# An Ensemble of Deep Learning Models for detecting Invasive Ductal Carcinoma from Histopathological Image

A thesis

submitted in partial fulfillment of the requirement for the Degree of

**Master of Technology in Computer Technology** of

Jadavpur University

By

**Arpita Das**

Registration No: 149859 of 2019-2020

Examination Roll No: M6TCT22026

Under the Guidance of

**Dr. Ram Sarkar**

Professor

Department of Computer Science and Engineering

Jadavpur University, Kolkata-700032

India

2022

# FACULTY OF ENGINEERING AND TECHNOLOGY
# JADAVPUR UNIVERSITY

## Certificate of Recommendation

This is to certify that the dissertation entitled ―An ensemble of Deep Learning Models for detecting Invasive Ductal Carcinoma from Histopathological Image‖ has been carried out by Arpita Das (University Registration No: 149859 of 2019-2020, Examination Roll No: (M6TCT22026) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Technology in Computer Technology. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree in any other University or Institute.

……

Dr. Ram Sarkar (Thesis Supervisor)

Professor

Department of Computer Science and Engineering

Jadavpur University, Kolkata-32

Countersigned

…….

Dr. Anupam Sinha

Professor & Head, Department of Computer Science and Engineering,

Jadavpur University, Kolkata-32.

…….

Prof. Chandan Mazumdar

Dean, Faculty of Engineering and Technology,

Jadavpur University, Kolkata-32.

# FACULTY OF ENGINEERING AND TECHNOLOGY
# JADAVPUR UNIVERSITY

## Certificate of Approval*

This is to certify that the thesis entitled ― An ensemble of Deep Learning Models for detecting Invasive Ductal Carcinoma from Histopathological Image‖ is a bonafide record of work carried out by Arpita Das in partial fulfillment of the requirements for the award of the degree of Master of Technology in Computer Technology in the Department of Computer Science and Engineering, Jadavpur University during the period of July 2019 to June 2022. It is understood that by this approval the undersigned does not necessarily endorse or approve any statement made, opinion expressed, or conclusion drawn therein but approves the thesis only for the purpose for which it has been submitted.

……..

Signature of Examiner 1

Date:

……..

Signature of Examiner 2

Date:

*Only in case the thesis is approved

# FACULTY OF ENGINEERING AND TECHNOLOGY
# JADAVPUR UNIVERSITY

## Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis entitled ― An ensemble of Deep Learning Models for detecting Invasive Ductal Carcinoma from Histopathological Image‖ contains a literature survey and original research work by the undersigned candidate, as part of her Degree of Master of Technology in Computer Technology.

All information has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Arpita Das

Registration No: 149859 of 2019-2020

Exam Roll No: M6TCT22026

Thesis Title: An ensemble of Deep learning models for detecting Invasive ductal carcinoma from Histopathological Image

……..

Signature with Date

# Acknowledgement

I would like to start by thanking the holy trinity for helping me deploy all the right resources and for shaping me into a better human being.

I would like to express my deepest gratitude to my advisor, **Dr. Ram Sarkar**, Professor, Department of Computer Science and Engineering, Jadavpur University for his admirable guidance, care, patience and for providing me with an excellent atmosphere for doing research.

I would like to thank **Dr. Anupam Sinha**, Professor & Head, Department of Computer Science and Engineering, Jadavpur University, for providing me with moral support at times of need.

I would like to thank **Ms. Ipsita Das** without whom this thesis could not be completed as he played a major role in helping me about technical aspect.

Most importantly none of this would have been possible without the love and support of my family, especially my mother for her constant motivation and my bambino sacrificed unbelievably.

This thesis would not have been completed without the inspiration and support of a number of wonderful individuals — my thanks and appreciation to all of them for being part of this journey and making this thesis possible.

……..
Arpita Das
Registration No: 149859 of 2019-2020
Exam Roll No: M6TCT22026
Department of Computer Science & Engineering
Jadavpur University

# Table of Contents

# List of Figures

# List of Tables

## Chapter 1

# Introduction

CANCER is one of the root causes of death worldwide. According to WHO (World Health Organization) 9.6 million deaths in 2018 were due to cancer [1]. Cancer is the term used when the cells in our body grow abnormally beyond their boundaries and slowly invade other parts of our body. The most common cases of cancer as maintained by the IARC (International Agency for Research on Cancer) are lung, breast, colorectal, prostate, skin and stomach. Breast cancer reported 2.09 million cases in the year 2018 and caused 627000 deaths globally [2]. The SEER (Surveillance, Epidemiology and End Results) estimates 268,000 cases and 41,760 deaths due to breast cancer in the year 2019 [3].
Breast cancer is the worst and the most common form of cancer occurring among women compared to other forms of cancer [4]. Breast cancer plays a significant role in affecting female mortality. Researchers are actively seeking to develop early detection methods of breast cancer. Several technologies contributed to the reduction in mortality rate from this disease, but early detection contributes most to preventing disease spread, breast amputation and death [5].

## 1.1. Types of Invasive Carcinoma in Breast cancer

Breast cancers that have spread into surrounding breast tissue are known as **invasive breast cancers**.

Most breast cancers are invasive, but there are different types of invasive breast cancer. The two most common are as follows:

- Invasive ductal carcinoma (IDC)
- Invasive lobular carcinoma (ILC)

**Invasive (infiltrating) ductal carcinoma**

This is the most common type of breast cancer. About 8 in 10 invasive breast cancers are invasive (or infiltrating) ductal carcinomas. IDC starts in the cells that line a milk duct in the breast. From there, the cancer breaks through the wall of the duct, and grows into the nearby breast tissues. At this point, it may be able to spread (metastasize) to other parts of the body through the lymph system and bloodstream.

## Invasive lobular carcinoma

About 1 in 10 invasive breast cancers is an invasive lobular carcinoma. ILC starts in the breast glands that make milk (lobules). Like IDC, it can spread (metastasize) to other parts of the body. Invasive lobular carcinoma may be harder to detect on physical exam and imaging, like mammograms, than invasive ductal carcinoma. And compared to other kinds of invasive carcinoma, it is more likely to affect both breasts. About 1 in 5 women with ILC might have cancer in both breasts at the time they are diagnosed.

Figure 1.1 shows all sub-types of breast cancers.

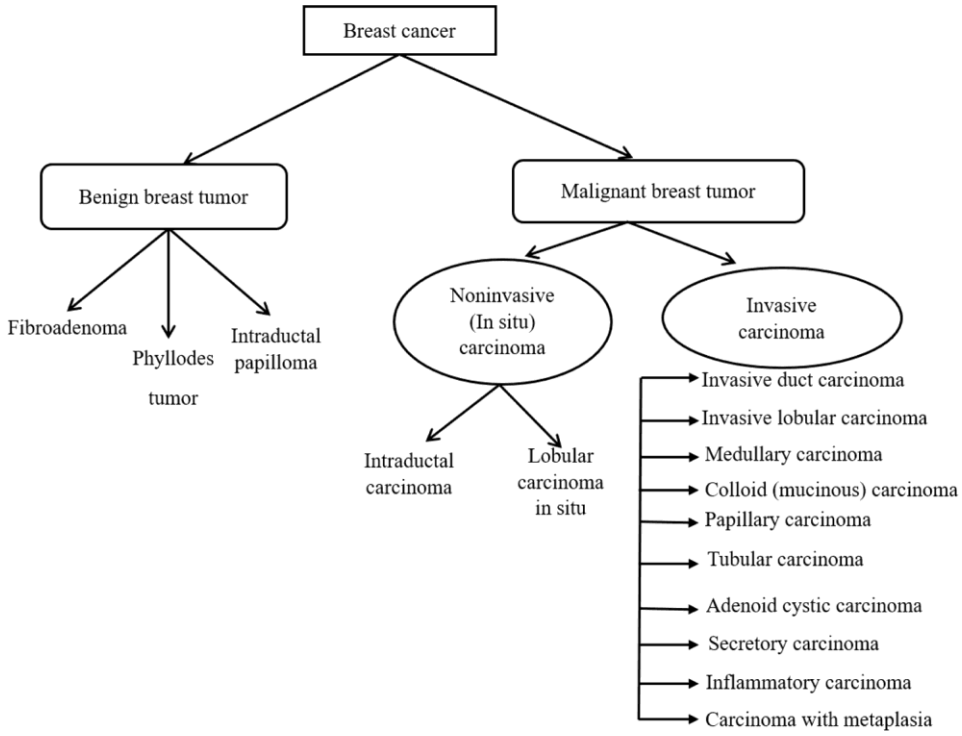



**Figure 1.1 sub-types of breast cancers**

## 1.2. History of Breast Cancer Detection

Egyptians identified breast cancer in 3,000 BC [8]. Then the Greeks, when a woman brought to Hippocrates, had a bloody discharge from the nipple and died. Hippocrates linked breast cancer due to menopause and called it hidden cancer because it did not appear on the skin. In 450 BC, Hippocrates diagnosed the hidden diseases of the patient by placing the mud on the entire body of the patient and the area that first dries out is the disease. It is the primitive process of thermal detection in the medical field [9]. Signs of breast cancer appear bitter in the mouth, loss of appetite, disturbed intelligence, dry eyes and nostrils, and loss of smell [10]. In the first century AD, a surgeon from the school of Alexandria pointed out that breast cancer is a huge swelling of harsh texture and uneven and grey to red. In 1913, radiography of breast cancer patients began in Germany. The study was carried out on 3,000 patients by surgeon Salmon [11]. In 1951 ultrasound was used as a research tool to detect breast tumor and identify it as benign or malignant. The other research was supported in 1952, when 21 cases of breast cancer were successfully identified. Through the results of this research, ultrasound was tested in the hospital as a diagnostic tool for breast cancer in 1954. In the 1960s, improvements were made to the internal structure of the ultrasound system and improvements in detection methods, including placing breasts on controlled temperature water for early detection of the tumor. Technological revolution after 1980 contributed to changes in the detection of the tumor and the flow of blood to the tumor. In the late 20th century, it was developed to use ultrasound to guide the needle biopsy in the breast area [12]. In 1957, Lawson used the thermal camera for the first time to diagnose breast cancer when he found the temperature difference between the tumor and the surrounding healthy area. When doctors and surgeons found Lausanne and Ghatmati in 1963 when they published research that the increase in skin temperature associated with breast cancer was associated with venous convection. In 1982, the Food and Drug Administration (FDA) approved the use of a thermal camera as a diagnostic aid to detect breast cancer. In 1996, a comparison between thermal images and X-rays for the diagnosis of a patient was conducted where the disease was detected by the thermal images disease, while it was not detected by X-rays [9].

## 1.3. Types of Breast Cancer Imaging

There are several types of imaging for breast cancer detection such as:

### A) Mammogram

Mammograms are the gold standard for breast cancer screening since 1960. However, there are many challenges affecting diagnosis using mammograms such as age, breast tissue density, and family history [13]. Mammograms can detect breast cancer in the early stages, reducing mortality by 25%. The doses of mammograms used in diagnosis affect patients over 70 years of age and cause rupture of weak tissue in the breast. They may also cause the formation of cancer in these vessels. Also, it is unable to detect cancer in younger women because of the density of breast tissue [14].

## B) Computerized Tomography (CT)

Computerized tomography takes X-rays of the breast from different angles as the patient enters in a closed machine, and a computer collects the image of the breast. The patient is injected into the vein of his hand with a substance to increase the contrast of the image [15]. Modern image reconstruction techniques have reduced 70% of the radiation and reduced the time it takes to take pictures [16]. However, there are disadvantages to this technique, including that some patients cannot hold breathe. This is in addition to the risk of radiation to the patient and its effect on pregnant women.

## C) Magnetic Resonance

IMAGING (MRI) MRI is a medical examination tool that uses radio waves and a field Magnetic. To show the tumor and calcifications clearly, the patient is injected with a substance into the bloodstream. MRI is often used to follow the response to chemotherapy for breast cancer patients before resorting to breast amputation [17]. Furthermore, when using MRI, the patient must be injected with gadolinium to show the details of the blood vessels in the breast. The syringe Gadolinium has the least effect on the sensitivity of iodine used in X-ray. However, Gadolinium affects allergic patients, so a supervision doctor is needed. MRI has many disadvantages such as its inability to detect breast cancer at an early stage and it is expensive too. Furthermore, women are not allowed to breastfeed for 48 hours. The device is also a closed space that causes anxiety in claustrophobic patients who are afraid of confined places.

