when compared to the variants with only one involution layer or no involution layers.

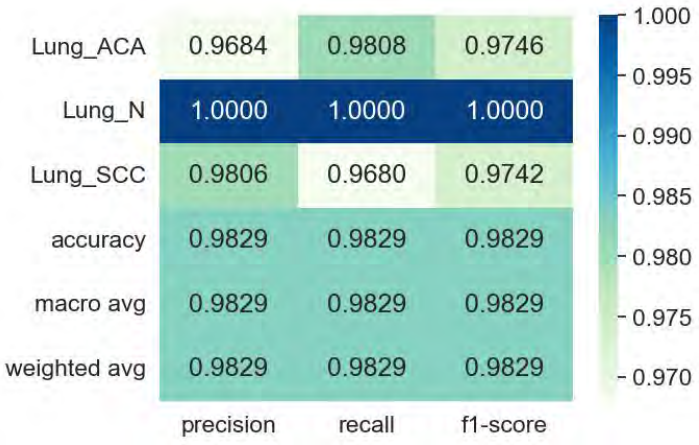


Figure 4.38: Classification Report of the ResInvolution Architecture with Two Involution Layers

As shown in Figure 4.38, There is a slight increase in precision for Lung_ACA and Lung_SCC, and the F1-scores for these classes also improve, suggesting a better balance between precision and recall. The accuracy metrics achieve a remarkable 0.9829, and the macro and weighted averages also reach this high level of 0.9829, demonstrating the effectiveness of incorporating two involution layers to enhance the model's predictive accuracy.

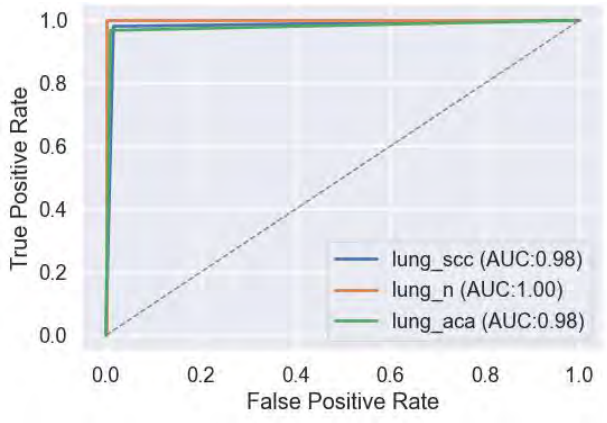


Figure 4.39: ROC Curve and AUC Score of the ResInvolution Architecture with Two Involution Layers

Based on the information provided in Figure 4.39, With the addition of a second involution layer, the AUC scores for Lung_ACA and Lung_SCC remain consistently high at 0.98, just like the one involution variant. It seems that using two layers

doesn't really improve the discriminative performance much compared to just using a single involution layer.

4.3.3.4 ResInvolution Architecture with Three Involution Layers

The inclusion of an additional involution layer significantly improves the model's performance, as illustrated in Figure 4.40. Compared to previous variants, it has the highest number of correct positive results for Lung_ACA (1227) and Lung_SCC (1226), with the lowest number of incorrect positive results. This enhancement highlights the efficacy of incorporating three involution layers to enhance the model's ability to accurately classify different types of lung cancer. This optimization improves both sensitivity and specificity to nearly optimal levels.

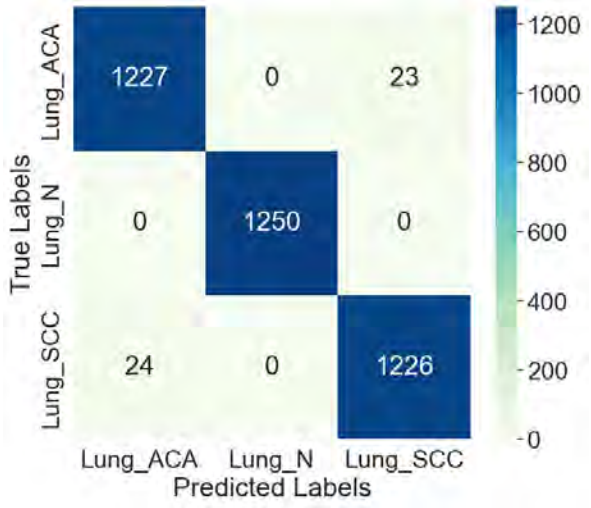


Figure 4.40: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Three Involution Layers

The classification report in Figure 4.41 showcases the exceptional performance of the ResInvolution architecture with three involution layers.

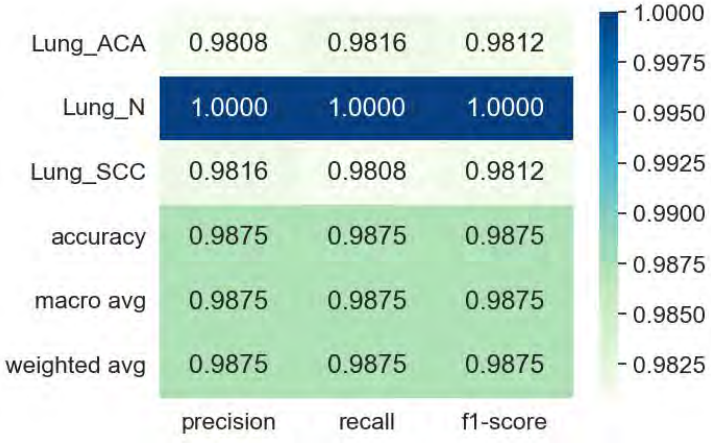


Figure 4.41: Classification Report of the ResInvolution Architecture with Three Involution Layers

It outperforms all other variants, achieving the highest precision and F1-scores across all classes. This version achieves exceptional metrics with macro and weighted averages both at 0.9875. This suggests that incorporating three involution layers enhances the model's ability to accurately differentiate between classes without compromising precision or recall.

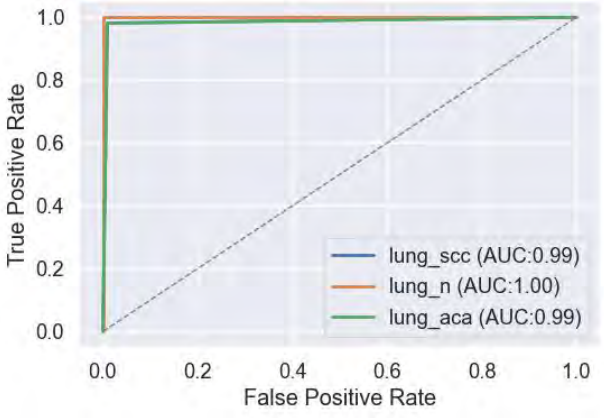


Figure 4.42: ROC Curve and AUC Score of the ResInvolution Architecture with Three Involution Layers

Figure 4.42 shows that the inclusion of three involution layers leads to a small but noticeable enhancement in the AUC scores for Lung_ACA and Lung_SCC, with both reaching an impressive 0.99. It is evident that incorporating a third layer significantly improves the model's ability to accurately distinguish between these conditions, surpassing previous versions and establishing itself as the top performer in terms of sensitivity and specificity. This configuration demonstrates the ideal equilibrium between intricacy and improvement in performance.

