Breast Cancer Classification from Histopathological Images Using Transfer Learning and Deep Neural Networks

Abdulrahman Aloyayri

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Computer Science at

Concordia University

Montreal, Quebec, Canada

April 2020

© Abdulrahman Aloyayri 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Abdulrahman Aloyayri**

Entitled: **Breast Cancer Classification from Histopathological Images Using Transfer Learning and Deep Neural Networks**

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____
Chair Dr. Thomas Fevens

_____
Examiner Dr. Thomas Fevens

_____
Examiner Dr. Tristan Glatard

_____
Supervisor Dr. Adam Krzyzak

Approved by _____
Chair of Department or Graduate Program Director

_____
Dr. Amir Asif, Dean
Gina Cody School of Engineering and Computer Science

Date _____

Abstract

Early diagnosis of breast cancer is the most reliable and practical approach to managing cancer. Computer-aided detection or computer-aided diagnosis is one of the software technology designed to assist doctors in detecting or diagnose cancer and reduce mortality via using the medical image analysis with less time. Recently, medical image analysis used Convolution Neural Networks to evaluate a vast number of data to detect cancer cells or image classification. In this thesis, we implemented transfer learning from pre-trained deep neural networks ResNet18, Inception-V3Net, and ShuffleNet in terms of binary classification and multiclass classification for breast cancer from histopathological images. We use transfer learning with the fine-tuned network results in much faster and less complicated training than a training network with randomly initialized weights from scratch. Our approach is applied to image-based breast cancer classification using histopathological images from public dataset BreakHis. The highest average accuracy achieved for binary classification of benign or malignant cases was **97.11%** for ResNet 18, followed by **96.78%** for ShuffleNet and **95.65%** for Inception-V3Net. In terms of the multiclass classification of eight cancer classes, the average accuracies for pre-trained networks are as follows. ResNet18 achieved **94.17%**, Inception-V3Net **92.76%** and ShuffleNet **92.27%**.

Acknowledgement

First and foremost, all praises to Allah for blessing, protecting, and guiding me throughout my studies. I could never have accomplished this without my faith.

Second, I would like to express my sincere gratitude to my supervisor Professor Adam Krzyzak for his continuous support, immense knowledge, continuous motivation, and patience. His unique personality as a supervisor and friend is the main reason behind the success of this research. The objectives of the research would not have achieved without the professional and experienced guidance and support of my supervisor.

I also extend my thanks to members of the examining committee, including Dr. Thomas Fevens and Dr. Tristan Glatard, for critically evaluating my thesis and providing me with valuable comments about my research. I would like to extend my thanks Halina Monkiewicz, the program advisor of Computer Science and Software Engineering, for her support, patience and motivation.

Third, I would like to express my sincere gratitude to the financial support from the Government of Saudi Arabia under the scholarship of Saudi Electronic University, which enabled me to undertake my studies, and I was lucky to obtain this opportunity. I am forever thankful to the Saudi Cultural Bureau and the Embassy of Saudi Arabia in Canada for their continued support and motivation during my journey.

A heartfelt deepest gratitude expresses to my first teachers (my dear mother and my father), my brothers, and my sisters for their unconditional love, prayers, and support. Without them, this journey would not have been possible, and to them, I dedicate this milestone.

Last but never least, to my lovely family here in Montreal (my wife Afnan, Ahmed and Faris) for their unwavering love and encouragement during the pursuit of my studies. Thank you for always believing in me and for reminding me to endure during the tough times. I consider myself lucky to have encountered family.

Contributions

The main contributions of this thesis are:

1. Implementation of transfer learning with the following models of pre-trained Deep Neural Networks: ResNet18, Inception-V3Net, and ShuffleNet and their application in breast cancer binary and multiclass image-based classification using BreakHis dataset of histopathological Images.

2. Implementation of the data augmentation based on training data only, to boost classification performance and to resolve the imbalanced class issue.

3. Evaluation of classifiers performance by different performance evaluation metrics such as confusion matrices, accuracy, precision, sensitivity, specificity and F1 score.

Table of Contents

List of Figures	viii
List of Tables	xii
List of Acronyms and Abbreviations	xiv
1 Introduction	1
1.1 Cancer and its Impact	1
1.2 Breast Cancer	2
1.3 Computer-Aided Cancer Detection	3
1.4 Structure of the Thesis	6
2 Literature Survey	7
2.1 Traditional Approaches to Medical Image Classification	7
2.2 Deep Learning Approaches to Medical Image Classification	9
2.3 Pre-trained Deep Neural Networks Approach	16
3 Methodology	22
3.1 Proposed Method	22
3.2 Dataset Description	24
3.2.1 Data Pre-Processing	27
3.2.2 Data Augmentation	28
3.3 Transfer Learning	30
3.4 Training Methodology	31
3.5 Dataset Experimental Protocol	31
3.6 Pre-trained Deep Neural Networks Types	32
3.6.1 ResNet18	32
3.6.2 Inception-V3Net	33
3.6.3 ShuffleNet	33
3.7 Evaluation metrics	33
4 Experiment Results and Discussion	35
4.1 Results for binary classification	35

4.1.1 Confusion Matrices	40
4.1.2 Classification Performance on Testing Data	43
4.2 Results for Multiclass Classification	46
4.2.1 Inception-V3Net	47
4.2.2 ResNet18	51
4.2.3 ShuffleNet	55
4.2.4 Confusion Matrices	59
4.3 Comparison with the State-of-the-Art Result	71
4.3.1 Binary Classification	72
4.3.2 Multiclass Classification	74
5 Conclusions and Future Work	75
5.1 Conclusions	75
5.2 Future Work	76
Bibliography	77

List of Figures

1.1 Common Women Cancers (Extracted from [21]).

1.2 Common Men Cancers (Extracted from [21]).

1.3 The Breast (Extracted from [12]).

1.4 Convolutional Neural Network workflow.

1.5 Architecture of Convolutional Neural Network.

1.6 Pooling Types: max pooling and average pooling.

2.1: Proposed Multi-view Features Fusion (MVFF) based CAD system (Extracted from [39]).

2.2 A typical ConvNet architecture with two feature stages (Extracted from [45]).

2.3 Schematic presentation of a proposal that used (Extracted from [9]).

2.4: Typical CNN architecture with specific stages (Extracted from [53]).

2.5 Architecture of the model used for DCNN (Extended from [1]).

2.6 Overview of the workflow they proposed (Extracted from [28]).

2.7: Hybrid CNN architecture (Extracted from [82]).

2.8: Feature extraction process (Extracted from [20]).

2.9 Illustration of the attention modules (Extracted from [75]).

2.10: Architecture for SE-ResNet module used in the experiment (Extracted from [32]).

2.11: The main framework of [25] proposed method (Extracted from [25]).

2.12: Diagram for breast cancer recognition using the IRRCNN model (Extracted from [4]).

2.13: An illustration of the transfer learning approach (Extracted from [62]).

2.14: Output of stain normalization methods (Extracted from [36]).

3.1 Methodology used in the implementation.

3.2 Methodology for Binary Classification Formulation.

3.3 Methodology for Multiclass Classification Formulation.

3.4 Slides of Adenosis breast cancer seen under different magnifications.

3.5 Slides of Fibroadenoma breast cancer seen under different magnifications.

3.6 Slides of Phyllodes Tumor breast cancer seen under different magnifications.

3.7 Slides of Tubular Adenoma breast cancer seen under different magnifications.

3.8 Slides of Ductal Carcinoma breast cancer seen under different magnifications.

3.9 Slides of Lobular Carcinoma breast cancer seen under different magnifications.

3.10 Slides of Mucinous Carcinoma breast cancer seen under different magnifications.

3.11 Slides of Papillary Carcinoma breast cancer seen under different magnifications.

3.12 Example of some images using data augmentation.

3.13 Transfer learning workflow (Extracted from [48]).

4.1: Performance stats for DNNs trained on 40x images.

4.2: Performance stats for DNNs trained on 100x images.

4.3: Performance stats for DNNs trained on 200x images.

4.4: Performance stats for DNNs trained on 400x images.

4.5: Confusion matrices for Inception-V3Net for different magnifications.

4.6: Confusion matrices for ResNet18 for different magnifications.

4.7: Confusion matrices for ShuffleNet for different magnifications.

4.8: Inception-V3 predictions on sample test images.

4.9: ShuffleNet predictions on sample test images.

4.10: ResNet18 predictions on sample test images.

4.11: Performance stats for Inception-V3Net trained on 40x images.

4.12: Performance stats for Inception-V3Net trained on 100x images.

4.13: Performance stats for Inception-V3Net trained on 200x images.

4.14: Performance stats for Inception-V3Net trained on 400x images.

4.15: Performance stats for ResNet18 trained on 40x images.

4.16: Performance stats for ResNet18 trained on 100x images.

4.17: Performance stats for ResNet18 trained on 200x images.

4.18: Performance stats for ResNet18 trained on 400x images.

4.19: Performance stats for ShuffleNet trained on 40x images.

4.20: Performance stats for ShuffleNet trained on 100x images.

4.21: Performance stats for ShuffleNet trained on 200x images.

4.22: Performance stats for ShuffleNet trained on 400x images.

4.23: Confusion matrices for Inception-V3Net trained on 40x images.

4.24: Confusion matrices for Inception-V3Net trained on 100x images.

4.25: Confusion matrices for Inception-V3Net trained on 200x images.

4.26: Confusion matrices for Inception-V3Net trained on 400x images.

4.27: Confusion matrices for ResNet18 trained on 40x images.

4.28: Confusion matrices for ResNet18 trained on 100x images.

4.29: Confusion matrices for ResNet18 trained on 200x images.

4.30: Confusion matrices for ResNet18 trained on 400x images.

4.31: Confusion matrices for ShuffleNet trained on 40x images.

4.32: Confusion matrices for ShuffleNet trained on 100x images.

4.33: Confusion matrices for ShuffleNet trained on 200x images.

4.34: Confusion matrices for ShuffleNet trained on 400x images.

List of Tables

1.1: Describe the pre-trained deep networks used in the thesis.

2.1: Comparison with some popular CNN using multiclass classification.

3.1: Describe the BreakHis dataset based on the image-level used in this experiment.

3.2: Parameters of data augmentation.

3.3 Properties of Training Methodology.

3.4: Confusion matrix for binary classification.

3.5: Evaluation metrics (Extracted from [30]).

4.1: Performance stats for DNNs trained on 40x images.

4.2: Performance stats for DNNs trained on 100x images.

4.3: Performance stats for DNNs trained on 200x images.

4.4: Performance stats for DNNs trained on 400x images.

4.5: Overall accuracy for multiclass classification.

4.6: Performance stats for Inception-V3Net trained on 40x images.

4.7: Performance stats for Inception-V3Net trained on 100x images.

4.8: Performance stats for Inception-V3Net trained on 200x images.

4.9: Performance stats for Inception-V3Net trained on 400x images.

4.10: Performance stats for ResNet18 trained on 40x images.

4.11: Performance stats for ResNet18 trained on 100x images.

4.12: Performance stats for ResNet18 trained on 200x images.

4.13: Performance stats for ResNet18 trained on 400x images.

4.14: Performance stats for ShuffleNet trained on 40x images.

4.15: Performance stats for ShuffleNet trained on 100x images.

4.16: Performance stats for ShuffleNet trained on 200x images.

4.17: Performance stats for ShuffleNet trained on 400x images.

4.18: Comparing the performance of similar DNNs from this thesis and literature in binary classification.

4.19: Comparing the performance of different DNNs from this thesis and literature in binary classification.

4.20: Comparing the performance of different DNNs from this thesis and literature in multiclass classification.

List of Acronyms and Abbreviations

WHO World Health Organization

MRI Magnetic Resonance Imaging

CT Computed Tomography

CAD Computer-aided detection

AI Artificial Intelligence

ML Machine learning

DL Deep Learning

SVM Support Vector Machines

CNN Convolutional Neural Network

ReLU Rectified Linear Unit

ConvNets Convolutional Networks

CSDCNN Class Structure-based Deep Convolutional Neural Network

BreakHis Breast Cancer Histopathological Image Classification

DAG Directed Acyclic Graph

Chapter 1: Introduction

1.1 Cancer and its Impact

Cancer is one of the most common causes of death worldwide. Cancer is a disease that unfortunately spread in cells and growing daily in the body. According to the World Health Organization (WHO), cancer is the second leading cause of death globally, and in 2018, it caused approximately 9.6 million deaths [74]. There are many types of cancer, such as breast cancer, skin cancer, lung cancer, and prostate cancer. According to the International Agency for Research on Cancer mentioned that the common diseases for females are breast cancer, and for males are lung cancer (see Fig. 1.1, 1.2).

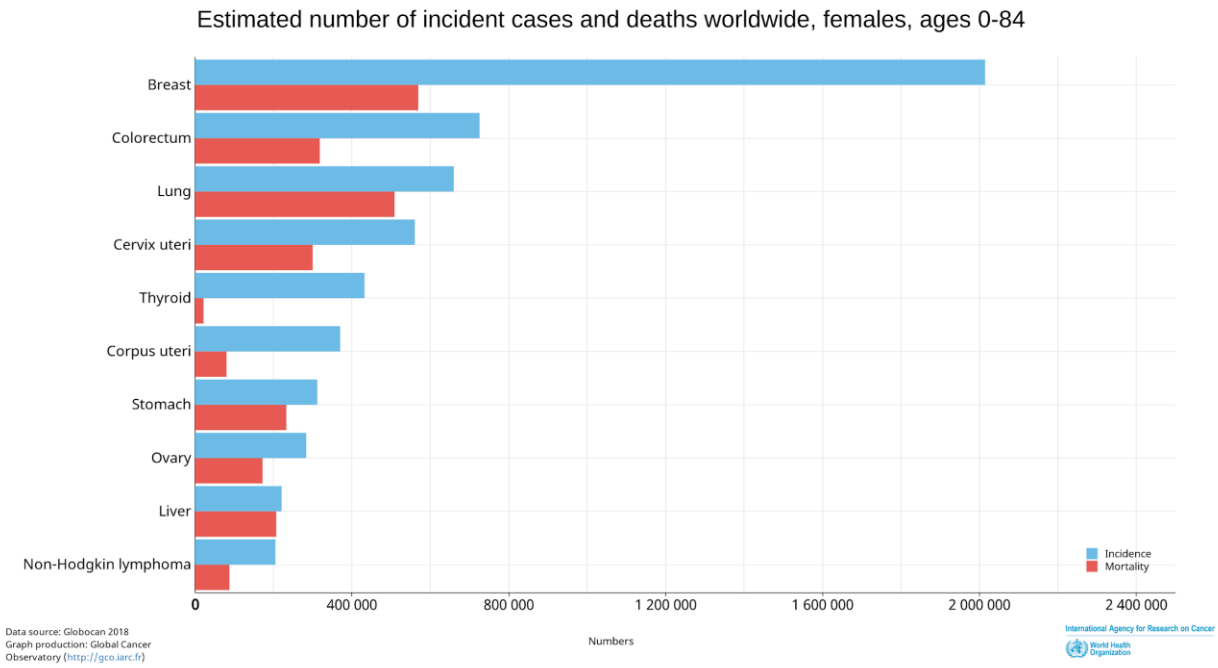


Figure 1.1: Common women cancers.

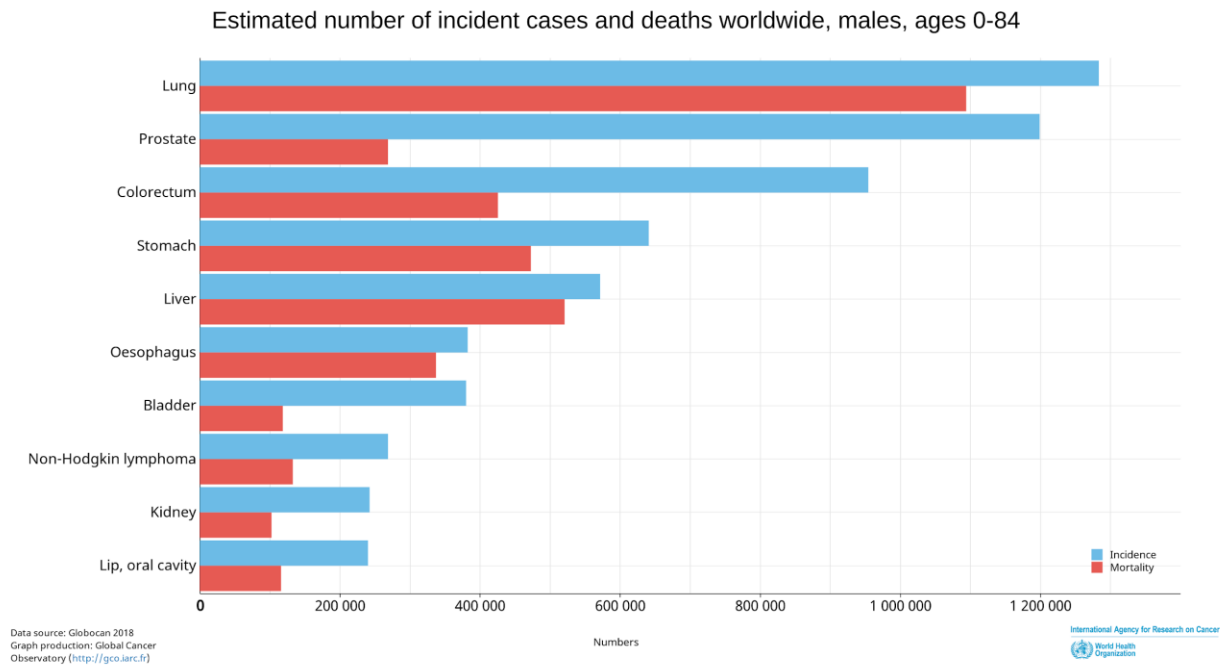


Figure 1.2: Common men cancers.

1.2 Breast Cancer

Breast Cancer is a widespread disease that many women are facing in their life; however, breast cancer can also appear in men. According to Breast Cancer Institute, breast cancer is one of the most deadly diseases that afflict women in the world [40]. Early diagnosis is the most dependable and reliable method for successfully managing cancer. In contrast, postponement of diagnosis may result in cancer spreading throughout the body and may be difficult to treat and control. Furthermore, late diagnosis leads reduced odds at successful treatment. The breast has many components such as connective tissue, fat, glands, and ducts, see Fig. 1.3 and they all may be affected by cancer. There are many methods for early diagnosis of breast cancer, such as self-examination at home, breast screening or visiting a doctor. These methods will reduce the rate of mortality and boost a chance for successful treatment. The most popular breast imaging tools are breast ultrasound, magnetic resonance imaging (MRI), computed tomography (CT), thermography, mammography, cytopathological and histopathological imaging. The most commonly used early breast cancer diagnostic modalities are ultrasound and mammography.

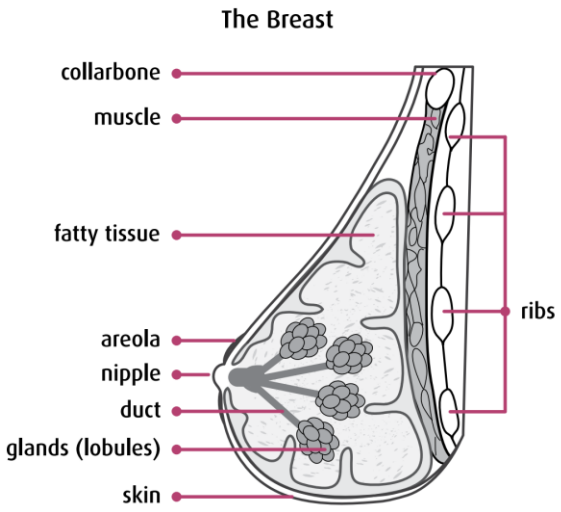


Figure 1.3: The structure of the breast.

1.3 Computer-Aided Cancer Detection

Computer-aided detection (CAD) software or computer-aided diagnosis of cancer is software technology aimed at assisting doctors in detecting or diagnosing cancer and reducing mortality by automatically processing medical imaging data. CAD is a software for medical image analysis, classification and grading of breast images to benign and malignant classes or stages. Recent years witness explosive growth of artificial intelligence (AI) and its applications. Among the most successful areas of AI is machine learning (ML) including deep learning (DL). There are many conventional classifiers used in ML, for instance, support vector machines (SVM) and decision trees. One of the most spectacular successes DL were obtained by the convolutional neural network (CNN) capable of automatically extracting features from images and classifying them with stunning accuracy [50],[64]. CNN has been applied to classification of images, speech, sound, text, videos and in designing self-driving cars. It is an effective technique to find patterns in data to recognize objects in different applications of computer vision and to extract features automatically to classify images (see Fig. 1.4).

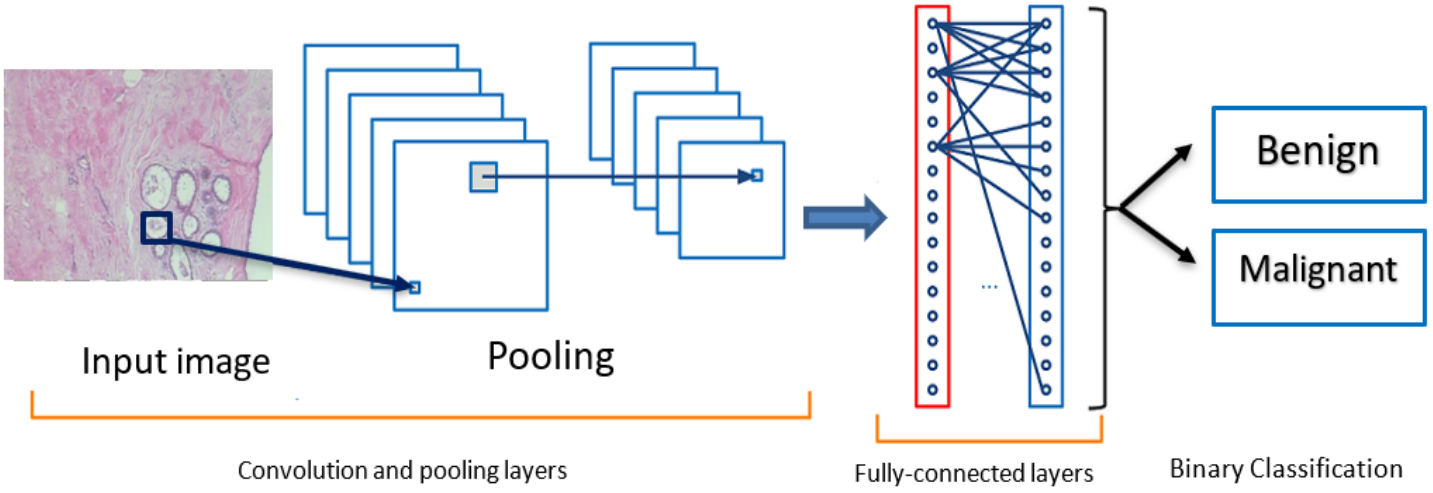


Figure 1.4: convolutional neural network workflow.

In terms of medical cancer applications, DL is using CNN to classify objects and require a large number of images and classify them to detect cancer cells. The concept of CNN may contain many hidden layers as learning from each other to obtain the features automatically of the objects. Moreover, filters can be applied in CNN to each training image with different pixel sizes, and the outcome used for the next layer. CNN consists of input layers, hidden layers, and output layers (see Fig. 1.5).

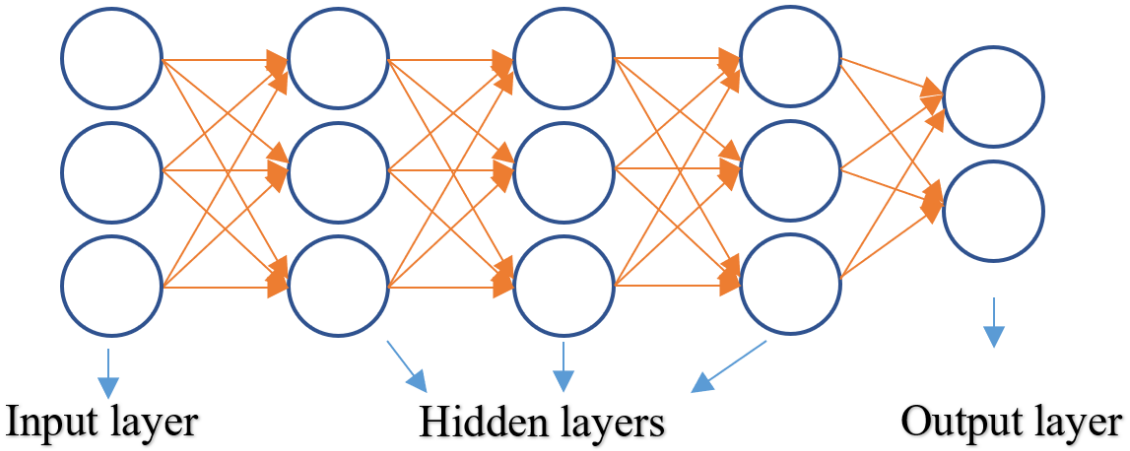


Figure 1.5: Architecture of the convolutional neural network.

CNN network consists of convolution layers and pooling layers. The convolution layer is uses a set of moving convolutional windows that convolve with the input image and extract features in the process. The pooling layer decreases spatial size of the convolutional layer. There are two types of pooling layers: max pooling and average pooling. Max pooling yields the max value of the spatial size; the average pooling returns the average value from the pooling window (see Fig. 1.6).

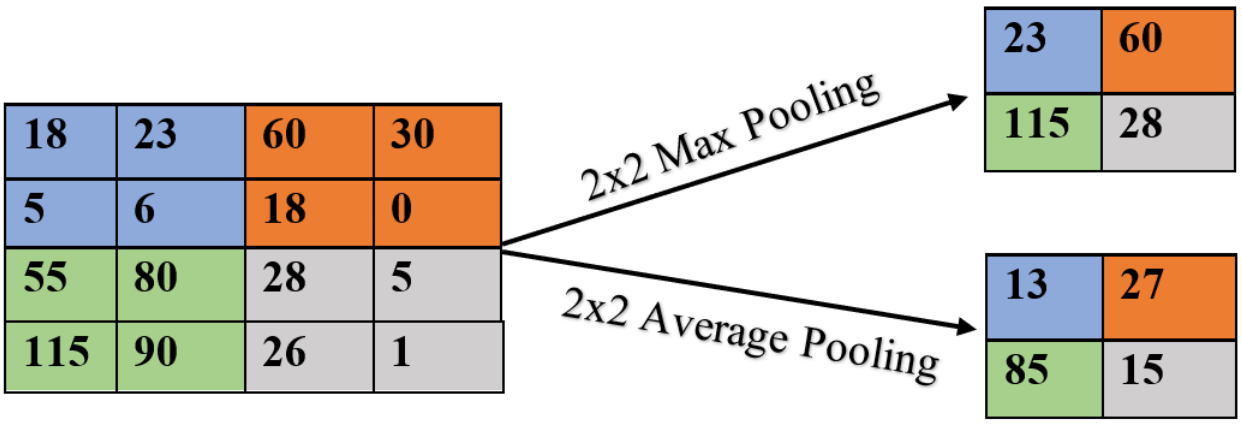


Figure 1.6: Max pooling and average pooling.

After learning features from layers, CNN performs classification using the fully connected output layer. There are many pre-trained deep neural networks implemented in different program languages such as AlexNet, GoogLeNet, Inception-V3Net, ResNet50, VGGNet19 and ShuffleNet which can be used to extract features automatically from natural images. Additionally, the tasks of pre-trained networks are classification and transfer learning with different models.

1.4 Structure of the Thesis

In this dissertation, chapter 2 focuses on the literature survey on breast cancer detection and classification using different deep learning models. Chapter 3 focuses on the methodology we used to classify breast cancer in terms of binary classification and multiclass classification from histopathological images using three pre-trained deep learning models. The following table describes the pre-trained deep neural networks classifiers used in this dissertation:

Network	Depth	Layers	Image input size
ResNet18	18	72	224-by-224
Inception-V3Net	48	316	299-by-299
ShuffleNet	50	173	224-by-224

Table 1.1: The pre-trained neural networks use in this thesis.

Chapter 4 shows the results of the three experiments models and compare them with state-of-the-art. Furthermore, explain more about the three models and the high accuracy achieved. Finally, chapter 5 presents the conclusions of the thesis and discusses future work.

Chapter 2: Literature Survey

This chapter focuses on the literature survey relevant to the main topic of this dissertation, i.e. classification of medical images. We briefly review breast cancer classification with different types of images and datasets. Many techniques have been published and improved to detect and classify breast cancer using deep neural networks with different architectures. The difficult task of medical image analysis and classification is to classify histopathological images based on the structure of the cells and complex morphology and texture. Many modern methods have been proposed to resolve the challenging issue of image classification such as deep learning models and pre-trained deep neural networks.

2.1 Traditional Approaches to Medical Image Classification

Traditional image classification approaches used standard features such as colour, texture, and standard classifiers such as support vector machines, random forests. In [8], the authors proposed two systems for the detection of melanoma in dermoscopy images using texture and colour features. The first system was used in their global implementation methods evolves in three sequential steps to classify skin lesions in the following: automatic segmentation, colour and texture features extracted, and train a classifier to perform binary classification, whereas the second system uses local features and the bag-of-features classifier. Paper [66] is concerned with the detection of granularity in dermoscopy images of skin lesions, which allow discriminating melanoma from non-malignant skin lesions using colour and texture features. Paper [81] uses MRI brain image classification based on a weighted-type fractional Fourier transform to extract spectra from each image based on the support vector machine.

Furthermore, [39] proposed a multi-view feature fusion-based computer-aided diagnosis system using the feature fusion technique of four views for the classification of mammograms. The main objective of the system is to evaluate deep convolutional neural network in the following tasks:

- (stage1) classify mammogram into two classes (normal and abnormal)
- (stage2) classification of abnormality type (mass or calcification)
- (stage3) classification of pathology (malignant or benign) using a fusion of four-views (R-CC, R-MLO, L-CC, and L-MLO), see Fig. 2.1.

They used two public datasets:

1. Digital Database for Screening Mammography with Curated Breast Imaging Subset (CBIS-DDSM)
2. Mammographic Imaging Analysis Society (MIAS).

The authors achieved the following results: AUC rates of 0.93 for stage-1, 0.932 for stage-2 and 0.84 for stage-3.