## D) Ultrasound

Ultrasound imaging based on echo or reflection of sound waves is considered safer and more effective than X-rays. Ultrasound was first used in 1940 by France and Germany in the medical field. Ultrasound can detect breast cancer successfully in women with dense tissue and it has no impact on health and is quick and comfortable [17]. However, the disadvantage of this technique is its inability to detect breast cancer at an early stage and it has a higher rate of false-positive results [18].

## E) Histological Images

Histological images are generated using a microscope and they allow for the study of the microanatomy of cells, tissues, and organs by examining the correlation between structure and function. To detect cancer, breast tissue is stained with hematoxylin and eosin. The diagnosis of breast cancer histology images with hematoxylin and eosin stained, however, is non-trivial, labor-intensive, and often leads to a disagreement between pathologists [19]. Furthermore, the process of generating the images themselves requires a microscope that is expensive to acquire and maintain.

## 1.4. Analysis of Histological Images

Advances in Medical Science have led to an adaptation of the latest technologies in Breast cancer diagnosis and its treatment. Several diagnosing methods are commonly used like CT scan, PET scan, MRI, Ultrasound, Mammograms, and Histopathological Analysis [20]. Histopathology Analysis uses both software and hardware to extract the important features of the tissue and then prepares them for the image analysis. The tissues are mounted on slides and examined under microscopes to detect any growth in cancer cells or genetic progression. With the advent of WSI (Whole-Slide Imaging) scanners, the tissue Histopathological slides can be digitized and generated into a computerized image.

The main motive of these digitized techniques is to obtain quantitative data like cell size, tissue abnormalities, and the uneven number of cells.

The steps involved in the Image Analysis using Histopathology include:

1. **Pre-Processing**: This is the stage where operations like low pass filtering, dilation and thresholding are carried out [21].

2. **Segmentation**: Significant regions are extracted along with the ROI of the image to separate the image from the background. Some common techniques include: HMM, ACM, and Watershed Algorithm [22].

3. **Feature Extraction and Classification**: The visual information of the image is obtained with the help of Feature Extraction. These are the input to be fed for classification. Classifiers like ANN, CNN learn during the training phase and then classify the cancerous nuclei into diverse classes during the testing.

These methods can be time-consuming and restrictions like human error, bad image quality, and misdiagnosis can cost someone's life. CAD (Computer-Aided Diagnosis) is one recent context in radiology. CAD helps in improving the performance of the pathologists and the radiologists in finding out the cancerous tissue. A needle might not be able to extract as much information as a computer-aided system can detect an abnormal tissue with improved accuracy and precision. Provided with the efficacious image processing steps a high level of efficiency can be achieved in cancer diagnosis with the help of automated diagnosis. Fig. 2 illustrates the overall process in CAD. As a huge dataset of patients are dealt with worldwide, applying conventional machine learning algorithms would not be a good recommendation. This is where Deep Learning plays a key role. Deep learning handles huge datasets perfectly and can also extract high-level features without any domain interference or hard-core feature extraction. Deep learning takes a long time to train but the testing phase is faster than a machine learning algorithm which is the reason it is preferred for handling complex problems like image classification, speech recognition and NLP easily.

Figure 1.2 shows the flow chart of generic CAD system in breast cancer detection.

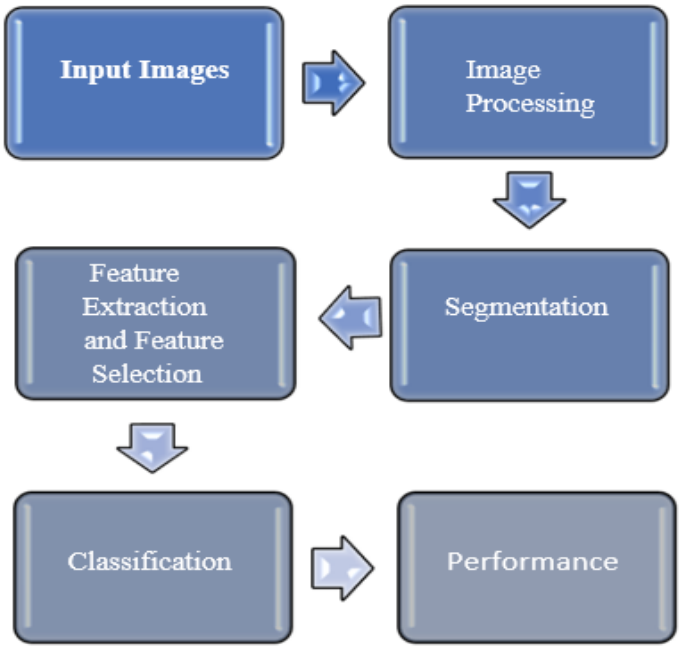



**Figure 1.2 Generic overall process of CAD in breast cancer diagnosis [23]**

## 1.5. Challenges

There is a considerable amount of literature on the application of medical image processing techniques to process Breast Cancer Histopathological Image (BCHI). However, there exists a few challenges which are given below.

- The lack of standard datasets makes it difficult to evaluate and compare various methods. A standard dataset would provide various researchers a common platform facilitating appropriate comparison.
- There are no standard metrics to evaluate the performance of the color normalization methods.
- The heterogeneous characteristics of malignant samples make it difficult to model the patterns to differentiate them from benign samples.
- CNN-based methods for histopathological image classification extracts features from the entire image and may fail to focus on the regions of interest (ROI) such as nuclei, gland, and mitotic cells, which contribute largely to the decision of classifying images as malignant and benign. Hence, there is scope for incorporating an attention mechanism in CNN to enable the model to focus on a potential ROI.[7]

## 1.6. Motivations for Automatic Diagnostic Systems

The need for automated diagnostic systems to detect breast cancer rose due to the high percentage of human errors in assessing and detecting breast cancer [52].

In addition, the automated diagnostic system detects breast cancer at a very early stage when they are too small to be detected using standard medical procedures. In fact, early breast cancers detected using automated screening and diagnosing systems are relatively easy to heal at this stage.

Furthermore, the recent work in [69] indicates that the automated procedure is more sensitive than the manual one by a large margin, where the manual procedure achieved a 68% sensitivity ratio against a 100% ratio for the automated procedure.

## 1.7. Scope of the work

- Data augmentation can be done for handling imbalanced datasets.
- The different pre-trained models can be used to train. An ensemble approach of pre-trained models can be used for better performance.
- Few additional layers can be added with pre-trained models for better accurate results.
- Different loss functions can be used. The ensemble approach of loss functions can be also used for getting good accuracy.

## 1.8. Organization of the Thesis

In *Chapter1*, a general introduction is presented along with the idea of Breast Cancer and CAD system detecting breast cancers and, types of breast cancer imaging. This chapter ends with the challenges and motivation of the present work done.

In *Chapter 2*, Literature Survey is presented in detail.

In *Chapter 3*, Methods and Methodologies applied to implement the present work are described in detail with some examples to understand an ideas described therein.

In *Chapter 4*, the experimental results are depicted in tabular format to get an idea of the results obtained in the different voting procedures.

In *Chapter 5*, the present work is concluded along with the feature scope of the present model to improve its performance of it.

## Chapter 2

# Literature Survey

Deep Learning in the last years has attracted the great interest of the research community and has reported great success in image analysis and in relatively challenging tasks such as image analysis, medical imaging, object detection, and recognition, and many others [24] [25].

From the literature survey, it can be observed that a large number of research work have been done concerning the detection and classification of histological images [26].

Over the last few years, researchers have suggested various methods for breast cancer diagnosis in histopathology images. New innovative methodologies are being proposed that include traditional machine learning and deep learning architectures. Many research works have emphasized on feature extraction. Textural and statistical features are often used.

Doyle et al. [27] made use of 64 Gabor filter features, 15 statistical gray-level features, graph-based features, 16 second-order statistical features, and 24 nuclear features per each of three color channels (hue, saturation, and intensity) the in HSI color space utilizing three distinct window sizes.

Dundar et al. [28] created an automatic classification system for classifying breast microscopic tissues into actionable subtypes (atypical ductal hyperplasia and ductal carcinoma in situ) and usual ductal hyperplasia. Statistical features, like the perimeter, mean gray-level intensity and the ratio of major to minor axis of the best-fitting ellipse, were utilized to model the cell size, shape and nucleoli appearance, respectively (histological descriptors).

Niwas et al. [29] proposed a method where they extracted first-order statistical and second-order statistical features.

With the help of the Log–Gabor complex wavelet bases, the color texture features of the segmented nucleus were assessed. The aforementioned techniques include integrating various forms of intricately handcrafted features to reflect the visual content of breast cancer histopathological images. In this context, extracting 59 image features using a Local Binary Pattern (LBP) based feature descriptor for breast mammography yielded an accuracy of 84% on a Support Vector Machine (SVM) classifier with a polynomial kernel [30]. Even statistical features were investigated in computer-aided breast cancer detection by Yasiran et al. [31].

Narayanan et al. [32] came up with a novel deep CNN architecture. Images were downsized from 50 × 50 to 48 × 48. In the pre-processing stage, a color constancy technique and histogram equalization were applied. The pre-processed images were then fed into the CNN. The proposed architecture was made out of five convolutional layers

followed by a fully connected (FC) layer and a softmax layer. Images that were pre-processed using color constancy techniques resulted in better performance metrics compared with images pre-processed utilizing Histogram Equalization.

Debelee et al. [33] extracted features using CNN and employed Principal Component Analysis for dimensionality reduction. k-Nearest Neighbors (k-NN) was utilized to classify mammograms as normal or abnormal.

Debelee et al. [34] extracted features from images using pre-trained Inception-V3 and their proposed modified adaptive K-means (MAKM) method. They collated images from the local and public datasets. GLCM and Gabor (Texture features) from ROIs and CNN-based extracted features were fused and fed into five classifiers (SVM, k-NN, NB, MLP and RF) to quantify the descriptive power of the features.

Rahman et al. [35] propounded a multilayered CNN. The IDC dataset has a significant class imbalance. An equal number of images were picked out from both the IDC(+) and IDC(-) classes in order to eradicate the class imbalance problem. The selection of images was performed randomly to avoid bias. Overfitting was mitigated by applying various data augmentation, viz; random 10-degree rotation, shifting, zooming and flipping horizontally and vertically. The implemented architecture achieved an accuracy of 89%.

Romano and Hernandez [36] also rectified the class imbalance present in the IDC(+) dataset. The images were normalized, and data augmentation was carried out in order to prevent overfitting. The images were augmented utilizing an array of random transformations so that the classifier would never twice view the exact same image. The proposed CNN architecture consists of two convolutional layers, a new pooling layer, called accept– reject pooling, dropout layers and FC layers. After feeding the images into the CNN, the model yielded the best performance with an accuracy of 85.41%.

Cruz-Roa et al. [37] proposed a three-layer CNN architecture for the automatic detection of IDC tissue regions in whole slide images (WSI). Their CNN has 16, 32 and 128 neurons in the first and second convolutional-pooling layers and the FC layer, respectively.

Wang et al. [38] extended the work of the Cruz-Roa et al. [37] architecture to four different architectures. They also concluded that data augmentation was not efficacious in the automatic detection of breast cancer with the IDC(+) dataset. Segmenting irregularities were expunged. There were a plethora of images that were not 50 × 50 and had to be removed. All the redundant images with 90% black space or 90% white space were removed.

Instead of giving equal significance to all the patches, the authors in [39] presented an attention mechanism that allowed the network to focus on the relevant features of patches. A weighted representation of all the constituent patches of an image was used for learning. Sanyal et al. [40] came up with a novel hybrid ensemble framework made of many fine-tuned CNNs, such as supervised feature extractors and XGBoost as a top-level classifier. They used high-resolution histopathology images for patch-wise detection.
Recently, Chapala and Sujatha [41] utilized pre-trained ResNet-50 and ResNet-34 mechanisms. The dataset was split into three different ratios of training and testing

datasets, such as 90%–10%, 75%–25% and 50%–50%. ResNet-50 outperformed ResNet-34 in terms of accuracy. In [42], Debelee et al. presented a detailed review of DL techniques used in breast cancer image analysis. They reviewed recent publicly available databases of breast cancer images and discussed the modalities that are used for breast cancer imaging.