4.3.3.5 ResInvolution Architecture with Four Involution Layers

It is evident from Figure 4.43 that the trend goes through a slight reversal when four involution layers are implemented. True positives for Lung_ACA go up marginally to 1236, whereas false negatives for Lung_SCC go up to 51 and true positives for Lung_SCC go down to 1199. These differences suggest that adding too many layers may cause diminishing returns. This matrix indicates, in contrast to the three-layer arrangement, that although more complexity can improve classification up to a point, it might eventually result in a decrease in overall performance efficiency.

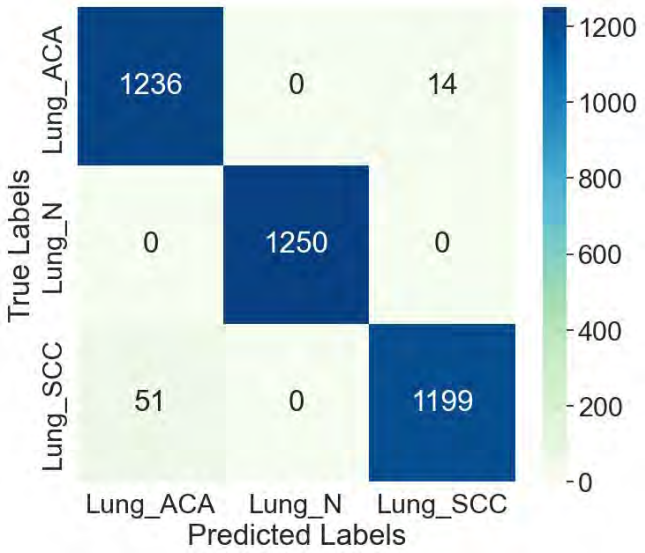


Figure 4.43: Confusion Matrix Generated based on the prediction of ResInvolution Architecture with Four Involution Layers

Based on Figure 4.44, it appears that the addition of four involution layers does not maintain the same level of improvement seen in the three-layer version. Although the precision and F1-score for Lung_SCC remain relatively high, there is a slight decrease observed when comparing it to the three-layer variant. There may be a possibility of reaching a point of diminishing returns when adding more complexity beyond three layers. The macro and weighted averages have decreased to 0.9827 from the highest performance achieved in the three-layer configuration.

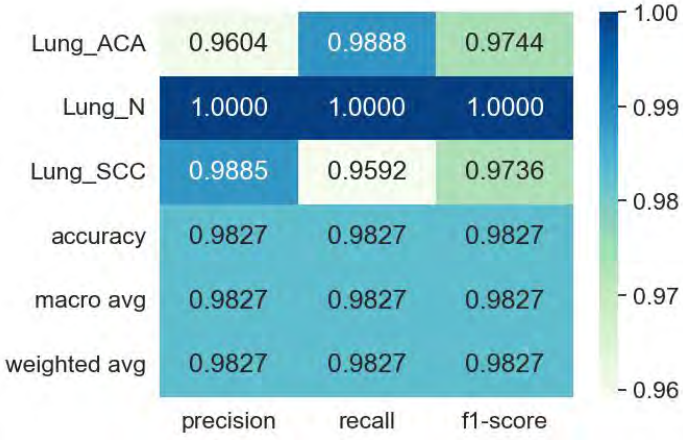


Figure 4.44: Classification Report of the ResInvolution Architecture with Four Involution Layers

Based on Figure 4.45, the addition of four involution layers does not lead to any further improvement in the AUC scores. Both Lung_ACA and Lung_SCC decrease slightly to 0.98, which is consistent with the scores observed with one and two involution variants. It appears that there may be a case of diminishing returns, where adding more layers does not lead to better performance and instead results in a slightly lower effectiveness compared to the three-layer setup.

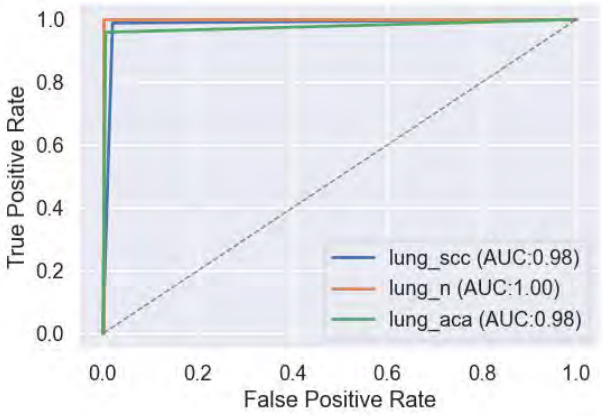


Figure 4.45: ROC Curve and AUC Score of the ResInvolution Architecture with Four Involution Layers

4.3.4 Comparison Analysis

The comprehensive analysis in Table 4.1 provides a consolidated view of the ResInvolution model's variants across three datasets: NCT-CRC-HE-100K, GasHisSDB, and LC25000 (Lung Cancer).

For the NCT-CRC-HE-100K dataset, the variant with three involution layers (Three Inv.) consistently outperforms other versions. It achieves an accuracy of 96.93%, precision of 96.80%, recall of 97.01%, and an F1-score of 96.90%, with an AUC score of 0.98. This variant demonstrates the highest performance metrics across all the models tested on this dataset, suggesting that three involution layers provide the optimal balance between complexity and predictive accuracy. On the GasHisSDB dataset, the three involution layers variant again shows superior performance with an accuracy of 91.57%, precision of 91.48%, recall of 90.93%, and an F1-score of 91.18%, along with an AUC score of 0.91. These results indicate that the model's ability to accurately identify relevant features and make precise predictions is maximized with three involution layers. For the LC25000 dataset, the variant with three involution layers achieves the highest accuracy of 98.75%, precision of 98.75%, recall of 98.75%, and an F1-score of 98.75%, with an AUC score of 0.99. This model version demonstrates outstanding performance, indicating that it effectively distinguishes between different classes and maintains high predictive accuracy with the optimal number of involution layers.

The analysis across all three datasets consistently highlights the variant with three involution layers as the best-performing model. This version balances complexity and performance, achieving the highest accuracy, precision, recall, and F1-score across different datasets. While the four involution layers variant also performs well, it shows slightly diminishing returns compared to the three involution layers variant. This pattern suggests that incorporating three involution layers in the ResInvolution model architecture maximizes its effectiveness, providing a robust and efficient solution for accurate predictions and feature identification across various datasets. This balance is important for practical applications where computational efficiency is as critical as predictive performance.