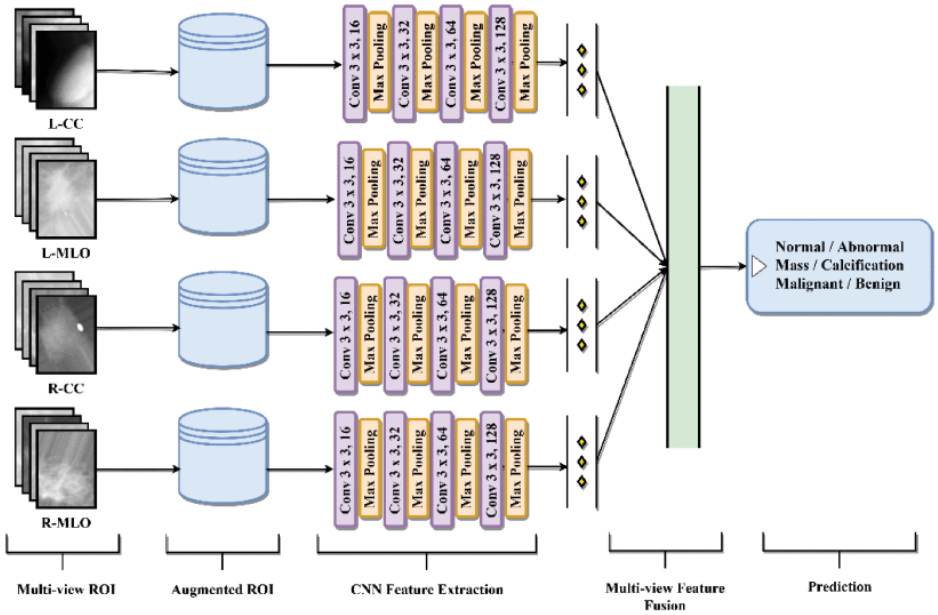


Figure 2.1: Proposed Multi-view Features Fusion (MVFF) based CAD system.

Lastly, in [22], the authors presented a discriminative feature fusion system for image classification. Also, they present a logistic regression-based fusion method that takes advantage of the different cues, and they develop a new marginalized kernel to use in the output of the regression model.

2.2 Deep Learning Approaches to Medical Image Classification

- **Convolutional Neural Network Approach**

Deep learning has been used in many domains and achieved high performance in applications such as natural language processing, speech recognition and computer vision. Convolutional Neural Network (CNN) is an algorithm technique for deep learning classification that learns from images, sounds, text, videos. CNN's have obtained extraordinary performance in different subjects of computer vision, such as image classification, image segmentation, face recognition. Paper [45] stated that Convolutional Networks (ConvNets) is biologically-inspired flexible architecture that can learn invariant features. The authors considered how deep learning- artificial intelligence can learn and extract features automatically. Furthermore, they mentioned that ConvNets are training with several stage architectures, and each layer has input and output called feature maps, which represents extracted features, see Fig. 2.2 Convolutional layer, pooling layer, and fully connected layer are the main three layers in ConvNets architecture. The severe problem facing deep learning is that many data are unlabeled. Also, they proposed that unsupervised learning with sparse predictive decomposition [38], and ConvNets require a large number of data for training.

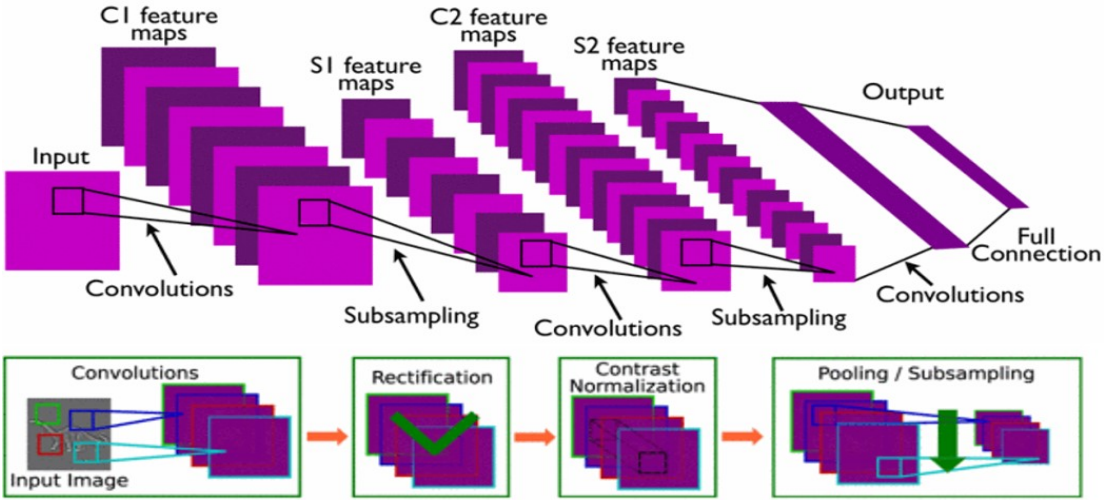


Figure 2.2: A typical ConvNet architecture with two feature stages and an example of a feature.

Paper [9] developed a deep learning model for magnification-independent breast cancer histopathology images classification based on CNN with different sizes of kernels filters such as 3x3, 5x5, and 7x7. The following papers used breast cancer histopathology image analysis and convolutional neural networks for detection of regions of interest [16], segmentation [67], and mitosis detection [14]. In their experiments, they used BreakHis public dataset [65] of breast cancer histopathological images using different magnifying factors (40×,100×,200×,400×). They proposed two architectures: single task CNN and multi-task CNN with preprocessing and data augmentation. Single task CNN is used to predict malignancy, and multi-task CNN architecture models are used to predict both malignancy and image magnification levels simultaneously based on data. Also, for pre-processing, they cropped the images and fixed the size of images in order to be appropriate for CNN, and they achieved an 83.25% recognition rate [9]. Fig. 2.3 below presents more details of the proposal they used in their experiments.

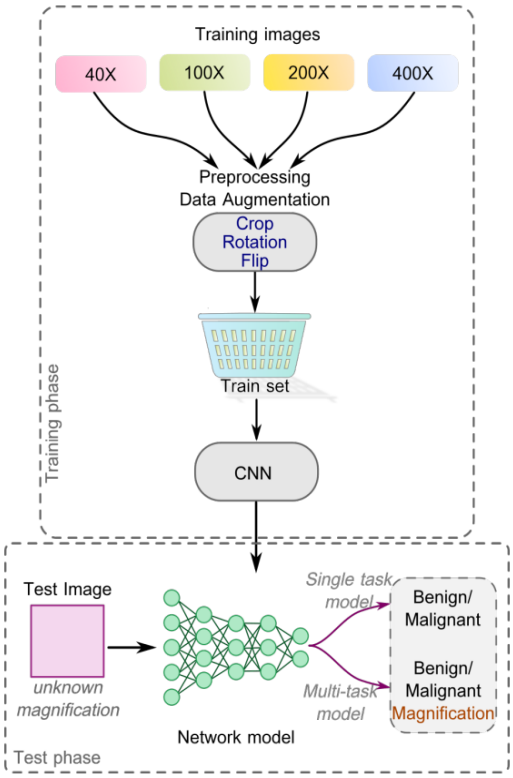


Figure 2.3: Schematic presentation of the algorithm used in [9].

Paper [55] designed a system for breast cancer classification using histopathological images and the main objective was to achieve higher performance accuracy from minimal labelling. Their approach includes labelled data and unlabeled data integrated into the training set. Also, they used two selection strategies: an entropy-based strategy and a confidence-boosting strategy. Likewise, [53] pointed out that multiclass breast cancer classification using a convolutional neural network of histopathological images. They used BreakHis public dataset, which 7,909 breast cancer histopathology images with subclasses of benign and subclasses malignant. Fig. 2.4 presents the convolutional neural network architecture and approach.

In [64], the authors proposed a breast cancer histopathological image classification system using convolutional neural networks and applied it for the BreakHis dataset. In their experiments,

they extracted patches of images for training data convolutional neural networks and integrated of these patches for final classification with two different image patch sizes 32×32 and 64×64. BreakHis contains microscopic biopsy images of benign and malignant breast tumours collected in 2014. They split the data as follows: 30% testing and 70% for training. They implemented a new approach similar to AlexNet's pre-training deep neural network with different fusion techniques for patient-level classification and image-level classification.

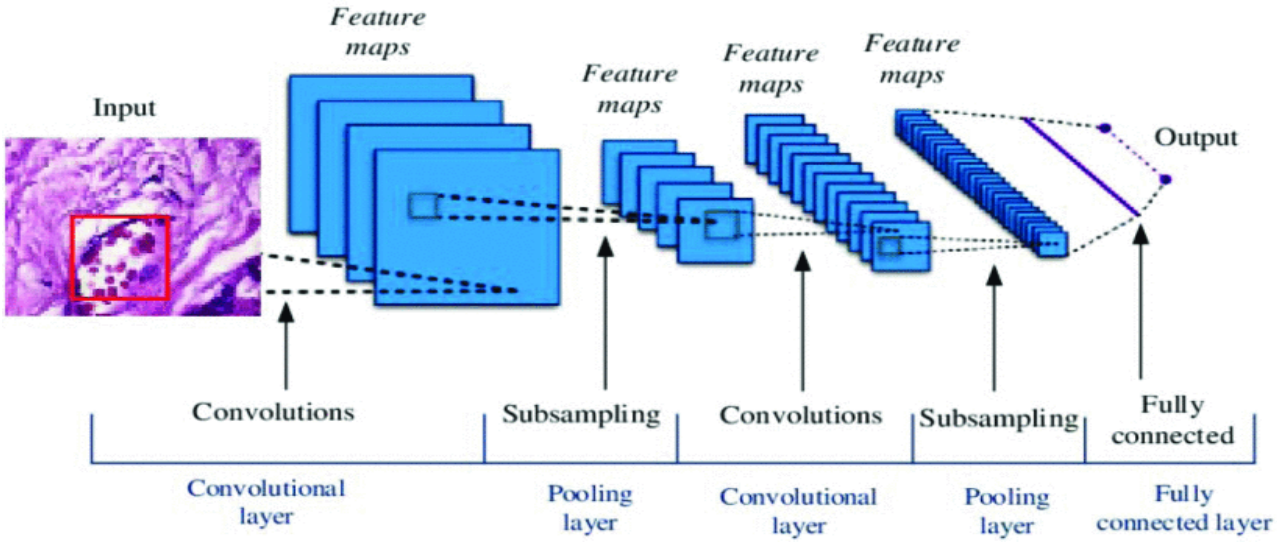


Figure 2.4: Typical CNN architecture with specific stages.

- **Deep Convolutional Neural Networks**

In [1] the authors discussed the breast cancer classification using histopathological images via using Deep Convolutional Neural Networks to solve the issue for eight cancers of either Benign or Malignant type. In their methodology, they applied deep convolutional neural networks and showed their effectiveness in the classification of images. Fig. 2.5 explains the architecture of the implemented model, which used Deep Convolutional Neural Networks with an ensemble learning method with the TensorFlow Framework with Backpropagation training and ReLu activation function. In [46], the authors introduced the cancer diagnostic system using a tandem of classifiers and tested it on digitized histopathological images. In their experiments, they tested their system on colon and breast cancer using different datasets concerning four classes: healthy tissue and cancerous of grades 1, 2 or 3.

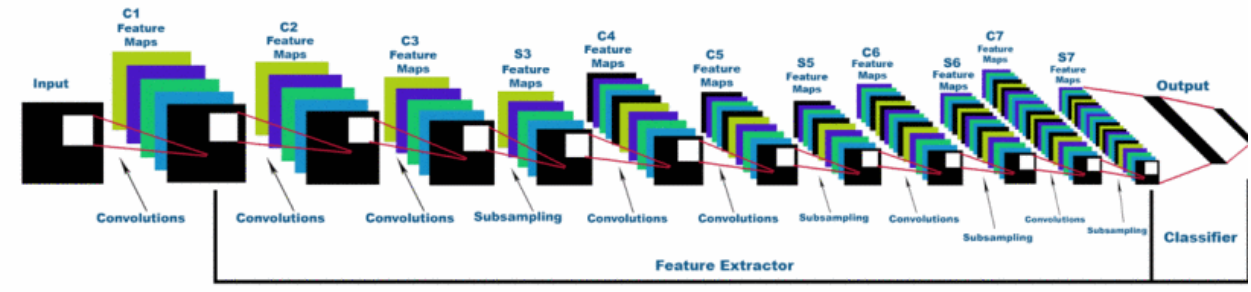


Figure 2.5: Architecture of the model used for DCNN.

They used public dataset BreakHis [65] and tested the model with 600 images. They used SoftMax function as the output of their model, which transfers the probability of each class given the input. The test images were classified into the class with the highest probability. Consequently, they achieved an inter-class classification accuracy of 91.5%.

Work [28] proposed a breast cancer multiclass classification system with a structured deep learning model and trained and tested it on histopathological images from the BreakHis Database. The authors proposed a class structure-based **deep convolutional neural network** (DCNN). In their experiment they tested automated breast cancer classifier for multiclass classification and achieved extremely high accuracy. Fig. 2.6 describes the proposed system. Furthermore, they used

the data augmentation to increase the size of the dataset and improve the performance of the system and to resolve the issue of imbalance. There are many ways to achieve data augmentation, e. g., random combination of intensity variation, rotation, flips and translation. They used the real dataset for testing and validation, but they applied augmentation for training data only. The average accuracy for all magnification factors of the patient level multiclass classification is 93.8%, and for image-level multiclass classification is 93.3%.

Paper [82] proposed breast cancer histopathology images classification by assembling multiple compact Convolutional Neural Networks. In their experiments, the hybrid model with adopting the multi-model assembling scheme achieved higher performance based on multiple compact hybrid CNNs. They implement a system that reduces the workload of pathologists as well as improve the quality of diagnosis. Fig. 2.7 proposed hybrid CNN architecture. Two model branches integrated to extract more critical information, and the channel pruning module is embedded to compact the network.

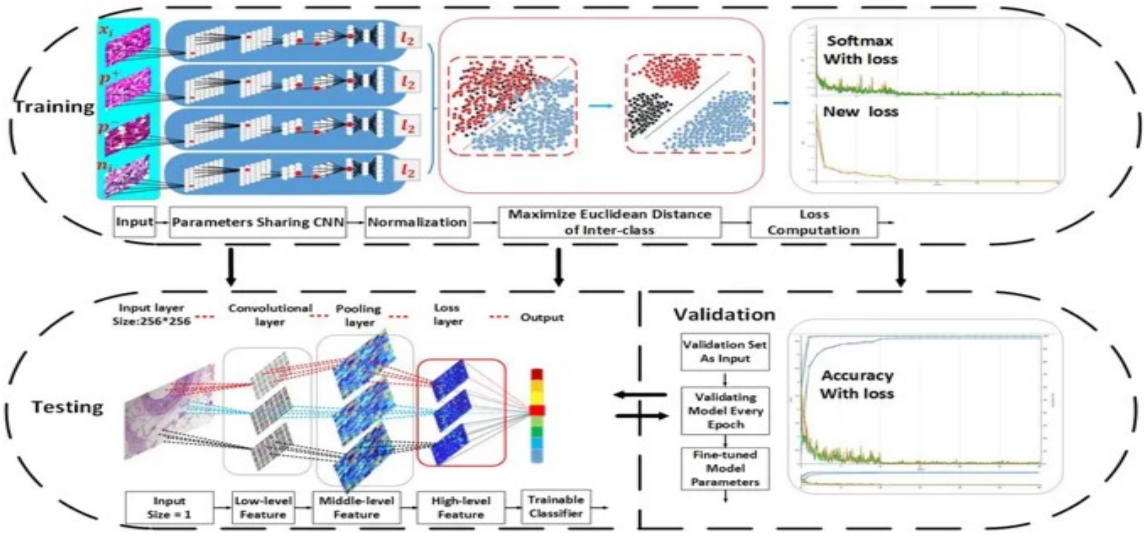


Figure 2.6: Overview of the workflow proposed in [82].

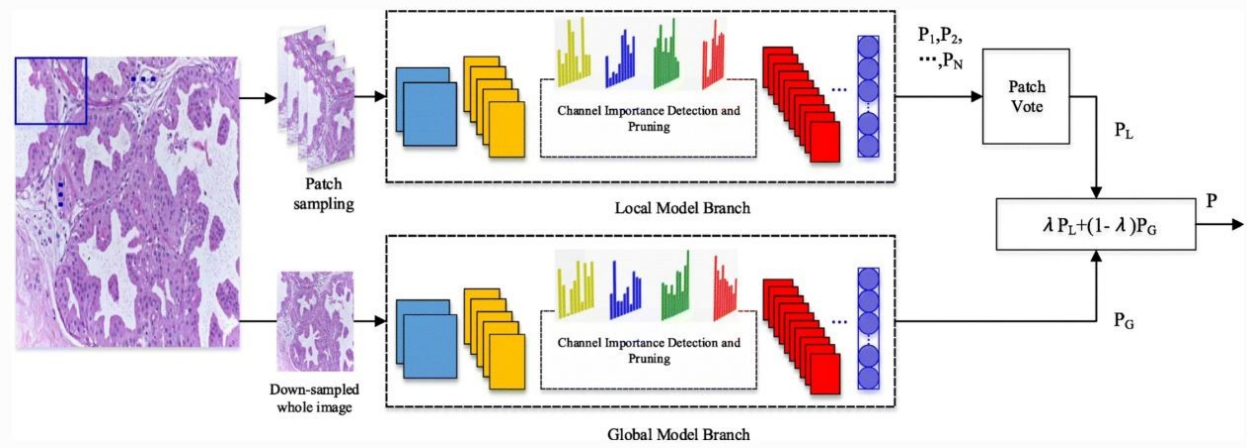


Figure 2.7: Hybrid CNN architecture.

In paper [20], the authors proposed a new system based on local binary patterns histograms on histopathology images that explicitly aware of the heterogeneity of local texture patterns through heterogeneity-based weighting. They used homogeneity and the second moment (variance) of local neighbourhoods based on heterogeneity information, so that makes better capture in histopathology images. Besides, they implement three datasets in their experiments KimiaPath24 dataset, IDC datasets, and BreakHis datasets and on KimiaPath24, they achieved better performance than on other datasets. Fig. 2.8 shows the image from the KimiaPath24 dataset divided into square patches, and then sample local binary patterns values extracted for each block. The feature extraction process is applied to each colour channel separately, and then these histograms are concatenated into the final feature vector.

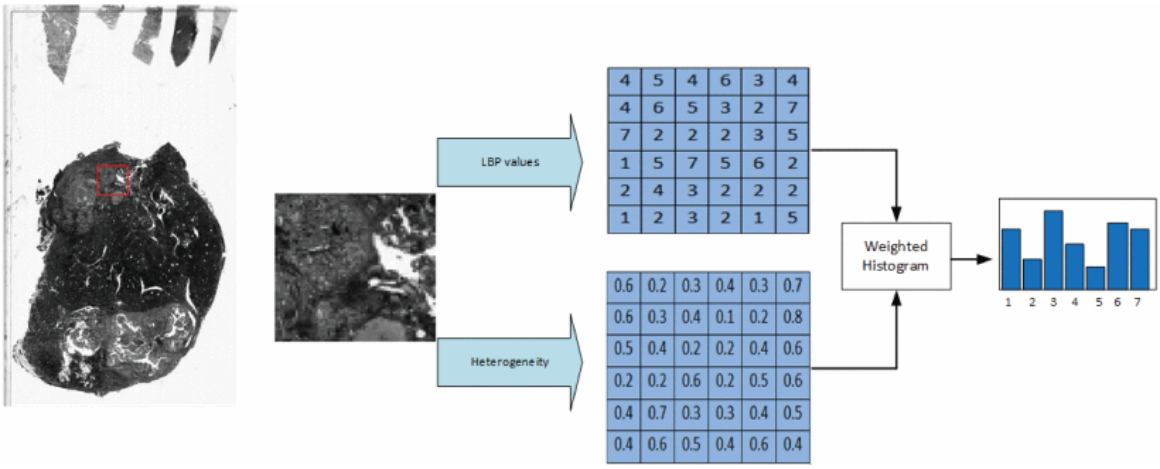


Figure 2.8: Feature extraction process for histopathology image thumbnail from the KimiaPath24 dataset.

2.3 Pre-trained Deep Neural Networks Approach

Next, [52] presented multiclass classification breast cancer classification using deep learning convolutional neural networks using an existing deep neural network DenseNet. DenseNet is a pre-trained convolutional neural network and has 201 layers. They used BreakHis public database and implemented the patient and image classification. They achieved 95.4% of the accuracy for image classification and 96.48% for patient classification in the multiclass breast cancer classification. They modified the DenseNet model to deal with breast cancer histopathological images using transfer learning. They define transfer learning as a fine-tuning CNN model pre-trained from the original image dataset to medical image tasks. DenseNet is a pre-trained deep neural network that can work well with medical images to classify cancer. In [27] proposed breast cancer histopathological image classification using a sequential framework that utilizes multi-layered in-depth features that extracted from fine-tuned DenseNet pre-trained convolutional neural network. They applied the transfer learning of pre-trained DenseNet for breast histopathology image classification, the classifier at each layer trained on features extracted from that individual convolution layer of DenseNet.

Moreover, [75] showed that deep attentive feature learning for histopathology image classification and they implemented a new deep learning-based CNN architecture. They used a deep learning-based approach for image classification and patient classification to detect benign and malignant breast cancer on the BreakHis dataset. Besides, they adopted VGG19 as the base model and insert the attention modules at different positions. VGG-19 is a pre-trained convolutional neural network and has 19 layers with an input size of 224x224. Furthermore, they designed two attention modules, the channel-wise attention module, and the spatial attention module. Based on that, they integrated the two modules, and they used VGG19 neural network Fig. 2.9 for more details on the two modules they applied in their experiments. The average accuracy of image classification is 91.75%, and the average accuracy for patient classification is 98.05%.

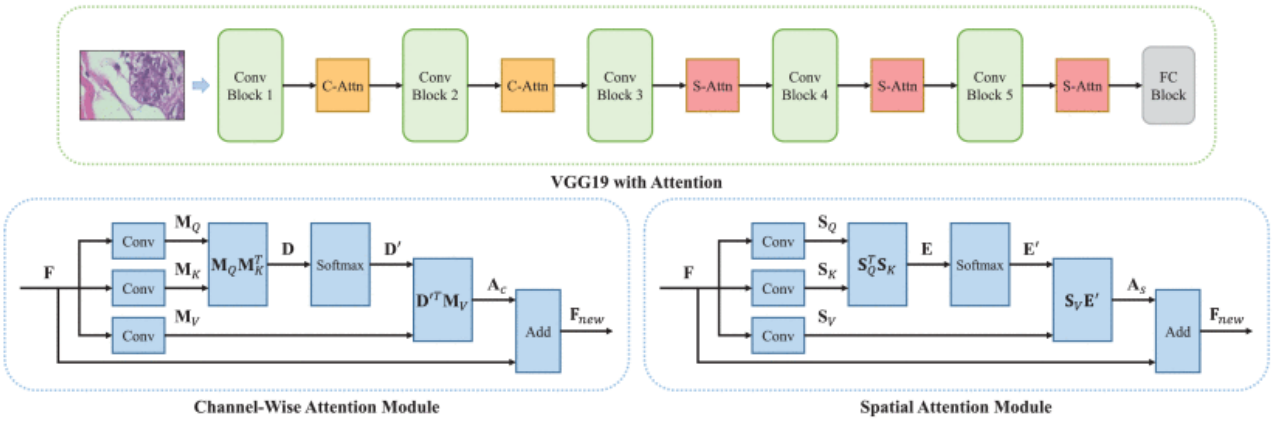


Figure 2.9: Illustration of the attention modules, which placed at different positions of the VGG19 network.

Paper [32] discusses breast cancer histopathological image classification using convolutional neural networks with a small SE-ResNet module. In their experiments, they implemented a convolutional neural network with a small SE-ResNet module for breast cancer histopathological image on BreakHis public dataset. The SE-ResNet module includes a convolutional layer, a small SE-ResNet module, and fully connected layer, to improve on the combination of residual module and Squeeze-and-Excitation block, and achieves the similar performance with fewer parameters as they used BreakHis dataset for automatic classification for multiclass classification and binary classification. Fig. 2.10 shows the architecture for the benign and malignant classification of breast cancer histopathological images using in the implementation of the SE-ResNet module. In the end, they achieved accuracy between 90.66% and 93.81% for the multiclass classification.

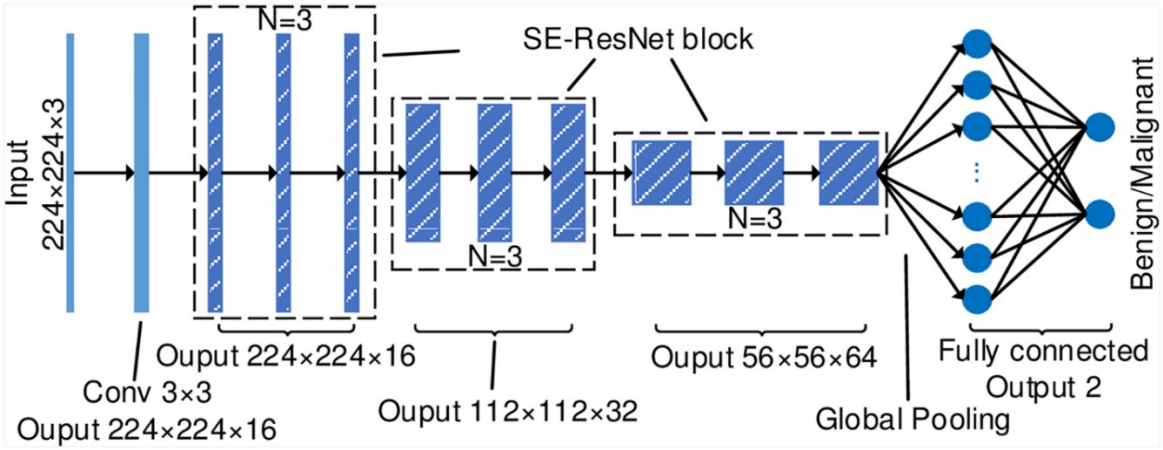


Figure 2.10: Architecture of SE-ResNet module used in the experiment.

According to [31], there are different techniques on deep learning based on breast cancer detection. The objective of their experiment is to compare two model networks of deep learning techniques, which are VGG16 and ResNet50. They used Image Retrieval in Medical Application dataset, which is a mammography image and, in their methodology, started with input images, pre-processed data, and image resize to fit in the training network system. Next, their data is classified based on normal tumour or abnormal tumour, and the performance evaluated by using measures such as precision, recall, and accuracy. It turns out that VGG16 produces a better result, with 94% compared to ResNet50 with 91.7% in terms of accuracy.

Paper [25] adapted the VGG16 pre-trained neural network as the base model for hashing. The authors presented a densely-connected multi-magnification framework in order to generate the discriminative binary codes by exploiting the histopathological images with multiple magnification factors. Figure 2.11 shows the framework of their experiments. The components of the system are: (a) the architecture of densely-connected include learning from low-magnification images and learning from the high-magnification images (b) learning component from high-magnification images, (c) learning component from low-magnification images.

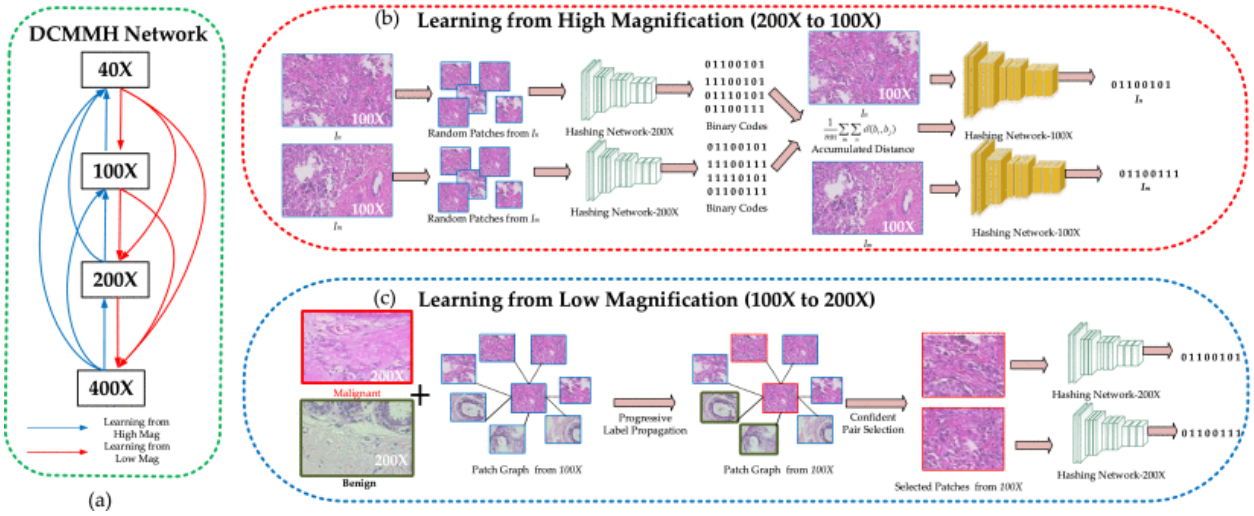


Figure 2.11: The main framework of the method proposed [25].

In [4] the authors published a new model for breast cancer classification from histopathological images based on deep neural networks. They combined the Inception Network (Inception) and Residual Network (ResNet) coming up with the new convolutional neural network (IRRCNN). They applied their system to breast cancer classification on two public datasets: BreakHis and Breast Cancer Classification Challenge 2015. They achieved the best performance to date on both datasets in image, patch and patient based classification. They developed a new system for the pre-trained neural network and utilized the power of ResNet and Inception obtaining impressive performance. Figure 2.12 below shows structure of their system for breast cancer recognition using the IRRCNN model. The upper part of the figure shows the steps used for training and the lower part of the figure displays the steps used for testing of their model.

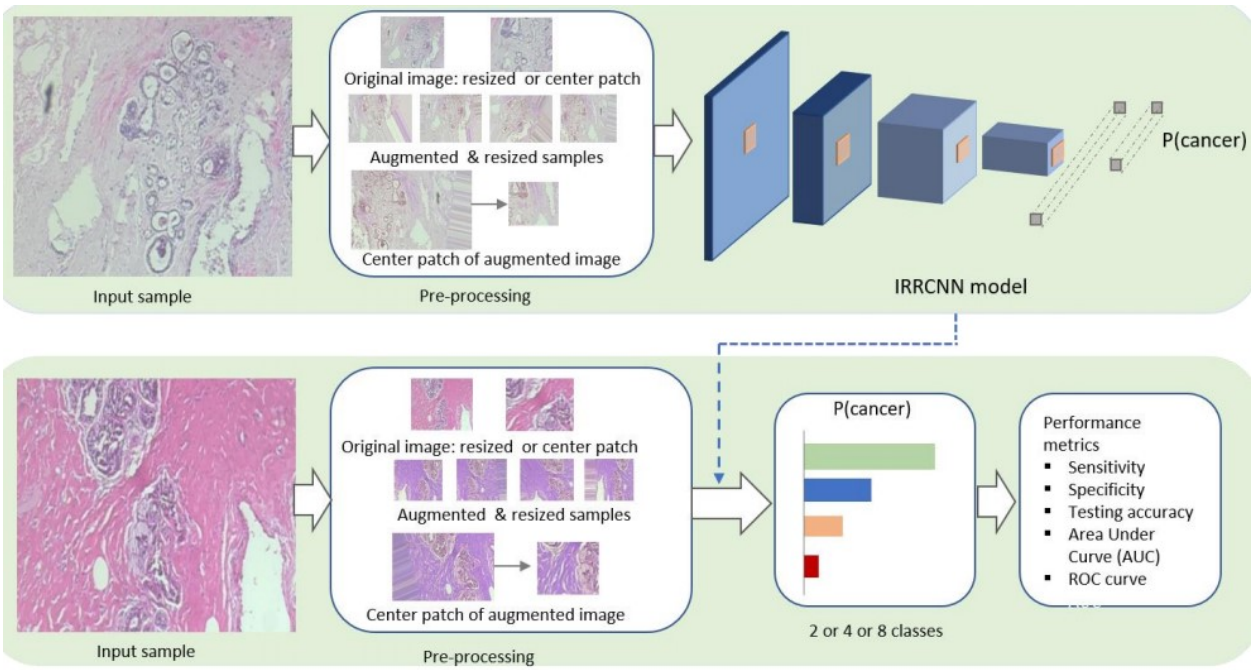


Figure 2.12: Diagram for breast cancer recognition using the IRRCNN model.