Roy et al. [43], used ensemble learning by stacking and extracting the textural features from the histopathological images for classifying them into IDC + and IDC - categories. The study also conducted an in-depth comparative analysis of different machine learning classifiers in classifying breast cancer histopathological image patches inferring that the CatBoost (CB). The Conventional CAD methods have limited abilities to perform the required task yielding lesser accuracy and increased false-positive rates. In this study a wide range of machine learning methods is implemented, the most efficient among which is the CatBoost method, but that too yielded less accurate results.

Litjens et al. [44] specify the potential of deep learning as tools for great performance in the analysis and the diagnosis of histopathology. This makes an appearance from the need to minimize the human workload and error by improving diagnostic efficiency significantly. Their results lead them to the fact that deep learning can substantially improve the efficacy of breast cancer recognition procedures and automated systems.

Araújo et al. [45] proposed an architecture based on convolutional neural networks that are oriented to wholeslide histology images. Image dataset used was composed by high-resolution, uncompressed and annotated H&E stain images of breast histology. The authors used the pre-processing method proposed by [29] for obtaining the optimal stain vectors has been evaluated on slides with various stain combinations satisfactorily and the results reported 83.3% accuracy and 95.6% sensitivity for the classes of carcinoma and non-carcinoma.

Chen et al. [46] constructed a novel deep cascaded convolutional neural network to detect mitosis which is an indicator of aggressiveness in the invasive breast carcinoma situation. Initially, they detected the mitosis candidates by employing a fully convolutional neural network and maintaining high sensitivity rates. Then, they further discriminate the mitosis sections through knowledge transferred from the cross-domain. Their results in 394,275 training samples with 6.7% mitosis, 67.8% random selected negative samples and 24.8% false positives, were F1 score 0.788 and a recall of 0.772.

Bardou et al. [47] point out the superiority of convolutional neural networks by comparing two different machine learning approaches for classifying cases with benign tumors and cases with malignant tumors. In order to improve the accuracy of the convolutional neural network, they implemented different data augmentation techniques. Their results reached an accuracy of 98.33%.

The pre-trained architectures namely ResNet50 and DenseNet-161 were used to extract the features and to detect IDC by Celik et al. [48]. The experiment was conducted on 277,524 image patches of 50X50 pixels. An accuracy of 91% was reported on DenseNet-161 and 94% on ResNet-50 respectively.

Shallu et al. [49] utilized a pre-trained network namely VGG16, VGG19, and ResNet50 for feature extraction. An accuracy of 92% with logistic regression classifier for VGG16 was reported.

Saha et al. [50] designed a DL architecture with handcrafted features to detect mitotic cells from BCHI. The architecture comprises five convolution layers, four max-pooling layers, four ReLU, and two fully connected layers. The morphology, texture, and intensity features were extracted. They reported a precision, recall, and F1-score of 92%, 88%, and 90% respectively.

Beevi et al. [51] reported a method to detect mitosis using transfer learning technique. The features were extracted by combining a pre-trained CNN with a random tree classifier. An F1-score of 94% was reported.

## Chapter 3

# Methods and Methodologies

The theoretical foundation of this work is explained in this section. The concept of convolutional neural network (CNN), different pre-trained CNN models, different types of loss functions are illustrated.

## 3.1. Convolutional neural network (CNN)

A CNN or also known as ConvNet (used alternatively in the text) usually takes an input image, and assigns learnable weights with biases to different aspects in the image subsequently differentiating one picture from the other. CNN uses convolution operation in place of simple matrix multiplication in at least one of their layers. It is mainly used in an unstructured dataset (e.g., image and video). 2D-convolutional kernels are used by 2D-CNN for the prediction of segmentation maps of a single slice. 2D-CNN can leverage features from only spatial dimensions (height and width). Since 2D-CNN takes only a single slice as input, they intrinsically fail to extract context information from adjacent slices. From a practical perspective, voxel information from adjoining slices might contain enough information for the classification tasks. On the other hand, 3D-CNN can preserve temporal dimensions by predicting the volumetric patch of histological image data [52]. Although 3D-CNN possesses the ability to anchorage interslice context information which leads to improved performance but comes with a computational cost resulting in the increased number of parameters to be used by the 3D-CNNs. The various architecture of CNN are available (e.g., VGGNet, Xception, ResNet, DenseNet, etc.) and can be used to build models for histological image analysis.

## 3.2. A Typical Architecture of 3D CNN

A typical architecture of CNN may include four basic components:
(1) local receptive field,
(2) sharing weights,
(3) pooling and
(4) fully connected (fc) layers.

Deep CNN architecture is constructed by stacking several convolutional layers and pooling layers and one or so fully connected layers at the end of the network [53,54].
While 1D CNN can extract spectral features from the data, 2D CNN can extract spatial features from the input data. However, 3D CNNs can take advantage of both 1D and 2D CNNs by extracting both spectral and spatial features simultaneously from the input

volume. These 3D CNN features are very useful in analyzing the volumetric data in medical imaging. The mathematical formulation of 3D CNN is very similar to 2D CNN with an extra dimension added.

The basic architecture of 3D CNN is shown in Figure 3.1. Features are extracted from 3D images and classified as shown in Figure 3.1.

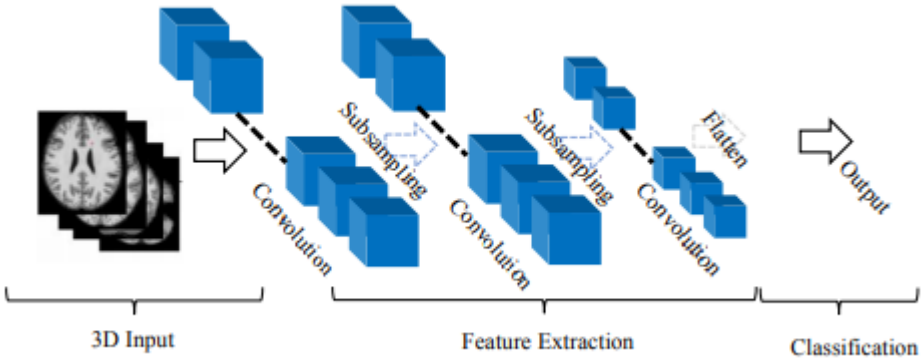



**Figure 3.1 Typical architecture of 3D CNN [55]**

The mathematical background of 3D CNN is discussed briefly.

## 3.2.1. Convolutional Layer:

The basic definition, principle, and working equation of 3D CNN is quite similar to 2D CNN. Only an extra dimension of depth is added to the working equation of 2D CNN. Suppose 3D CNN of input $x$ has a dimension of $M \times N \times D$ with $i,j,k$ as iterators. The kernel $\omega$ with dimensions $n1 \times n2 \times n3$ has iterator $a, b, c$. $l$ is denote the $l^{\text{th}}$, where $l = 1$ is the first layer and $l = L$ is the last layer.$yl$ and $b^l$ are denoted as the output and the bias unit the $l^{th}$ layer. To compute the nonlinear input $x^l_{i,j,k}$to $(i,j,k)$ $^{th}$ unit in layer $l$ , The weight contribution from the previous layer added up as follows:

$$x^l_{i,j,k} = \sum_a \sum_b \sum_c \omega_{a,b,c}\, y^{l-1}_{(i+a)(j+b)(k+c)} + b^l . \qquad (3.1)$$

The output of the (i,j)$^{\text{th}}$ until in the $\ell$ $^{th}$ convolutional layer is given as follows:

$$y^l_{i,j,k} = f\left(x^l_{i,j,k}\right). \qquad (3.2)$$

## 3.2.2. Pooling Layer:

Each feature map in the convolutional layer of 3D CNN can be a pooling layer. There are two kinds of pooling. If the pooling layer averages across the group of input voxels, it is called average pooling, while if it obtains a maximum of the input voxels, it is called maximum pooling. The output of the pooling layer will be the input of the next layer. Since a small shift in the input image results in a shift in activation function, the pooling

layer also introduces some translational invariance to the 3D CNN. To lower the sampling effect of pooling, the pooling layer can be removed by increasing the number of strides in the preceding CNN layer [56]. This will not result in any significant depreciation of the performance. However, by doing this, overlapping is happened in the CNN layer that precedes the pooling layer. This is simply equivalent to the pooling operation where only the top-left features are considered.

### 3.2.3. Dropout regularization:

Deep neural networks with a large number of parameters are very dominant learning systems. Multiple deep nonlinear hidden layers allow them to learn complex relationships between input and outputs. However, with the limited training data, these complex relationships introduce sampling noise, which appears in training data sets but not in real test datasets even if both are drawn from the same distribution. This scenario leads to overfitting and there have been several strategies [57] to tackle the problem, such as early stopping of the training epochs and weight penalties (L1 and L2 regularizations, soft weight sharing, and pooling). Ensemble models of several CNNs with different configurations on the same dataset are known for their overfitting. However, this leads to extra computational and maintenance costs for training several models. Moreover, training a large network requires large datasets, but the availability of such datasets in the field of medical imaging is very rare. Even if one can train large networks with a versatile setting of parameters, testing these networks is not feasible in a real-time situation due to the nature of medical imaging systems. In the case of ensemble models, a CNN model can also simulate multiple configurations just by probabilistically dropping out edges and nodes. Dropout is a kind of regularization technique to reduce overfitting by temporarily dropping a unit out of the network [58]. This simple idea shows a significant improvement in CNN performance.

### 3.2.4. Batch normalization:

Unit variance of the input of each hidden layer dynamically changes during training because the parameters in the previous layer update at each training epoch. If these changes are large, the search for an optimal hyperparameter becomes difficult for the network and may be computationally expensive to reach an optimal value. This problem can be solved by an algorithm called batch normalization, which was proposed by two researchers [59]. Batch normalization allows the use of a higher learning rate and thereby achieves the optimal value in less time. It facilitates the smooth training of deeper network architectures in less time. The normalization of data from a particular batch is about finding the mean and variance of the data points from mini-batch and normalizing them to have a zero mean and unit variance.

In backward pass, the CNN adjusts its weights and parameters according to the output by calculating the error through some loss functions, $e$ (other names are cost function and error function) and backpropagating the error with some rules towards the input. The loss is calculated by taking the partial derivative of $e$ w.r.t., which is the output of each neuron in that layer, such as $\frac{\partial e}{\partial y^{l}_{i,j,k}}$ for the output, $y^{l}_{i,j,k}$ of $(i,j,k)$ $^{th}$ unit in layer $l$. The chain rule allows us to write and add up the contribution of each variable as follows:

$$\frac{\partial e}{\partial x_{i,j,k}^{l}} = \frac{\partial e}{\partial y_{i,j,k}^{l}} \frac{\partial f(y_{i,j,k}^{l})}{\partial x_{i,j,k}^{l}}$$

$$= \frac{\partial e}{\partial y_{i,j,k}^{l}} f'\left(x_{i,j,k}^{l}\right)\#. \qquad (3.3)$$

Weights in the previous convolutional layer can be updated by backpropagating the error to the previous layer according to the following equation:

$$\frac{\partial e}{\partial y_{i,j,k}^{l-1}} =$$

$$\sum_{a=0}^{n1-1} \sum_{b=0}^{n2-1} \sum_{c=0}^{n3-1} \frac{\partial e}{\partial x_{(i-a)(j-b)(k-c)}^{l}} \frac{\partial x_{(i-a)(j-b)(k-c)}^{l}}{\partial y_{i,j,k}^{l-1}} \qquad (3.4)$$

$$= \sum_{a=0}^{n1-1} \sum_{b=0}^{n2-1} \sum_{c=0}^{n3-1} \frac{\partial e}{\partial x_{(i-a)(j-b)(k-c)}^{l}} \omega_{a,b,c}. \qquad (3.5)$$

Equation (3.5) allows us to calculate the error for the previous layer. Further, the above eq. makes sense for those points which are n times away from each side of the input data. This situation can be avoided by simply padding with zeros to the end of each side of the input volume.