Table 4.1: Comparison of the variants of Proposed Model

Dataset	Variants of the Architecture	Trainable Parameter	Accuracy	Precision	Recall	F1-score	AUC Score
NCT-CRC-HE-100K	No Involution	187,785	90.61%	93.55%	90.27%	90.99%	0.94
	One Involution	187,870	94.34%	94.74%	94.45%	94.33%	0.96
	Two Involution	187,955	95.90%	96.36%	95.52%	95.85%	0.97
	Three Involution	188,040	96.93%	96.80%	97.01%	96.90%	0.98
	Four Involution	188,128	96.31%	96.33%	96.37%	96.31%	0.97
GasHisSDB (80 x 80 subset)	No Involution	185,986	88.15%	87.84%	87.45%	87.63%	0.87
	One Involution	186,071	87.52%	87.09%	88.41%	87.32%	0.88
	Two Involution	186,156	87.12%	89.42%	84.61%	85.90%	0.85
	Three Involution	186,241	91.57%	91.48%	90.93%	91.18%	0.91
	Four Involution	186,326	90.46%	89.87%	90.81%	90.22%	0.91
LC25000 (Lung Cancer)	No Involution	186,243	97.60%	97.61%	97.62%	97.63%	0.98
	One Involution	186,328	97.95%	97.95%	97.95%	97.95%	0.98
	Two Involution	186,413	98.29%	98.29%	98.29%	98.29%	0.98
	Three Involution	186,498	98.75%	98.75%	98.75%	98.75%	0.99
	Four Involution	186,583	98.27%	98.27%	98.27%	98.27%	0.98

The comparison between the proposed ResInvolution model and the existing architectures mentioned in Table 4.2, reveals significant differences in terms of trainable parameters and performance metrics across three different datasets: GasHisSDB, NCT-CRC-HE-100K, and LC25000 (Lung Classes).

In this study, we have reconstructed the proposed models with the improved accuracy mentioned in some research articles. The DCNN model was developed to classify colon cancer using the colon_n and colon_a classes of LC25000 [37]. In a related study, researchers have suggested using this model to classify skin cancer by utilizing the HAM10000 dataset [38]. In addition, the VGG16 and ResNet50 have demonstrated superior performance compared to other pre-trained CNN models on the GasHisSDB dataset [39]. The authors have conducted a thorough analysis of various pre-trained CNN architectures using three distinct sets of the GasHisSDB dataset. During the reconstruction process, we strictly adhered to the model specifications outlined in the articles. Following the successful construction, we have utilized the identical hyperparameter settings employed in the ResInvolution architecture to train and evaluate the models for comparison.

The DCNN model showcases impressive performance, with its trainable parameters ranging from 671,682 to 675,273. When evaluating the GasHisSDB dataset, the DCNN model achieves impressive results. It achieves an accuracy of 88.90%, precision of 88.39%, recall of 88.63%, and an F1-score of 88.50%. The metrics show a significant improvement on the NCT-CRC-HE-100K dataset, achieving an accuracy of 96.85%, precision of 96.77%, recall of 96.84%, and an F1-score of 96.80%. The LC25000 dataset achieves the highest performance, with impressive accuracy, precision, recall, and F1-score, all at 98.00%. In comparison, ResInvolution outperforms DCNN across all datasets while maintaining a significantly lower number of trainable parameters, emphasizing its efficiency and effectiveness.

The VGG16 and ResNet50 model, characterized by its high complexity with over 14 million trainable parameters, exhibits excellent performance metrics. On the GasHisSDB dataset, VGG16 achieves an accuracy of 93.33%. The performance remains high on the NCT-CRC-HE-100K dataset, with an accuracy of 98.01%. The model performs best on the LC25000 dataset, with an accuracy of 99.68%. Similarly, ResNet50 achieves an accuracy of 91.72% on the GasHisSDB dataset. The performance remains high on the NCT-CRC-HE-100K dataset, with an accuracy of 97.24%. The model performs best on the LC25000 dataset, with an accuracy of 99.23%. In contrast, the ResInvolution model achieves competitive results with significantly fewer parameters. On the GasHisSDB dataset, ResInvolution records an accuracy of 91.57%. For the NCT-CRC-HE-100K dataset, it achieves an accuracy of 96.93%. On the LC25000 dataset, the accuracy is 98.75%. While VGG16 and ResNet50 consistently outperform ResInvolution in terms of raw accuracy, the difference is marginal in many cases. Given the vast reduction in parameter count with ResInvolution, the trade-off between complexity and performance makes ResInvolution an attractive alternative, especially in resource-constrained environments.

Table 4.2: Performance Comparison of Proposed ResInvolution with Existing Architectures

Model	Dataset	Trainable Parameter	Accuracy	Precision	Recall	F1-score
DCNN [37] [38]	GasHisSDB	671,682	88.90%	88.39%	88.63%	88.50%
	NCT-CRC-HE-100K	675,273	96.85%	96.77%	96.84%	96.80%
	LC25000 (Lung Classes)	672,195	98.00%	98.02%	98.00%	98.00%
VGG16 [39]	GasHisSDB	14,731,074	93.33%	93.23%	93.06%	93.14%
	NCT-CRC-HE-100K	14,733,129	98.01%	98.00%	98.00%	98.00%
	LC25000 (Lung Classes)	14,720,835	99.68%	99.68%	99.68%	99.68%
ResNet50 [39]	GasHisSDB	23,571,458	91.72%	91.51%	91.54%	91.52%
	NCT-CRC-HE-100K	23,700,489	97.24%	97.26%	97.24%	97.24%
	LC25000 (Lung Classes)	23,571,458	99.23%	99.23%	99.23%	99.23%
ResInvolution Three Involution (Proposed)	GasHisSDB	186,241	91.57%	91.48%	90.93%	91.18%
	NCT-CRC-HE-100K	188,040	96.93%	96.80%	97.01%	96.90%
	LC25000 (Lung Classes)	186,498	98.75%	98.75%	98.75%	98.75%

The ResInvolution model, with its minimal parameter count, strikes an impressive balance between efficiency and performance. Although both DCNN and VGG16 have impressive accuracy, ResInvolution proves that exceptional performance can be attained with a much lighter model. This makes it a great option for practical applications with limited computational resources. In a Federated Learning Environment, distributed learning can be seamlessly performed and clients can communicate with each other with minimal buffer. In addition, clients with limited computing power can also seamlessly contribute. We can see the difference in trainable parameters from the line graph illustrated in Figure 4.46.