- **Transfer Learning Pretrained Deep Neural Networks**

The effect of layer-wise fine-tuning in the magnification-dependent classification of breast cancer histopathological images was discussed in [62]. A common problem with medical data is the imbalance problem, which makes training of CNN a challenging task. The authors of [62] used pre-trained AlexNet CNN and tested their system on BreakHis public dataset. The BreakHis is an imbalanced dataset, which creates inevitable problems. The use of transfer learning mitigates the issue of data inadequacy, and it results in faster training than learning from scratch. Figure 2.13 below illustrates the transfer learning approach used in [62], where the network is trained on a simple dataset of natural images. In their experiment, they used transfer learning, that on the last three layers of the pre-trained network to make the pre-trained suitable for their dataset.

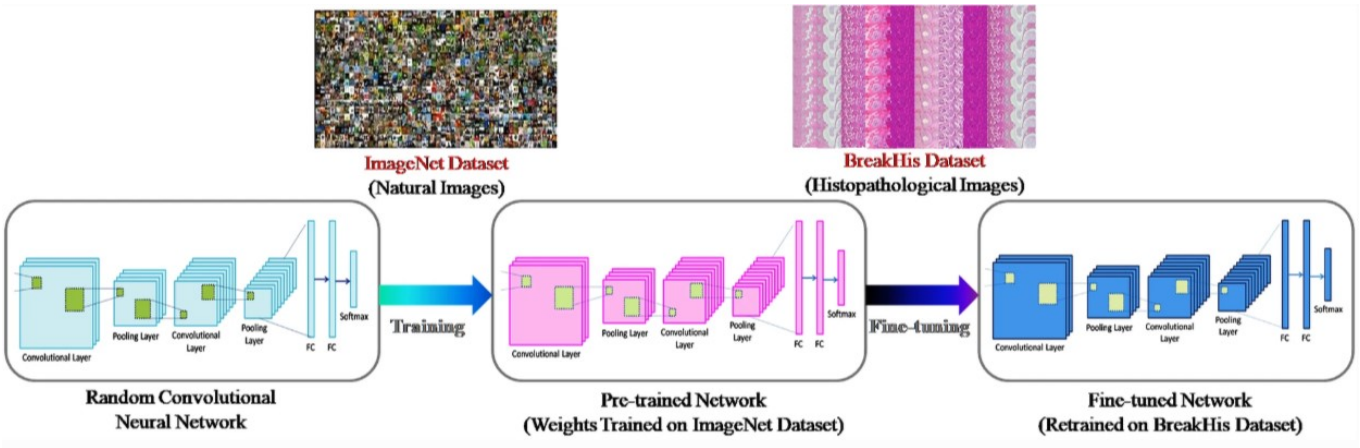
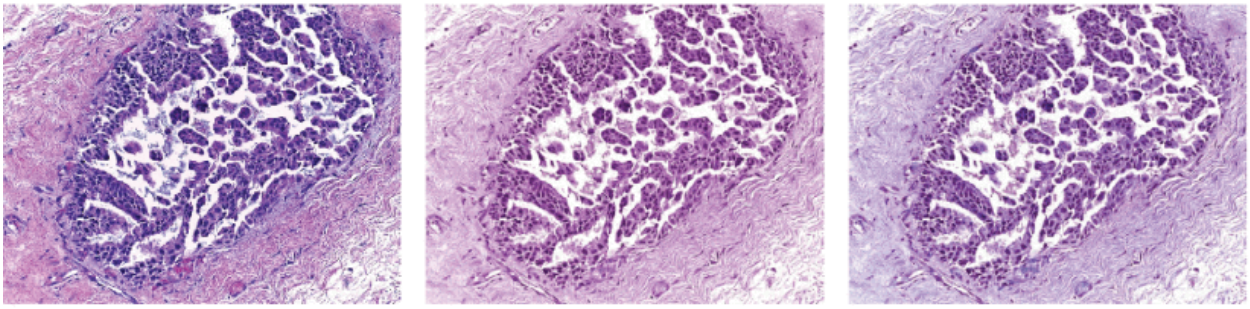


Figure 2.13: An illustration of the transfer learning approach.

Lastly, in [36], the authors proposed a breast cancer diagnosis system with transfer learning and global pooling. They used public dataset ICIAR 2018 Grand Challenge on Breast Cancer Histology (BACH) Images with four categories, namely benign, normal, in-situ, and invasive

carcinoma, evenly distributed. In their data pre-processing, they used the popular stain normalization techniques of Macenko of [47] and Reinhard [57]. They implemented different pre-trained deep learning models, which are Inception-V3NET, InceptionResNetV2, Xception, VGG16, and VGG19, and the results show that Xception achieved the best performance and the average classification accuracy was 92.50%.



(a) (b) (c)

Figure 2.14: a) Original image H#E stained image, b) Macenko-normalized image, c) Reinhard-normalized image.

Chapter 3: Methodology

3.1 Proposed Method

In this chapter, we present the main contributions and the proposed method used in the task of classification of breast cancer using histopathological images for binary and multiclass classification. There have been numerous studies on applications of deep learning to the classification of breast cancer from histopathological images. In this thesis, we use a publicly available Database called BreakHis [65]. Our framework uses the following steps:

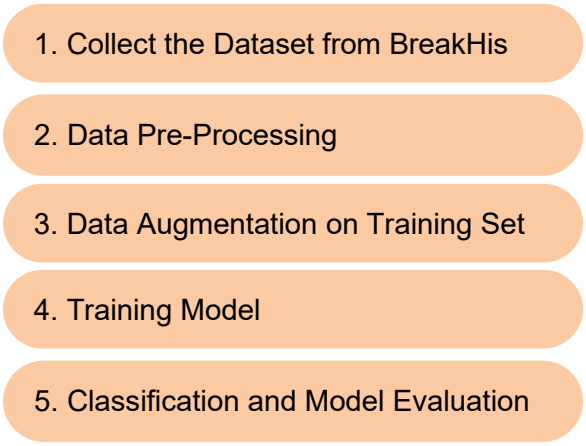


Figure 3.1: Breast cancer classification framework

First, to obtain the database and to have access to the microscopic biopsy images, one must request and fill the form online from the Laboratory of Vision, Robotics, and Imaging of the Federal University of Parana, Brazil [64]. For each input image, we used some pre-processing techniques to improve its quality. The next Section explains in detail pre-processing used in the implementation. After pre-processing, we dived the data randomly on 70% for the training set and 30% for the testing set. Next, we applied data augmentation for the training dataset only. Augmentation involved image operations such as resizing, rotation and reflection images. After training the parameters of the network it classifies the test set and performance metrics such as classification accuracy and confusion matrices are computed. In the next chapter, we present the results using different evaluation metrics.

In this dissertation, as mentioned above, we used BreakHis Database based on image-level, which has two main classes: benign and malignant, and each class has four sub-classes with different magnification factors (40X, 100X, 200X, 400X). The following Fig. 3.2 shows the framework of binary classification.

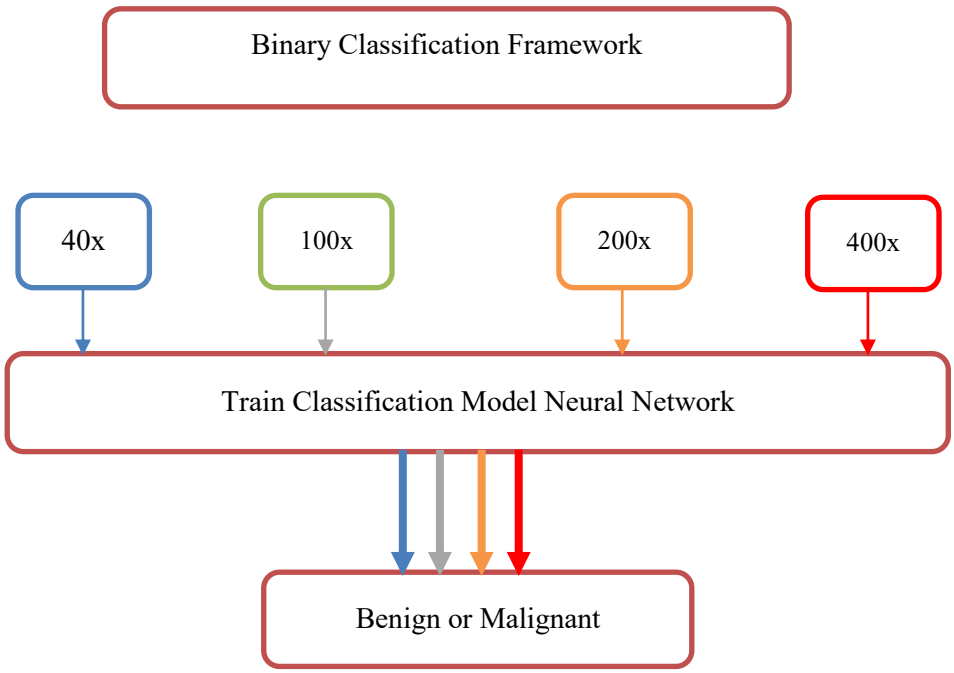


Figure 3.2: Methodology for Binary Classification Formulation.

As explained in Fig. 3.2 the system for binary classification of images from the dataset involves images having different magnification factors that use the methodology presented in Fig. 3.1. Fig. 3.3 shows the implementation for sub-classes with different magnification factors (40X, 100X, 200X, 400X). After training, the test data used to evaluate system performance.

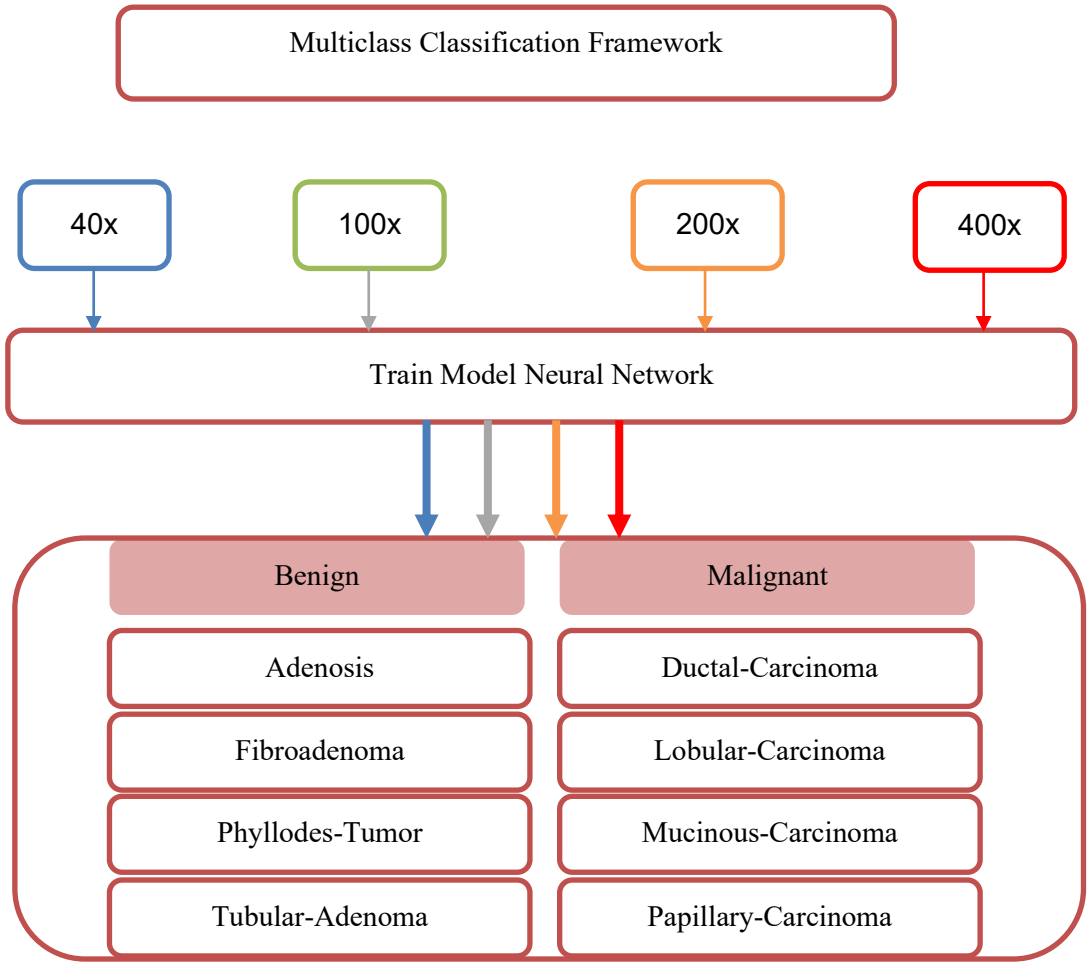


Figure 3.3: Methodology for Multiclass Classification Formulation.

3.2 Dataset Description

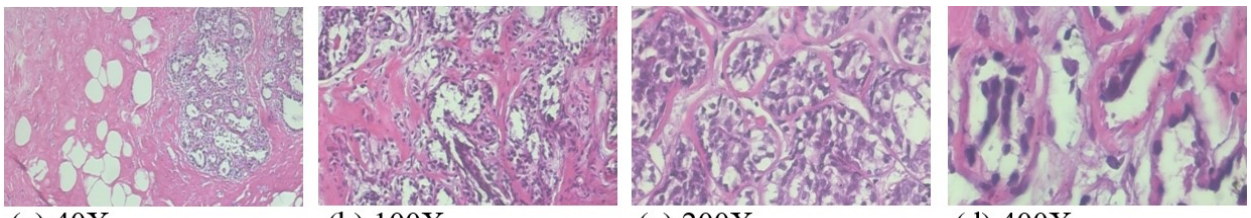
In this thesis, we use the Breast Cancer Histopathological Image Classification (BreakHis) Database from the Laboratory of Vision Robotics and Imaging at the Federal University of Parana, Brazil [64]. The dataset BreakHis is divide into two main groups: benign tumours and malignant tumors with different magnification factors (40X, 100X, 200X, 400X). The images in the dataset are RGB images of resolution 700Xx460 pixels in 3-channels. There are 2,480 benign samples and 5,429 malignant samples from 82 patients. Additionally, benign images represent four classes: Adenosis (A), Fibroadenoma (F), Phyllodes Tumor (PT) and Tubular Adenoma (TA). Malignant images represent four classes: Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous

Carcinoma (MC) and Papillary Carcinoma (PC). The following Table describes the distribution of images in the BreakHis dataset according to sub-types of cancer.

Types	Sub-Types	40X	100X	200X	400X	Total
Benign	**A**	114	113	111	106	**444**
	F	253	260	264	237	**1014**
	PT	109	121	108	115	**453**
	TA	149	150	140	130	**569**
Malignant	**DC**	864	903	896	788	**3451**
	LC	156	170	163	137	**626**
	MC	205	222	196	169	**792**
	PC	145	142	135	138	**560**
Total		**1995**	**2081**	**2013**	**1820**	**7909**

Table: 3.1: The distribution of images from the BreakHis dataset by cancer categories.

The following figures show cancer samples at different magnifications.



(a) 40X (b) 100X (c) 200X (d) 400X

Figure 3.4: Slides of Adenosis breast cancer under different magnifications.

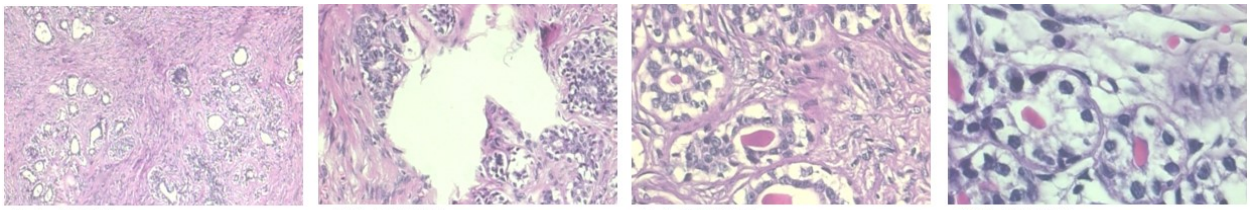


Figure 3.5: Slides of Fibroadenoma breast cancer under different magnifications.

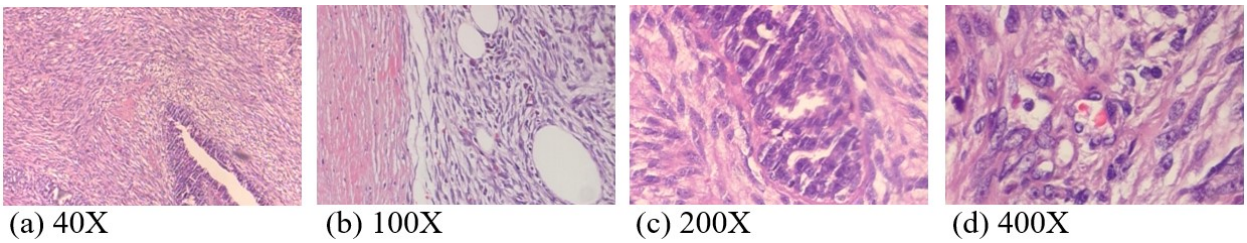


Figure 3.6: Slides of Phyllodes Tumor breast cancer seen different magnifications.

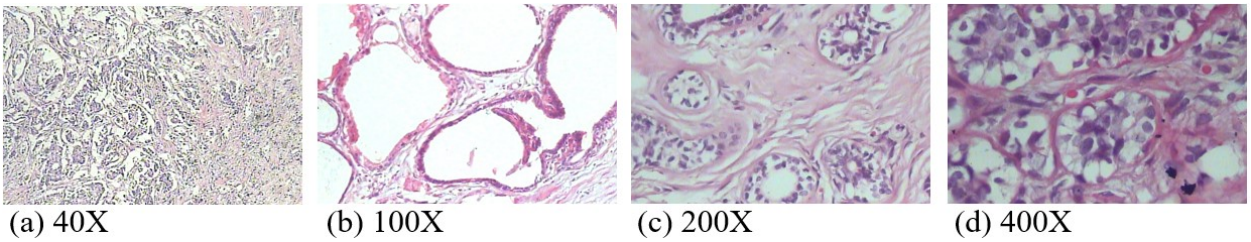


Figure 3.7: Slides of Tubular Adenoma breast cancer under different magnifications.

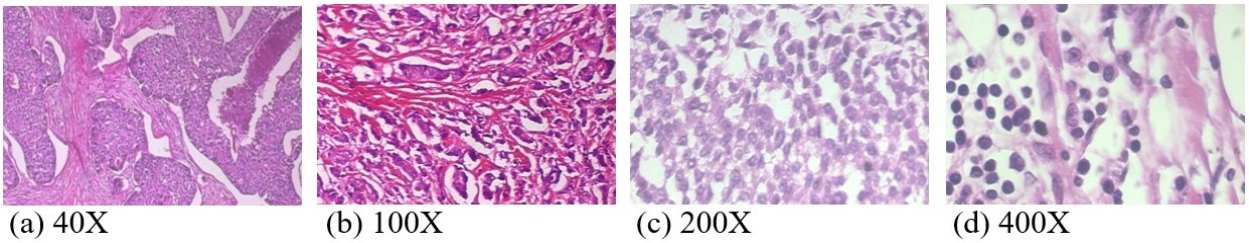


Figure 3.8: Slides of Ductal Carcinoma breast cancer under different magnifications.

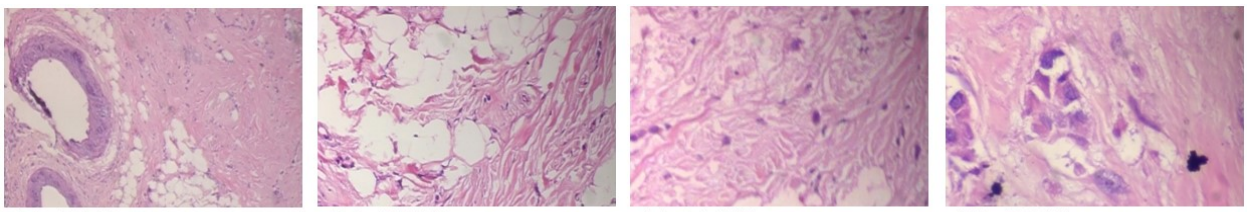


Figure 3.9: Slides of Lobular Carcinoma breast cancer under different magnifications.

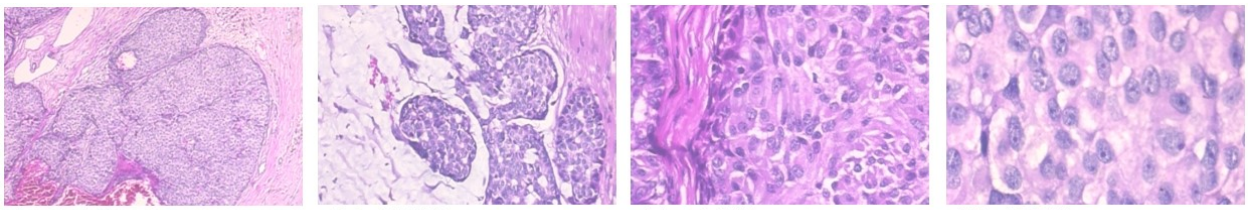


Figure 3.10: Slides of Mucinous Carcinoma breast cancer under different magnifications.

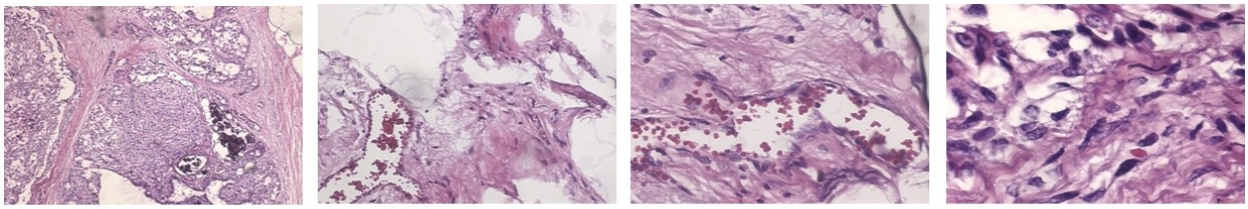


Figure 3.11: Slides of Papillary Carcinoma breast cancer under different magnifications.

3.2.1 Data pre-processing

Before we could use the data, we had to size normalize it to make it consistent with the requirements of different networks. Resizing of input images was accomplished via data rescaling and cropping to ensure that image sizes fit each pre-trained deep neural network. The networks used in this thesis are pre-trained deep neural network ResNet18, ShuffleNet and Inception-V3Net, and they all require different size input images. Specifically, ResNet18 uses 224x224 images; Inception-V3Net uses 299x299 images, and lastly, ShuffleNet uses 224x224 images.

3.2.2 Data Augmentation

CNN requires a considerable amount of data for learning its parameters. A standard technique for expanding the training data set is augmentation [35]. Augmentation helps in improving system performance, reduces the chance of overfitting and data imbalance. There are many techniques for data augmentation. They include random reflection, rotations and horizontal or vertical translations. We applied data augmentation only to the training set, and the test set has not been augmented. The transformations used in augmentation of our training set are presented in Table 3.2.

Parameters	Value
Range of Rotation	90°, 180°
Random Reflection	True
Range of horizontal shear	True
Range of vertical shear	True
Range Scale	True

Table 3.2: Transformations used in data augmentation.

Fig. 3.12 below shows examples of data augmentation by means of rotations by 90°and 180° and by flipping.

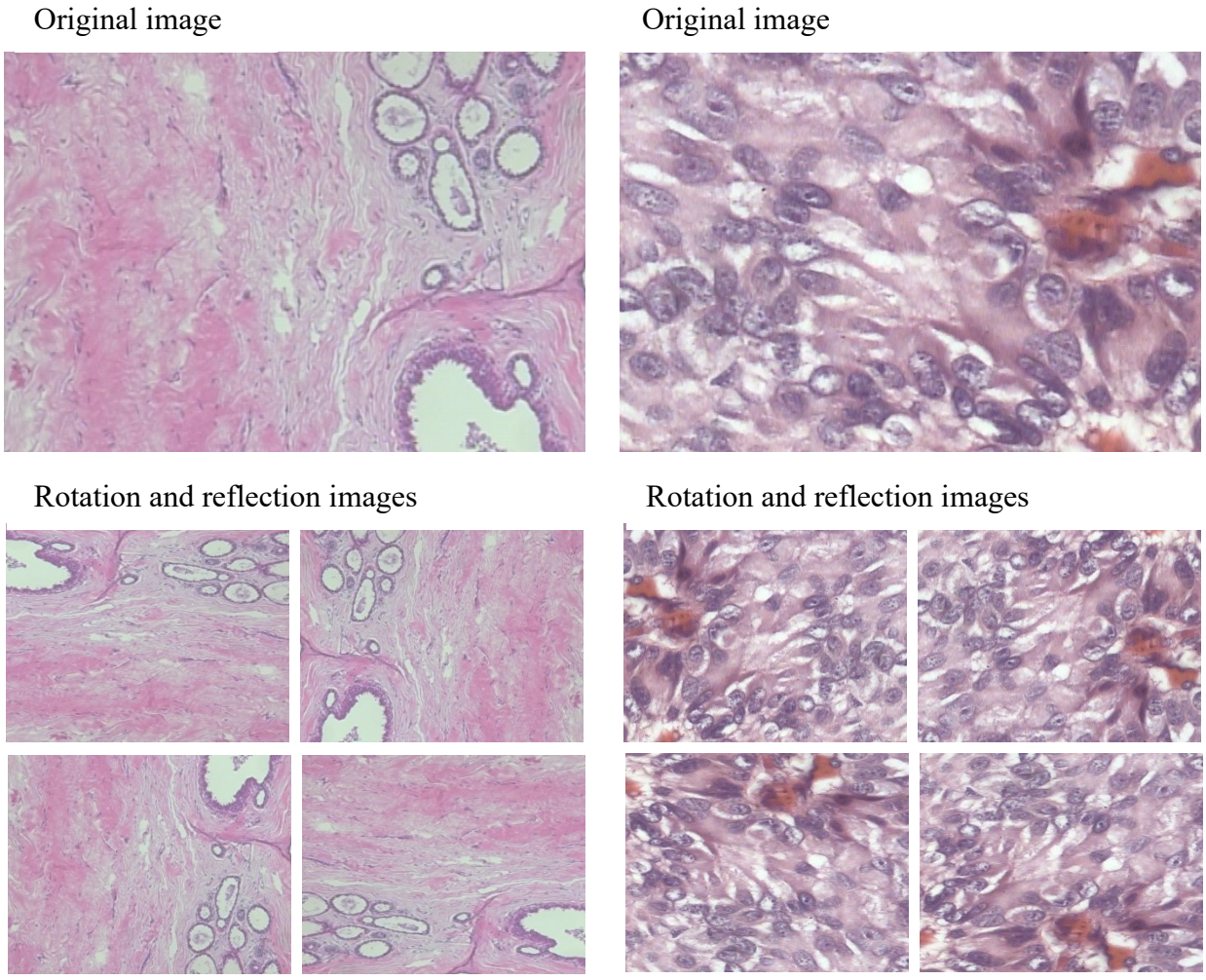


Figure 3.12: Illustration of data augmentation.

3.3 Transfer Learning

Considering the complexity of medical imaging and in particular of histopathological images, transfer learning (TL) helps to boost performance classification. Transfer learning is a method that can be applied in the learning of deep neural networks. It allows knowledge gained in solving one type of problem to be transferred to solving related problems. Typically the first layers of deep nets are trained on large datasets of images, e.g., on ImageNet, which contains about 14 million images from 20,000 categories and then the last layers of the nets are trained on the specific problem at hand, e.g., on breast cancer classification and grading. Fine-tuning of the deep net in combination with transfer learning allows for much faster and more efficient learning than training all network weights (often millions of them), starting with randomly initialized weights. Transfer learning has found its place in numerous applications such as image recognition, medical image classification, and speech recognition. Fig. 3.13 shows the transfer learning workflow and explains how this technique can be used in deep learning. The first segment of our deep network consists of deep network pre-trained on ImageNet Database. The final three output layers of our classifier are as follows: fully connected layer, SoftMax layer and output classification layer. The final layers are trained on the BreakHis dataset of histopathological images. In other words, we extract all layers from the pre-trained network except for the last three layers. The fully connected output layer has the number of outputs consistent with the number of classes in the dataset. Furthermore, we increase the value of the Weight Learn Rate Factor (WLRF) and the Bias Learn Rate Factor (BLRF) to boost the learning rate of the output layers weights. TL with a pre-trained deep network model can learn general features from a source dataset that does not exist in the current dataset [34].

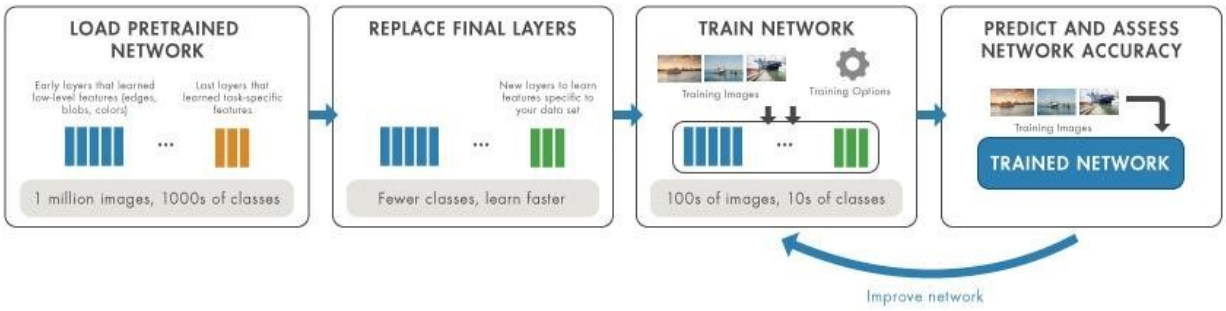


Figure 3.13: Transfer Learning workflow.

3.4 Training Methodology

After implementing transfer learning in the pre-trained network and setting both values for WLRF and BLRF to 20, training of the network requires adjusting sizes of input images depending on the network used. The Adaptive Moment Estimation algorithm (Adam [41]) was used in network optimization, and values of important parameters are listed in Table 3.3.

Property	Value
Initial learning rate	1e-4 (0.0001)
Learn Rate Schedule	Piecewise
Squared Gradient Decay Factor	0.99
Epoch	15
Mini Batch Size	128

Table 3.3: Properties of Training Methodology.