## 3.3. Different types of Pre-trained Model

Some pre-trained models have been used in this work such as below:

### 3.3.1. VGG-19

VGG is a convolutional neural network that has a depth of 19 layers. The VGG-19 network is also trained using more than 1 million images from the ImageNet database. Naturally, you can import the model with the ImageNet trained weights. This pre-trained network can classify up to 1000 objects. The network was trained on 224x224 pixels colored images [60]. Here is a brief info about its size and performance:

- **Size:** 549 MB
- **Top-1:** Accuracy: 71.3%
- **Top-5:** Accuracy: 90.0%
- **Number of Parameters:** 143,667,240
- **Depth:** 26

Figure 3.2 shows the architecture of VGG19.which has depth of 19 layers.

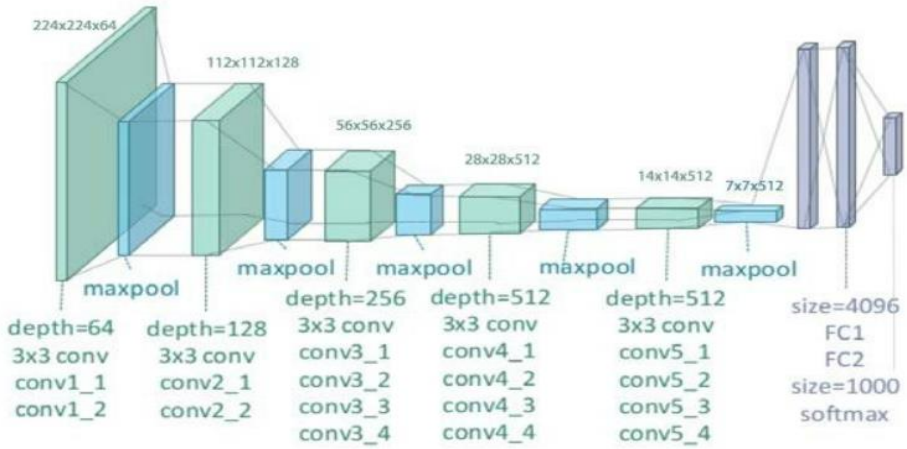



**Figure 3.2 An Illustration of the VGG-19 Network[60]**

### 3.3.2. ResNet (Residual Network)

ResNet is a convolutional neural network that has a depth of 50 layers. This model is also trained on more than 1 million images from the ImageNet database. Just like VGG-19, it can classify up to 1000 objects and the network was trained on 224x224 pixels colored images [60]. Here is brief info about its size and performance:

- **Size:** 98 MB
- **Top-1:** Accuracy: 74.9%
- **Top-5:** Accuracy: 92.1%
- **Number of Parameters:** 25,636,712

Figure 3.3 shows the illustration of ResNet architecture. this architecture introduced the concept called Residual Blocks.

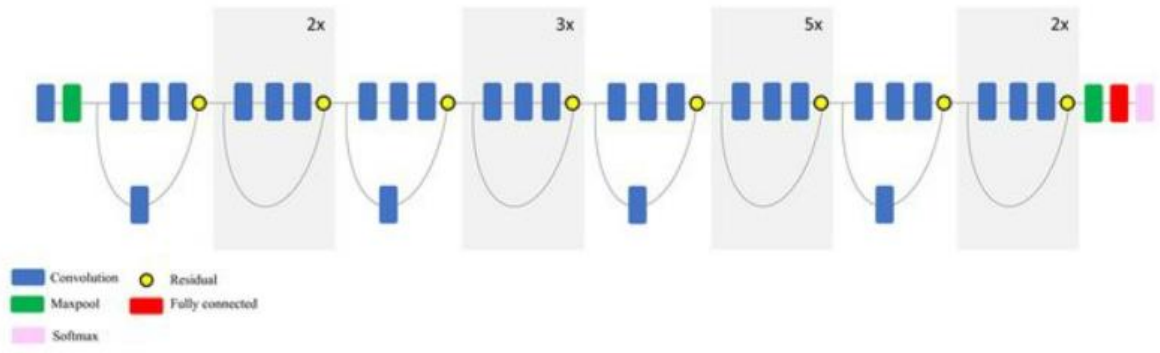



**Figure 3.3 An Illustration of the ResNet Network[60]**

### 3.3.3. Xception

The Xception is an extension of inception architecture that replaces the standard inception model with depth-wise separable convolutions. it slightly outperforms the ImageNet dataset compared to VGGNet and ResNet [60].

- **Size:** 88 MB
- **Top-1:** Accuracy: 79%
- **Top-5:** Accuracy: 94.5%
- **Number of Parameters:** 22,910,488
- **Depth:** 126

Figure 3.3 shows the workflow of Xception architecture. The data first goes through the entry flow, then through the middle flow which is repeated eight times, and finally through the exit flow.

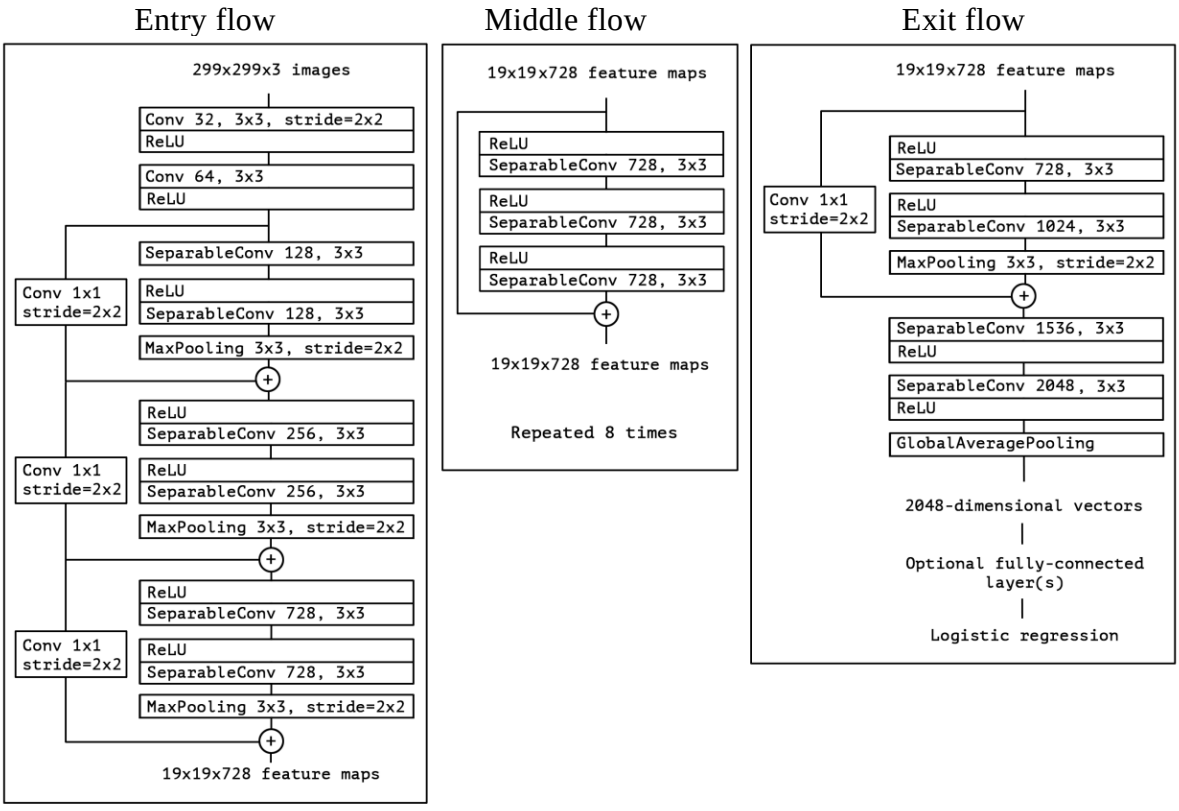



**Figure 3.4 An Illustration of the Xception Network[60]**

### 3.3.4. DenseNet201

DenseNet is one of the new discoveries in neural networks for visual object recognition. DenseNet is quite similar to ResNet with some fundamental differences. ResNet uses an additive method (+) that merges the previous layer (identity) with the future layer, whereas DenseNet concatenates (.) the output of the previous layer with the future layer[61].

- **Size:** 80 MB
- **Top-1:** Accuracy: 77.3%
- **Top-5:** Accuracy: 93.6%
- **Number of Parameters:** 20,242,984

Figure 3.5 shows the illustration of DenseNet architecture. each layer has direct access to the gradients from the loss function and the original input image. Traditional feed-forward neural networks connect the output of the layer to the next layer after applying a composite of operations.

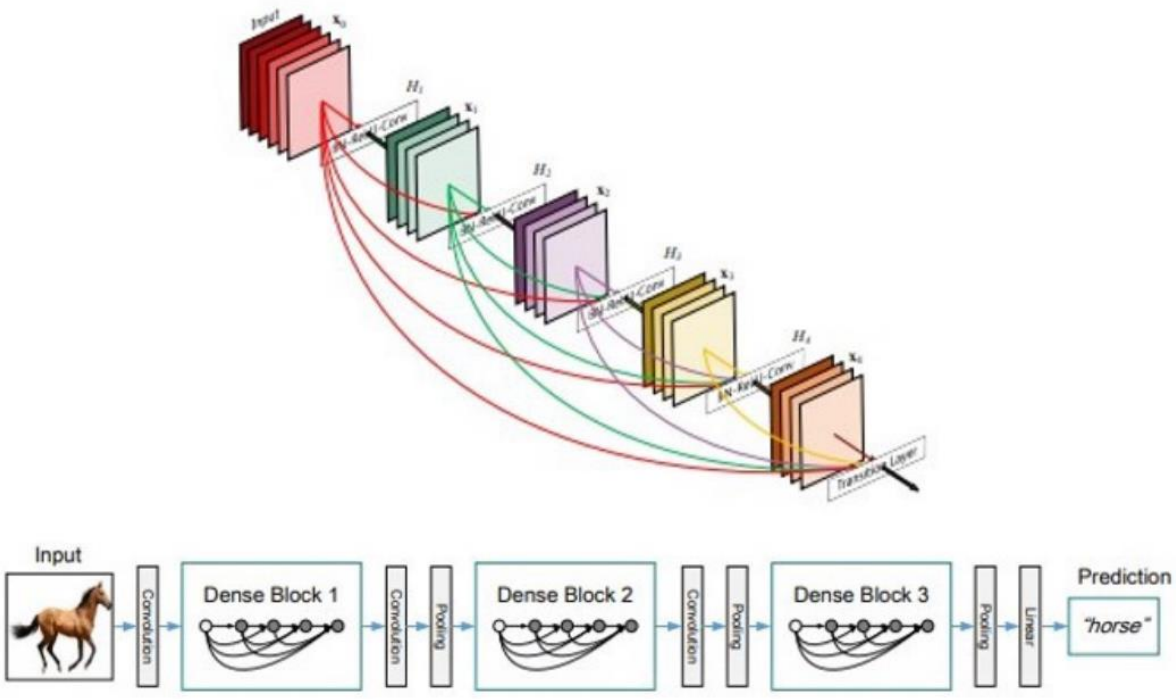



**Figure 3.5 DenseNet with 5 layers with a growth rate of k = 4[61]**

## 3.4. Loss Functions

The central components of machine learning algorithms are loss functions indicating which sample and how much it affects the model training. In other words, loss functions attribute to each sample a value, which indicates how much that sample is involved in the optimization problem. For example, If the employed loss function assigns a very large value to an outlier sample, the outlier may negatively impact the model's parameters [62].

Sometimes, even though that the learned classifier is able to classify the training data well, it fails to estimate unseen data (test), and it leads a high generalization error although the training error is low. A reason for this failure is often known as the overfitting problem, which means that the classifier fits on the training data and lose its ability of generalization. A solution for better generalization is to use a loss function resulting in a more generalized classifier.

The goal of a learning method is the ability to classify unseen data. Thus, the classifier should be robust to data perturbance. Sometimes training and testing data may be sampled from different processes, which are similar to some extent but are not identical. A more difficult scenario can happen in noisy environments where outliers corrupt the training data, testing data or both. To cope with very noisy environments, an effective approach is the use of a robust loss function.