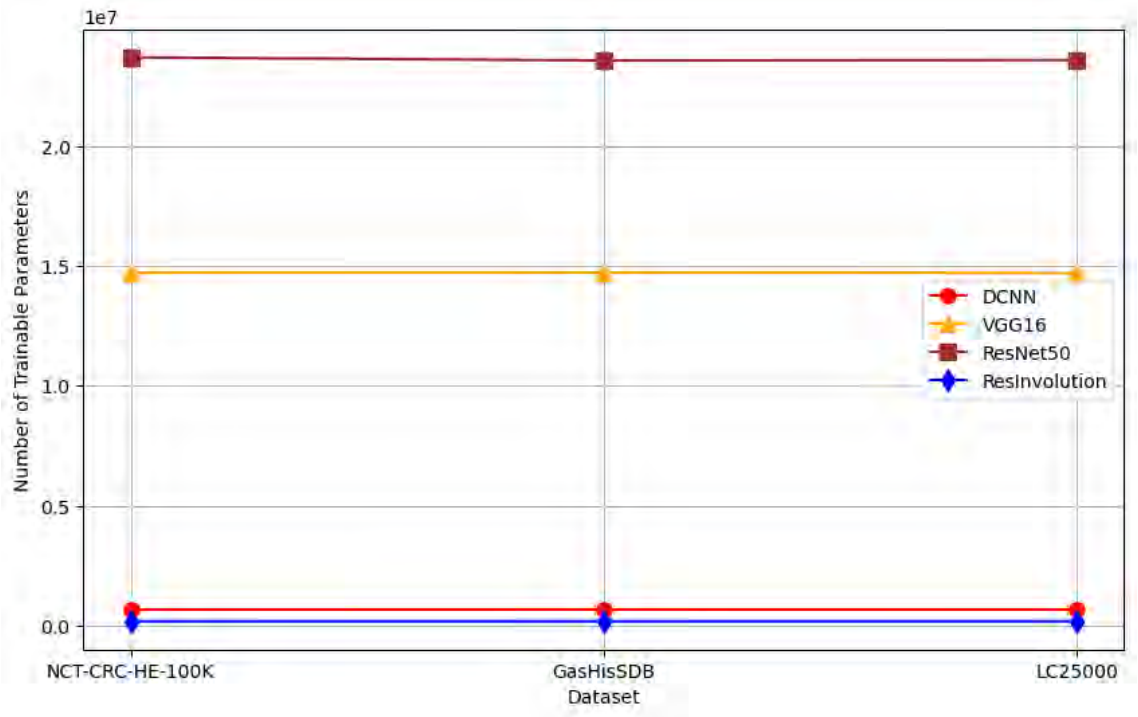


Figure 4.46: Comparison of Trainable Parameters for Different Model Across Datasets

4.4 Performance Analysis of the proposed ResInvolution Architecture in the Federated Learning Environemnt

For this experiment, the training data was evenly distributed among ten clients in the federated learning framework, as per the specific class distribution that was given earlier. Every client was provided with data from all classes, distributed evenly based on their respective quantities in the dataset. The distribution ensured that every client had an equally distributed sample of each class, allowing for a thorough and unbiased evaluation across the network. The ResInvolution architecture's performance was assessed by tracking the validation accuracy of the global model after each communication round, covering a total of fifty epochs. This method enabled the determination of the point in time when the overall model reached its peak

validation accuracy, offering a definitive measure of the model's performance in a distributed learning setting.

4.4.1 Performance Evaluation on the NCT-CRC-HE-100K Dataset

In this experiment, the training data has been evenly distributed among the ten clients. Every client has been given ADI, BACK, DEB, LYM, MUC, MUS, NORM, STR, and TUM classes, with 780, 792, 863, 866, 667, 1015, 657, 783, and 1073 images from each class respectively. We analyzed the communication round where the global model achieved its highest validation accuracy after fifty communication rounds to evaluate its performance.

Figure 4.47 depicts the learning curve of the global model. The graph showcases a highly encouraging and efficient training progression. At the start, the accuracy rises significantly, rapidly ascending to its peak within the initial epochs. The rapid progress indicates that the model can swiftly adjust and acquire the necessary generalized features from diverse datasets. This demonstrates the model's ability to effectively handle the distinct data distribution among clients in the Federated Learning environment. After the initial increase, the accuracy plateaus and remains consistently high for the rest of the training period. This suggests that the model has reached stable and optimal performance early in the training process. Conversely, the loss curve begins with a steep decline, reaching almost zero levels early on in the training, and stays consistently low throughout the training period. The decrease and stabilization of the loss enhance the effectiveness of the model's learning process, ensuring rapid error minimization and maintaining minimal fluctuation. The quick convergence of both metrics' accuracy at a high level and loss at a low level suggests that the model doesn't need prolonged training beyond the initial epochs to reach and sustain its optimal performance. This pattern is especially beneficial in a Federated Learning context, where achieving efficient convergence is vital because of the computational and communication overhead involved in training across multiple decentralized nodes.

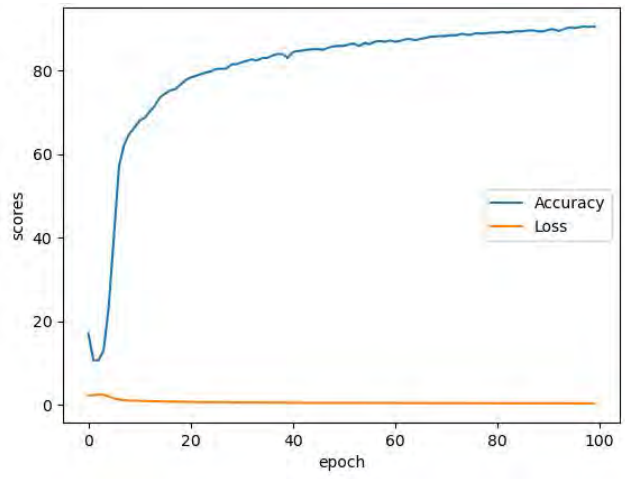


Figure 4.47: Learning Curves of the Global Model

In the Federated Learning environment, the confusion matrix shown in Figure 4.48 provides valuable insights into the performance of the ResInvolution architecture across various classes. The matrix displays impressive diagonal values, which suggest a significant number of accurate predictions for each class. It is worth noting that classes such as ADI and BACK have shown impressive accuracy. ADI achieved 2522 correct predictions out of its total samples, while BACK achieved 2010 correct predictions. However, there is a clear issue with misclassification in certain classes, such as DEB and STR, where the number of false positives and false negatives is particularly noticeable. It appears that the model is able to accurately identify the majority of classes, but it encounters difficulties with certain specific ones. This could be attributed to differences in data distribution among various clients in the federated setup. These differences can result in inconsistencies in how the model learns features that are specific to classes that are predicted with less accuracy. The overall matrix demonstrates impressive accuracy in classifying labels, but it also emphasizes the difficulties encountered in federated learning. In this approach, data is not evenly distributed among clients, which can result in suboptimal learning outcomes for specific classes. This analysis is crucial as it highlights the need to focus on possible areas for improvement in training strategies or data allocation in federated scenarios.

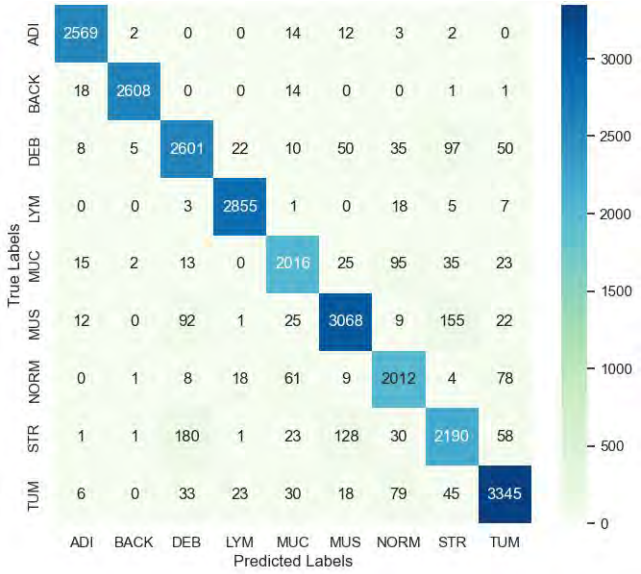


Figure 4.48: Confusion Matrix Generated based on the prediction of the proposed ResInvolution architecture in the Federated Learning environemnt

The ResInvolution architecture's classification report in the Federated Learning environment, shown in Figure 4.49, gives a full breakdown of its accuracy, recall, and F1-scores for each class. Despite the slightly lower performance metrics for each class compared to a traditional centralized learning environment, the results are still impressive, with most values above 0.88.