3.5 Dataset Experimental Protocol

There are two possible ways for using data for training and testing of the classifier: one is the image-based approach, and the other is a patient-based approach. In the image-based approach, the data is divided into the training set and the testing set using only image labels. This may result in placing data from the same patient in both training and testing sets, which makes the train and test data correlated and may result in overly optimistic performance indicators. In the patient-based approach, the train and test data contain images from different patients and are thus uncorrelated. The performance indicators may be lower than in the former case, but they are more realistic and have better predictive value. There are different studies that implemented the dataset in terms of the image-based approach, such as [25, 59, 62]; in contrast, paper [27] they implemented in their

experiment the patient-based approach. In this thesis, we adopted the image-based approach to learning our deep net classifiers.

In our experiments with the BreakHis dataset, we used the image-based approach and divided the dataset by randomly selecting without replacement 70% of the data for the training set and 30% for the testing set. We train and test the data for each magnification separately. We did not use cross-validation, and we implemented data augmentation on the training set to balance the classes and to avoid overfitting. Furthermore, we implemented dropout in each pre-trained neural network to prevent the neural network from overfitting. The optimizer algorithm is Adaptive Moment Estimation (Adam), and we set the learning rate to 0.0001, decay factor set to 0.99, and the batch size to 128. We judged the performance of our neural network classifiers using standard metrics such as classification accuracy, precision, sensitivity, specificity, and F1 score.

3.6 Pre-trained Deep Neural Networks Types

Deep Neural Networks are one of the most successful techniques used in medical imaging. In this dissertation, we applied three pre-trained deep neural network classifiers. They are ResNet18, Inception-V3Net and ShuffleNet. These networks are capable of learning robust and useful features automatically from images. Next we briefly describe the three nets.

3.6.1 ResNet18

ResNet is a short name for Residual Neural Network, which is a convolutional neural network that is trained on more than a million images from the ImageNet database and can classify images into thousand object categories [29]. ImageNet is built upon the hierarchical structure provided by WordNet and, in its completion, aims to contain in the order of 50 million images [18]. ResNet18 has 44 MB size and consists of 18 deep layers and accepts input image 224x224 resolution. The Directed Acyclic Graph (DAG) is an effective system that uses to build a deep neural network from scratch, or it can use pre-trained neural networks such as ResNet18,

Inception-V3Net, and ShuffleNet. DAG network architecture for ResNet18 has 79 connection layers.

3.6.2 Inception-V3Net

Inception-V3Net is a convolutional neural network that trained on more than a million images from the ImageNet database and can classify images into thousand object categories [69]. The Inception-V3Net network has 48 deep layers and can classify images into 1000 object categories. The Inception-V3Net network has 89MB size and accepts inputs of size is 299x299. The DAG network architecture for Inception-V3Net consists of 350 connected layers.

3.6.3 ShuffleNet

ShuffleNet is a convolutional neural network pre-trained on more than a million images from the ImageNet database and can classify images into thousand object categories [80]. The shuffle network has 50 deep layers and can classify images into 1000 object categories. The shuffle network accepts input images of resolution 224x224 and of 6.3MB size. The DAG network architecture for the Shuffle network consists of 188 connects layers.

3.7 Evaluation Metrics

The confusion matrix is an excellent tool to measure the performance of the binary classifier. Tables 3.4 and Table 3.5 below show the confusion matrices for binary classification and the evaluation matrices used in the experiment.

	Class1 Predicted	Class2 Predicted
Class1 Actual	TP	FP
Class2 Actual	FN	TN

Table 3.4: Confusion matrix for binary classification.

• True Positive (TP): Observation is positive and predicted to be positive.

• False Negative (FN): Observation is positive but predicted negative.

• True Negative (TN): Observation is negative and predicted to be negative.

• False Positive (FP): Observation is negative but predicted positive.

Evaluation metrics	Formula
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$
Precision **Positive Predictive Value**	$\frac{TP}{TP + FP}$
Sensitivity (Recall) **True Positive Rate**	$\frac{TP}{TP + FN}$
Specificity **True Negative Rate**	$\frac{TN}{FP + TN}$
F1 Score	$\frac{2 * Sensitivity * Precision}{Sensitivity + Precision}$

Table 3.5: Evaluation metrics.

Chapter 4: Experiment Results and Discussion

In this chapter will present the results of experiments with three pre-trained in transfer learning deep neural networks system applied to breast cancer classification involving histopathological images. The system is trained and tested on image data from the publicly available BreakHis dataset of 7909 real samples (images) from 82 patients, divided into two subsets of 2,480 benign samples and 5,429 malignant samples. The following Sections will describe the result for binary classification, multiclass classification and compare the result with other studies.

4.1 Results for Binary Classification

In this Section, we present classification results for ResNet18, ShuffleNet and Inception-V3Net deep neural network classifiers (DNN) tested on images from BreakHis with different magnifying factors 40x, 100x, 200x and 400x and compare their performance. In Table 4.1 and Fig. 4.1, we compare the performance of three DNNs on images with magnification 40. For all performance metrics, the best performance is achieved by ResNet18. Next, we compare the performance of DNNs on images with magnification 100 (Table 4.2 and Fig. 4.2), 200 (Table 4.3 and Fig. 4.3) and 400 (Table 4.4 and Fig. 4.4), respectively. The results are mixed. For the 100x magnifying factor, the performances of ResNet 18 and ShuffleNet vary according to different metrics. ResNet 18 and ShuffleNet obtained the same performance accuracy; for precision and specificity, ShuffleNet received the highest rate; however, ResNet18 achieved the highest rate for sensitivity and F1 score. For the 200x case, ResNet18 obtained the highest performance in terms of accuracy, precision, specificity and F1 measure. In terms of sensitivity, ShuffleNet scored higher than ResNet18 and Inception-V3Net. Eventually, Table 4.4 and Fig. 4.4 compare the three pre-trained networks for 400x. For 400x case, ShuffleNet obtained the highest performance in terms of accuracy, F1 score and matched the performance of ResNet18 for sensitivity. Inception-V3Net exhibited higher performance than ShuffleNet and ResNet18 in terms of precision and specificity.

Comparison of performance of DNNs trained on 40x images			
Evaluation Metrics	**ResNet18**	**ShuffleNet**	Inception-V3Net
Accuracy	**99.18%**	98.64%	97.01%
Precision	**99.45%**	98.91%	97.27%
Sensitivity	**98.91%**	98.37%	96.74%
Specificity	**99.46%**	98.91%	97.28%
F1 Score	**99.18%**	98.64%	97.00%

Table 4.1: Performance stats for DNNs trained on 40x images.

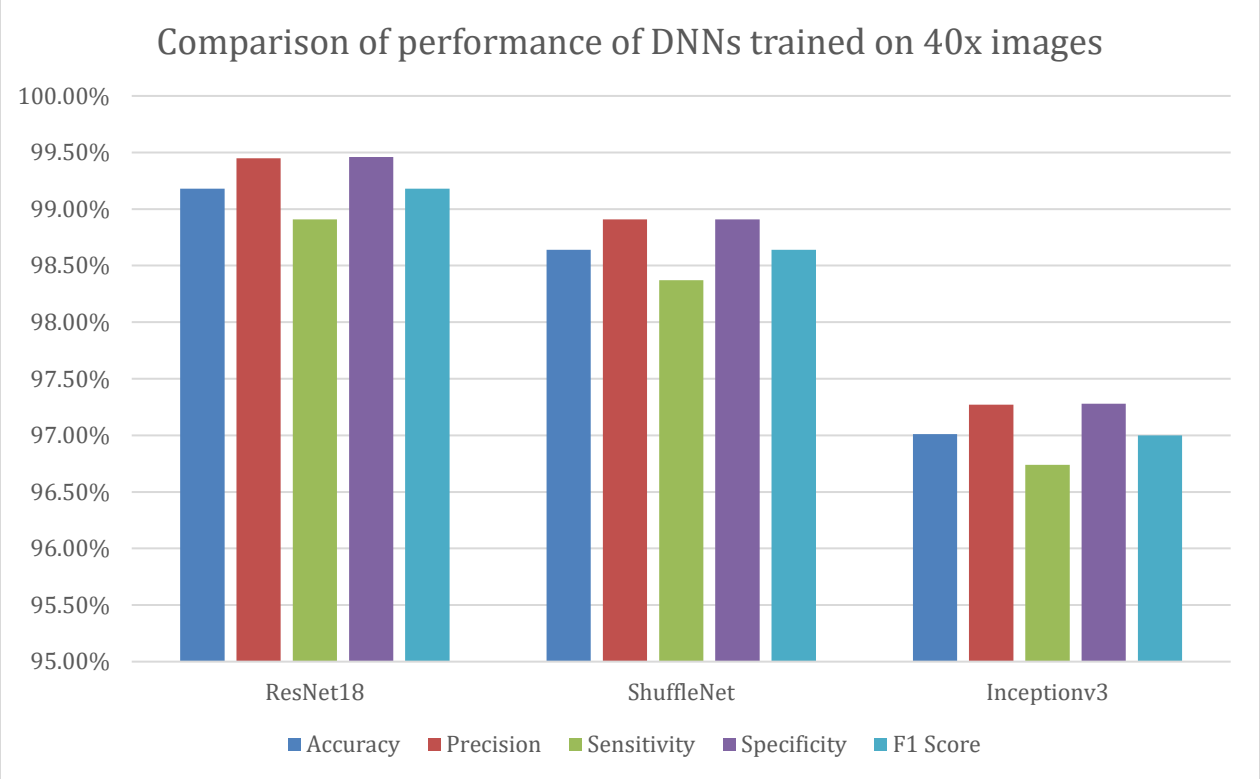


Figure 4.1: Performance stats for DNNs trained on 40x images.

Comparison of performance of DNNs trained on 100x images			
Evaluation metrics	**ResNet18**	**ShuffleNet**	Inception-V3Net
Accuracy	**97.67%**	**97.67%**	96.11%
Precision	98.42%	**98.94%**	96.84%
Sensitivity	**96.89%**	96.37%	95.34%
Specificity	98.45%	**98.96%**	96.89%
F1 Score	**97.65%**	97.64%	96.08%

Table 4.2: Performance stats for DNNs trained on 100x images.

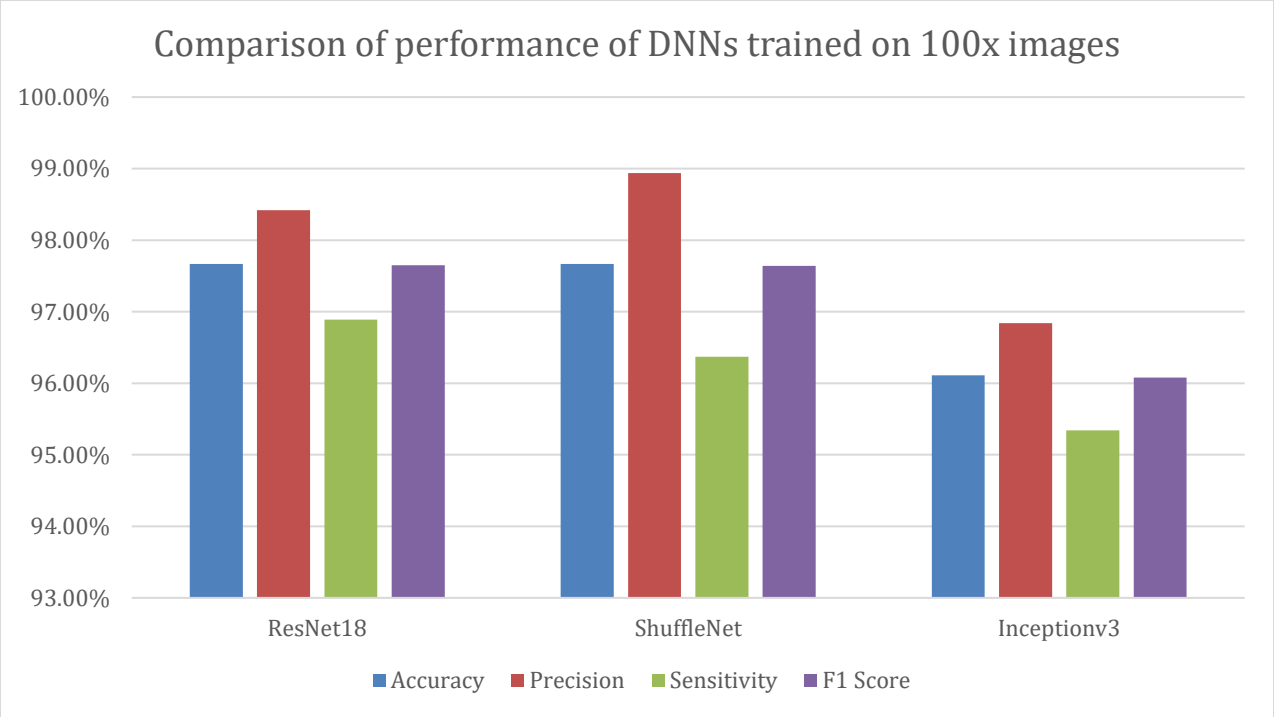


Figure 4.2: Performance stats for DNNs trained on 100x images.

Comparison of performance of DNNs trained on 200x images			
Evaluation metrics	**ResNet18**	**ShuffleNet**	**Inception-V3Net**
Accuracy	**97.59%**	96.52%	95.45%
Precision	**98.90%**	95.31%	96.20%
Sensitivity	96.26%	**97.86%**	94.65%
Specificity	**98.93%**	95.19%	96.26%
F1 Score	**97.56%**	96.57%	95.42%

Table 4.3: Performance stats for DNNs trained on 200x images.

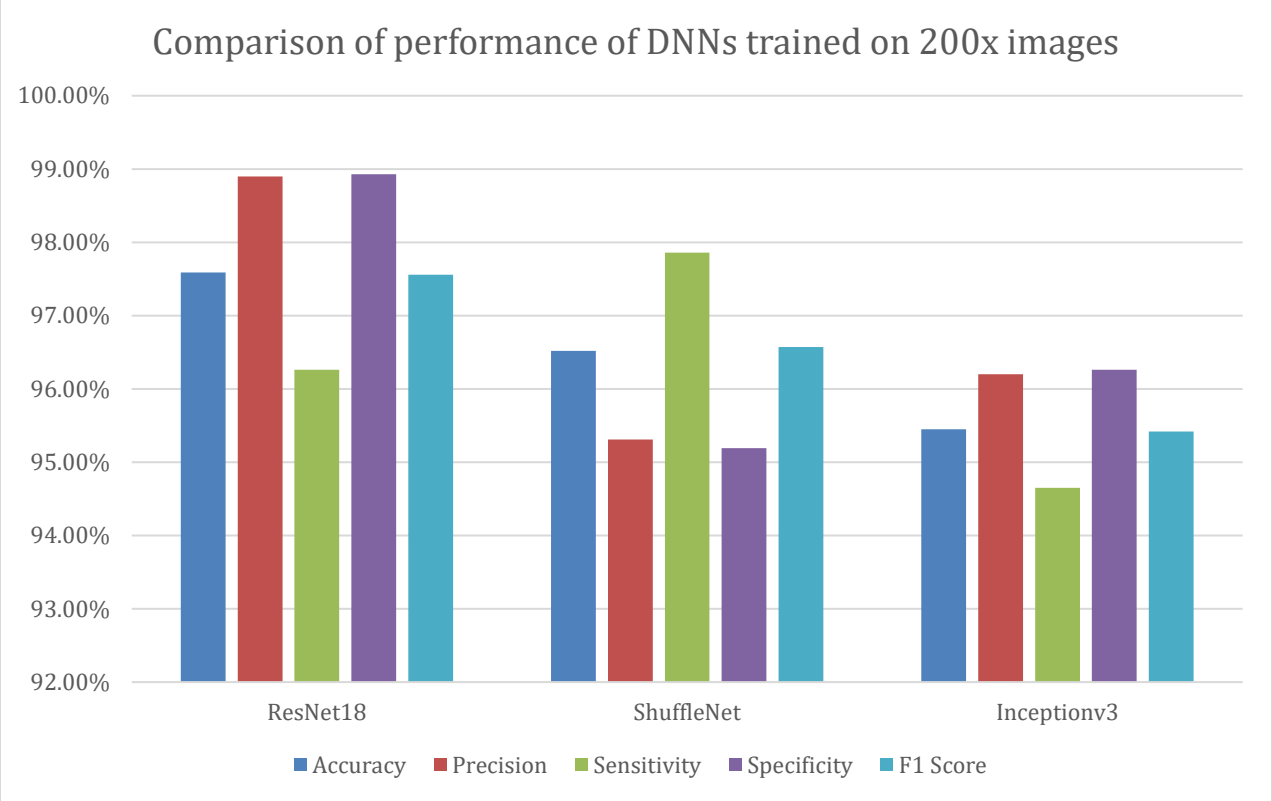


Figure 4.3: Performance stats for DNNs trained on 200x images.

Comparison of performance of DNNs trained on 400x images			
Evaluation metrics	**ResNet18**	**ShuffleNet**	**Inception-V3Net**
Accuracy	94.03%	**94.32%**	94.03%
Precision	93.79%	94.32%	**95.32%**
Sensitivity	**94.32%**	**94.32%**	92.61%
Specificity	93.75%	94.32%	**95.45%**
F1 Score	94.05%	**94.32%**	93.95%

Table 4.4: Performance stats for DNNs trained on 400x images.

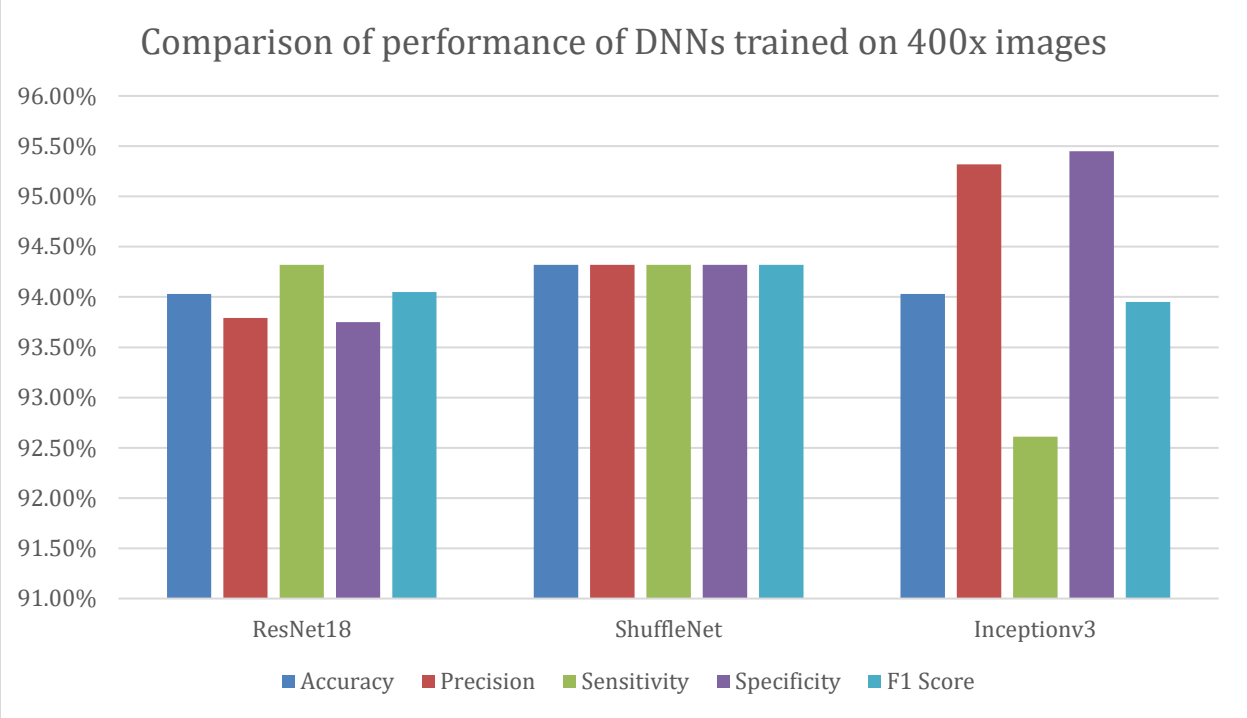


Figure 4.4: Performance stats for DNNs trained on 400x images.

4.1.1 Confusion Matrices

Figures show the confusion matrices for Inception-V3Net, ResNet18 and ShuffleNet DNNs with different magnifying factors of 40x, 100x, 200x and 400x.

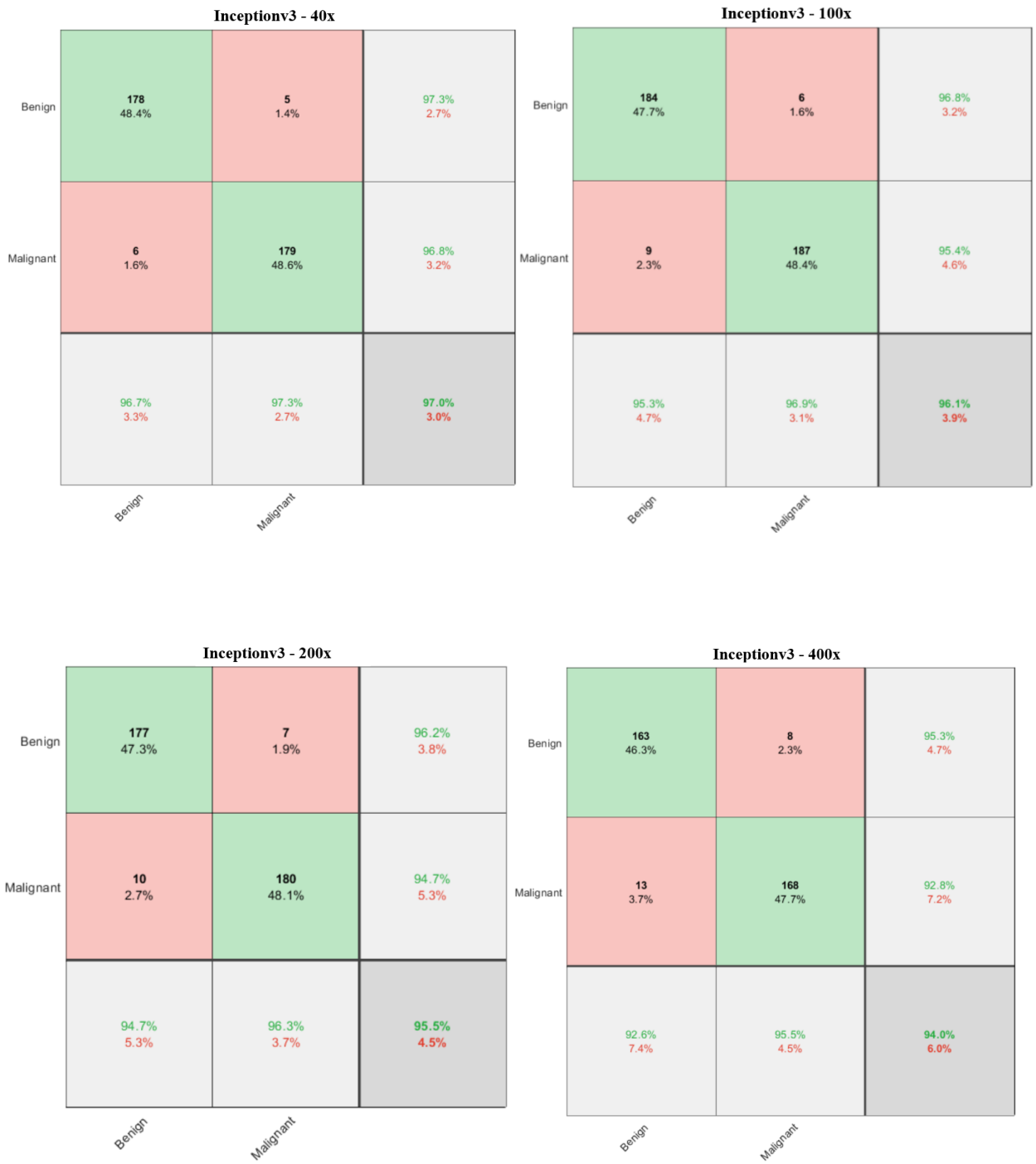


Figure 4.5: Confusion matrices for Inception-V3Net for different magnifications.

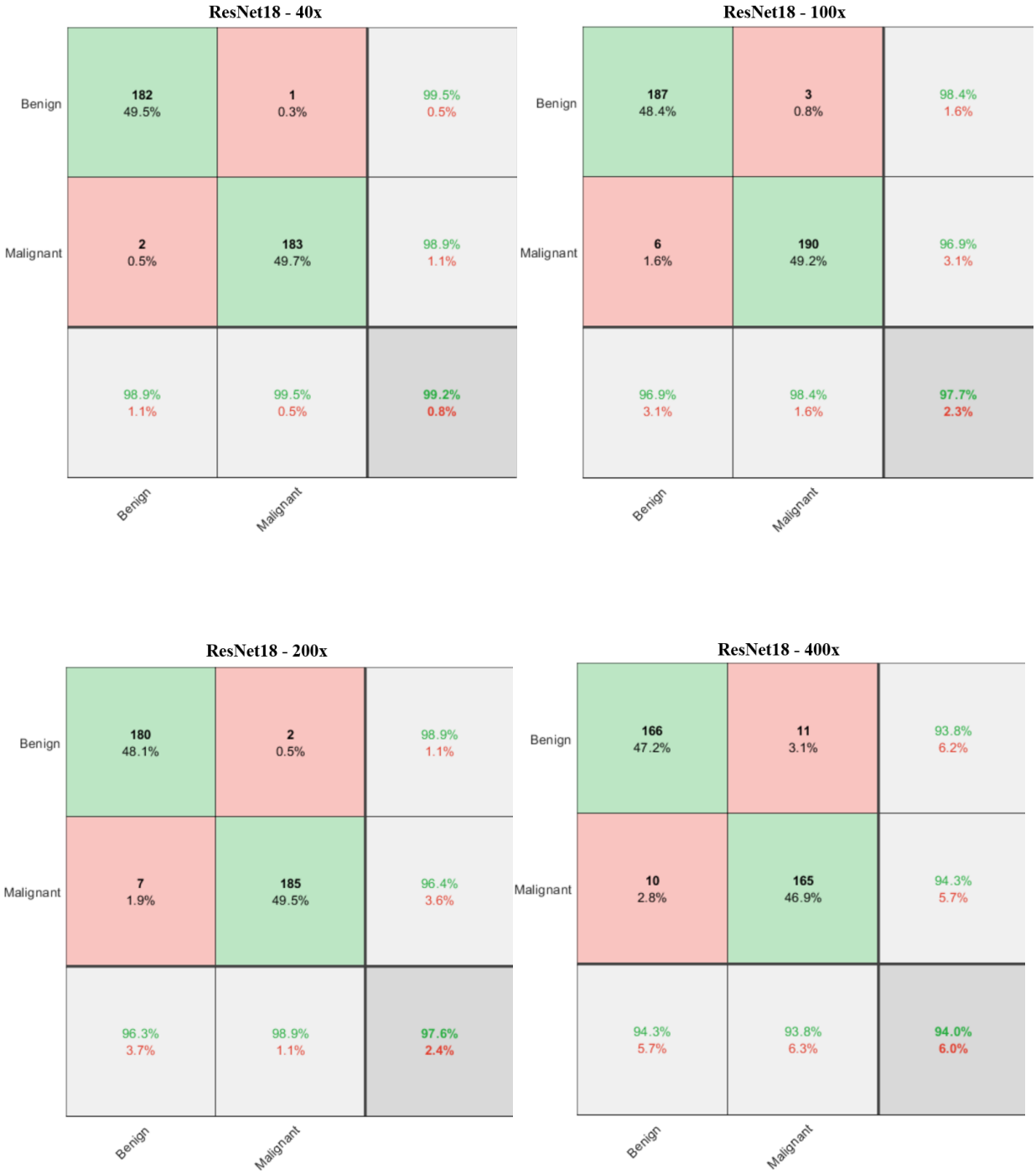


Figure 4.6: Confusion matrices for ResNet18 for different magnifications.

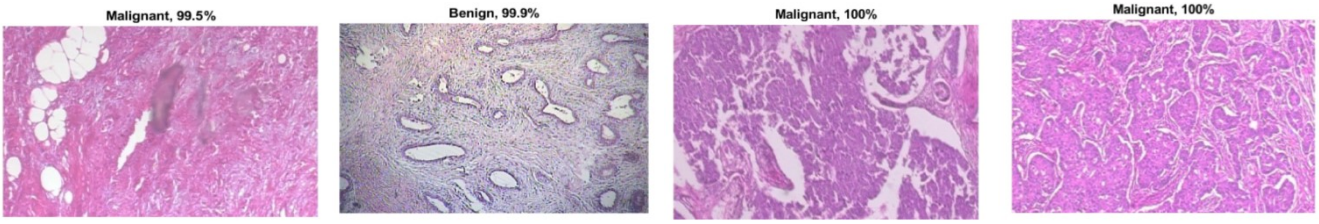
Figure 4.7: Confusion matrices for ShuffleNet for different magnifications.

4.1.2 Classification Performance on Testing Data

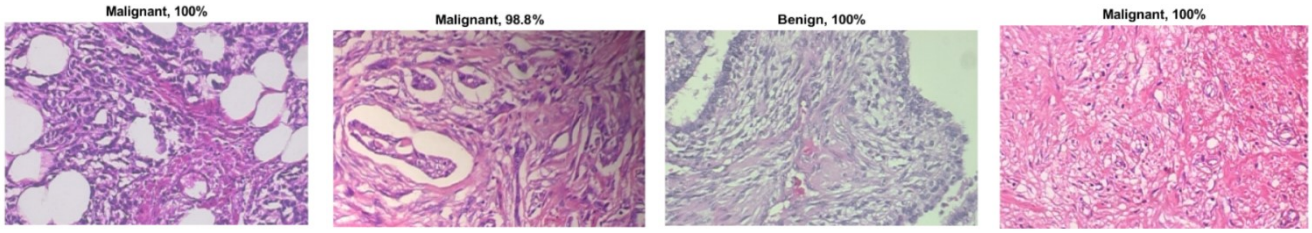
In this Section will display four sample testing images for different magnifying factors (40X, 100X, 200X, and 400X) with predicted labels and predicted probabilities for the three DNNs.

Inception-V3Net Deep Neural Networks:

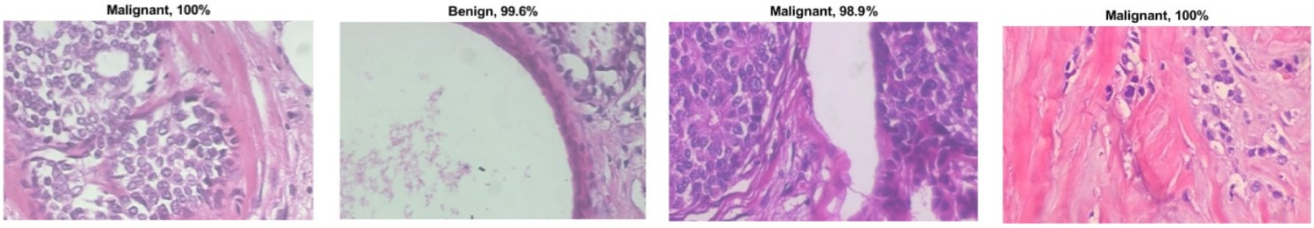
40x



100x



200x



400x

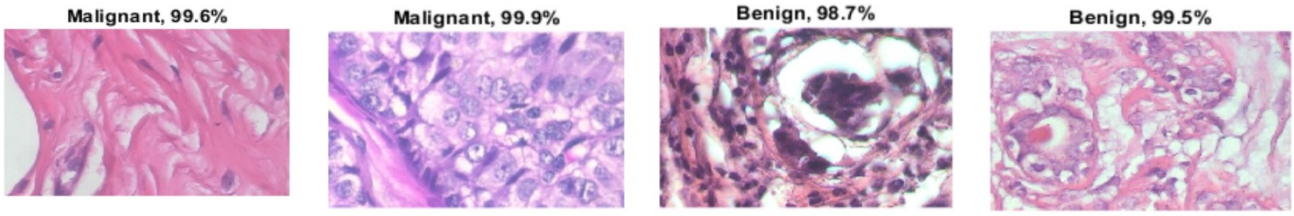


Figure 4.8 : Inception-V3 predictions on sample test images.