## 3.5. Ensemble approach of deep learning models

Two major types of ensembles are used namely:

- Weighted Average ensemble
- Stacking ensemble with meta learners

### 3.5.1. Weighted Average Ensemble:

Ensemble technique that weights the prediction of each ensemble member, combining the weights to calculate a combined prediction. Weight optimization search is performed with randomized search based on the Dirichlet distribution on a test dataset [63].

Figure 3.6 shows workflow of weighted average ensemble method. Dirichlet distribution based randomized search is used for performing weight optimization.

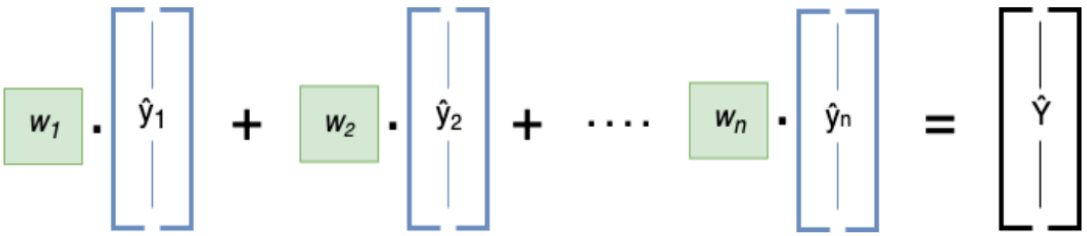



**Figure 3.6 Weighted Average ensemble [63]**

This method weights the contribution of each ensemble member based on their performance on a hold-out validation dataset. Models with better contributions receive a higher weight.

### 3.5.2. Stacking

Stacking is based on training a Meta-Learner on top of pre-trained Base-Learners.

In Stacking there are two types of learners called Base Learners and a Meta Learner. Base Learners and Meta Learners are the normal machine learning algorithms like Random Forests, SVM, Perceptron etc. Base Learners try to fit the normal data sets whereas Meta learner fit on the predictions of the base Learner [63].

Figure 3.7 shows workflow of stacking ensemble method. Meta Learner tries to find the optimal combination of base learners.

.

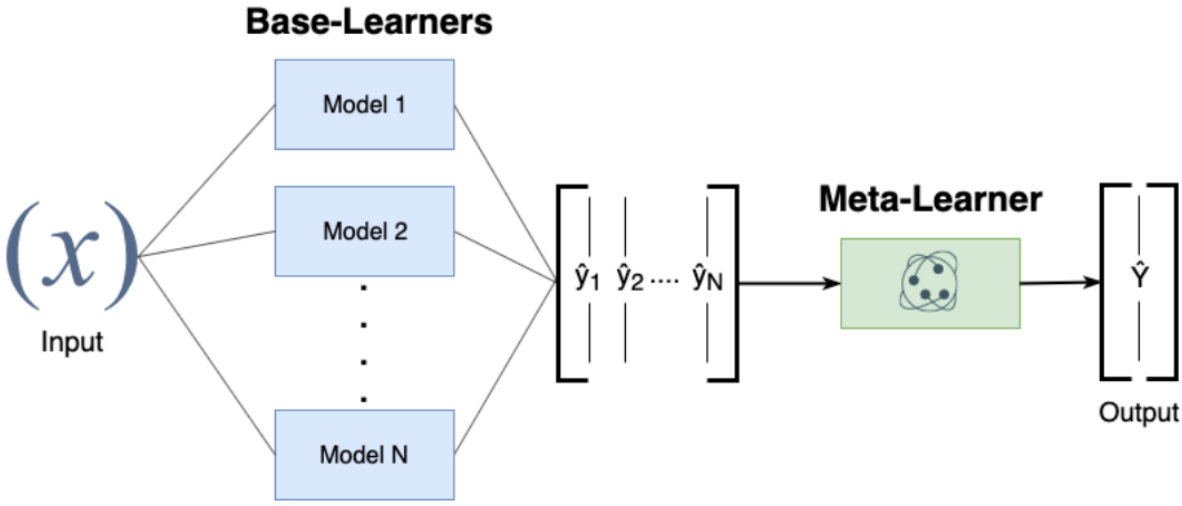



**Figure 3.7 Stacking ensemble [63]**

Stacking Technique involves the following steps:

1. Split the training data into 3 disjoint sets
2. Train several Base Learners on the first part
3. Test the Base Learners on the second part and make predictions
4. Using the predictions from (3) as inputs, the correct responses from the output, train the higher-level learner.

## 3.6. The Proposed Loss Function

In this work, proposed Ensemble loss function is a combination of MAE, Hinge and Cross-Entropy loss functions which are leading to a more robust, generalized and improved probability estimator respectively. Using these three loss functions, It is assumed that the classifier performance will increase significantly [64].

Let (x, y) be a sample where x ∈ Rd is the input and y ∈ {0, 1}C is one-hot encoding of the label (C is the number of classes).

Let θ be the parameters of neural network classifier with a top softmax layer so that the probability estimates are $\hat{y}$ = softmax(f(x; θ ))

Let $\{L_j(y, \hat{y})\}_{j=1}^{k}$ denote K single loss functions.

In addition to finding the optimal parameter of the neural network the goal is to find the best weights, $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$, to combine K base loss functions in order to generate a better application-tailored loss function. A further constraint needs to be added to avoid yielding near zero values for all $\lambda_j$weights. The proposed loss function is defined as below.

$$L = \sum_{j=1}^{k} \lambda_j L_j(y, \hat{y}), \qquad \sum_{j=1}^{k} \lambda_j = 1 \tag{3.6}$$

In particular, the ensemble loss function is a linear combination of three loss functions: 1) binary cross-entropy, 2) Mean absolute error (MAE), and 3) hinge. The ensemble loss function inherits the advantages of each.

1. **Binary Cross-Entropy** is a very popular loss in neural networks. The minimization of the average value of the cross-entropy loss function over data means maximization of the conditional log-likelihood of the data.
2. **Mean absolute error (MAE)** is less sensitive to outliers since it is a bounded function. The MAE loss, therefore, leads to robustclassifiers.
3. **Hinge** is margin enhancing which leads to a better generalization.

In this study, Binary cross-entropy, MAE, and hinge are used for the ensemble and weights for these loss functions are used 0.7, 0.15, and 0.15 respectively while combining.

**Table 3.1 Weights of each loss function in the ensemble loss**

| Loss Function | Weights |
|---|---|
| Binary Cross-Entropy | 0.7 |
| Mean Absolute Error | 0.15 |
| Hinge | 0.15 |

In Figure 3.8, it shows model and the way, weights of neural network are updated through backpropagation.

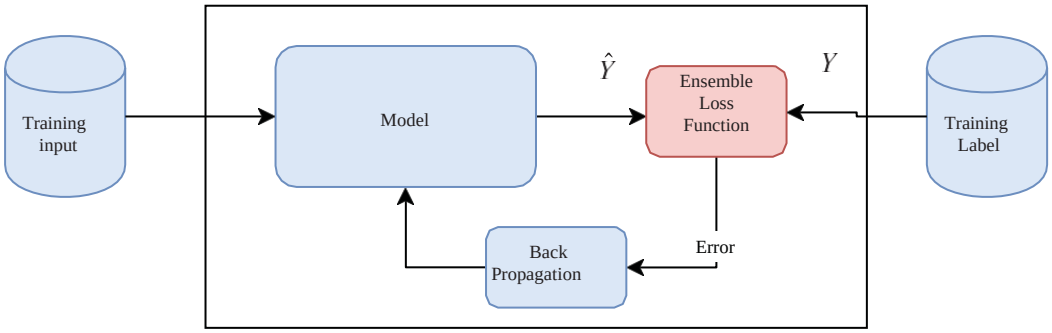



**Figure 3.8 Model Workflow structure with ensemble loss function**

## 3.7. The Proposed Model

The goal of combining various pre-trained models is to observe improvement in the performance of predicting benign and malignant images.

For classifying histopathological images into IDC + and IDC - categories, four different deep CNNs DenseNet, ResNet, Xception and VGG19 are trained on histological images.

In this work, different scenarios have been experimented such as:

First scenario – Above mentioned four deep CNNs are trained individually. Output is passed to a fully connected layer that performs the classification using a softmax activation function. The loss function binary cross entropy is used and back propagation is implemented by using Adam algorithm.

Second scenario – Mentioned four pre-trained models are trained individually. Output is passed to a fully connected layer that performs the classification through a softmax activation function. In this scenario, loss function Hinge is used and back propagation is implemented by using Adam algorithm.

Third scenario – Similarly, all four deep CNN architectures are trained individually having a fully connected layer that performs the classification using a softmax activation with loss function MAE and back propagation is implemented by using Adam algorithm.

Forth Scenario – In this scenario, all four deep CNN models are trained having a fully connected layer which performs the classification employing a softmax activation. The proposed ensemble loss function and back propagation is implemented by using Adam algorithm. As using the ensemble loss function, the accuracy of models is improved, ensemble loss function is used for further experiment.

Fifth Scenario – In this scenario, one additional convolutional layer is added with every pre-trained model followed by ReLU activation and batch normalization layer. Output is

passed to a fully connected layer that performs the classification by through a softmax activation function. The proposed ensemble loss function and back propagation is implemented by using Adam algorithm.

Sixth Scenario – In this scenario, two additional convolutional layer is added with every pre-trained model. All layers except the final fully connected layer are followed by ReLU activation and batch normalization layer. Output is passed to a fully connected layer that performs the classification by means of a softmax activation function. The proposed ensemble loss function and back propagation is implemented by using Adam algorithm.

Above mentioned scenarios are made to observe the performance of hybrid models and based on the best accuracy, hybrid models are finalized for ensemble approach.

In this work, Both Stacking and Weighted Average ensemble are performed and based on better performance, the most effective model has been chosen.

## 3.8. The hyperparameter settings of the proposed model

Below hyperparameters are used for fitting models. Callback function ReduceLROnPlateau is used for reducing the learning rate while training. It reduces the learning rate when a metric has stopped improving. Models often benefit from reducing the learning rate if no improvement is seen for a number of epochs.

In Table 4.2, hyperparameters are set as below to fit models.

**Table 3.2 Hyperparameter Settings**

| Variable | Setting |
|---|---|
| Epochs | 150 |
| Batch Size | 100 |
| Learning rate schedule | Adam |
| Learning rate | max - 0.0010 and min - 0.00001 |
| Loss function | Proposed loss function |