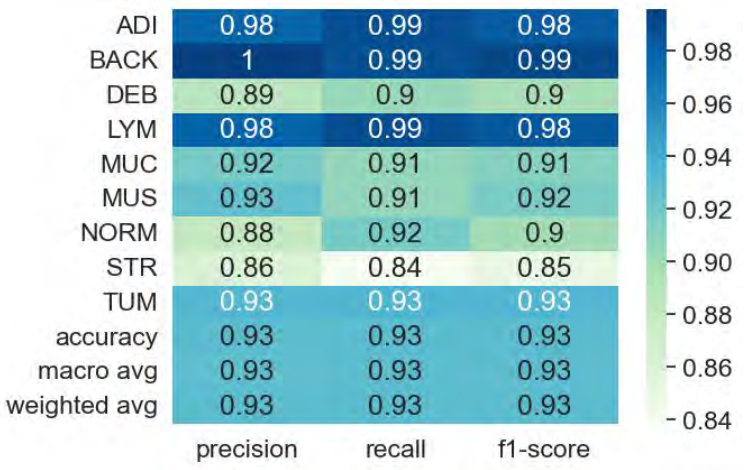


Figure 4.49: Claasification Report of the proposed ResInvolution architecture in the Federated Learning environemnt

It is worth mentioning that classes like MUC and MUS exhibit exceptional precision and recall, which demonstrates the model's remarkable ability to identify these classes even in distributed training conditions. Nevertheless, STR and TUM classes exhibit slightly lower scores, which indicates the difficulties in distinguishing between classes in federated learning settings. The average scores, both macro and weighted, are approximately 0.93. This indicates a strong overall performance but also highlights the importance of optimizing the handling of class imbalances or feature distribution among different clients. This report highlights the trade-offs involved in federated learning. While it offers privacy and data locality, it also presents challenges in maintaining consistent model performance across different classes.

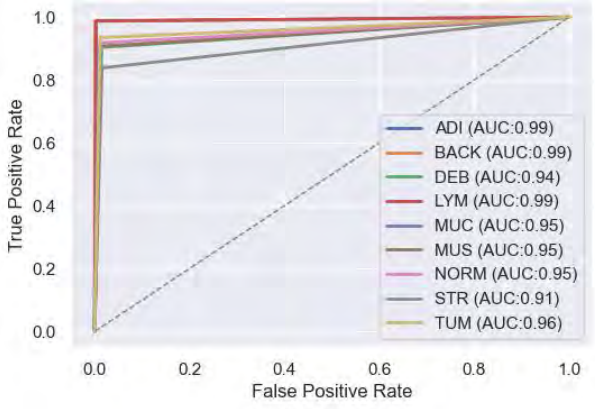


Figure 4.50: ROC Curve and AUC Score of the proposed ResInvolution Architecture in the Federated Learning environment

The ROC curve and AUC scores in the Federated Learning environment for the ResInvolution architecture provided in Figure 4.50 show how well it can tell the difference between different classes. Most classes have remarkably high AUC scores, with values like 0.99 for BACK and MUC. These scores indicate exceptional model performance in effectively distinguishing between the positive and negative classes within these categories. Nevertheless, certain classes, such as STR and TUM, exhibit

slightly lower AUC values, indicating potential areas for improving the model's performance. The model's decreased capacity to effectively generalize across distributed data segments, particularly when data relevant to these classes is scarce or unevenly distributed among different clients, could be the cause of these lower scores. The ROC curves demonstrate the overall strong performance of the model, but they also reveal the variations in sensitivity and specificity across different classes. This indicates the difficulties in training a highly accurate model when working with a decentralized dataset. Understanding the limitations and areas for improvement is crucial when deploying deep learning models in a federated environment. It particularly focuses on enhancing data consistency and model robustness.

4.4.2 Performance Evaluation on GasHisSDB Dataset

In this experiment, the entire set of training data has been equitably disseminated among the ten clients. Each individual client has been provided with normal and abnormal classes, with 4,682 and 9,548 images allocated to them, accordingly. The weight of the communication round in which the global model achieved the highest validation accuracy after fifty communication rounds was determined to assess the performance.

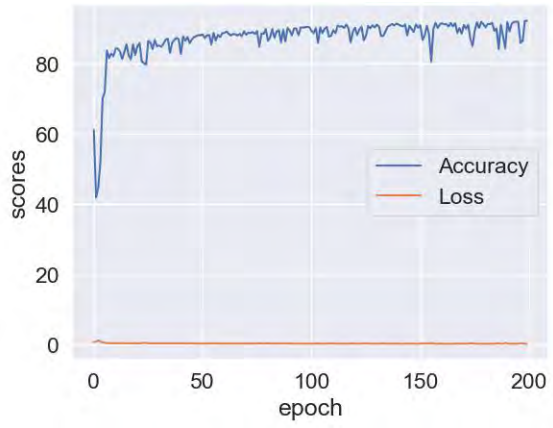


Figure 4.51: Learning Curves of the Global Model

The graph in 4.51 illustrates the learning curve, showing how the global model rapidly stabilizes its accuracy after an initial drop in the early epochs. The rapid stabilization of the model effectively grasps the crucial features of the distributed datasets. The accuracy remains consistently high, with only minor fluctuations near the peak throughout the training process, indicating strong and reliable performance. Meanwhile, the loss curve demonstrates a rapid decrease to almost zero at the beginning of the training. It remains consistently low, suggesting the model successfully reduces error as the training progresses. The pattern of both curves indicates that the model achieves optimal performance early on and continues to maintain it, without any notable improvements from further training epochs.