ShuffleNet Deep Neural Networks

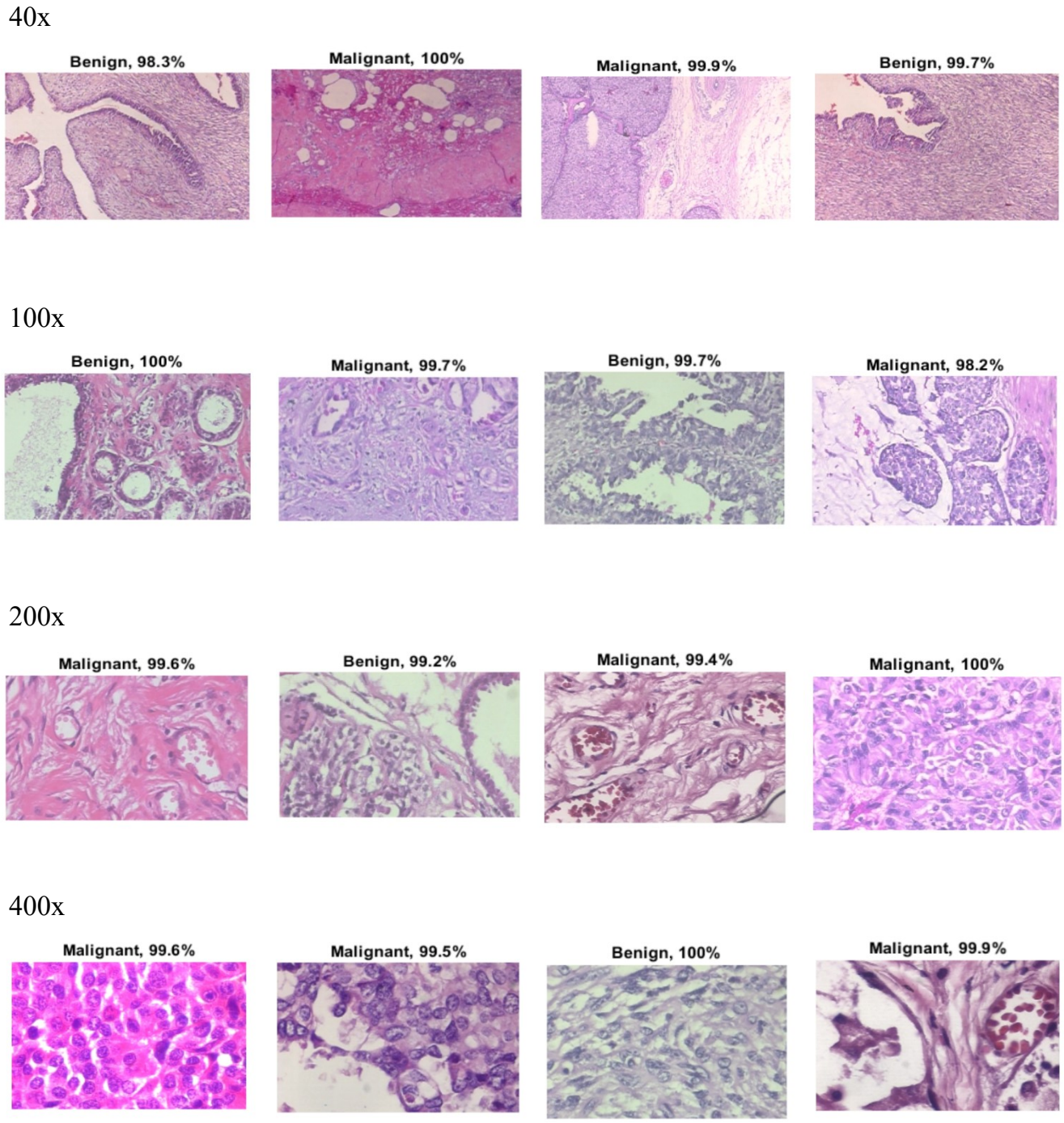


Figure 4.9: ShuffleNet predictions on sample test images.

ResNet18 Deep Neural Networks

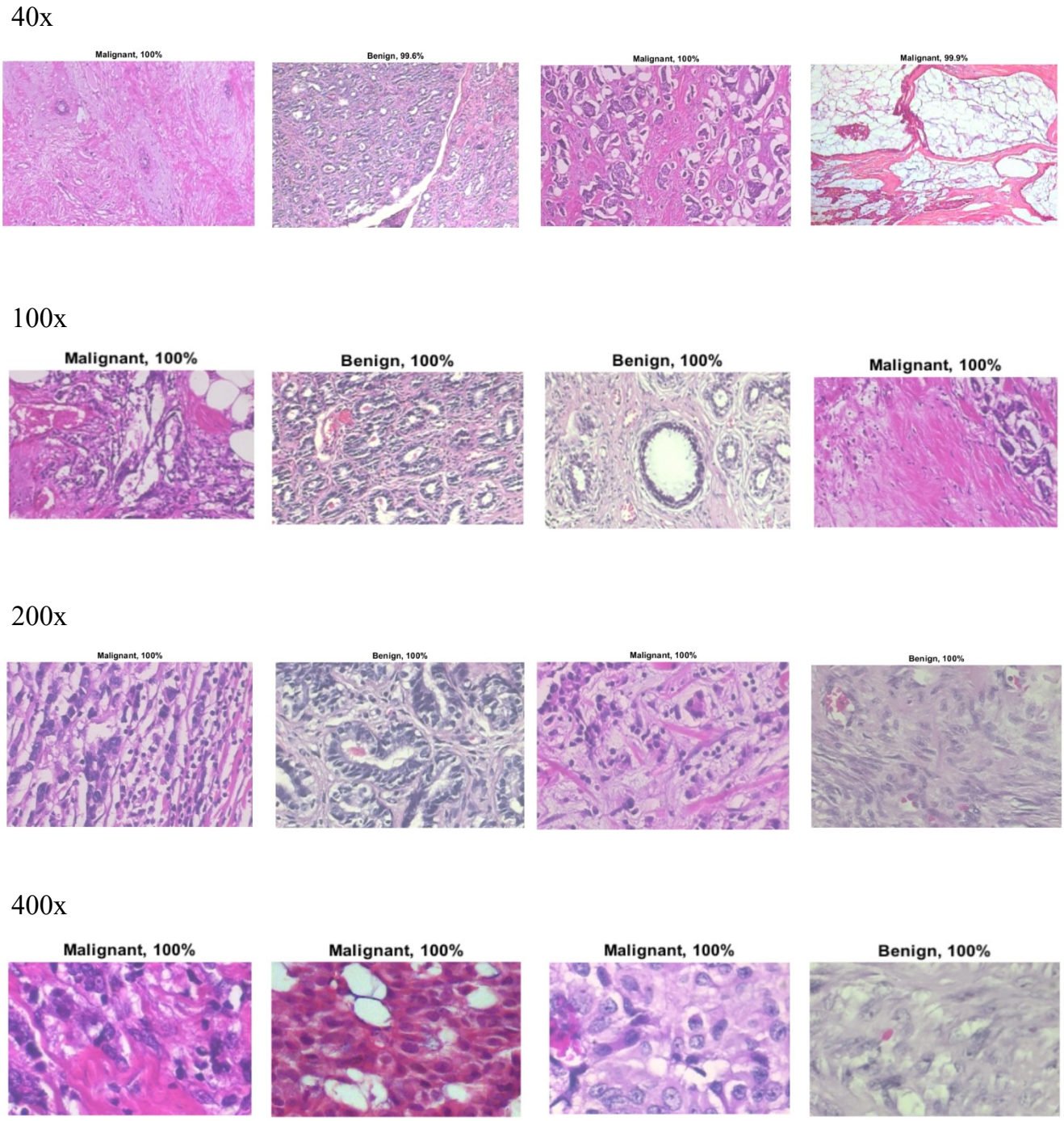


Figure 4.10: ResNet18 predictions on sample test images.

4.2 Results for Multiclass Classification

In this Section, we compare the performance of the three pre-trained deep neural network ResNet18, ShuffleNet, and Inception-V3Net on images with four different magnifying factors 40x, 100x, 200x, and 400x in the multiclass classification task. The following Tables and Figures describe the outcomes for the multiclass classification of the three pre-trained DNNs. Table 4.5 describes the overall accuracy for eight classes (Adenosis (A), Fibroadenoma (F), Phyllodes Tumor (PT) Tubular Adenoma (TA), Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC) and Papillary Carcinoma (PC)). The results demonstrate that ResNet18 achieved the highest accuracy rate on images with magnifying factors 40x, 100x, and 200x. On 400x data, Inception-V3Net and ShuffleNet obtained higher accuracy than ResNet18. In order to compare our results with [2] we show in Section 4.3.4 the confusion matrices for all DNNs.

Overall Accuracy for Multiclass Classification				
Neural Networks	40X	100X	200X	400X
Inception-V3Net	92.05%	92.28%	93.36%	**93.36%**
ResNet18	**95.08%**	**94.12%**	**94.53%**	92.97%
ShuffleNet	92.05%	92.65%	91.02%	**93.36%**

Table 4.5: Overall accuracy for multiclass classification.

Next, we compare the multiclass classification performance of Inception-V3Net, ResNet18, and ShuffleNet on images with different magnification factors using the following evaluation metrics: accuracy, precision, recall or sensitivity, and F1-score. In Section 4.3.1, Tables 4.6, 4.7, 4.8, 4.9, and Figures 4.11, 4.12, 4.13, 4.14, we compare the performance of Inception-V3Net for different classes and magnification factors. Similar results for ResNet18 are presented in Section 4.3.2 and for ShuffleNet in Section 4.3.3.

4.2.1 Inception-V3Net

Comparison of performance of Inception-V3Net trained on 40x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	99.24%	94%	**100%**	97%
Ductal Carcinoma	98.11%	**100%**	85%	92%
Fibroadenoma	**98.86%**	97%	94%	95%
Lobular Carcinoma	**96.59%**	85%	88%	87%
Mucinous Carcinoma	**96.97%**	86%	91%	88%
Papillary Carcinoma	**96.59%**	85%	88%	87%
Phyllodes Tumor	**99.24%**	97%	97%	97%
Tubular Adenoma	**98.48%**	94%	94%	94%
Overall Accuracy	**92.05%**			

Table 4.6: Performance stats for Inception-V3Net trained on 40x images.

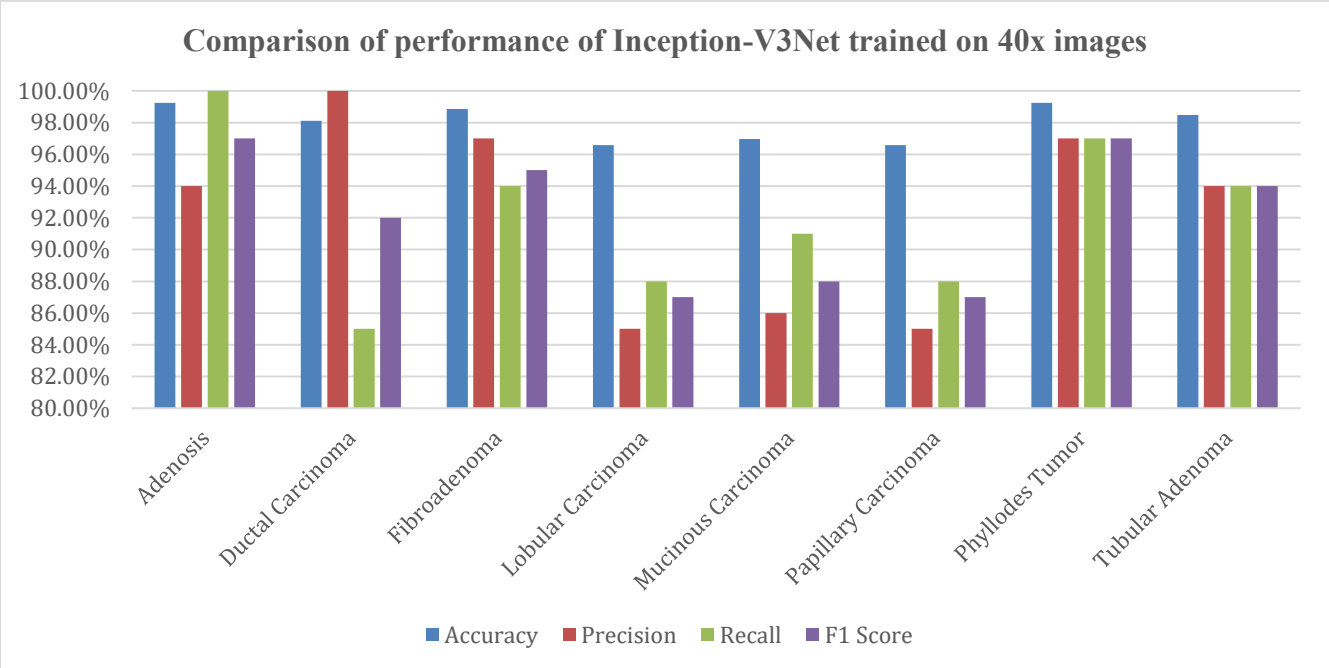


Figure 4.11: Performance stats for Inception-V3Net trained on 40x images.

Comparison of performance of Inception-V3Net trained on 100x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**99.26%**	97%	97%	97%
Ductal Carcinoma	**95.59%**	79%	88%	83%
Fibroadenoma	**98.53%**	92%	97%	94%
Lobular Carcinoma	**97.06%**	88%	88%	88%
Mucinous Carcinoma	**98.16%**	97%	88%	92%
Papillary Carcinoma	**97.79%**	94%	88%	91%
Phyllodes Tumor	**98.53%**	94%	94%	94%
Tubular Adenoma	99.63	**100%**	97%	99%
Overall Accuracy	**92.28%**			

Table 4.7: Performance stats for Inception-V3Net trained on 100x images.

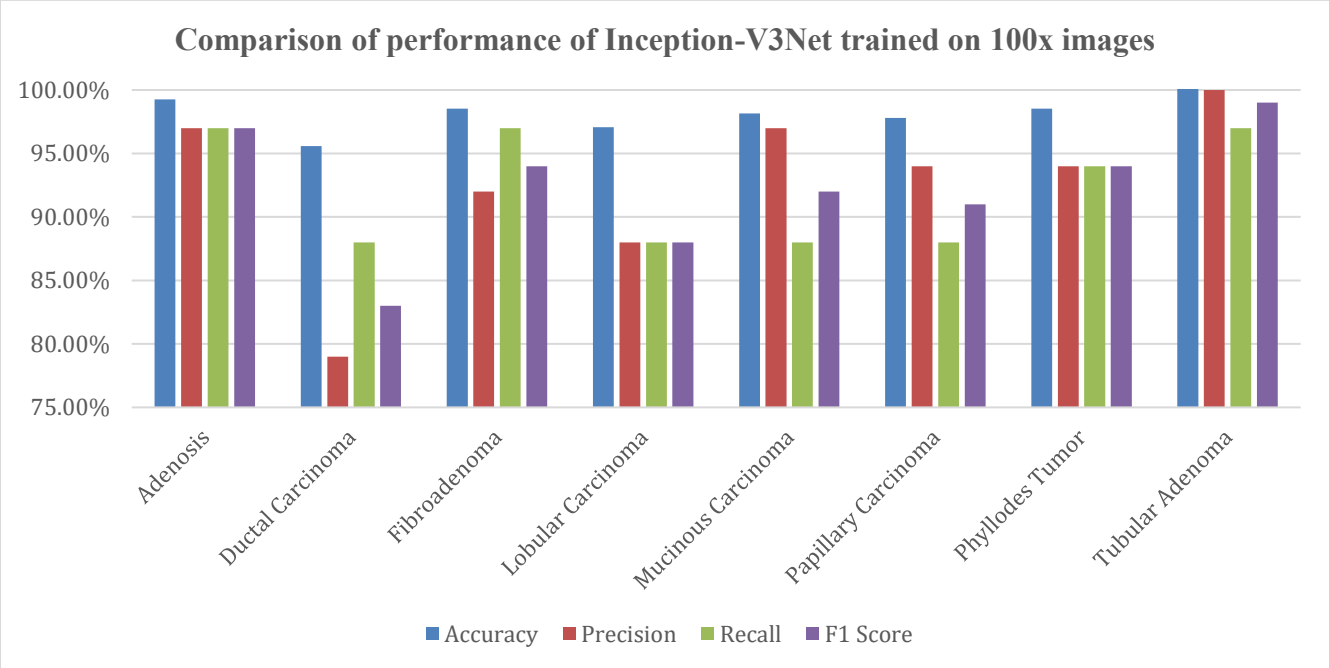


Figure 4.12: Performance stats for Inception-V3Net trained on 100x images.

Comparison of performance of Inception-V3Net trained on 200x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**99.22%**	97%	97%	97%
Ductal Carcinoma	**97.27%**	93%	84%	89%
Fibroadenoma	**98.05%**	94%	91%	92%
Lobular Carcinoma	**98.44%**	94%	94%	94%
Mucinous Carcinoma	**98.44%**	94%	94%	94%
Papillary Carcinoma	**98.44%**	97%	91%	94%
Phyllodes Tumor	**98.44%**	91%	97%	94%
Tubular Adenoma	98.44%	89%	**100%**	94%
Overall Accuracy	**93.36%**			

Table 4.8: Performance stats for Inception-V3Net trained on 200x images.

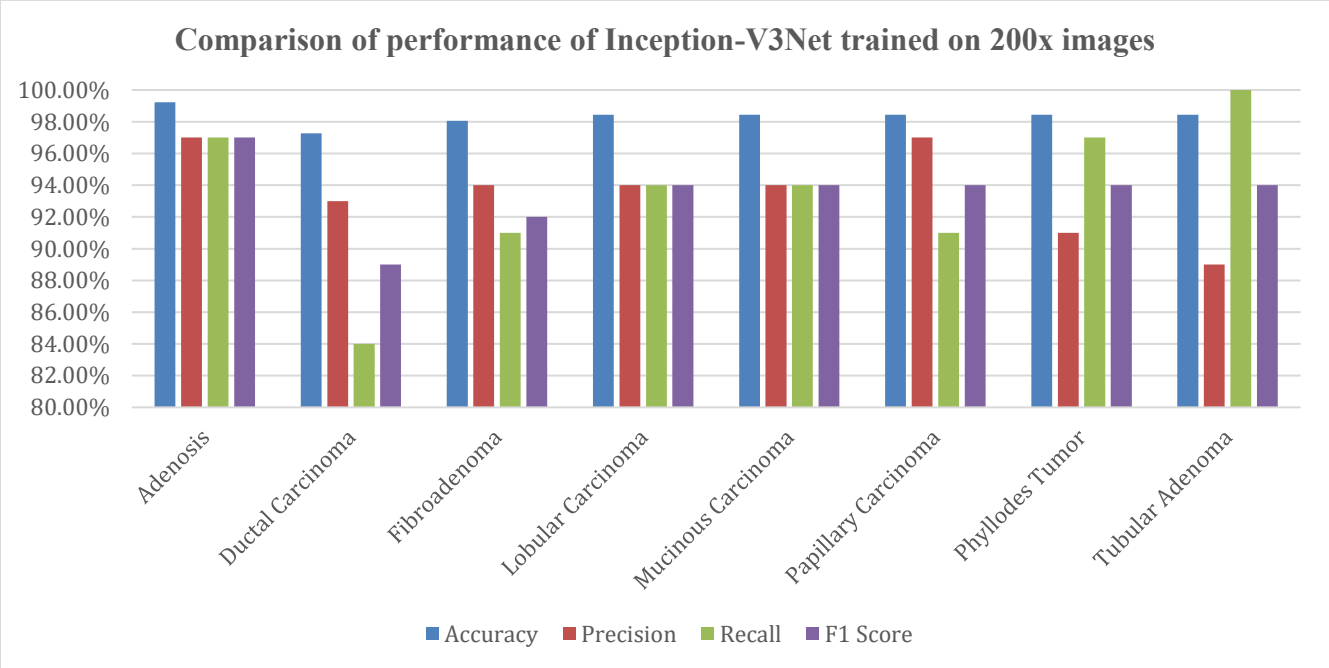


Figure 4.13: Performance stats for Inception-V3Net trained on 200x images.

Comparison of performance of Inception-V3Net trained on 400x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	99.22%	94%	100%	97%
Ductal Carcinoma	97.27%	88%	91%	89%
Fibroadenoma	98.05%	91%	94%	92%
Lobular Carcinoma	97.66%	91%	91%	91%
Mucinous Carcinoma	98.44%	97%	91%	94%
Papillary Carcinoma	98.44%	97%	91%	94%
Phyllodes Tumor	97.66%	91%	91%	91%
Tubular Adenoma	98.44%	94%	94%	94%
Overall Accuracy	**93.36%**			

Table 4.9: Performance stats for Inception-V3Net trained on 400x images.

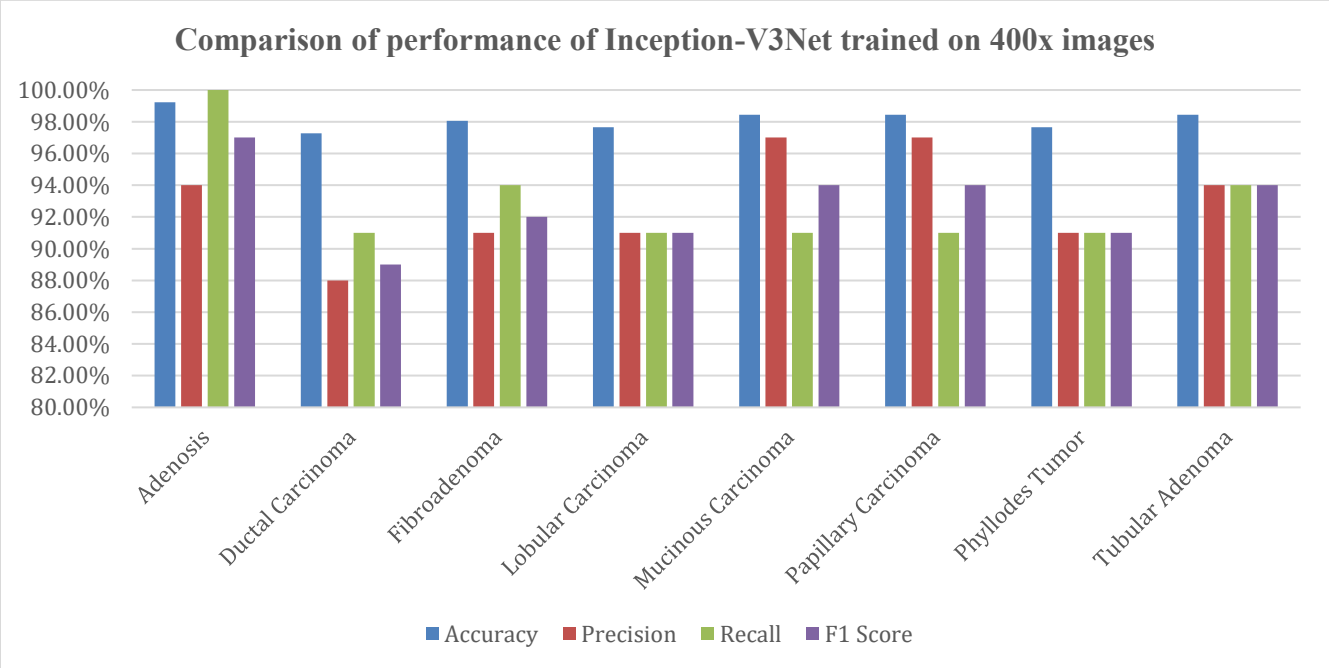


Figure 4.14: Performance stats for Inception-V3Net trained on 400x images.

4.2.2 ResNet18

Comparison of performance of ResNet18 trained on 40x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	99.24%	94%	**100%**	97%
Ductal Carcinoma	**97.35%**	86%	94%	90%
Fibroadenoma	**99.24%**	97%	97%	97%
Lobular Carcinoma	**98.11%**	94%	91%	92%
Mucinous Carcinoma	**98.48%**	97%	91%	94%
Papillary Carcinoma	**99.24%**	97%	97%	97%
Phyllodes Tumor	99.24%	**100%**	94%	97%
Tubular Adenoma	**99.24%**	97%	97%	97%
Overall Accuracy	**95.08%**			

Table 4.10: Performance stats for ResNet18 trained on 40x images.

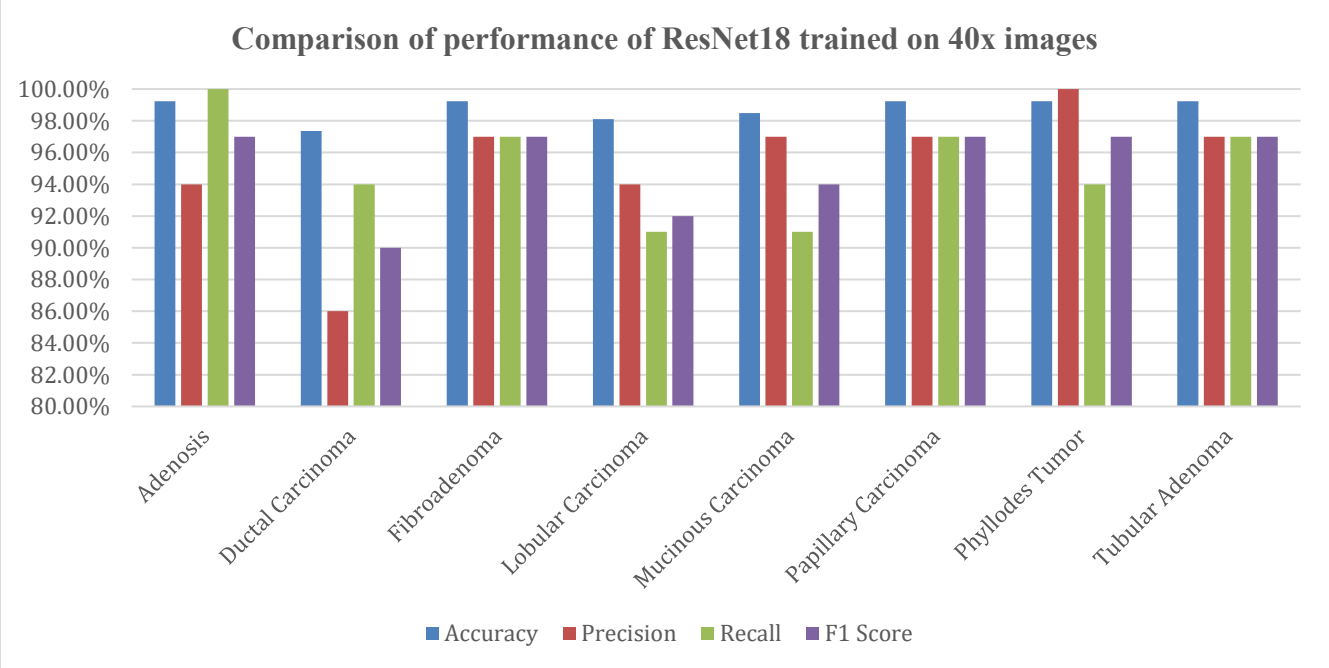


Figure 4.15: Performance stats for ResNet18 trained on 40x images.

Comparison of performance of ResNet18 trained on 100x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**99.26%**	97%	97%	97%
Ductal Carcinoma	**98.16%**	94%	91%	93%
Fibroadenoma	**98.53%**	94%	94%	94%
Lobular Carcinoma	**98.53%**	94%	94%	94%
Mucinous Carcinoma	**98.53%**	94%	94%	94%
Papillary Carcinoma	**98.53%**	92%	97%	94%
Phyllodes Tumor	**98.16%**	94%	91%	93%
Tubular Adenoma	**98.53%**	94%	94%	94%
Overall Accuracy	**94.12%**			

Table 4.11: Performance stats for ResNet18 trained on 100x images.

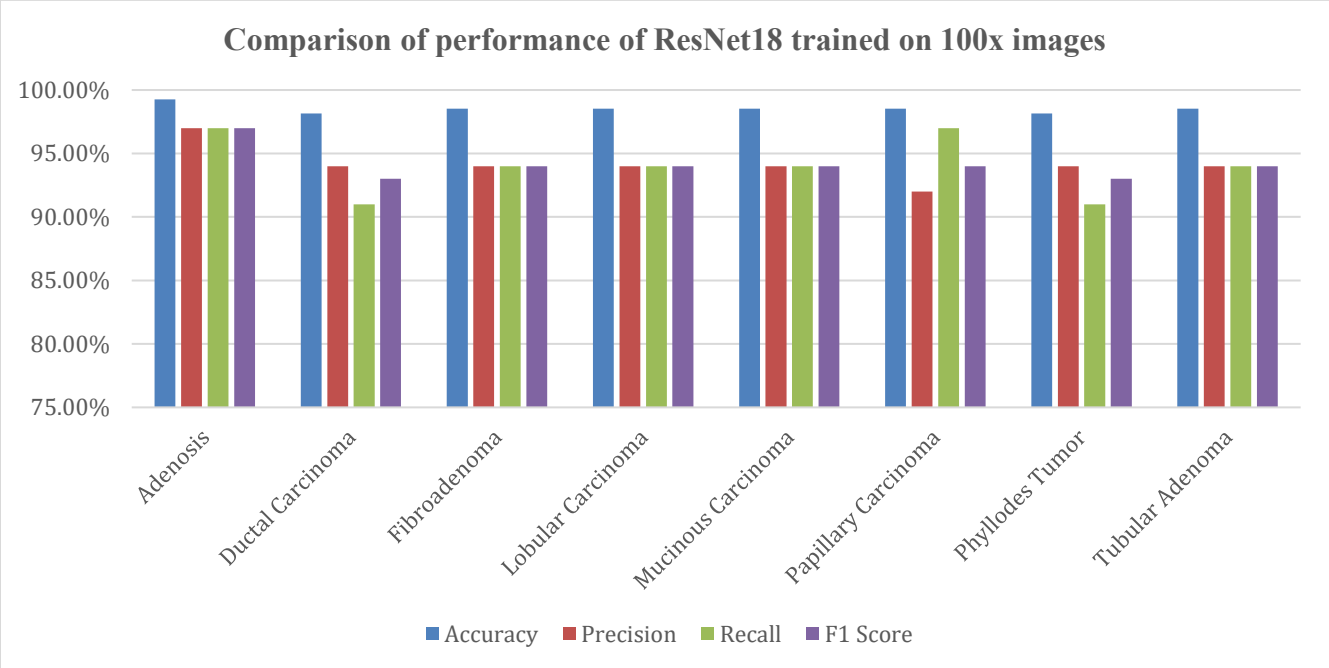


Figure 4.16: Performance stats for ResNet18 trained on 100x images.