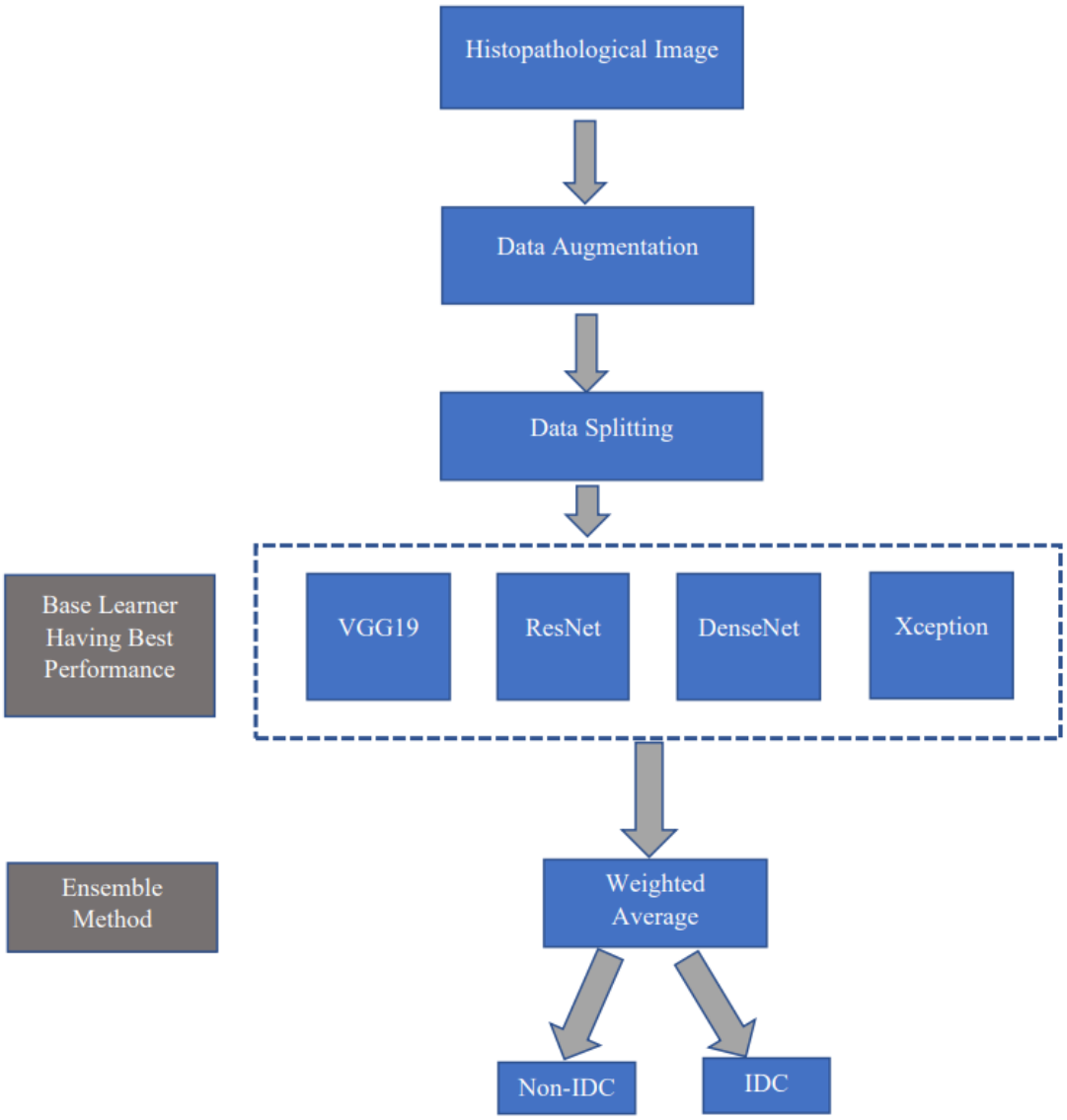



**Figure 3.9 Block Diagram of Proposed Weighted Average Ensemble Model**

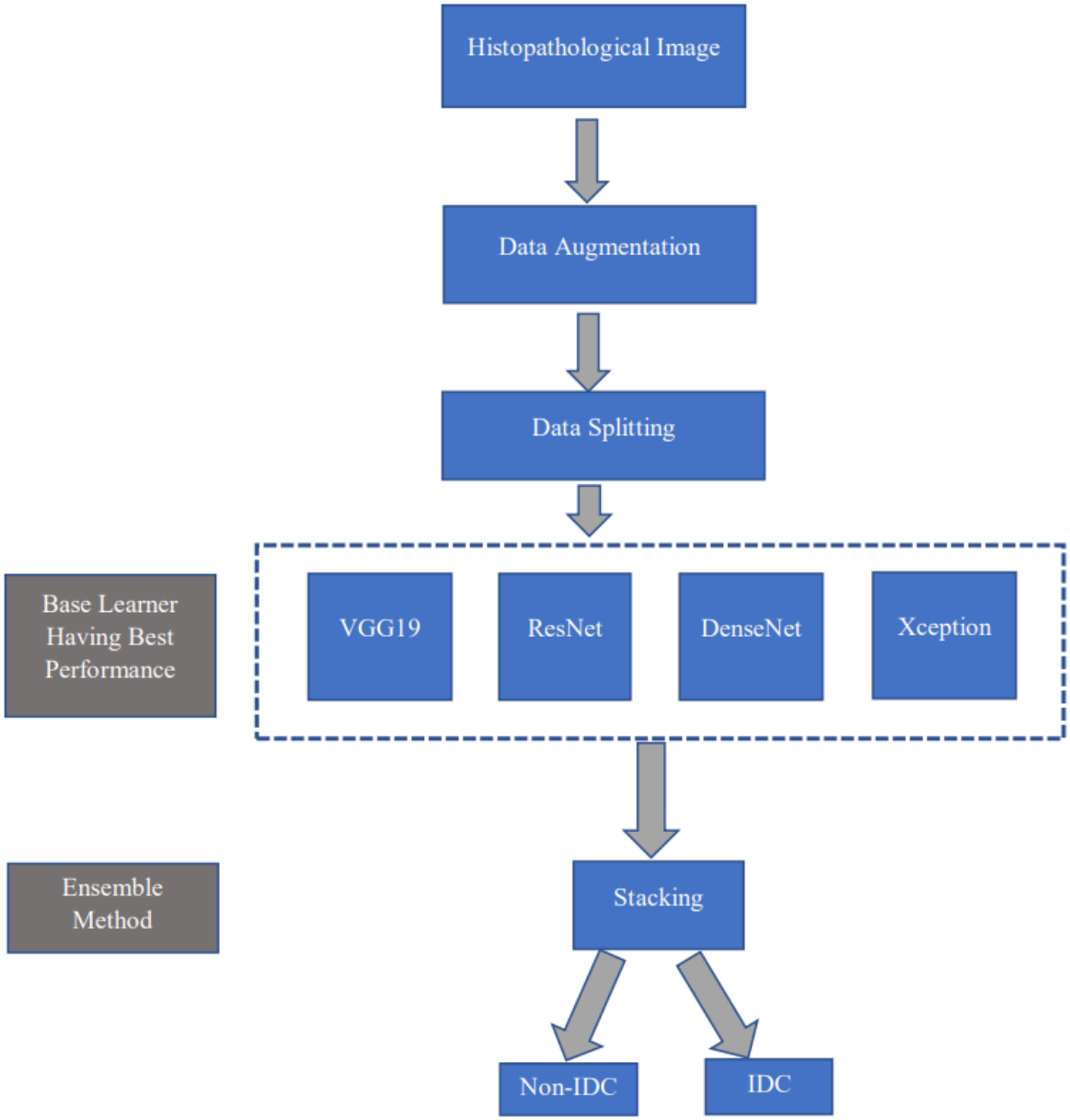



**Figure 3.10 Block Diagram of Proposed Stacking Ensemble Model**

## 3.9. Individual models based on the best accuracy

Below are the individual models based on best accuracy.

### 3.9.1. **MODEL 1**: Base model ResNet with an additional convolutional layer and Ensemble loss function

ResNet152V2 is used here and one layer is added with it as below.

**Table 3.3 MODEL 1 Network Structure**

| Layer (type) | Output Shape |
|---|---|
| resnet152v2 (Functional) | (None, 3, 3, 2048) |
| conv2d (Conv2D) | (None, 3, 3, 64) |
| max_pooling2d_3 (MaxPooling2D) | (None, 1, 1, 64) |
| dropout (Dropout) | (None, 1, 1, 64) |
| flatten (Flatten) | (None, 64) |
| batch_normalization (BatchNormalization) | (None, 64) |
| dropout_1 (Dropout) | (None, 64) |
| dense (Dense) | (None, 2) |

### 3.9.2. **MODEL 2:** Base model DenseNet with an additional convolutional layer and Ensemble loss function

DenseNet201 is used here and one layer is added with it as below.

**Table 3.4 MODEL2 Network Structure**

| Layer (type) | Output Shape |
|---|---|
| densenet201 (Functional) | (None, 2, 2, 1920) |
| conv2d (Conv2D) | (None, 2, 2, 64) |
| max_pooling2d (MaxPooling2D) | (None, 1, 1, 64) |
| dropout (Dropout) | (None, 1, 1, 64) |
| flatten (Flatten) | (None, 64) |
| batch_normalization (BatchNormalization) | (None, 64) |
| dropout_1 (Dropout) | (None, 64) |
| dense (Dense) | (None, 2) |

**3.9.3. MODEL 3**: Base model VGG19 with two additional convolutional layer and Ensemble loss function

VGG19 is used here and one layer is added with it as below.

**Table 3.5 MODEL 3 Network Structure**

| Layer (type) | Output Shape |
|---|---|
| vgg19 (Functional) | (None, 2, 2, 512) |
| conv2d_4 (Conv2D) | (None, 2, 2, 64) |
| max_pooling2d_1 (MaxPooling2D) | (None, 1, 1, 64) |
| dropout_5 (Dropout) | (None, 1, 1, 64) |
| conv2d_5 (Conv2D) | (None, 1, 1, 128) |
| dropout_6 (Dropout) | (None, 1, 1, 128) |
| flatten_1 (Flatten) | (None, 128) |
| batch_normalization_1 (BatchNormalization) | (None, 128) |
| dropout_7 (Dropout) | (None, 128) |
| dense_1 (Dense) | (None, 2) |

**3.9.4. MODEL 4:** Base model Xception with an additional convolutional layer and Ensemble loss function

Xception is used here and one layer is added with it as below.

**Table 3.6 MODEL 4 Network Structure**

| Layer (type) | Output Shape |
|---|---|
| xception (Functional) | (None, 3, 3, 2048) |
| conv2d_4 (Conv2D) | (None, 3, 3, 64) |
| max_pooling2d (MaxPooling2D) | (None, 1, 1, 64) |
| dropout (Dropout) | (None, 1, 1, 64) |
| flatten (Flatten) | (None, 64) |
| batch_normalization_4 (BatchNormalization) | (None, 64) |
| dropout_1 (Dropout) | (None, 64) |

# Chapter 4

# Results and Discussion

## 4.1. Dataset Description

### 4.1.1. Original Dataset

In this work, the IDC image dataset publicly available in Kaggle [65] is selected as the original dataset. The dataset originated from pioneering research published in Ref. [66]. The image dataset consists of 277,524 patches of size 50 X 50 px images extracted from hundreds of IDC whole slide images. Each image patch was individually labeled with a positive or negative IDC class. This same dataset was also used in another published research to verify their custom neural network model [67].
This dataset of 277,524 .png files are downloaded from Kaggle which includes 198,738 negative and 78,786 positive images. These image files are organized in a hierarchical structure of patient ID > label (0 or 1) > list of image patches. There are 279 patient IDs in the file structure. The name convention of each image file is in a format of "PID_xX_yY_classC.png", such as "10260_idx5_x1101_y1501_class0.png", where:
PID: patient ID
X: x-coordinate of slide image where the patch was cropped from
Y: y-coordinate of slide image where the patch was cropped from
C: IDC label. 0 for negative and 1 for positive IDC
It was reported that the original slide images were scanned at 40x microscope magnification. However, after further examination, the resolution of these downloaded image patches is much lower than 40x. As indicated by the other research, the resolution of these original 40x images were likely downscaled by 16 times to be around 2.5x primarily due to concerns of image file size. In this study, these patches are treated as scanned by 2.5x and scale down our own test images to the same resolution.

Figure 4.1 shows randomly selected samples of IDC negative and positive patches from this dataset.