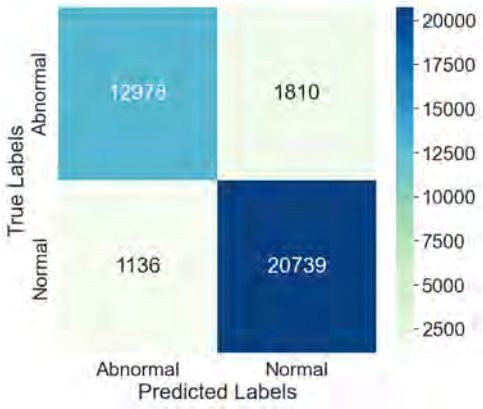


Figure 4.52: Confusion Matrix Generated based on the prediction of the proposed ResInvolution architecture in the Federated Learning environment

The confusion matrix shown in Figure 4.52 demonstrates the effectiveness of a model in differentiating between Normal and Abnormal conditions. Based on the matrix, the model accurately detected 19,466 Abnormal cases as true positives and 13,699 Normal cases as true negatives. The model demonstrates a strong capability to accurately classify both conditions. Unfortunately, there were 1,089 instances where Normal cases were mistakenly identified as Abnormal, and 2,409 cases where Abnormal cases were overlooked and labeled as Normal. A marginal decline in the distributed system's performance can be observed in predicting abnormal classes, while the model's prediction for the normal class marginally improved as the dataset contained more images in the normal class.

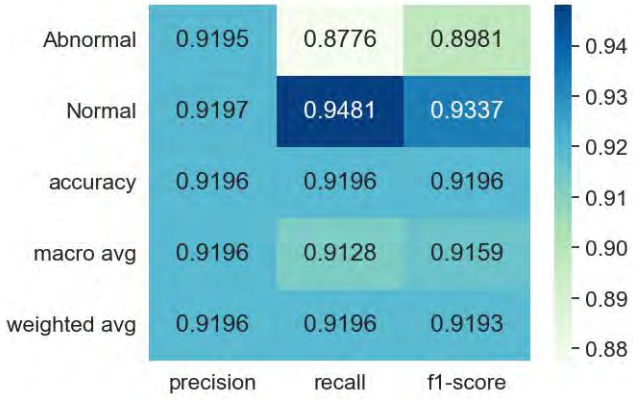


Figure 4.53: Claasification Report of the proposed ResInvolution architecture in the Federated Learning environemnt

The classification report depicted in Figure 4.53 demonstrates a consistent and high-performing model with precision, recall, and F1-scores all aligning closely across two distinct categories: 'Normal' and 'Abnormal'. For the 'Normal' category, all three metrics stand at 0.88, indicating that the model is quite adept but slightly less effective in identifying and classifying normal conditions compared to abnormal ones. In contrast, the 'Abnormal' category shows stronger metrics, with precision, recall, and F1-score all at 0.92, reflecting the model's enhanced capability to accurately

identify and confirm abnormal conditions without much error. The overall accuracy of the model is 0.90, which is substantiated by equally high macro and weighted averages of 0.90 across all metrics. This high level of accuracy and uniformity across the metrics suggests that the model is well-calibrated and performs robustly in distinguishing between normal and abnormal states, making it reliable for practical applications where distinguishing such conditions is critical.

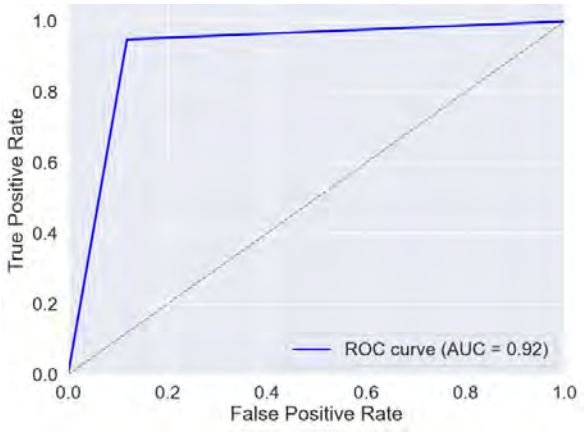


Figure 4.54: ROC Curve and AUC Score of the proposed ResInvolution Architecture in the Federated Learning environment

Figure 4.54 presents the ROC curve and AUC score of the proposed ResInvolution architecture in a Federated Learning environment, achieving an AUC of 0.92. This indicates a strong ability to differentiate between classes, demonstrating high overall performance. The ROC curve and AUC score of the ResInvolution architecture with three involution layers achieve an AUC of 0.91 in a traditional learning environment. While both models exhibit excellent performance, the slightly higher AUC in the Federated Learning environment suggests that the proposed architecture benefits from the collaborative learning approach, enhancing its classification capabilities. The comparison highlights that Federated Learning can provide a marginal improvement in the model's discriminative power, making it a valuable approach for distributed data scenarios.

4.4.3 Performance Evaluation on the Lung Classes of LC25000 Dataset

In this experiment, the training data has been evenly distributed among the ten clients. Every client has been given lung_aca, lung_n, and lung_scc classes, with 375 images from each class assigned to them. We analyzed the communication round where the global model achieved its highest validation accuracy after fifty communication round to evaluate its performance.

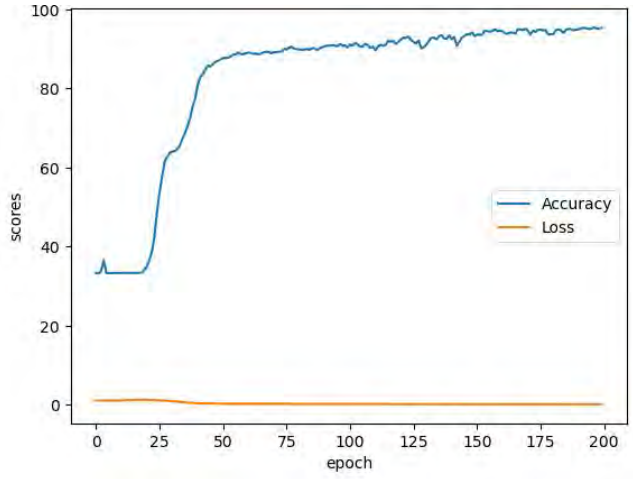


Figure 4.55: Learning Curves of the Global Model

The learning curves shown in Figure 4.55 demonstrate the progress of the global model over 200 epochs, with a specific focus on accuracy and loss. The model rapidly achieves a high level of accuracy, nearing 90%, early on in the training process. This suggests that it effectively captures the dominant features present in the dataset. The accuracy of the model reaches a stable state, indicating that it effectively generalizes the training data without requiring additional complexity or adjustments. Concurrently, the loss metric demonstrates a substantial decrease, reaching a point close to zero. This indicates that the model has effectively reduced errors between the predicted outputs and the actual labels. The consistent and minimal loss during the training suggests that the model has reached its optimal convergence. The consistent accuracy and minimal loss observed during training suggest that the model has been finely tuned to meet the desired performance standards, without the need for further improvements in accuracy.