Comparison of performance of ResNet18 trained on 200x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**98.83%**	94%	97%	95%
Ductal Carcinoma	**98.83%**	97%	94%	95%
Fibroadenoma	**98.44%**	94%	94%	94%
Lobular Carcinoma	**98.05%**	94%	91%	92%
Mucinous Carcinoma	**97.66%**	91%	91%	91%
Papillary Carcinoma	**98.83%**	94%	97%	95%
Phyllodes Tumor	**98.83%**	94%	97%	95%
Tubular Adenoma	99.61%	**100%**	97%	98%
Overall Accuracy	**94.53%**			

Table 4.12: Performance stats for ResNet18 trained on 200x images.

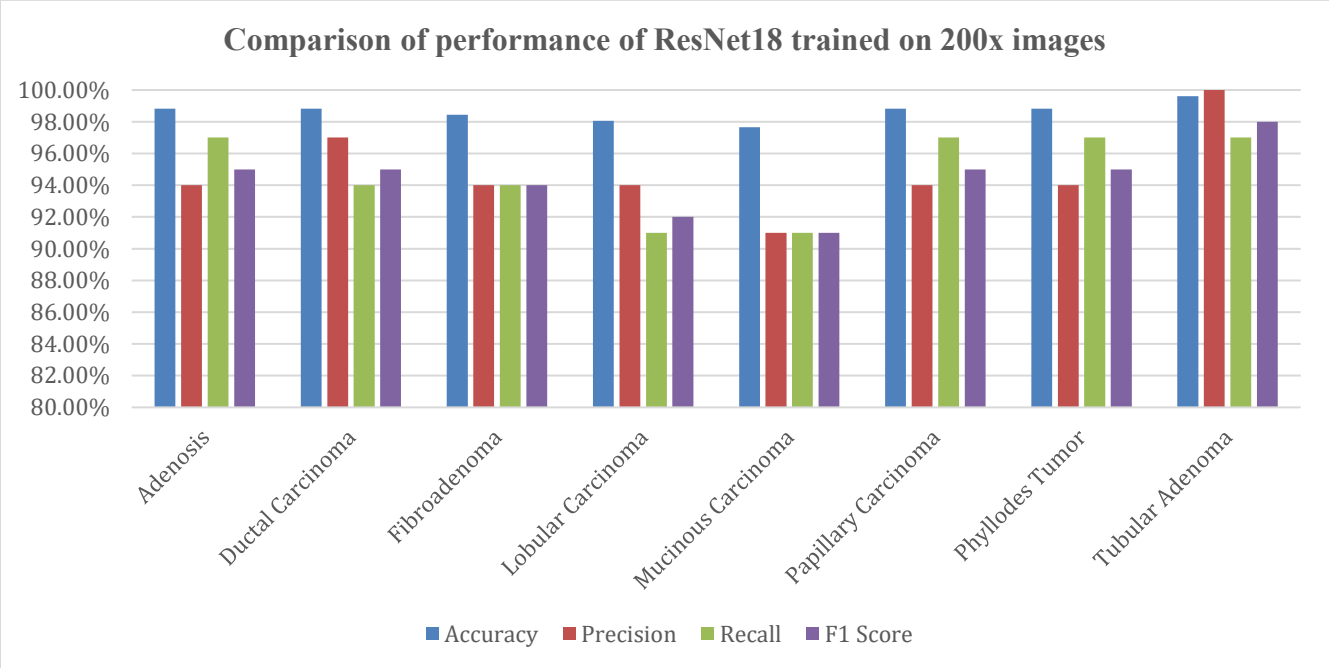


Figure 4.17: Performance stats for ResNet18 trained on 200x images.

Comparison of performance of ResNet18 trained on 400x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**98.83%**	94%	97%	95%
Ductal Carcinoma	**97.66%**	91%	91%	91%
Fibroadenoma	**97.66%**	91%	91%	91%
Lobular Carcinoma	**98.33%**	97%	94%	95%
Mucinous Carcinoma	**98.05%**	94%	91%	92%
Papillary Carcinoma	**98.44%**	91%	97%	94%
Phyllodes Tumor	**98.05%**	94%	91%	92%
Tubular Adenoma	**98.44%**	94%	94%	94%
Overall Accuracy	**92.97%**			

Table 4.13: Performance stats for ResNet18 trained on 400x images.

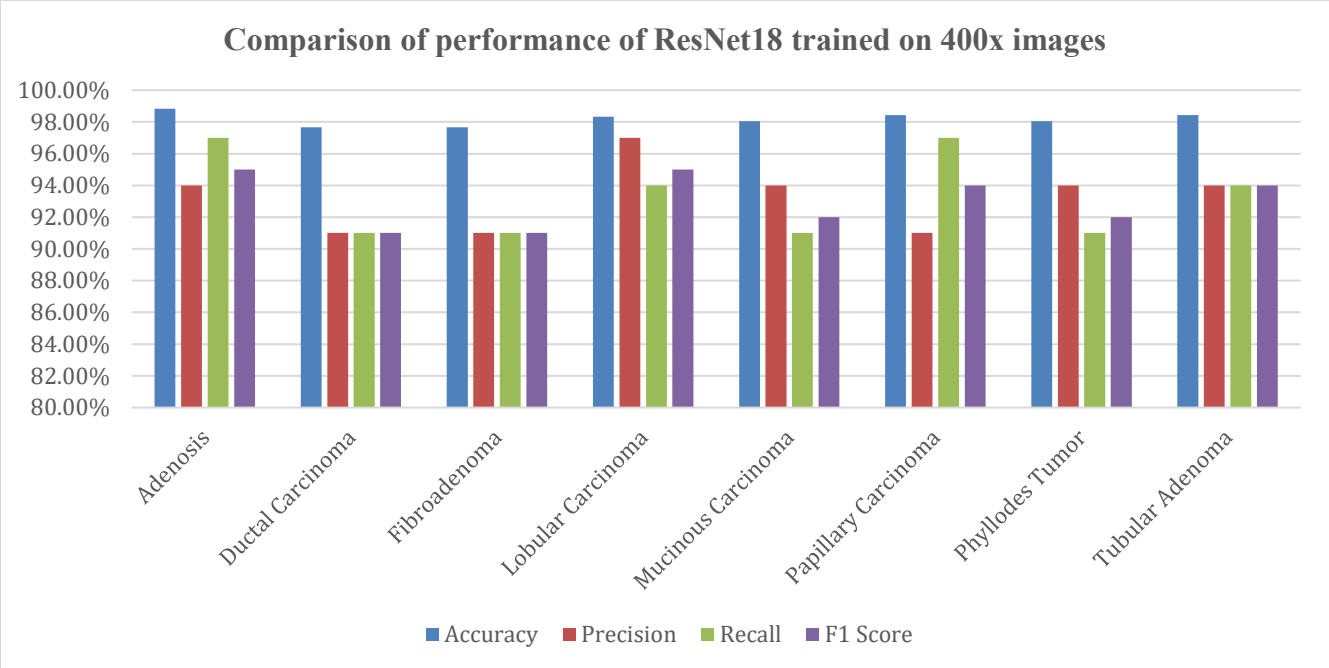


Figure 4.18: Performance stats for ResNet18 trained on 400x images.

4.2.3 ShuffleNet

Comparison of performance of ShuffleNet trained on 40x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**99.24%**	97%	97%	97%
Ductal Carcinoma	**98.48%**	97%	91%	94%
Fibroadenoma	**98.11%**	91%	94%	93%
Lobular Carcinoma	**98.11%**	97%	88%	92%
Mucinous Carcinoma	**96.97%**	88%	88%	88%
Papillary Carcinoma	**97.73%**	91%	91%	91%
Phyllodes Tumor	**97.73%**	89%	94%	91%
Tubular Adenoma	**97.73%**	89%	94%	91%
Overall Accuracy	**92.05%**			

Table 4.14: Performance stats for ShuffleNet trained on 40x images.

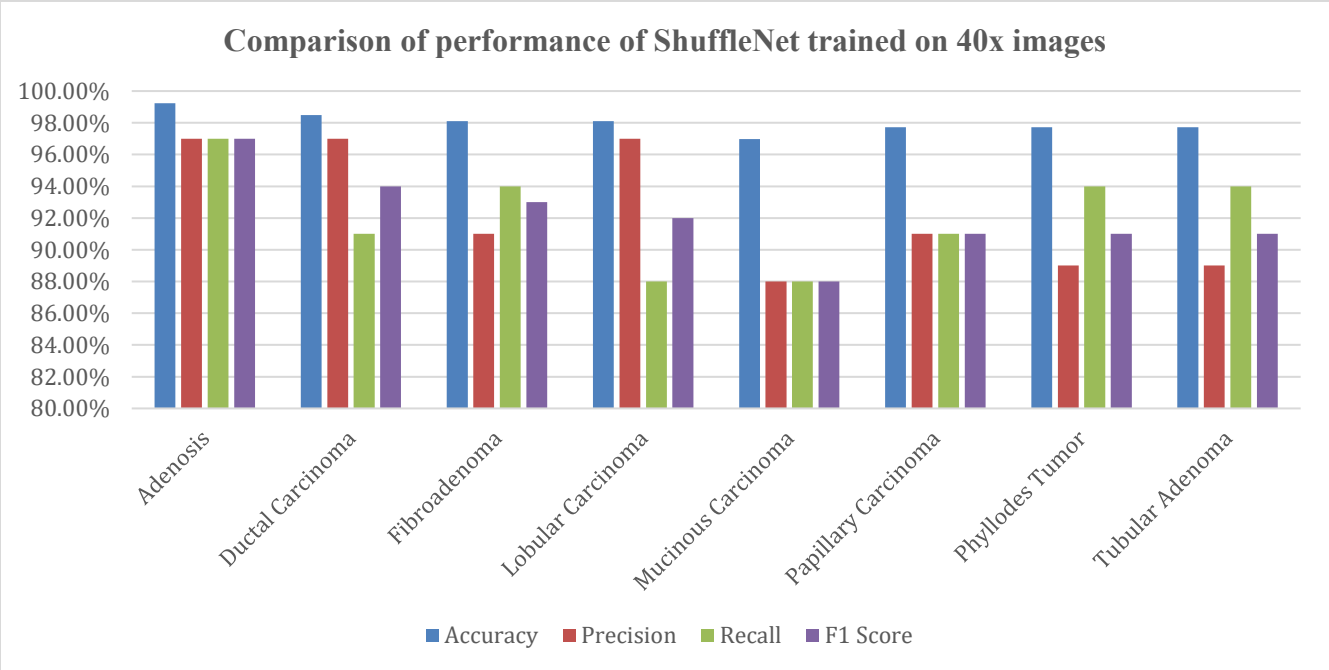


Figure 4.19: Performance stats for ShuffleNet trained on 40x images.

Comparison of performance of ShuffleNet trained on 100x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**98.9%**	94%	97%	96%
Ductal Carcinoma	**98.16%**	94%	91%	93%
Fibroadenoma	**98.53%**	92%	97%	94%
Lobular Carcinoma	**98.16%**	94%	91%	93%
Mucinous Carcinoma	98.53%	**100%**	88%	94%
Papillary Carcinoma	**96.69%**	84%	91%	87%
Phyllodes Tumor	**98.16%**	94%	91%	93%
Tubular Adenoma	**98.16%**	91%	94%	93%
Overall Accuracy	**92.65%**			

Table 4.15: Performance stats for ShuffleNet trained on 100x images.

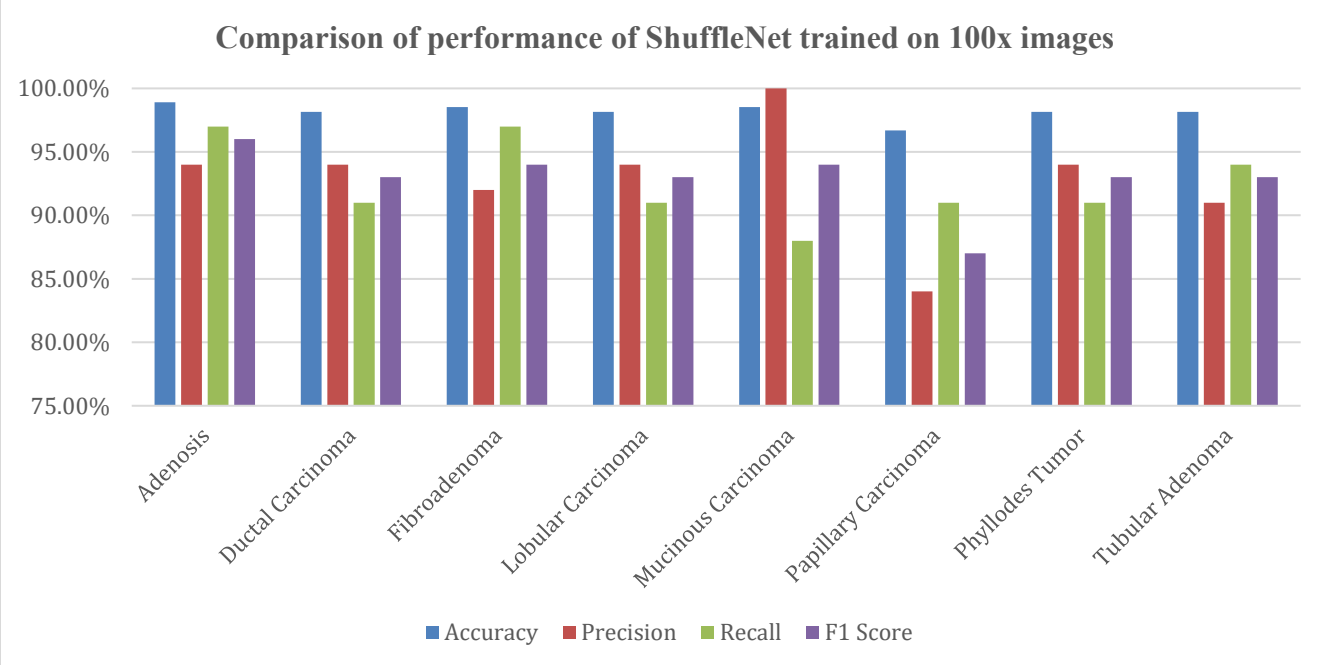


Figure 4.20: Performance stats for ShuffleNet trained on 100x images.

Comparison of performance of ShuffleNet trained on 200x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**98.83%**	97%	94%	95%
Ductal Carcinoma	**97.27%**	90%	88%	89%
Fibroadenoma	**98.44%**	94%	94%	94%
Lobular Carcinoma	**97.27%**	90%	88%	89%
Mucinous Carcinoma	**97.27%**	88%	91%	89%
Papillary Carcinoma	**97.27%**	86%	94%	90%
Phyllodes Tumor	98.83%	**100%**	91%	95%
Tubular Adenoma	**96.88%**	85%	91%	88%
Overall Accuracy	**91.02%**			

Table 4.16: Performance stats for ShuffleNet trained on 200x images.

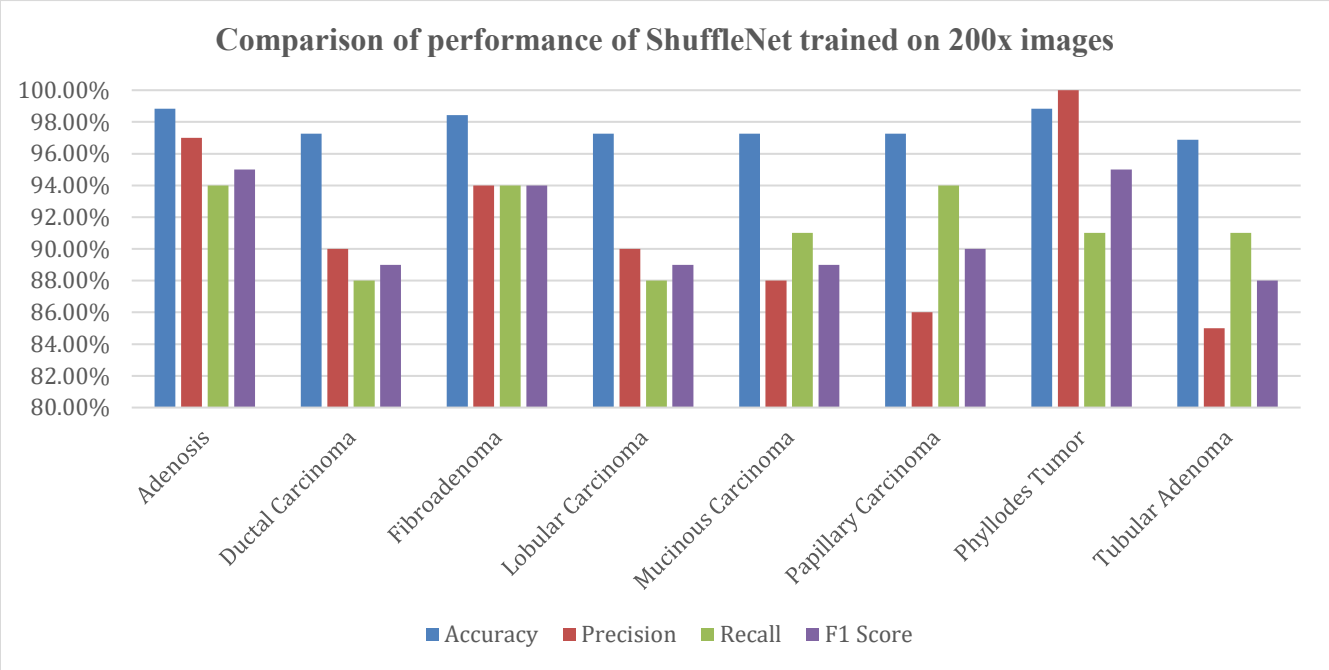


Figure 4.21: Performance stats for ShuffleNet trained on 200x images.

Comparison of performance of ShuffleNet trained on 400x images				
Class Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Adenosis	**98.44%**	91%	97%	94%
Ductal Carcinoma	**97.27%**	90%	88%	89%
Fibroadenoma	**98.05%**	89%	97%	93%
Lobular Carcinoma	99.22%	**100%**	94%	97%
Mucinous Carcinoma	**98.05%**	94%	91%	92%
Papillary Carcinoma	**98.44%**	94%	94%	94%
Phyllodes Tumor	**98.05%**	94%	91%	92%
Tubular Adenoma	**96.88%**	88%	88%	88%
Overall Accuracy	**93.36%**			

Table 4.17: Performance stats for ShuffleNet trained on 400x images.

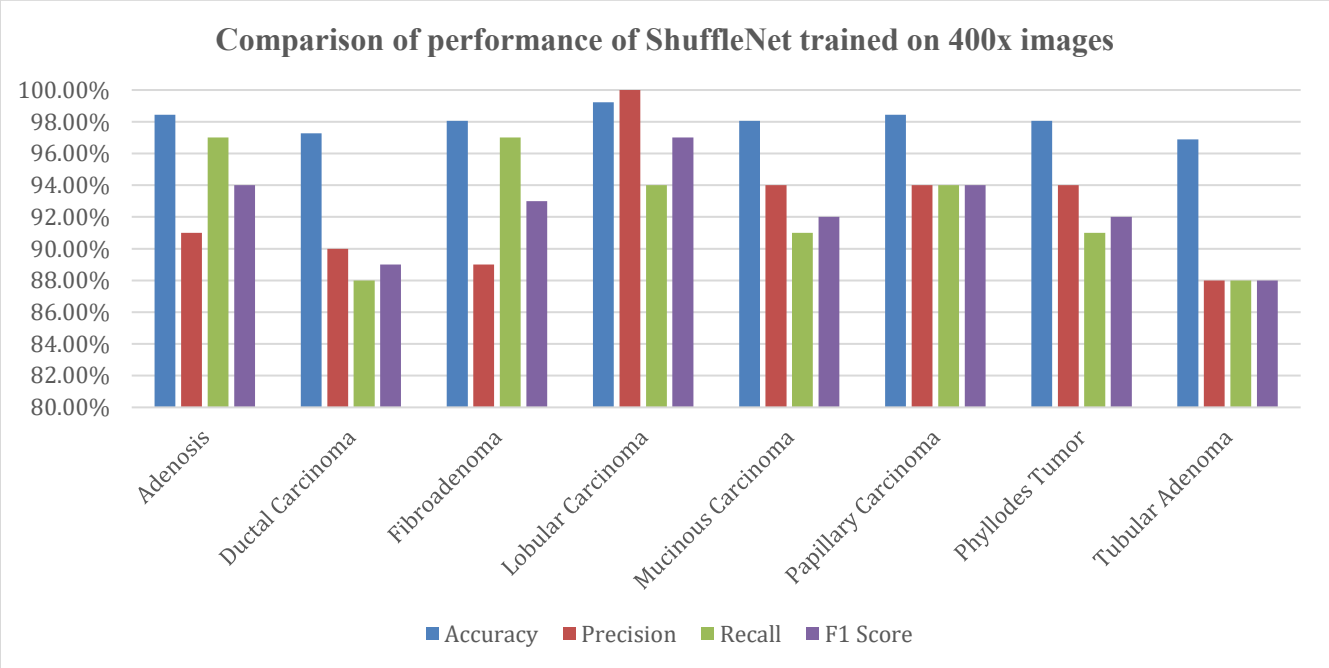


Figure 4.22: Performance stats for ShuffleNet trained on 400x images.

4.2.4 Confusion Matrices

Next, we present confusion matrices for InceptationV3Net, ResNet18, and ShuffleNet for images with different magnifying factors 40x, 100x, 200x, and 400x.

Inception-V3Net for 40x

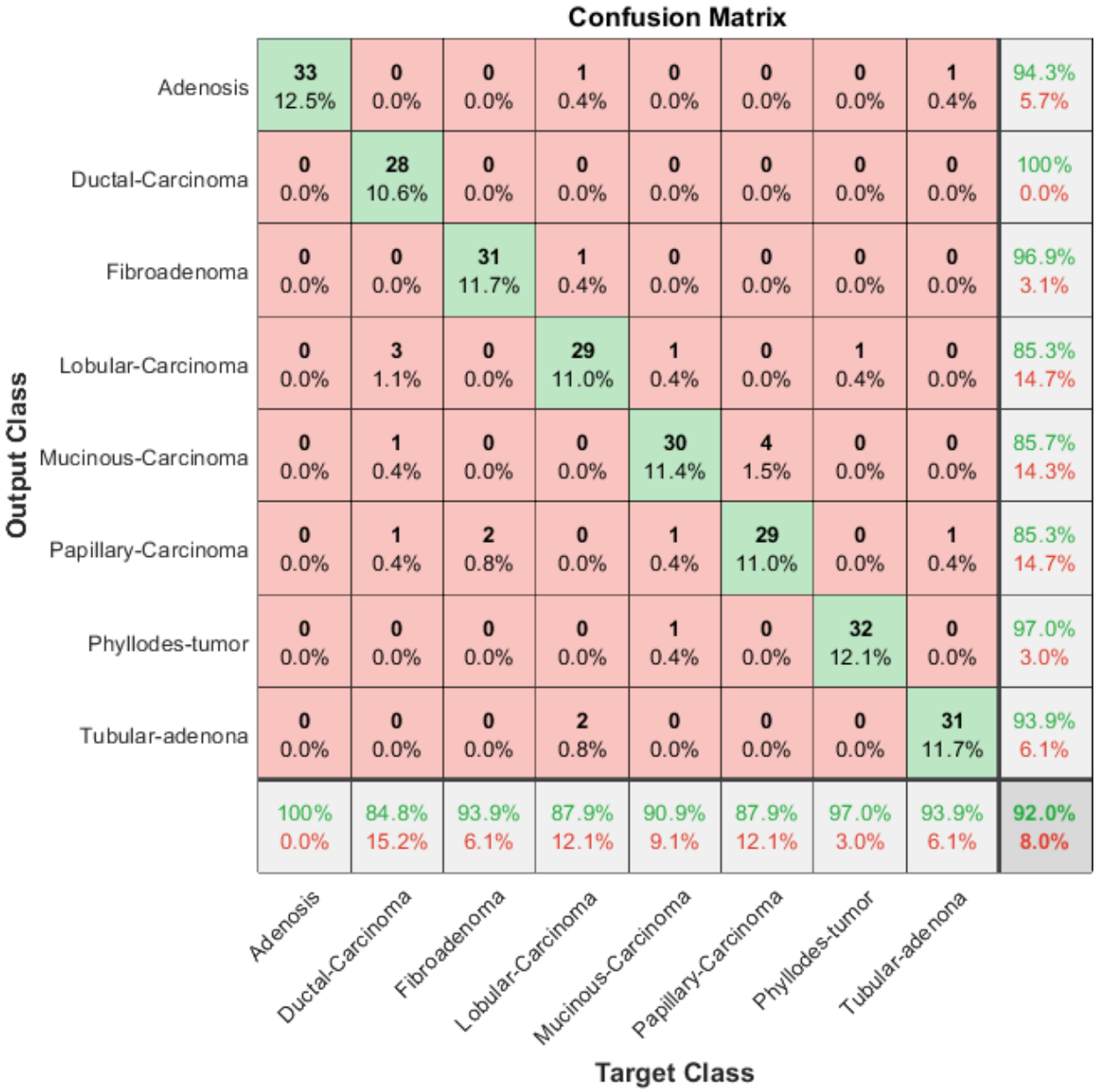


Figure 4.23: The confusion matrices for Inception-V3Net for 40x.

Inception-V3Net for 100x

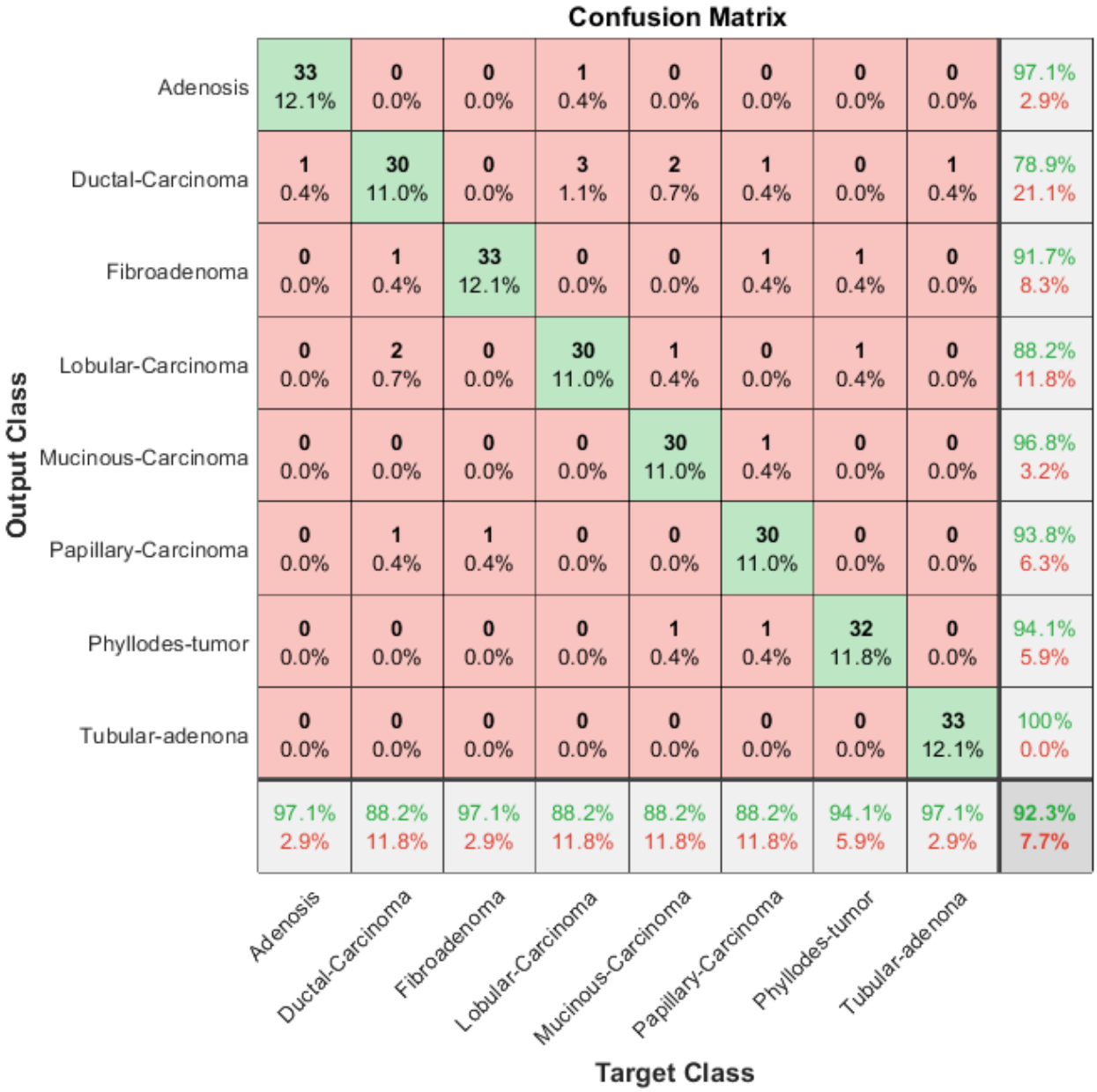


Figure 4.24: Confusion matrices for Inception-V3Net trained on 100x images.