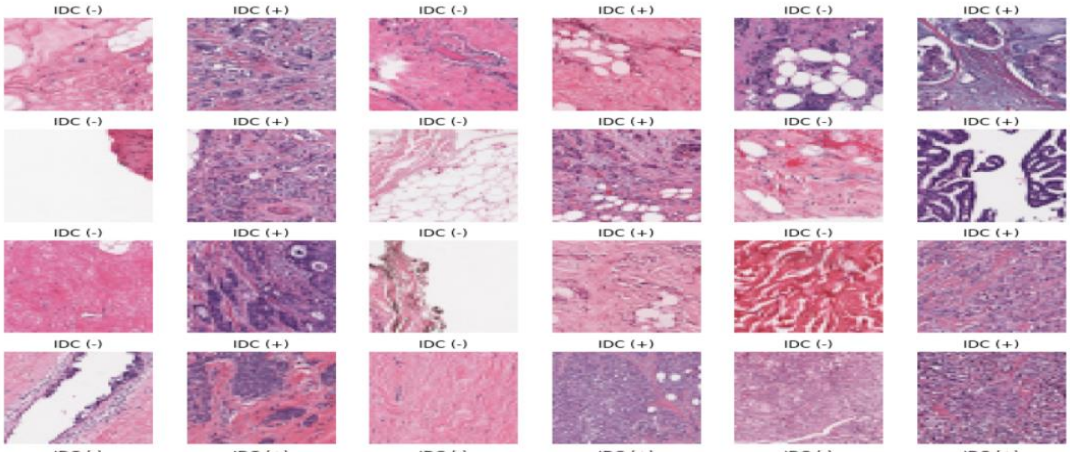
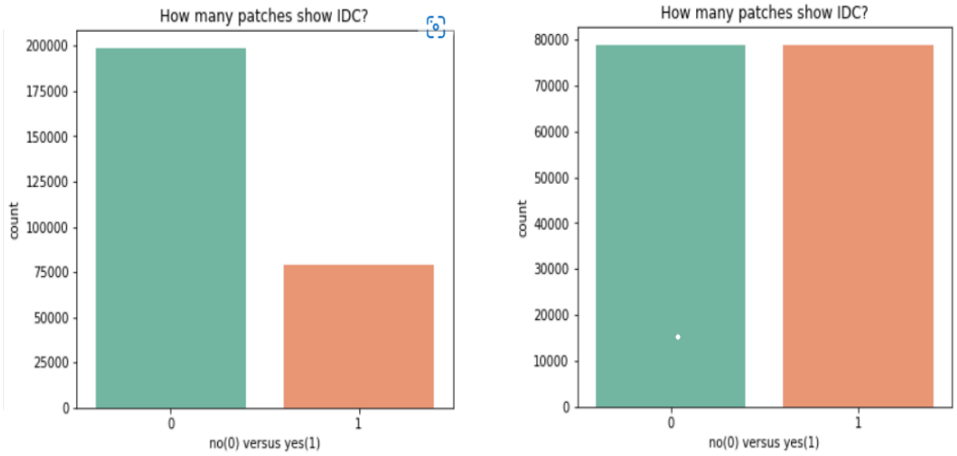



**Figure 4.1 Examples of image patches (50 X 50) from datasets.**

### 4.1.2. Data augmentation

The original dataset is heavily imbalanced. Each of the 279 patients has a various number of positive and negative image patches. Altogether, there are 198,738 negative patches (72%) and only 78,786 positive ones (28%). An imbalanced dataset is likely to introduce bias in the learning of a ML model. Therefore, It is decided to augment the original dataset to balance out the number of positive and negative IDC patches. 78,786 images are have been taken as sample size. Hence, total number of 78,786 negative (50%) and 78,786 positive sample patches (50%) are present in the dataset.

In Figure 4.2, it shows the distribution of IDC and Non-IDC patches. a) shows that dataset is highly imbalanced. b) shows after balancing the dataset.




**a)** **b)**

**Figure 4.2 distributions of IDC negative vs positive patches a)imbalanced dataset b)balanced dataset**

There are currently many different techniques for image augmentation, such as scaling, flipping, padding, rotation, translation. In this work, a large portion of positive patches have been rotated by 90 and normalizes as a part of data augmentation.

### 4.1.3. Dataset Split

The augmented dataset is split into training (60%), validation (20%) and test (20%) sets.

Figure 4.3 shows, augmented dataset is split into 94543 training image patches, 31514 validation image patches, and 31515 test image patches.

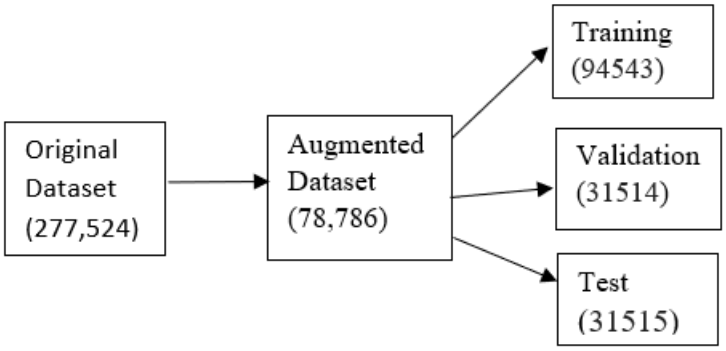



**Figure 4.3 Dataset splits for Models.**

## 4.2. Evaluation Metrics

Result of the present work are measured in terms of accuracy and based on best accuracy individual models are chosen for ensemble and evaluation of final ensembled model has been done using precision,recall and f-measure.

- TruePositives = Both True and predicted labels are positive.
- FalsePositives = True label is negative but predicted label is positive
- FalseNegatives = True label is positive but predicted label is negative
- TrueNegatives = Both True and predicted labels are positive.

- Accuracy represents the number of correctly classified data instances over the total number of data instances.

  Accuracy = (TruePositives + TrueNegatives) /
  (TruePositives + TrueNegatives + FalsePositives + FalseNegatives)

- Precision quantifies the number of positive class predictions that actually belong to the positive class. It can be evaluated as

  Precision = TruePositives / (TruePositives + FalsePositives)

- Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

Recall = TruePositives / (TruePositives + FalseNegatives)

- F-Measure provides a single score that balances both the concerns of precision and recall in one number.

F-Measure = (2 * Precision * Recall) / (Precision + Recall)

## 4.3. Experimental Outcomes and Analysis

In this section, the experimental results are carried out using this model.
In the present work, four pre-trained models are selected for IDC classification and accuracy are measured according to scenarios which are mentioned in section 3.9.
For above mentioned scenarios, accuracies are measured and based on best accuracy, best base learner models have been selected and ensemble methods are used on the selected base learner and evaluated evaluation metrics such as precision, accuracy, recall and f-measure.

**First Scenario** - Only base model with loss function Binary Cross-entropy.

In Table **Results for First scenario,** accuracy is measured for all pre-trained models. Loss function binary cross-entropy is used and Its weight is 1.

**Table 4.1 Results for First scenario**

| Model | Accuracy(%) |
|---|---|
| ResNet (base model) | 87.58 |
| DenseNet (base model) | 90.19 |
| VGG19 (base model) | 81.76 |
| Xception (base model) | 80.25 |

**Second Scenario** - Only base model with loss function Hinge.

In Table **Results for Second scenario,** accuracy is measured for all pre-trained models. Loss function Hinge is used and Its weight is 1.

**Table 4.2 Results for Second scenario**

| Model | Accuracy(%) |
|---|---|
| ResNet (base model) | 87.58 |
| DenseNet (base model) | 90.19 |
| VGG19 (base model) | 81.76 |
| Xception (base model) | 80.25 |

**Third Scenario** - Only base model with loss function MAE.

In Table **Results for Third scenario,** accuracy is measured for all pre-trained models. Loss function MAE is used and Its weight is 1.

**Table 4.3 Results for Third scenario**

| Model | Accuracy(%) |
|---|---|
| ResNet (base model) | 85.69 |
| DenseNet (base model) | 89.75 |
| VGG19 (base model) | 87.52 |
| Xception (base model) | 86.28 |

**Forth Scenario** - Only base model with loss function Ensemble loss function.

In Table **Results for Forth scenario,** accuracy is measured for all pre-trained models. The ensemble loss function is used and it is observed that models trained with ensemble loss function perform better. Hence proposed loss function will be used for next two experiments.

**Table 4.4 Results for Forth scenario**

| Model | Accuracy(%) |
|---|---|
| ResNet (base model) | 95.01 |
| DenseNet (base model) | 94.73 |
| VGG19 (base model) | 92.95 |
| Xception (base model) | 94.69 |

**Fifth Scenario** - Base model with one additional convolutional layer and Ensemble loss function is used.

In Table **Results for Fifth scenario,** accuracy is measured for all pre-trained models with one additional convolutional layer after the base model. The ensemble loss function is used.

**Table 4.5 Results for Fifth scenario**

| Model | Accuracy(%) |
|---|---|
| ResNet (base model) with One Convolutional layer | **96.9** |
| DenseNet (base model) with One Convolutional layer | **97.41** |
| VGG19 (base model) with One Convolutional layer | 92,60 |
| Xception (base model) with One Convolutional layer | **97.9** |

**Sixth Scenario** - Base model with two additional convolutional layer and Ensemble loss function is used.

In Table **Results for Sixth scenario,** accuracy is measured for all pre-trained models with two additional convolutional layer after base model. The ensemble loss function is used.

**Table 4.6 Results for Sixth scenario**

| Model | Accuracy(%) |
|---|---|
| ResNet (base model) with Two Convolutional layer | 96.85 |
| DenseNet (base model) with Two Convolutional layer | 97,25 |
| VGG19 (base model) with Two Convolutional layer | **96.35** |
| Xception (base model) with Two Convolutional layer | 97.62 |

**Final Model:** Models based on best Accuracy as defined in chapter 3

In Table **Results for Final Model**, Models are selected having best accuracy from the above six tables. Best accuracies are highlighted in bold. Then those selected models are ensembled using stacking ensemble and weighted average ensemble method and accuracy, precision, recall and f-measure is measured for all models.

**Table 4.7 Results for Final Model**

| Model | Accuracy(%) | Precision | Recall | f-measure |
|---|---|---|---|---|
| MODEL 1 | 96.91 | 0.9706 | 0.9706 | 0.9706 |
| MODEL 2 | 97.43 | 0.9727 | 0.9727 | 0.9727 |
| MODEL 3 | 96.31 | 0.9614 | 0.9614 | 0.9614 |
| MODEL 4 | 97.93 | 0.9785 | 0.9785 | 0.9785 |
| StackEnsemble | 93.23 | 0.9312 | 0.9377 | 0.9345 |
| DirichletEnsemble | 98.11 | 0.9805 | 0.9805 | 0.9803 |

After several experiments, it is concluded that the weighted Average ensemble performs better and provides the best accuracy.

## Chapter 5

# Conclusion and Future Direction

Breast cancer is a deadly disease among women that kills more than half a million people every year. However, the chances of survival and longevity increase substantially with early detection. According to the American Cancer Society, the 5-year survival rate for detection in the localized stage is 99%, i.e., 99 out of 100 patients detected in the localized stage survive 5 years or beyond. On the other hand, the survival rate drops to 88% if the cancer is regional. Hence, early detection can go a long way in treatment, and image classification using ML and DL techniques can expedite this process.

In this work, it is observed the impact and contribution of different deep CNN models and loss functions. Different classifier models along with different loss functions were investigated, and various combinations were used to produce the highest accuracy. An ensemble of different classifier models was designed after observing the correlations in their individual predictive capacities. Finally, best ensembled model achieved an accuracy of 98.11%. However, further improvements can still be made.

In spite of this achievement, there are still some areas for improvement. First and notable, high-dimensional feature extraction can be used for detecting IDC. Therefore, the application of a feature selection algorithm would be a worthy choice to increase the performance of the proposed approach.

Feature selection based on optimization algorithms can also be employed to improve the predictive power. It is also noteworthy that the dataset is highly imbalanced. This can be rectified by using Generative Adversarial Networks (GANs) to create IDC(+) histopathology images.

The results obtained in this work are encouraging preliminary to continue with further research in different directions. One continuation path might be improving this work in order to manage possible scalability problem that presents some problems to solve, such as different standard datasets, different model structure or variations in structure. This will lead to ease in the process of CAD system in breast cancer diagnosis in near future.

# References


[1] J. Ferlay, I.Soerjomataram , M. Ervik , R. Dikshit , S. Eser ,C. Mathers et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11Lyon, France: International Agency for Research on Cancer; 2013

[2] F. Bray, J. Ferlay, I. Soerjomataram, R.L.Siegel,L.A. Torre, and A. Jemal, 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 68(6), pp.394-424.

[3] Siegel, L. Rebecca, Kimberly D. Miller, and J. Ahmedin. "Cancer statistics, 2019." CA: a cancer journal for clinicians69.1 (2019): 7-34.

[4] Chandra Churh Chatterjee and Gopal Krishna "A Novel method for IDC Prediction in Breast Cancer Histopathology images using Deep Residual Neural Networks" , ICCT, Sep 28-29, 2019

[5] MOHAMMED ABDULLA SALIM AL HUSAINI, MOHAMED HADI HABAEBI, SHIHAB A. HAMEED, MD. RAFIQUL ISLAM AND TEDDY SURYA GUNAWAN "A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks" P-RIGS18-003-0003

[6] American cancer socity, "Types of Breast Cancer," https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer.

[7] Rashmi R, Keerthana Prasad,Chethana Babu K Udupa "Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review", https://doi.org/10.1007/s10916-021-01786-9

[8] R. Lakhtakia, ''A brief history of breast cancer: Part I: Surgical domination reinvented,'' Sultan Qaboos Univ. Med. J., vol. 14, no. 2, pp. 166–169, 2014.