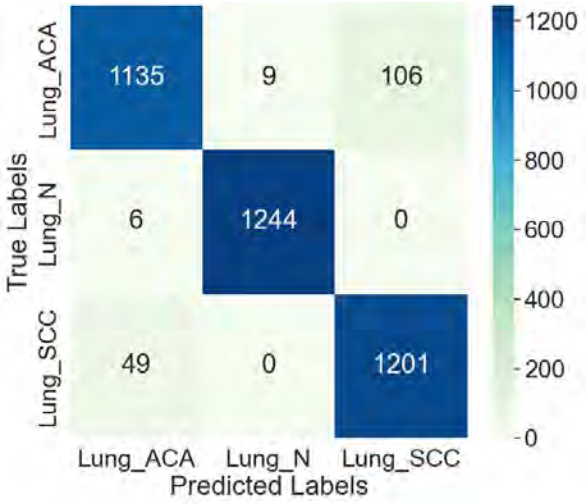


Figure 4.56: Confusion Matrix Generated based on the prediction of the proposed ResInvolution architecture in the Federated Learning environment

The confusion matrix in Figure 4.56 demonstrates a competent performance, per-

haps with noticeable reductions when compared to the Traditional Learning environment. The presence of 106 false positives and 9 false negatives in Lung_ACA suggests that there may be issues with the specificity and sensitivity of the data. These issues are likely due to the variability and non-uniformity of the distributed data. Similarly, Lung_SCC exhibits 49 false positives, indicating potential challenges in accurately differentiating this class in federated conditions. The variation in results can be explained by the uneven distribution of data among different clients, which may not accurately capture the differences between classes as well as centralized datasets. .

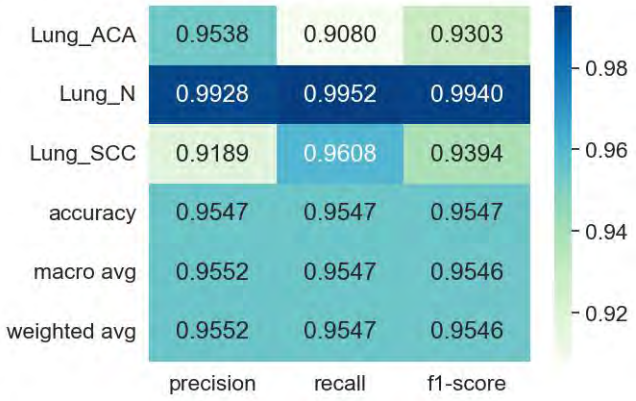


Figure 4.57: Claasification Report of the proposed ResInvolution architecture in the Federated Learning environemnt

The classification report in Figure 4.57 shows that there have been slight decreases in precision and recall for Lung_ACA and Lung_SCC. The precision values for Lung_ACA and Lung_SCC are 0.9538 and 0.9189, respectively. These reductions can arise from difficulties in model training when each node in a federated network has different data characteristics, resulting in less effective learning generalizations compared to a cohesive, centralized dataset. The overall accuracy and F1-scores experience a slight decrease, potentially due to the decreased frequency of synchronization between the federated nodes compared to continuous learning updates in traditional environments.

In the Federated Learning environment, the AUC scores for Lung_ACA and Lung_SCC from Figure 4.58 are slightly lower (0.94 and 0.96) compared to Traditional Learning. This drop indicates a decrease in the model's ability to differentiate between classes, possibly because of decentralized learning and limited aggregation of learning updates across distributed environments. The ROC curves show less steepness for these classes, indicating a compromised discriminative performance. This could be due to the presence of noise and outliers in the distributed datasets.

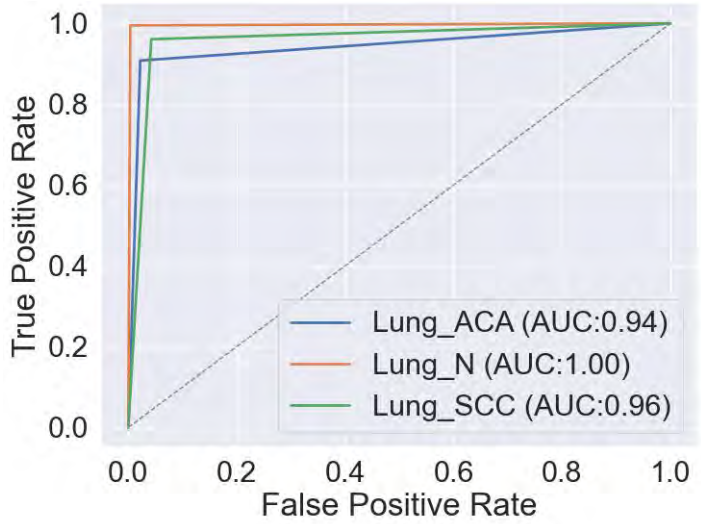


Figure 4.58: ROC Curve and AUC Score of the proposed ResInvolution Architecture in the Federated Learning environment

4.5 Discussion

The proposed model combines involution features with residual features, in an attempt to combine different types of features extracted by the given models so that better classification results can be achieved. In the process, residual blocks have been utilized alongside different combinations of involution layers to experiment with the performance of them. In addition, a federated learning framework has been proposed with the best-performing model and the performance has been compared against the centrally learned models.

In the result analysis, we see a consistent pattern of result related to the number of involution layers. The model with no involution layer usually performs the worst with the only exception of the GasHisSDB Dataset. Adding more involution layers gradually improves the performance until the third layer. More layers in a deep learning architecture enable the extraction of lower-level features, which is necessary for samples that have more intricate features. The detailed patterns that can be seen in the histopathological images indicate the necessity of accessing such lower-level information, and having more involution layers enables us to achieve that. However, increasing the number of involution layers to four ultimately stagnates the improvement of the result with reduced accuracy due to loss of information in the feature maps. Consequently, our observation is that the combination of three involution layers is a perfect setup for our proposed model using the given dataset. However, the setup might need to be changed with different structures in the rest of the architecture or with different datasets. In addition, in the comparsion study with some existing architectures the ResInvolution model with three involution layers, with its minimal parameter count, strikes an impressive balance between efficiency and performance. This architecture showcases the ability to achieve high performance with a much lighter model, making it a great option for practical applications with limited computational resources.

Due to the effectiveness of the feature extractors, the classifier portion of the model is designed to be very basic, consisting of only the output layer with a number of neurons corresponding to the number of classes. Hence, the model remains lightweight and is a perfect material for federated learning environment. In the federated test, the models perform rather competitively compared to the centrally learned models, sometimes even scoring better compared to them.

Based on this analysis, we have implemented our proposed three involution-based ResInvolution architectures in a federated learning environment. We then thoroughly analyzed the performance of these architectures on three selected datasets.

Chapter 5

Conclusion

In this research, we present a groundbreaking innovation in histopathological image classification, ResInvolution, an Involution-Based ResNet Architecture. The breakthrough involved is critical in the field. The breakthrough was the integration of involution layers into the ResNet architecture to enhance the adaptable feature extraction of the model without affecting its computational efficiency. The proposal drastically reduces the dimensionality of the trainable parameters in the model, thereby creating a lightweight model that is easy to deploy in context.