Inception-V3Net for 200x

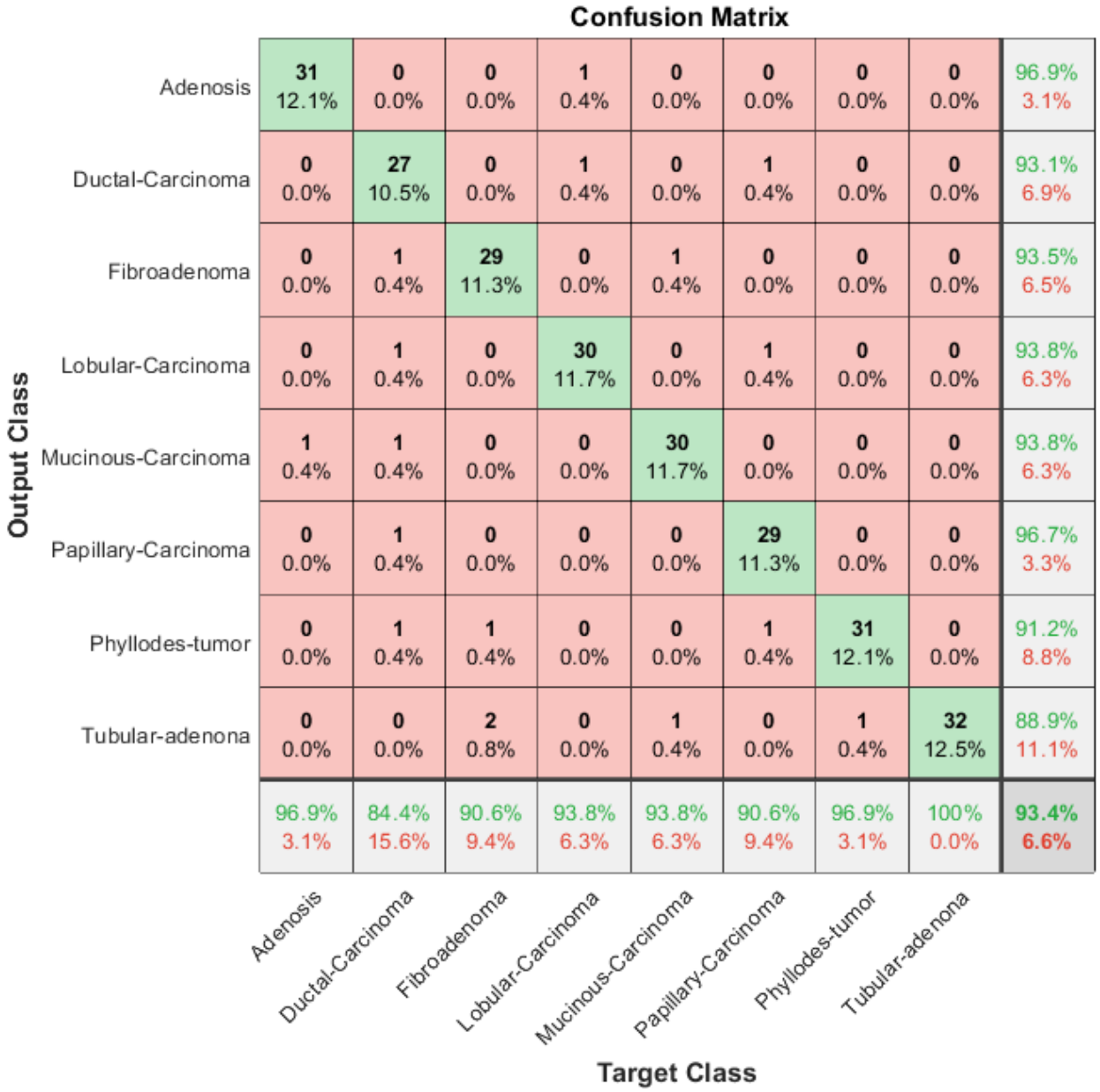


Figure 4.25: Confusion matrices for Inception-V3Net trained on 200x images.

Inception-V3Net for 400x

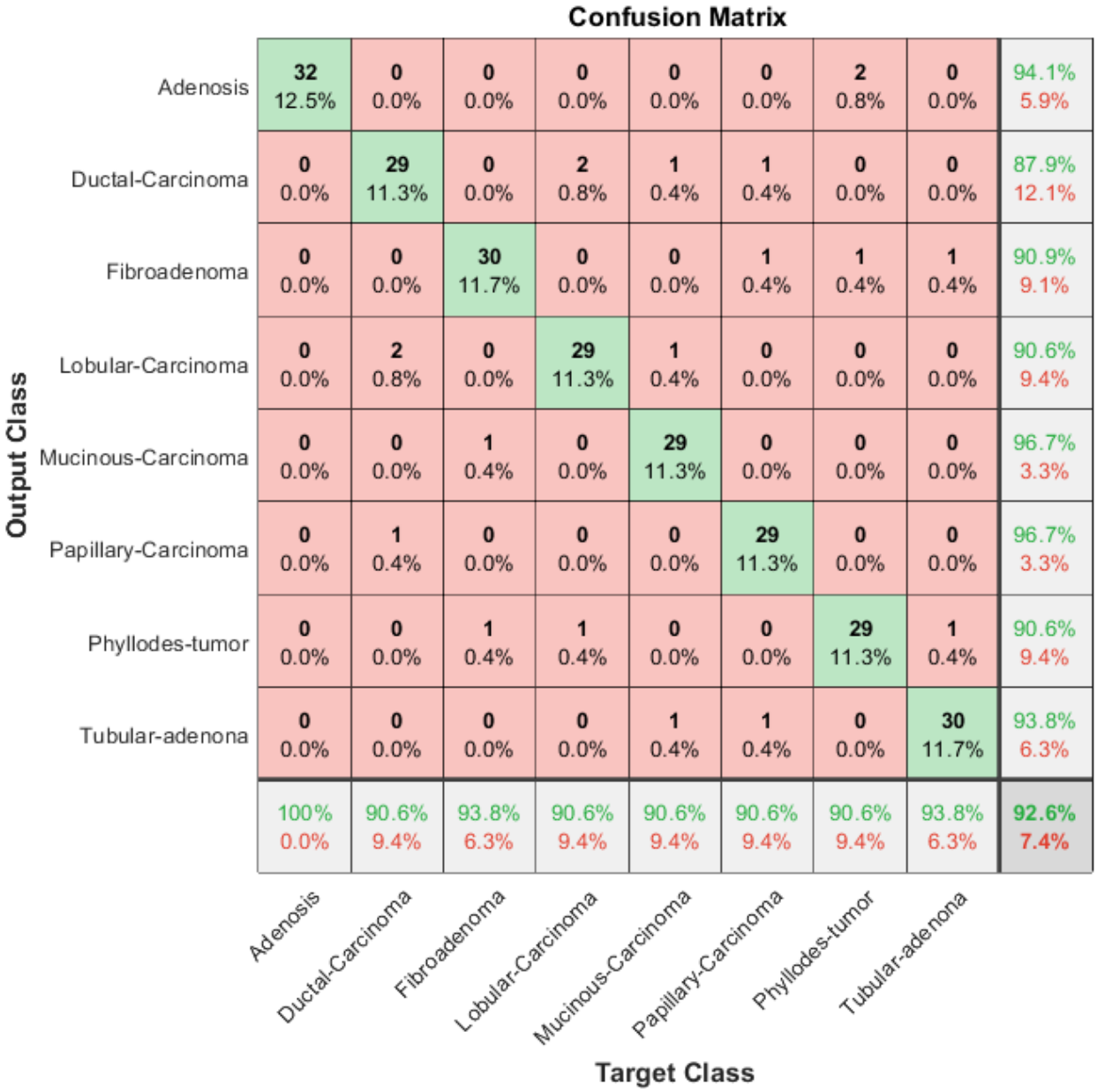


Figure 4.26: Confusion matrices for Inception-V3Net trained on 400x images.

ResNet18 for 40x

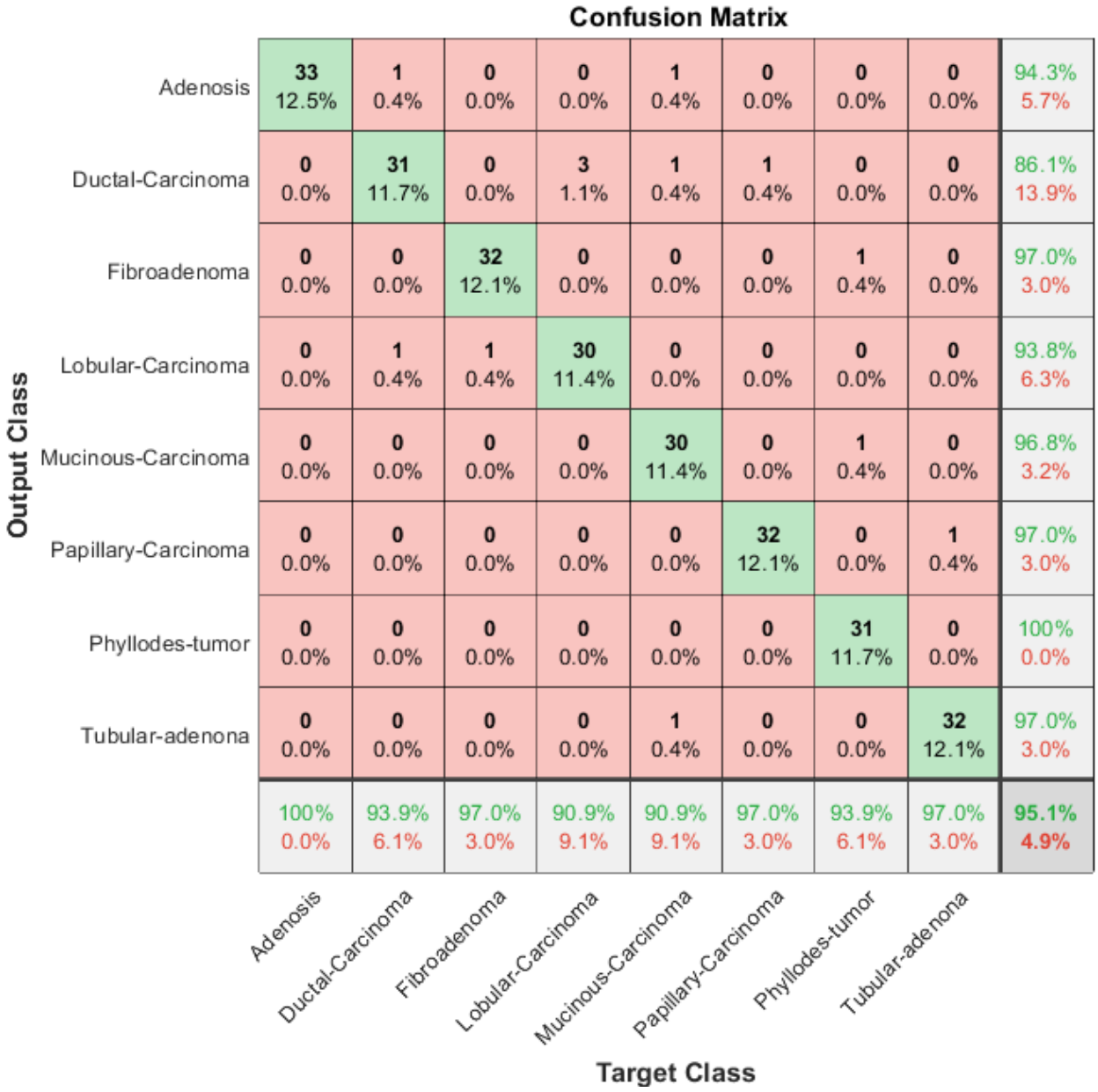


Figure 4.27: Confusion matrices for ResNet18 trained on 40x images.

ResNet18 for 100x

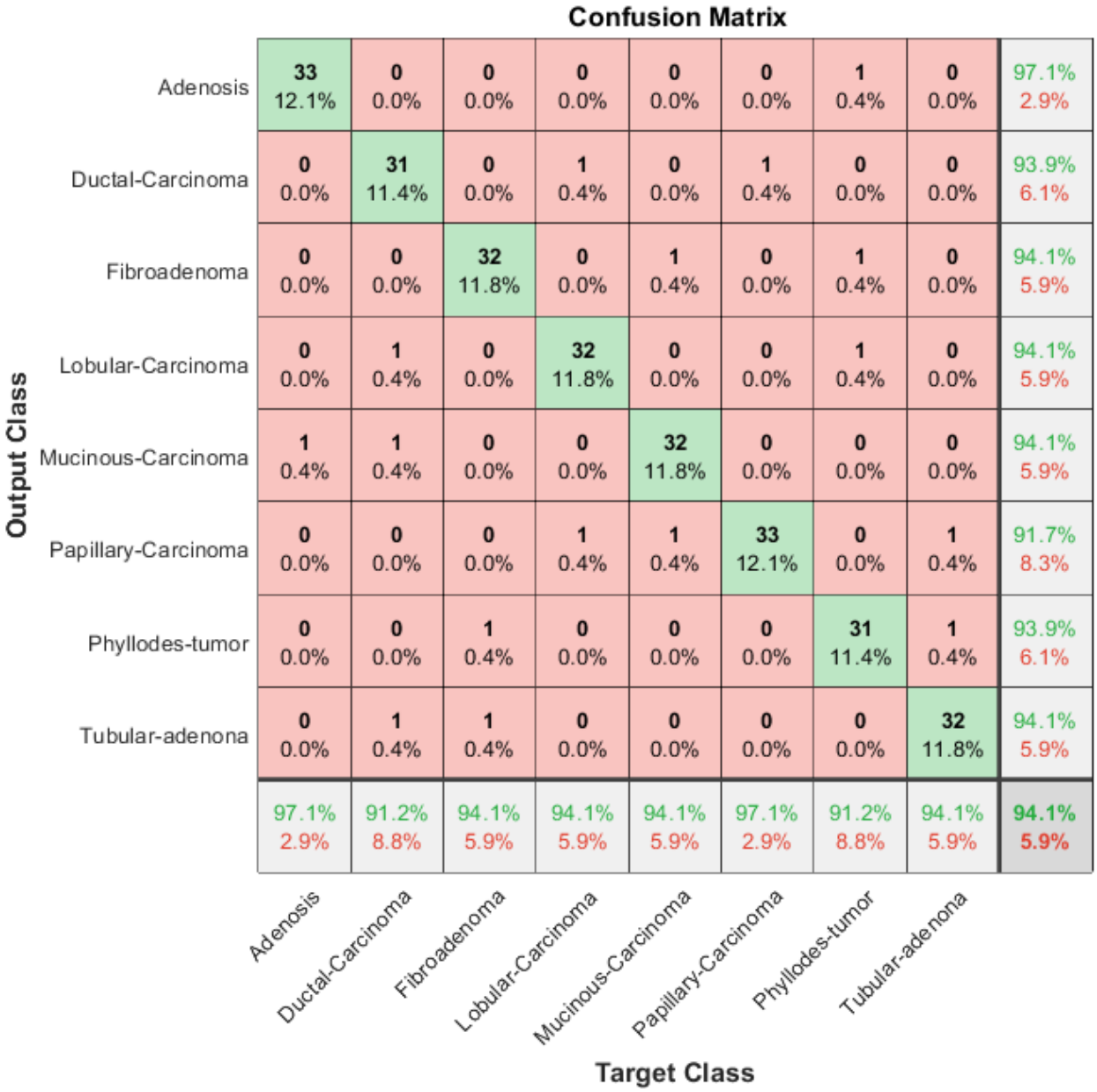


Figure 4.28: Confusion matrices for ResNet18 trained on 100x images.

ResNet18 for 200x

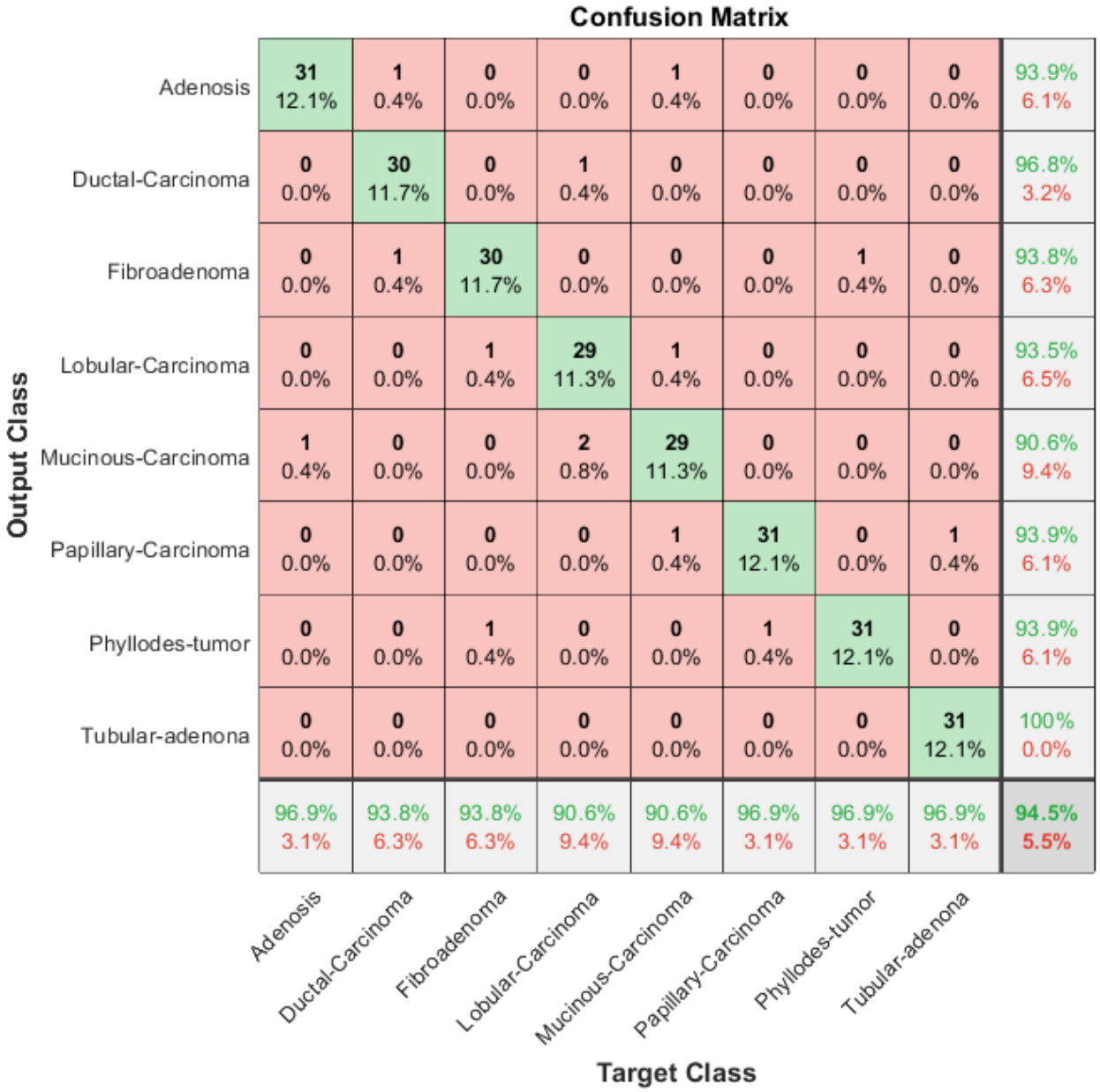


Figure 4.29: Confusion matrices for ResNet18 trained on 200x images.

ResNet18 for 400x

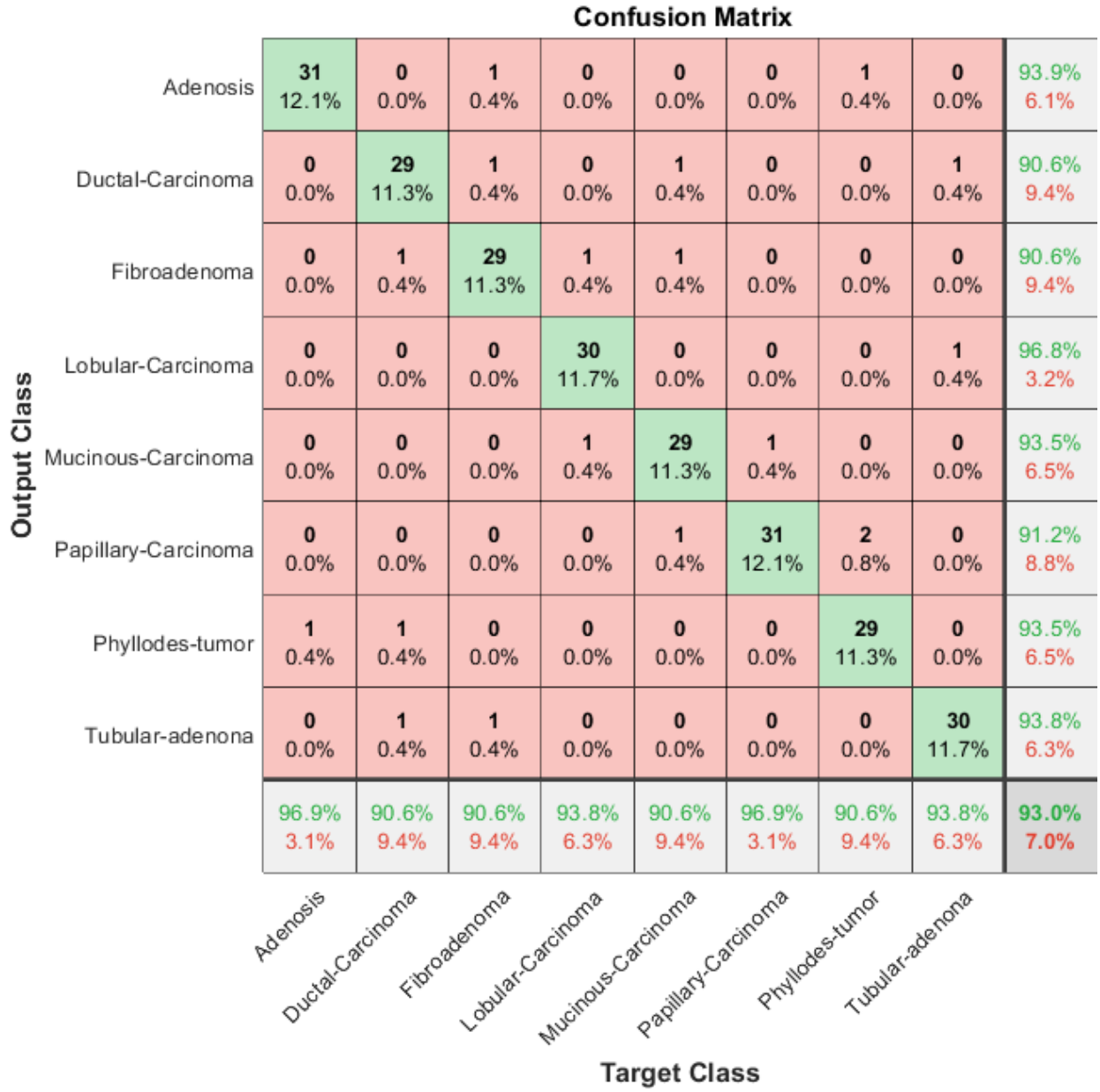


Figure 4.30: Confusion matrices for ResNet18 trained on 400x images.

ShuffleNet for 40x

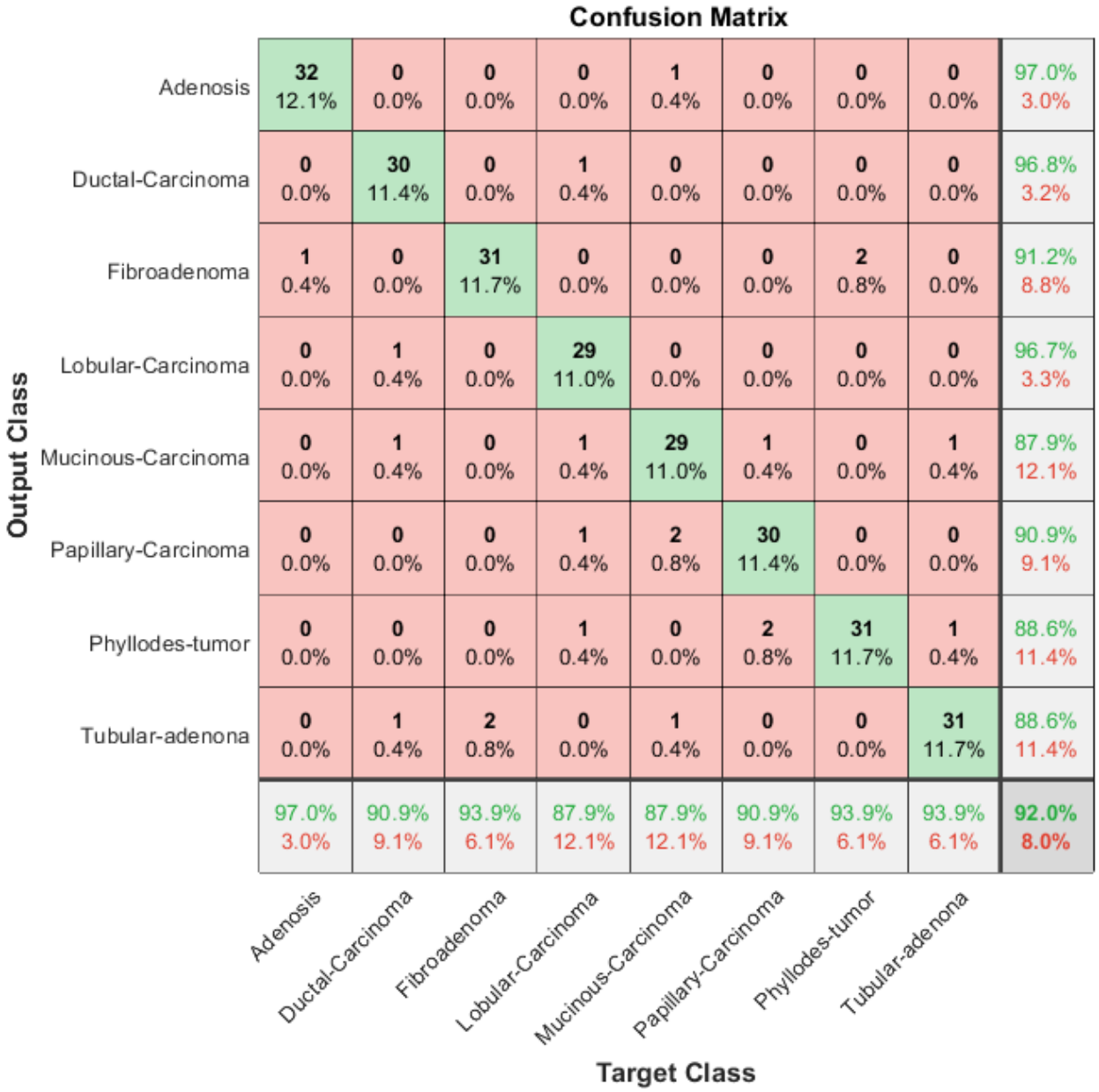


Figure 4.31: Confusion matrices for ShuffleNet trained on 40x images.

ShuffleNet for 100x

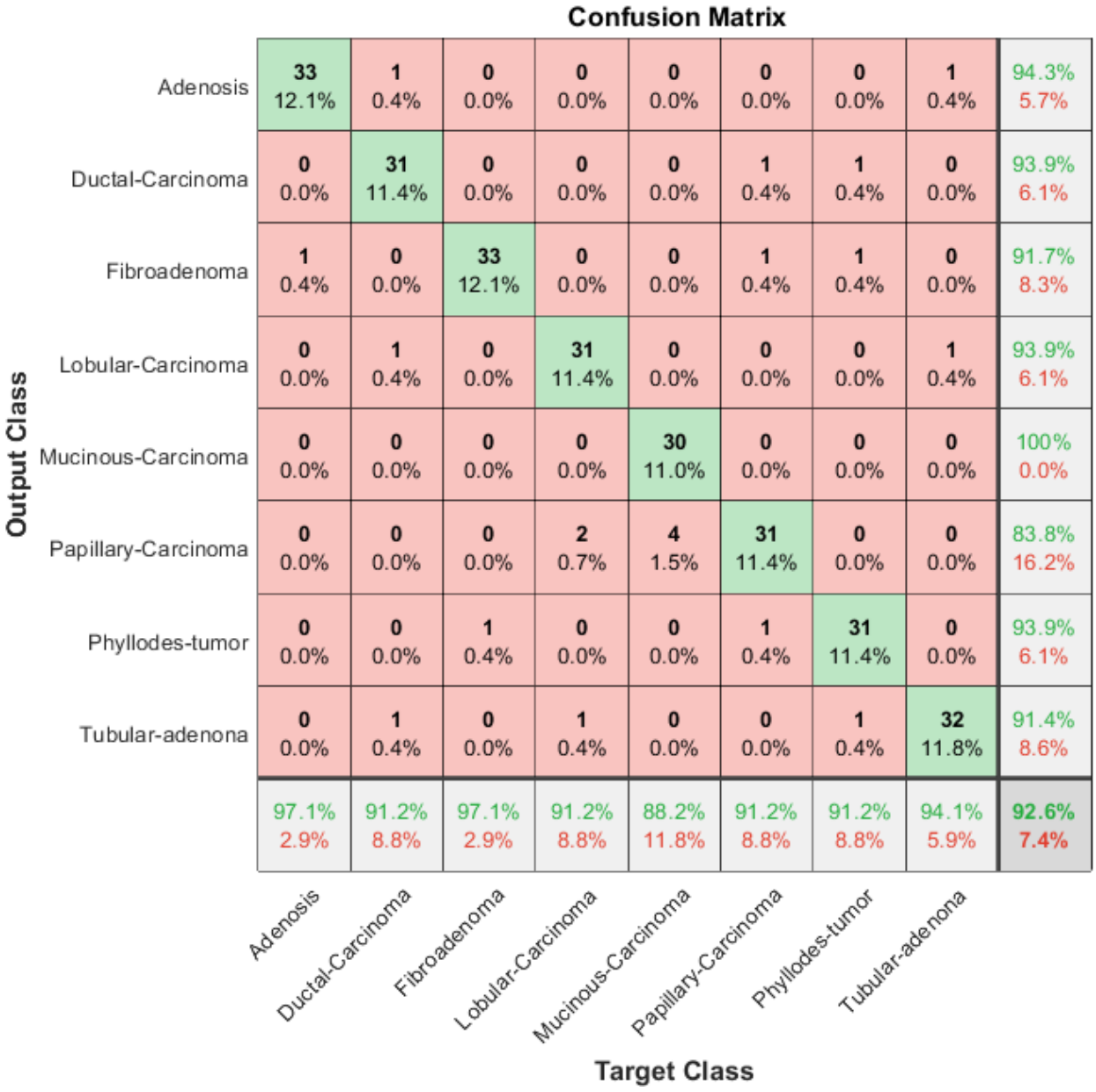


Figure 4.32: Confusion matrices for ShuffleNet trained on 100x images.