[9] B. W. C. Amalu, ''A review of breast thermography,'' Int. Acad. Clin. Thermol., p. 112, 2003. [Online]. Available: https://www.iact-org.org/articles/ articles-review-btherm.html

[10] R. Lakhtakia, ''A brief history of breast cancer: Part I: Surgical dom ination reinvented,'' Sultan Qaboos Univ. Med. J., vol. 14, no. 2, pp. 166–169, 2014.

[11] H. Koch, ''Mammography as a method for diagnosing breast cancer,'' Radiologia Brasileira, vol. 49, no. 6, p. 7, Dec. 2016, doi: 10.1590/0100- 3984.2016.49.6e2.

[12] P. J. Dempsey, ''The history of breast ultrasound,'' J. Ultrasound Med., vol. 23, no. 7, pp. 887–894, Jul. 2004, doi: 10.7863/jum.2004.23.7.887.

[13] D. A. Kennedy, T. Lee, and D. Seely, ''A comparative review of thermography as a breast cancer screening technique,'' Integrative Cancer Therapies, vol. 8, no. 1, pp. 9–16, Mar. 2009, doi: 10.1177/1534735408326171.

[14] X. Yao, W. Wei, J. Li, L. Wang, Z. Xu, Y. Wan, K. Li, and S. Sun, ''A comparison of mammography, ultrasonography, and far-infrared thermography with pathological results in screening and early diagnosis of breast cancer,'' Asian Biomed., vol. 8, no. 1, pp. 11–19, Feb. 2014, doi: 10.5372/1905- 7415.0801.257.

[15] A. Hossam, H. M. Harb, and M. A. E. K. Hala, ''Performance analysis of breast cancer imaging techniques,'' Int. J. Comput. Sci. Inf. Secur., vol. 15, no. 5, pp. 48–56, 2018.

[16] S. P. Power, F. Moloney, M. Twomey, K. James, O. J. O'Connor, and M. M. Maher, ''Computed tomography and patient risk: Facts, perceptions and uncertainties,'' World J. Radiol., vol. 8, no. 12, p. 902, 2016, doi: 10. 4329/wjr.v8.i12.902.

[17] A. Bhide, S. Datar, and K. Stebbins, ''Case histories of significant medical advances: Gastrointestinal endoscopy,'' Harvard Bus. School Accounting Manage. Unit Working Paper 20-005, Jul. 2020. [Online]. Available: https://ssrn.com/abstract=3429986 or http://dx.doi.org/ 10.2139/ssrn.3429986

[18] S. N. Prasad and D. Houserkova, ''The role of various modalities in breast imaging,'' Biomed. Papers, vol. 151, no. 2, pp. 209–218, 2007, doi: 10.5507/bp.2007.036.

[19] H. H. Aghdam and E. J. Heravi, Guide to Convolutional Neural Networks: A Practical Application to Traffic Sign Detection and Classification. Cham, Switzerland: Springer, 2017.

[20] Bonnema, Jorien, et al. "Ultrasound-guided aspiration biopsy for detection of nonpalpable axillary node metastases in breast cancer patients: new diagnostic method." World journal of surgery 21.3 (1997): 270-274.

[21] Ponraj, D. Narain, et al. "A survey on the preprocessing techniques of mammogram for the detection of breast cancer." Journal of Emerging Trends in Computing and Information Sciences 2.12 (2011): 656-664.

[22] M. A., Aswathy,and M. Jagannath. "Detection of breast cancer on digital histopathology images: Present status and future possibilities." Informatics in Medicine Unlocked 8 (2017): 74-79.

[23] Puspanjali Mohapatra, Baldev Panda, Samikshya Swain, "Enhancing Histopathological Breast Cancer Image Classification using Deep Learning", ISSN: 2278-3075, Volume-8 Issue-7, May, 2019

[24] Q. Li, et al. "Medical image classification with convolutional neural network." 2014 13th international conference on control automation robotics & vision (ICARCV). IEEE, 2014.

[25] G. Litjens, et al. "A survey on deep learning in medical image analysis." Medical image analysis 42 (2017): 60-88

[26] N. Brancati et al. "A Deep Learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images." DOI 10.1109/ACCESS.2019.2908724

[27] Doyle, S.; Agner, S.; Madabhushi, A.; Feldman, M.; Tomaszewski, J. Automated grading of breast cancer histopathology using spectral clusteringwith textural and architectural image features. In Proceedings of the 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, 14–17 May 2008.

[28] Dundar, M.M.; Badve, S.; Bilgin, G.; Raykar, V.; Jain, R.; Sertel, O.; Gurcan, M.N. Computerized classification of intraductal breast lesions using histopathological images. IEEE Trans. Biomed. Eng. 2011, 58, 1977–1984.

[29] Niwas, S.I.; Palanisamy, P.; Zhang, W.; Isa, N.A.M.; Chibbar, R. Log-gabor wavelets based breast carcinoma classification using least square support vector machine. In Proceedings of the 2011 IEEE International Conference on Imaging Systems and Techniques, Batu Ferringhi, Malaysia, 17–18 May 2011.

[30] Kral, P.; Lenc, L. LBP features for breast cancer detection. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016.

[31] Yasiran, S.S.; Salleh, S.; Mahmud, R. Haralick texture and invariant moments features for breast cancer classification. AIP Conf. Proc. 2016.

[32] Narayanan, B.N.; Krishnaraja, V.; Ali, R. Convolutional Neural Network for Classification of Histopathology Images for Breast Cancer Detection. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019.

[33] Debelee, T.G.; Amirian, M.; Ibenthal, A.; Palm, G.; Schwenker, F. Classification of mammograms using convolutional neural network based feature extraction. In Proceedings of the International Conference on Information and Communication Technology for Develoment for Africa, Bahir Dar, Ethiopia, 25–27 September 2017; Springer: Berlin, Germany, 2017; pp. 89–98.
[34] Debelee, T.G.; Gebreselasie, A.; Schwenker, F.; Amirian, M.; Yohannes, D. Classification of mammograms using texture and cnn based extracted features. J. Biomim. Biomater. Biomed. Eng. 2019, 42, 79–97.
[35] Rahman, M.J.U.; Sultan, R.I.; Mahmud, F.; Ahsan, S.A.; Matin, A. Automatic System for Detecting Invasive Ductal Carcinoma Using Convolutional Neural Networks. In Proceedings of the TENCON 2018–2018 IEEE Region 10 Conference, Jeju, Korea, 28–31 October 2018.
[36] Romano, A.M.; Hernandez, A.A. Enhanced Deep Learning Approach for Predicting Invasive Ductal Carcinoma from Histopathol ogy Images. In Proceedings of the 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 25–28 May 2019.
[37] Cruz-Roa, A.; Basavanhally, A.; González, F.; Gilmore, H.; Feldman, M.; Ganesan, S.; Shih, N.; Tomaszewski, J.; Madabhushi, A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In Medical Imaging 2014: Digital Pathology; Gurcan, M.N., Madabhushi, A., Eds.; SPIE: Bellingham, WA, USA, 2014.
[38] Wang, J.L.; Ibrahim, A.K.; Zhuang, H.; Ali, A.M.; Li, A.Y.; Wu, A. A Study on Automatic Detection of IDC Breast Cancer with Convolutional Neural Networks. In Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 12–14 December 2018.
[39] Sanyal, R.; Jethanandani, M.; Sarkar, R. DAN: Breast Cancer Classification from High-Resolution Histology Images Using Deep Attention Network. In Advances in Intelligent Systems and Computing; Springer: Singapore, 2020; pp. 319–326.
[40] Sanyal, R.; Kar, D.; Sarkar, R. Carcinoma type classification from high-resolution breast microscopy images using a hybrid ensemble of deep convolutional features and gradient boosting trees classifiers. IEEE/ACM Trans. Comput. Biol. Bioinform. 2021.
[41] Chapala, H.; Sujatha, B. ResNet: Detection of Invasive Ductal Carcinoma in Breast Histopathology Images Using Deep Learning. In Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2–4 July 2020.
[42] Debelee, T.G.; Schwenker, F.; Ibenthal, A.; Yohannes, D. Survey of deep learning in breast cancer image analysis. Evol. Syst. 2020, 11, 143–163.
[43] Roy, S., Das, S., Kar, D., Schwenker, F., & Sarkar, R. (2021). Computer Aided Breast Cancer Detection using Ensembling of Texture and Statistical Image Features. Sensors.
[44] G. Litjens, et al. "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis." Scientific reports 6 ,2016: 26286.
[45] T.Araújo, G. Aresta, E. Castro., J., Rouco, P. Aguiar, C. Eloy, , ... & A. Campilho, Classification of breast cancer histology images using convolutional neural networks. PloS one, 2017 12(6), e0177544
[46] H. Chen et al. "Mitosis detection in breast cancer histology images via deep cascaded networks." Thirtieth AAAI Conference on Artificial Intelligence. 2016
[47] D. Bardou, K, Zhang, & S. M. Ahmad, Classification of breast cancer based on histology images using convolutional neural networks.
[48] Celik, Y., Talo, M., Yildirim, O., Karabatak, M., Acharya, U.R.: Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. Pattern Recognition Letters 133, 232–239 (2020)
[49] Mehra, R., et al.: Breast cancer histology images classification: Training from scratch or transfer learning? ICT Express 4(4), 247–254 (2018)
[50] Saha, M., Chakraborty, C., Racoceanu, D.: Efcient deep learning model for mitosis detection using breast histopathology images. Computerized Medical Imaging and Graphics 64, 29–40 (2018)
[51] Beevi, K.S., Nair, M.S., Bindu, G.: Automatic mitosis detection in breast histopathology images using convolutional neural network based deep transfer learning. Biocybernetics and Biomedical Engineering 39(1), 214–223 (2019)
[52] Sarraf S, Tofighi G, et al (2016) Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri. bioRxiv
[53] Ker, J.; Singh, S.P.; Bai, Y.; Rao, J.; Lim, T.; Wang, L. Image Thresholding Improves 3-Dimensional Convolutional Neural Network Diagnosis of Different Acute Brain Hemorrhages on Computed Tomography Scans. *Sensors* **2019**, *19*, 2167, doi:10.3390/s19092167.
[54] Pezeshk, A.; Hamidian, S.; Petrick, N.; Sahiner, B. 3-D Convolutional Neural Networks for Automatic Detection of Pulmonary Nodules in Chest CT. *IEEE J. Biomed. Heal. Informatics* **2019**, *23*, 2080–2090, doi:10.1109/JBHI.2018.2879449.
[55] Satya P. Singh, Lipo Wang, Sukrit Gupta, Haveesh Goli, Parasuraman Padmanabhan 1,2,* and Balázs Gulyás - 3D Deep Learning on Medical Images: A Review
[56] Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings; 2015.
[57] Kukačka, J.; Golkov, V.; Cremers, D. Regularization for Deep Learning: A Taxonomy. *arXiv Prepr. arXiv1710.10686* **2017**.
[58] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Salakhutdinov, R. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*; 2014; Vol. 15;.
[59] Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015; International Machine Learning Society (IMLS), 2015; Vol. 1, pp. 448–456.
[60] 4 Pre-Trained CNN Models to Use for Computer Vision with Transfer Learning | by Orhan Gazi Yalçın | Towards Data Science
[61] Understanding and visualizing DenseNets | by Pablo Ruiz | Towards Data Science
[62] Hajiabadi H, Molla-Aliod D, Monsefi R. On extending neural networks with loss ensembles for text classification. In: Proceedings of the Australasian language technology association workshop; 2017.
[63] GitHub - jcborges/DeepStack: DeepStack: Ensembling Keras Deep Learning Models into the next Performance Level
[64] Hajiabadi H, Molla-Aliod D, Monsefi R. Combination of loss functions for robust breast cancer prediction.
[65] P. Mooney, Breast Histopathology Images, Kaggle, 2018. https://www.kaggle.com /paultimothymooney/breast-histopathology-images/data.
[66] A. Cruz-Roa, A. Basavanhally, F. Gonzalez, H. Gilmore, M. Feldman, S. Ganesan, et al., Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. SPIE Medical Imaging, 2014, 9041, 904103- 904103-15 (2014.
[67] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases, J. Pathol. Inf. I (2016) 29.