Comprehensive experiments were conducted on three discrete histopathological image datasets: LC25000, GasHisSDB, and NCT-CRC-HE-100K. In both traditional learning and federated learning environment, the results demonstrate that the ResInvolution model exhibits a remarkable level of performance. The GasHisSDB, NCT-CRC-HE-100K, and LC25000 datasets yielded high accuracy rates of 91.57%, 96.93%, and 98.75%, respectively, when using the ResInvolution model in the traditional learning mode. The same metrics declined by a negligible amount when in a federated learning mode. Overall, the ResInvolution model outperforms existing architectures in terms of efficiency and effectiveness. Given the relatively low number of trainable parameters, the ResInvolution model outperformed and thus compared to other state-of-the-art architectures. This confirms the advantages provided by the involution layers, which develop the convolutional kernels according to the input features. Thus, the involution can help the model focus on the most informative parts of the image meaningfully. Moreover, the ResInvolution model was successfully integrated into the federated environment, proving its contribution to privacy in the medical image analysis field. The federated learning approach implies the decentralized training of the ML model across various organizations. In the healthcare domain, the model's privacy mechanisms are discussed in terms of avoiding data sharing. Despite a relatively simple structure, the model demonstrated high accuracy and performance indicators in the described setting, proving its viability for real-world use in tasks that require.

Overall, this research makes significant contributions in developing a unique and effective architecture for classifying histopathological images and its successful implementation in a federated learning setting. These advancements offer significant potential for improving medical diagnostics and pave the way for further exploration of lightweight, adaptive models in various domains. Future research could

investigate potential improvements to the ResInvolution architecture and its potential for use in various medical imaging tasks, further establishing its significance in advancing healthcare technologies.

Bibliography

[1] J. G. Elmore, G. M. Longton, P. A. Carney, *et al.*, "Diagnostic concordance among pathologists interpreting breast biopsy specimens," *Jama*, vol. 313, no. 11, pp. 1122–1132, 2015.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[3] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of pathology informatics*, vol. 7, no. 1, p. 29, 2016.

[4] A. Esteva, B. Kuprel, R. A. Novoa, *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[5] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint arXiv:1605.07678*, 2016.

[6] A. Blasimme and E. Vayena, "Biomedical big data: New models of control over access, use and governance," *Journal of Bioethical Inquiry*, vol. 14, no. 4, 2017.

[7] D. Li, J. Hu, C. Wang, *et al.*, "Involution: Inverting the inherence of convolution for visual recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 321–12 330.

[8] J. Lee, S. Narang, J. J. Martinez, G. Rao, and A. U. K. Rao, "Associating spatial diversity features of radiologically defined tumor habitats with epidermal growth factor receptor driver status and 12-month survival in glioblastoma: methods and preliminary investigation," *Journal of Medical Imaging*, vol. 2, no. 4, p. 041 006, 2015. DOI: 10.1117/1.JMI.2.4.041006. [Online]. Available: https://doi.org/10.1117/1.JMI.2.4.041006.

[9] S. Leng, L. Yu, L. Chen, J. C. R. Giraldo, and C. H. McCollough, "Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain," in *Medical Imaging 2012: Physics of Medical Imaging*, N. J. Pelc, R. M. Nishikawa, and B. R. Whiting, Eds., International Society for Optics and Photonics, vol. 8313, SPIE, 2012, p. 83131M. DOI: 10.1117/12.912126. [Online]. Available: https://doi.org/10.1117/12.912126.

[10] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," en, *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, 2022.

[11] N. Rieke, J. Hancox, W. Li, *et al.*, "The future of digital health with federated learning," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–7, 2020.

[12] A. L. Meirelles, T. Kurc, J. Kong, R. Ferreira, J. H. Saltz, and G. Teodoro, "Building efficient cnn architectures for histopathology images analysis: A case-study in tumor-infiltrating lymphocytes classification," *Frontiers in Medicine*, vol. 9, p. 894 430, 2022.

[13] P. A. Winata and I. Roysida, "Implementation of a faster r-cnn algorithm for identification of metastatic tissue using lymphoma histopathological images," *Journal of Soft Computing Exploration*, vol. 4, no. 2, 2023.

[14] B. R. Devassy and J. K. Antony, "Histopathological image classification using cnn with squeeze and excitation networks based on hybrid squeezing," *Signal, Image and Video Processing*, vol. 17, no. 7, pp. 3613–3621, 2023.

[15] A. De, N. Mishra, and H.-T. Chang, "An approach to the dermatological classification of histopathological skin images using a hybridized cnn-densenet model," *PeerJ Computer Science*, vol. 10, e1884, 2024.

[16] K. Gupta and N. Chawla, "Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained cnn," *Procedia Computer Science*, vol. 167, pp. 878–889, 2020.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] G. Liang and H. Wang, "I-cnet: Leveraging involution and convolution for image classification," *IEEE Access*, vol. 10, pp. 2077–2082, 2021.

[19] S. Vyas, A. N. Patra, and R. M. Shukla, "Histopathological image classification and vulnerability analysis using federated learning," *arXiv preprint arXiv:2310.07380*, 2023.

[20] S. Nazir and M. Kaleem, "Federated learning for medical image analysis with deep neural networks," *Diagnostics*, vol. 13, no. 9, p. 1532, 2023.

[21] F. R. da Silva, R. Camacho, and J. M. R. Tavares, "Federated learning in medical image analysis: A systematic survey," *Electronics*, vol. 13, no. 1, p. 47, 2023.

[22] N. Mouhni, A. Elkalay, M. Chakraoui, A. Abdali, A. Ammoumou, and I. Amalou, "Federated learning for medical imaging: An updated state of the art," *Ing. Syst. D'Inf*, vol. 27, pp. 143–150, 2022.

[23] G. N. Gunesli, M. Bilal, S. E. A. Raza, and N. M. Rajpoot, "Feddropoutavg: Generalizable federated learning for histopathology image classification," *arXiv preprint arXiv:2111.13230*, 2021.

[24] W. Hu, C. Li, X. Li, *et al.*, "Gashissdb: A new gastric histopathology image dataset for computer aided diagnosis of gastric cancer," *Computers in biology and medicine*, vol. 142, p. 105 207, 2022.

[25] J. N. Kather, N. Halama, and A. Marx, "100,000 histological images of human colorectal cancer and healthy tissue," *Zenodo10*, vol. 5281, p. 6, 2018.

[26] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (lc25000)," *arXiv preprint arXiv:1912.12142*, 2019.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, Springer, 2016, pp. 630–645.

[28] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

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

[30] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[31] K. Bonawitz, V. Ivanov, B. Kreuter, *et al.*, "Practical secure aggregation for privacy-preserving machine learning," in *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.

[32] A. Tharwat, "Classification assessment methods," *Applied computing and informatics*, vol. 17, no. 1, pp. 168–192, 2020.

[33] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *European conference on information retrieval*, Springer, 2005, pp. 345–359.

[34] Y. Sasaki *et al.*, "The truth of the f-measure," *Teach tutor mater*, vol. 1, no. 5, pp. 1–5, 2007.

[35] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.

[36] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[37] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, p. 100 036, 2021.

[38] M. I. Hasan, M. S. Ali, M. H. Rahman, M. K. Islam, *et al.*, "Automated detection and characterization of colon cancer with deep convolutional neural networks," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[39] W. Hu, H. Chen, W. Liu, *et al.*, "A comparative study of gastric histopathology sub-size image classification: From linear regression to visual transformer," *Frontiers in Medicine*, vol. 9, p. 1 072 109, 2022.