ShuffleNet for 200x

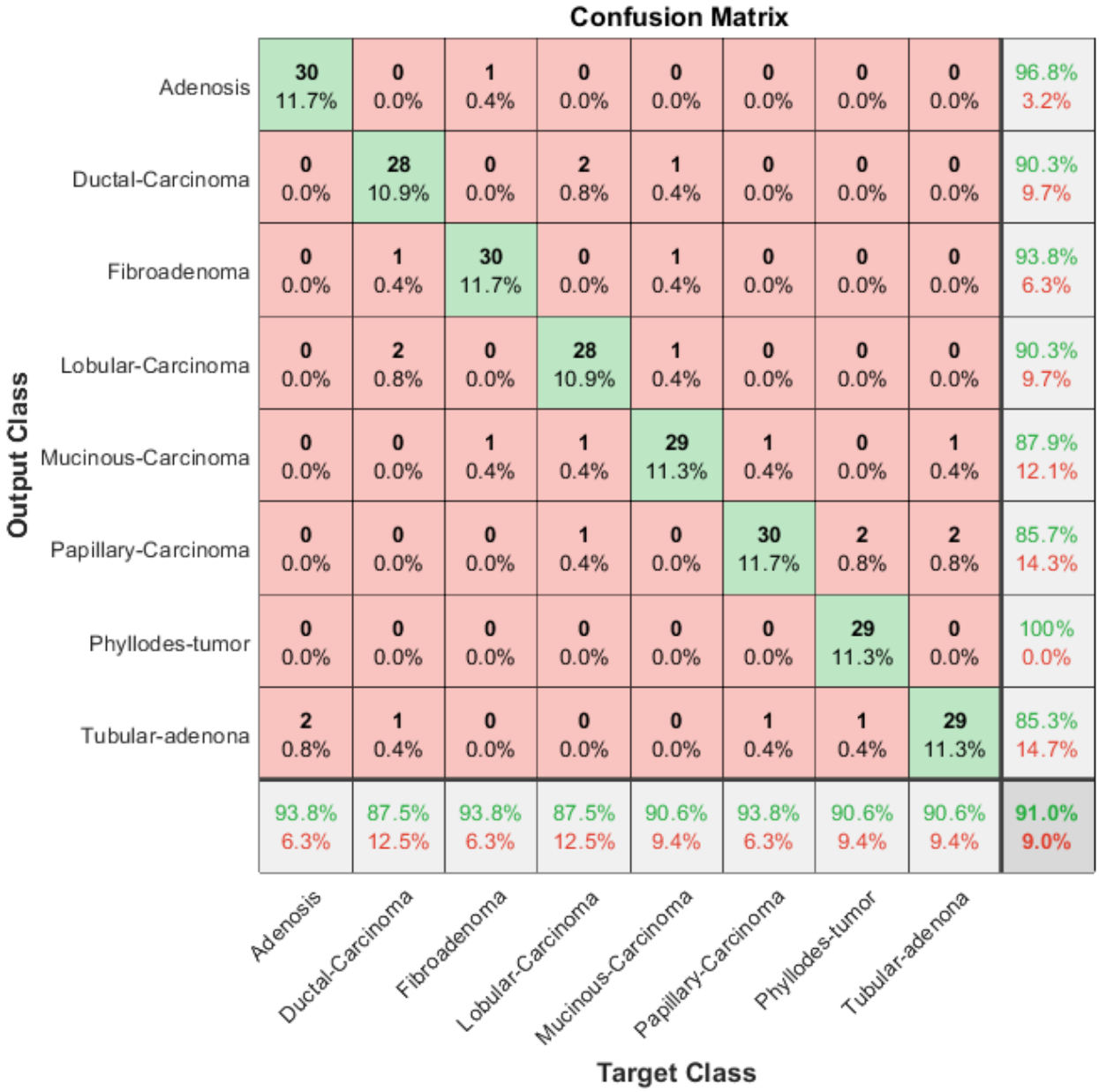


Figure 4.33: Confusion matrices for ShuffleNet trained on 200x images.

ShuffleNet for 400x

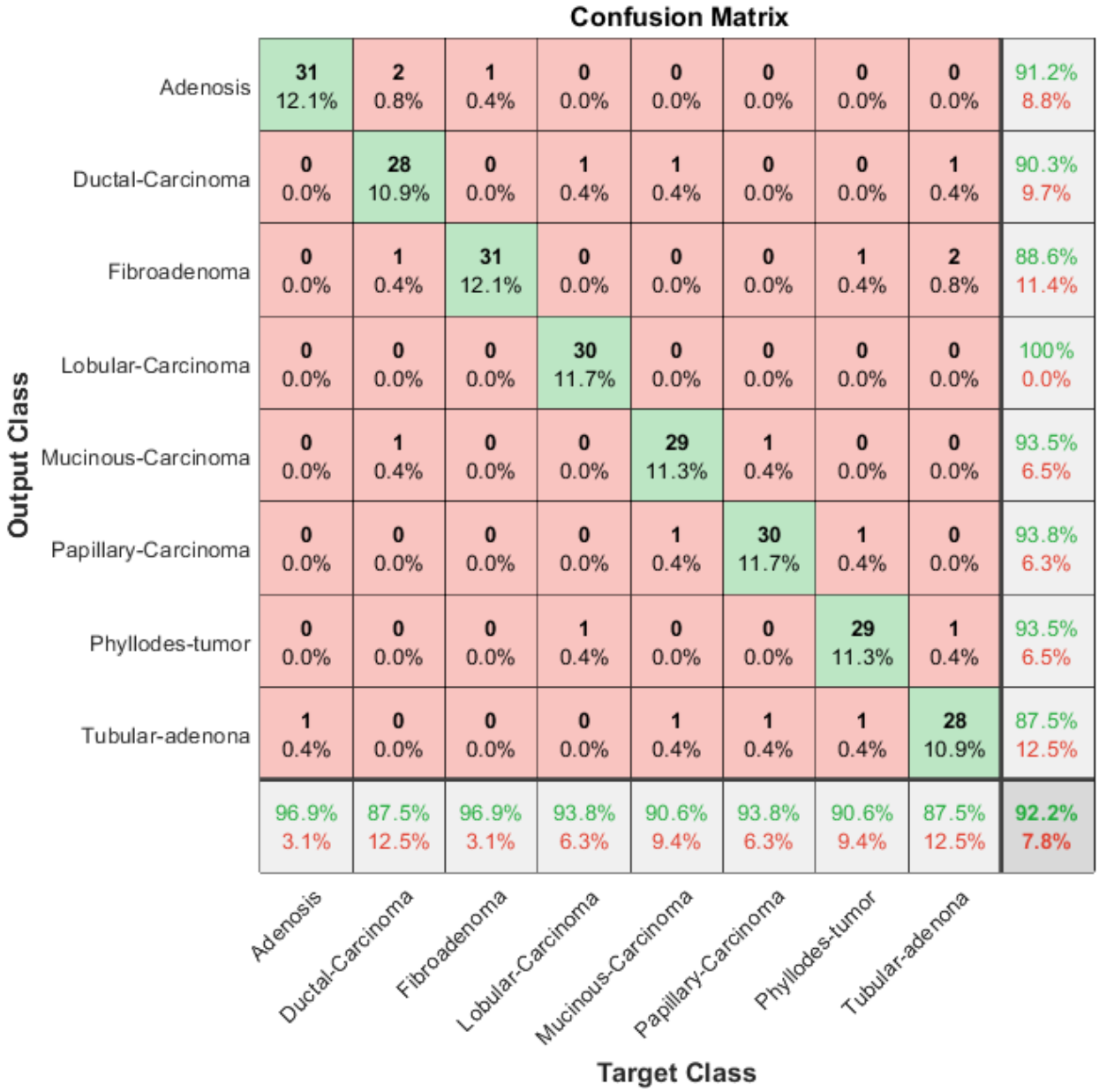


Figure 4.34: Confusion matrices for ShuffleNet trained on 400x images.

To summarize the results above, **ResNet18** achieved a higher accuracy rate in terms of magnifying factors 40x, 100x, and 200x. Besides, with magnifying factor 400x, **Inception-V3Net** and **ShuffleNet** obtained better results of accuracy than the deep neural network ResNet18. Besides, ResNet18 obtained 95.08% with 40x magnifying factor, 94.12% with 100x magnifying factor, 94.53% with 200x magnifying factor, and 92.97% with 400x magnifying factor. In contrast, **Inception-V3Net** and **ShuffleNet** obtained 93.36% with 400x magnifying factor. However, for overall average accuracy**, ResNet18** achieved the best performance of 94.18%, followed by Inception-V3Net with 92.76% and ShuffleNet with 92.27%.

4.3 Comparison with the State-of-the-Art Result

Several studies, such as [59,76,79], applied deep learning models to diagnostics of breast cancer using real histopathological images from the BreakHis dataset [65]. Most of the previous studies focused on binary classification, although some considered multiclass classification as well. All of the studies used different pre-trained deep neural networks. In this thesis, we applied three pre-trained (by means of transfer learning) neural networks ResNet18, Inception-V3Net and ShuffleNet in image-based binary classification (benign or malignant) and multiclass classification (eight classes) on real images from publicly available BreakHis dataset [65]. Table 4.18 compares different approach used the similar pre-trained neural network, Table 4.19 compares different studies for binary classification, and Table 4.20 for multiclass classification.

4.3.1 Binary Classification

In this Section, we compare the performance of our DNNs with state-of-the-art results in the literature in the binary classification task. Table 4.18 compares the performance of pre-trained DNNs similar to ours used in different studies with our DNNs implemented in this thesis. Our **ResNet18** achieved the best overall average accuracy. Also, for the following magnifying factors (40x,100x,200x), **ResNet18** achieved the best performance for all magnifications except for 400x, where **[76]** achieved the best accuracy. Our Inception-V3Net neural network and the one from [76] exhibited similar overall average accuracy.

Comparison of different approaches using similar neural networks for binary classification						
Study	**Year / Neural Network**	**Accuracy**				**Overall Average Accuracy**
		40x	**100x**	**200x**	**400x**	
[59]	2019 – Inception-V2Net	91.05%	88.93%	88.76%	87.42%	89.04%
[76]	2019 – Inception-V3Net	95.7%	95.7%	95.7%	**95.7%**	95.7%
[59]	2019 – ResNet-50	94.01%	93.34%	95.04%	94.96%	94.33%
[79]	2019 – ResNet-50	91.2%	91.7%	92.6%	88.9%	91.1%
ResNet18	2019-2020	**99.18%**	**97.67%**	**97.59%**	94.03%	**97.12%**
Inception-V3Net	2019-2020	97.01%	96.11%	95.45%	94.03%	95.65%

Table 4.18: Comparing the performance of similar DNNs from this thesis and literature in binary classification.

Furthermore, different data split or training parameters such as optimization algorithm, learning rate, batch size, decay factor and epoch will make a difference in the performance results. For instance, [76] used Stochastic Gradient Descent for optimization algorithm and a batch size of 32. However, in [79] the authors set the learning rate of the Stochastic Gradient Descent optimizer

to 0.001. In [59] they used ResNet-50 and they divided the data into three parts for training, validation and testing and they used Stochastic Gradient Descent optimizer with a learning rate of 0.0005.

In Table 4.19 below, we compare the performance of different DNNs from literature and this thesis. **ResNet18** achieved the best performance accuracy for images with magnifications 40x, 100x and 200x. On images with magnification 400x **[4]** achieved higher performance accuracy. **ShuffleNet** of this thesis's implementation achieved the second-highest result for magnifying factors 40x and 100x. Moreover, **ResNet18** and **[4]** have similar overall average accuracy, followed by **ShuffleNet.**

Comparison of different approaches using the BreakHis data for Binary Classification						
Study	**Year**	**Accuracy**				**Overall Average Accuracy**
		40x	**100x**	**200x**	**400x**	
[46]	2019	81.61%	84.47%	86.67%	83.15%	83.98%
[62]	2019	89.31%	85.75%	83.95%	84.33%	85.84%
[82]	2019	85.7%	84.2%	84.9%	80.1%	83.73%
[4]	2018	97.95%	97.57%	97.32%	**97.36%**	**97.7%**
[25]	2018	95.62%	95.03%	97.04%	96.31%	96.00%
[27]	2018	94.71%	95.9%	96.76%	89.11%	94.12%
ResNet18	2019-2020	**99.18%**	**97.67%**	**97.59%**	94.03%	97.12%
ShuffleNet	2019-2020	98.64%	**97.67%**	96.52%	94.32%	96.79%
Inception-V3Net	2019-2020	97.01%	96.11%	95.45%	94.03%	95.65%

Table 4.19: Comparing the performance of different DNNs from this thesis and literature in binary classification.

4.3.2 Multiclass Classification

In this Section, we compare the performance of our DNNs with state-of-the-art results in the literature in the multiclass classification task. In Table 4.20, we compare the performance of different DNNs form literature and this thesis. The net from **[4]** achieved the best overall average accuracy, and **ResNet18** achieved the second-highest accuracy overall for multiclass classification (Adenosis, Fibroadenoma, Phyllodes Tumor, Tubular Adenoma, Ductal Carcinoma, Lobular Carcinoma, Mucinous Carcinoma and Papillary Carcinoma).

Comparison of different approaches on BreakHis data for Multiclass Classification						
Study	**Year**	**Accuracy**				**Overall Average Accuracy**
		40x	**100x**	**200x**	**400x**	
[20]	2019	88.3%	88.3%	87.1%	83.4%	86.78%
[4]	2018	97.09%	97.57%	97.29%	97.22%	**97.2%**
[28]	2017	92.8%	93.9%	93.7%	92.9%	93.32%
[9]	2016	83.08%	83.17%	84.63%	82.10%	83.25%
[64]	2016	90%	88.4%	84.6%	86.1%	87.28%
ResNet18	2019-2020	95.08%	94.12%	94.53%	92.97%	94.18%
ShuffleNet	2019-2020	92.05%	92.65%	91.02%	93.36%	92.27%
Inception-V3NET	2019-2020	92.05%	92.28%	93.36%	93.36%	92.76%

Table 4.20: Comparing the performance of different DNNs from this thesis and literature in multiclass classification.

Chapter 5: Conclusions and Future Work

5.1 Conclusions

In this dissertation, we compared different pre-trained deep learning models in the task of breast cancer classification using histopathological images and an image-based approach. We used publicly available database BreakHis consisting of 7,909 histopathological images from 82 patients. The dataset contains 2,480 samples representing 4 benign classes (A,F,PT,TA) and 5,429 samples representing 4 malignant classes (DC,LC,MC,PC) with four different magnifying factors (40x,100x,200x,400x). We applied three pre-trained deep convolutional neural networks ResNet18, Inception-V3Net, and ShuffleNet to automatically extract features from images for binary and multiclass classification. We applied different techniques, including data augmentation for the training set and transfer learning. Data augmentation boosts the size of training data allowing to train the deep models better, and it also helps with balancing classes.

We applied several performance evaluation metrics such as accuracy, precision, sensitivity (recall), specificity, and F1-score for each magnifying factor. Moreover, we generated confusion matrices for binary and multiclass classification. Since the complexity of medical images, we applied the transfer learning technique based on replacing the final three layers of the three pre-trained networks, and the three layers must fine-tune for the new classification. We applied transfer learning to three pre-trained deep neural network ResNet18, ShuffleNet, and Inception-V3Net with different magnifying factors 40x, 100x, 200x, and 400x for binary classification (benign and malignant) and multiclass classification (eight classes). Regarding the average accuracy for binary classification for each pre-trained neural network, the ResNet 18 achieved the best overall average accuracy of 97.11%, followed by ShuffleNet 96.78%, and Inception-V3Net got 95.65%. In terms of multiclass classification, ResNet18 achieved the best performance for magnification factors 40x, 100x, 200x. For 400x magnification, Inception-V3Net and ShuffleNet performed better than ResNet18. To sum up, our DNNS achieved the following accuracies: ResNet18 94.17%, Inception-V3Net 92.76%, and ShuffleNet 92.27%. To compare the performance of our DNNs

with the state-of-the-art results for binary classification, our ResNet 18 achieved an excellent overall average accuracy, and for multiclass classification, ResNet18 achieved the second-highest performance.

5.2 Future Work

In this thesis, we implemented three pre-trained neural networks for image-based binary and multiclass classification of breast cancer histopathological images using the BreakHis dataset. In the future, it will interestingly be to extend our results to patient-based classification and to obtain higher accuracy than the present state-of-the-art techniques. Besides DNNs studied in this dissertation, we will try other promising models and create an ensemble of DNNs to achieve excellent performance on the BreakHis dataset and work on more challenging datasets. We are also planning to extend our system to make it capable of tackling challenging breast cancer grading problems.

Bibliography

[1] Adeshina S. A., Adedigba A. P., Adeniyi A. A., Aibinu A. M. (2018). "Breast Cancer Histopathology Image Classification with Deep Convolutional Neural Networks". *2018 14th International Conference on Electronics Computer and Computation (ICECCO)*, pages 206-212.

[2] Alberg A. J., Park J. W., Hager B. W., Brock M. V., Diener M. W. (2004). "The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests." *Journal of General Internal Medicine*, *19*(5), pages 460–465.

[3] Alirezazadeh P., Hejrati B., Monsef-Esfahani A., Fathi A. (2018). "Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images." *Biocybernetics and Biomedical Engineering*, *38*(3), pages 671–683.

[4] Alom M. Z., Yakopcic C., Nasrin M. S., Taha T. M., Asari, V. K. (2019). "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network." *Journal of Digital Imaging*, *32*(4), 605–617.

[5] Alyafeai Z., Ghouti L. (2020). "A fully-automated deep learning pipeline for cervical cancer classification." *Expert Systems with Applications*, *141*, 112951.

[6] Amin-Naji M., Aghagolzadeh A., Ezoji M. (2019). "Ensemble of CNN for multi-focus image fusion." *Information Fusion*, *51*, pages 201–214.

[7] Araújo T., Aresta G., Castro E., Rouco J., Aguiar P., Eloy C., … Campilho A. (2017). "Classification of breast cancer histology images using Convolutional Neural Networks." *Plos One*, *12*(6).

[8] Barata C., Ruela M., Francisco M., Mendonca T., Marques J. S. (2014). "Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features." *IEEE Systems Journal*, *8*(3), pages 965–979.

[9] Bayramoglu N., Kannala J., Heikkilä J., "Deep learning for magnification independent breast cancer histopathology image classification," *23rd International Conference on Pattern Recognition (ICPR 2016)*, Cancun, pages 2440-2445.

[10] Benhammou Y., Achchab B., Herrera F., Tabik S. (2020). "BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights." *Neurocomputing*, *375*, pages 9–24.

[11] Bhuiyan M. N. Q., Shamsujjoha M., Ripon S. H., Proma F. H., Khan F. (2019). "Transfer Learning and Supervised Classifier Based Prediction Model for Breast Cancer." *Big Data Analytics for Intelligent Healthcare Management*, pages 59–86.

[12] Breast Cancer Understanding your Diagnosis. *Canadian Cancer Society*. (2015 Revised:2019). *Available at: http://www.cancer.ca*

[13] Budak Ü., Cömert Z., Rashid Z. N., Şengür A., Çıbuk M. (2019). "Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images." *Applied Soft Computing*, *85*, 105765.

[14] Cireşan D. C., Giusti A., Gambardella L. M., Schmidhuber J. (2013). "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks." *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013 Lecture Notes in Computer Science*, pages 411–418.

[15] Coccia, M. (2020). "Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence." *Technology in Society*, 60, 101198, pages 1-11.

[16] Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shihd N, Tomaszewski J, Madabhushi, A. (2014). "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks." *Medical Imaging 2014: Digital Pathology*, pages 904103–904115.

[17] Dabeer S., Khan M. M., Islam S. (2019). "Cancer diagnosis in histopathological image: CNN based approach." *Informatics in Medicine Unlocked*, *16*, 100231.

[18] Deng J., Dong W., Socher R., Li L.-J., Li K., & Fei-Fei L. (2009). "ImageNet: A large-scale hierarchical image database." *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248-255.

[19] Devarriya D., Gulati C., Mansharamani V., Sakalle A., Bhardwaj A. (2020). "Imbalance breast cancer data classification using novel fitness functions in genetic programming." *Expert Systems with Applications*, *140*, 112866.

[20] Erfankhah H., Yazdi M., Babaie M., Tizhoosh H. R. (2019).” Heterogeneity-Aware Local Binary Patterns for Retrieval of Histopathology Images.” *Institute of Electrical and Electronics Engineers Access*, *7*, pages 18354–18367.

[21] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. *Available at: http://www.gco.iarc.fr/today*

[22] Fernando B., Fromont E., Muselet D., Sebban M. (2012). “Discriminative feature fusion for image classification.” *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 434–3441.

[23] Gandomkar, Z., Brennan, P. C., & Mello-Thoms, C. (2018). “MuDeRN: Multi-category classification of breast histopathological image using deep residual networks.” *Artificial Intelligence in Medicine*, *88*, pages 14–24.

[24] Gao, F., Yoon, H., Wu, T., & Chu, X. (2020). “A feature transfer enabled multi-task deep learning model on medical imaging.” *Expert Systems with Applications*, *143*, 112957.

[25] Gu Y., & Yang J. (2019). “Densely-Connected Multi-Magnification Hashing for Histopathological Image Retrieval.” *IEEE Journal of Biomedical and Health Informatics*, *23*(4), pages 1683–1691.

[26] Gu Y., Yang J. (2019). “Multi-level magnification correlation hashing for scalable histopathological image retrieval.” *Neurocomputing*, *351*, pages 134–145.

[27] Gupta V., Bhavsar A. (2018). “Sequential Modeling of Deep Features for Breast Cancer Histopathological Image Classification.” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* pages 2335-23357.

[28] Han Z., Wei B., Zheng Y., Yin Y., Li K., Li S., "Breast Cancer Multiclass Classification from Histopathological Images with Structured Deep Learning Model." *Scientific Reports,* 7, 4172 (2017), pages 1-10.

[29] He K., Xiangyu Z., Shaoqing R., Jian S. (2016) "Deep residual learning for image recognition." *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770-778.

[30] Hossin M. B., & Sulaiman M. N (2015). “A Review on Evaluation Metrics for Data Classification Evaluations.” *International Journal of Data Mining & Knowledge Management Process*, *5*(2), pages 01–11.

[31] Ismail N. S., & Sovuthy C. (2019). "Breast Cancer Detection Based on Deep Learning Technique." *2019 International UNIMAS STEM 12th Engineering Conference (EnCon)*, pages 89-92.

[32] Jiang Y., Chen L., Zhang H., Xiao X. (2019). "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module." *Plos One*, *14*(3).

[33] Kashif M., Malik K. R., Jabbar S., Chaudhry J. (2020). "Application of machine learning and image processing for detection of breast cancer." *Innovation in Health Informatics*, pages 145–162.

[34] Kassani S. H., Kassani P. H., Wesolowski M. J., Schneider K. A., Deters R. (2019) „Classification of histopathological biopsy images using ensemble of deep learning networks". *In Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering (CASCON '19)*, pages 92–99.

[35] Kassani, S. H., & Kassani, P. H. (2019). "A comparative study of deep learning architectures on melanoma detection." *Tissue and Cell*, *58*, pages 76–83.

[36] Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2019). "Breast Cancer Diagnosis with Transfer Learning and Global Pooling." *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 519-524.

[37] Kaushal C., Bhat S., Koundal D., Singla A. (2019). "Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images." *Innovation and Research in Biomedical Engineering*, *40*(4), pages 211–227.

[38] Kavukcuoglu K., Ranzato M., LeCun Y., "Fast inference in sparse coding algorithms with applications to object recognition", *arXiv preprint arXiv:1010.3467* (2008).

[39] Khan H. N., Shahid A. R., Raza B., Dar A. H., Alquhayz H. (2019). "Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network." *IEEE Access*, *7*, pages 165724–165733.

[40] Khuriwal N., Mishra N. (2018). "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm." *2018 IEEMA Engineer Infinite Conference (ETechNxT)*, pages 1-5.

[41] Kingma D. P., & Lei Ba, J. (2015). “Adam: A Method for Stochastic Optimization.” *2015 International Conference on Learning Representations (ICLR),* arXiv:1412.6980.

[42] Kumar A., Singh S. K., Saxena S., Lakshmanan K., Sangaiah A. K., Chauhan H., Shrivastava S, Singh R. K. (2020). "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer." *Information Sciences*, 508, pages 405–421.

[43] Kurmi Y., Chaurasia V., Ganesh N. (2019).” Tumor Malignancy Detection Using Histopathology Imaging.” *Journal of Medical Imaging and Radiation Sciences*, *50*(4), pages 514–528.

[44] Lecun Y., Bottou L., Bengio Y., & Haffner, P. (1998). “Gradient-based learning applied to document recognition.” *Proceedings of the IEEE*, *86*(11), pages 2278–2324.

[45] LeCun Y., Kavukcuoglu K., Farabet C., "Convolutional networks and applications in vision," *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, Paris, 2010, pages. 253-256.

[46] Lichtblau D., Stoean C. (2019). “Cancer diagnosis through a tandem of classifiers for digitized histopathological slides.” *Plos One*, *14*(1).

[47] Macenko M., Niethammer M., Marron J. S., Borland D., Woosley J. T., Guan X., … Thomas N. E. (2009). “A method for normalizing histology slides for quantitative analysis.” *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro,* pages 1107-1110.

[48] MATLAB. (2019). *Transfer Learning*. Deep learning approach to train new models faster by using pre-trained models. *Available at: http://www.mathworks.com/transfer-learning*

[49] Matos J. D., Britto A. D. S., Oliveira L. E. D., Koerich A. L. (2019). “Texture CNN for Histopathological Image Classification.” *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 580-583.

[50] Mckinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., … Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), pages 89–94.

[51] Nawaz M. A., Sewissy A. A., Soliman T. H. (2018). "Automated Classification of Breast Cancer Histology Images Using Deep Learning Based Convolutional Neural Networks". *IJCSNS International Journal of Computer Science and Network Security*, pages 152-160.

[52] Nawaz M., Sewissy A. A., Soliman T. H. A., "Multi-Class Breast Cancer Classification using Deep Learning Convolutional Neural Network." *International Journal of Advanced Computer Science and Applications (IJACSA)*, 9(6), (2018), pages 316-322.

[53] Nguyen P. T., Nguyen T. T., Nguyen N. C., Le T. T. (2019). "Multiclass Breast Cancer Classification Using Convolutional Neural Network." *2019 International Symposium on Electrical and Electronics Engineering (ISEE)*, pages 130-134.

[54] Pratiher S., Chattoraj S., Agarwal S., Bhattacharya S. (2018). "Grading Tumor Malignancy via Deep Bidirectional LSTM on Graph Manifold Encoded Histopathological Image." *2018 IEEE International Conference on Data Mining Workshops (ICDMW),* pages 674-681.

[55] Qi Q., Li Y., Wang J., Zheng H., Huang Y., Ding X., Rohde, G. K. (2019). "Label-Efficient Breast Cancer Histopathological Image Classification." *IEEE Journal of Biomedical and Health Informatics*, *23*(5), pages 2108–2116.

[56] Qiao J., Lv Y., Cao C., Wang Z., Li A. (2018) "Multivariate Deep Learning Classification of Alzheimer's Disease Based on Hierarchical Partner Matching Independent Component Analysis." *Frontiers in Aging Neuroscience*. 10:417.

[57] Reinhard E., Adhikhmin M., Gooch, B., Shirley P. (2001). "Color transfer between images." *IEEE Computer Graphics and Applications*, *21*(4), pages 34–41.

[58] Roy, K., Banik, D., Bhattacharjee, D., & Nasipuri, M. (2019). "Patch-based system for Classification of Breast Histology images using deep learning." *Computerized Medical Imaging and Graphics*, *71*, pages 90–103.

[59] Sabari, D. N., Saravanan, R., Anbazhagan, J., & Koduganty, P. (2019). "Comparison of Deep Feature Classification and Fine Tuning for Breast Cancer Histopathology Image Classification". *Communications in Computer and Information Science Recent Trends in Image Processing and Pattern Recognition*, pages 58–68.

[60] Saikia A. R., Bora K., Mahanta L. B., Das A. K. (2019). "Comparative assessment of CNN architectures for classification of breast FNAC images." *Tissue and Cell*, *57*, pages 8–14.

[61] Shallu S. & Mehra, R. (2018). "Breast cancer histology images classification: Training from scratch or transfer learning?." *Information & Communications Technology Express*, *4*(4), pages 247–254.

[62] Sharma S., & Mehra R. (2019). "Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image." *The Visual Computer*, pages 1-15

[63] Spanhol F. A., Oliveira L. S., Cavalin P. R., Petitjean C., Heutte L. (2017). "Deep features for breast cancer histopathological image classification." *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1868-1873

[64] Spanhol F. A., Oliveira L. S., Petitjean C., Heutte L., "Breast cancer histopathological image classification using Convolutional Neural Networks," *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pages 2560-2567.

[65] Spanhol F. A., Oliveira L. S., Petitjean C., Heutte, L. (2016). "A Dataset for Breast Cancer Histopathological Image Classification." *IEEE Transactions on Biomedical Engineering*, *63*(7), pages 1455–1462.

[66] Stoecker W. V., Wronkiewiecz M., Chowdhury R., Stanley R. J., Xu J., Bangert A., … Drugge, R. (2011). "Detection of granularity in dermoscopy images of malignant melanoma using color and texture features." *Computerized Medical Imaging and Graphics*, *35*(2), pages144–147.

[67] Su H., Liu F., Xie Y., Xing F., Meyyappan S., Yang L. (2015). "Region segmentation in histopathological breast cancer images using deep convolutional neural network." *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 55-58.

[68] Sudharshan P., Petitjean C., Spanhol F., Oliveira L. E., Heutte L., Honeine P. (2019). "Multiple instance learning for histopathological breast cancer image classification." *Expert Systems with Applications*, *117*, pages 103–111.

[69] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z., "Rethinking the inception architecture for computer vision." *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pages. 2818-2826.

[70] Toğaçar M., Ergen B., Cömert Z. (2020). "Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders." *Medical Hypotheses*, *135*, 109503.

[71] Toğaçar M., Özkurt K. B., Ergen B., Cömert Z. (2019). "BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer." *Physica A: Statistical Mechanics and Its Applications*, 123592.

[72] Vo D. M., Nguyen N. Q., Lee S. W. (2019). "Classification of breast cancer histology images using incremental boosting convolution networks." *Information Sciences*, *482*, pages 123–138.

[73] Wang P., Song Q., Li Y., Lv S., Wang J., Li L., Zhang H. (2020). "Cross-task extreme learning machine for breast cancer image classification with deep convolutional features." *Biomedical Signal Processing and Control*, *57*, 101789.

[74] World health organization cancer. (2018). *Fact Sheet-Cancer. Available at: http://www.who.int/health-topics/cancer*

[75] Wu P., Qu H., Yi J., Huang Q., Chen C., Metaxas D., "Deep Attentive Feature Learning for Histopathology Image Classification," *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pages 1865-1868.

[76] Xiang, Z., Ting, Z., Weiyan, F., & Cong, L. (2019). "Breast Cancer Diagnosis from Histopathological Image based on Deep Learning". *2019 Chinese Control and Decision Conference (CCDC)*, pages 4616-4619.

[77] Yang Z., Ran L., Zhang S., Xia Y., Zhang Y. (2019). "EMS-Net: Ensemble of Multiscale Convolutional Neural Networks for Classification of Breast Cancer Histology Images." *Neurocomputing*, *366*, pages 46–53.

[78] Yousefi M., Krzyżak A., & Suen C. Y. (2018). "Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning." *Computers in Biology and Medicine*, *96*, pages 283–293.

[79] Zhang X., Zhang Y., Qian B., Liu X., Li X., Wang X., … Wang, L. (2019). "Classifying Breast Cancer Histopathological Images Using a Robust Artificial Neural Network Architecture." *Bioinformatics and Biomedical Engineering Lecture Notes in Computer Science*, pages 204–215.

[80] Zhang X., Zhou X., Lin M., Sun J., "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pages 6848-6856

[81] Zhang, Y.-D., Chen, S., Wang, S.-H., Yang, J.-F., & Phillips, P. (2015). "Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine." *International Journal of Imaging Systems and Technology*, *25*(4), pages 317–327.

[82] Zhu C., Song F., Wang Y., Dong H., Guo Y., Liu J. (2019). “Breast cancer histopathology image classification through assembling multiple compact CNNs.” *BMC Medical Informatics and Decision Making*, *19*(1).

[83] Zilong H., Jinshan T., Ziming W., Kai Z., Ling Z., Qingling S., (2018) "Deep learning for image-based cancer detection and diagnosis − A survey," *Pattern Recognition* 83, pages134-149.