Chapter 4

Deep Feature Learning using Convolutional Neural Network

In the previous chapter, we have provided a basic overview of histopathological image classification approaches and thoroughly reviewed existing approaches. This chapter introduces a deep feature extractor and classification framework for CMT and HBC.

4.1 Introduction

Diagnoses of both human breast cancer (HBC) and canine mammary tumours (CMTs), are done by histopathological analysis of haematoxylin and eosin (H&E) stained tissue sections by skilled pathologists: a process that is very tedious and time-consuming. Deep learning-based approaches are recently being evaluated by computer scientists for the CAD of HBC. Recently, CNNs have been reported to be a powerful tool in the automated classification of human cancer histopathology images [8, 9]. CNNs automatically learn mid-and high-level abstractions obtained from RGB images, and generic descriptors extracted from CNNs are extremely effective in object recognition and image segmentation or target localization in natural images [42]. The recent success of CNNs for image classification has inspired us to use them for histopathology image clas-

sification. However, the dataset's small size presents a significant challenge for training a deep learning model. Deep CNN architectures with millions of parameters, such as VGG, Inception, and ResNet, have achieved cutting-edge results in many computer vision tasks. But, training these neural networks from scratch necessitates a large number of images because training on a small dataset results in overfitting or the inability to generalize knowledge. As a result, we used a different approach known as deep convolutional feature representation. To this end, deep CNNs trained on large and general datasets such as ImageNet (10M images, 20K classes) are used to extract unsupervised feature representations. In this study, a framework based on VGGNet-16, a popular CNN, has been utilized for the generation of a robust and reliable feature set. This step will reduce the risk of overfitting on the next stage of supervised learning significantly. Thereafter, different supervised classifiers were applied to the model to enable the learning of different patterns from these feature sets. Thus, the proposed framework has presented a fused model of VGGNet-16 with Support Vector Machines (SVM) and Random Forest (RF) for binary classification of H&E stained cancer images. The model was tested on a standard BreakHis and the introduced CMTHis dataset. To the best of our knowledge, this study presented for the first time CAD for histopathological diagnosis of CMTs. Besides this, the effects of data augmentation, stain normalization, and magnification on the performance of the proposed framework were also analyzed.

4.2 Motivation and Significant Contributions

Considering the importance of correct diagnosis in patient management, considerable efforts have been made in the past for developing robust, precise, and automated CAD systems for humans. However, in spite of higher incidences and mortality rates in dogs, to date, no efforts have been made to automate the diagnosis of CMTs to relieve the burden on veterinary oncologists so that they can focus more on the cases which are difficult to diagnose. This may be due to the lack of any publicly available dog mammary

tumour image dataset for automated analysis. Moreover, the extraordinary capability of representation learning possessed by CNN and its remarkable success in the field of computer vision motivated us to utilize its ability to address complex histopathological cancer classification. Another important concept of transfer learning made it suitable for several tasks. Besides, simplicity with the strong mathematical formulation of machine learning techniques inspired us to propose a novel framework that combines the strength of two different approaches for dealing with histopathology cancer classification. The identified research gaps during the literature survey are also one of the main factors that direct us to design an automated CAD to liberate the burden on pathologists.

The main contributions of this chapter are listed below:

1. In this study, we have introduced a CMTHis dataset comprising images captured from 44 clinical cases of CMTs.
2. Thereafter, we propose a simple and effective method for classifying H&E stained histological breast cancer images in the absence of adequate training data that integrates the strength of deep learning and machine learning.
3. To improve the robustness of the classifier, we use strong data augmentation and extracted deep convolutional features at various scales with VGGNet-16 pre-trained CNN.
4. On top of that, SVM and RF algorithms are employed to enable the learning of different patterns from the extracted feature set.
5. Unlike previous works, we intentionally avoid training neural networks on this amount of data to avoid suboptimal generalization.
6. Once the efficacy of the proposal is validated on the standard BreakHis dataset experimentally, it is applied for the CMTHis dataset. To the best of our knowledge, the reported results outperform the automated analysis of breast cancer images that have been published in the literature.

4.3 Theoretical Background

4.3.1 Transfer learning

Transfer learning is a prominent method in computer vision, as it enables us to build precise models in a time-saving manner and is commonly used for predictive modelling problems that use image data as input. It is described in detail in the background section.

Training CNNs from scratch needs a large number of images, or else the model will suffer from overfitting. A typical solution in these conditions is fine-tuning when only a part of the pre-trained neural network is fitted into a new dataset [26]. Fine-tuning is another approach used if a medium-sized dataset exists for the task. It uses a pre-trained CNN to initiate the network, and thereafter the training of several or all network layers is supervised using the new data for the task at hand [27]. Moreover, in many computer vision tasks, deep CNN architectures containing millions of parameters have achieved state-of-the-art results with fine-tuning [28, 29]. However, in our experiments, the fine-tuning approach did not perform well on the task of histopathological image classification. As a result, we used a different approach known as deep convolutional feature representation [164].

4.3.2 VGGNet-16 CNN

In 2014, the VGGNet architecture was proposed for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) to classify large scale images and also to locate learned objects within the image [115]. VGGNet substantially simplified the design of CNN by repeating 16 times the same smaller convolution filter configuration. All VGGNet filters were limited to 3×3, with stride and padding of 1, along with 2×2 max-pooling filters with a stride of 2. This model also showed that the depth of the network significantly improves classification performance. Figure 4.1 shows the architecture of VGGNet-16.

In this network, there are thirteen convolutional layers (Conv), five pooling layers, and three fully connected (dense) layers, totalling up to 21 layers, but it has only 16 weight layers. Conv 1 has 64 filters while Conv 2 has 128 filters, Conv 3 has 256 filters and Conv 4 & Conv 5 have 512 filters each. In this study, we propose our framework based on the VGGNet-16 architecture because it uses small size filters that are expected to be suitable for learning micro-textures, as compared to other CNN architectures such as AlexNet that uses larger filters to search for edges, macro-textures or other important object features, which is discussed later in the section. Instead of having different sizes of convolution and pooling layers, VGGNet-16 uses only one size for each of them, which is applied several times. Moreover, the architecture has 138 million parameters, approximately three times more than AlexNet (60 million), and similarly, it tries to detect 1000 image categories. Although the model is bigger, at the same time, it is easier to understand because of its uniform architecture.

As CNN architectures have matured, the design of the filter is one of the features that has changed with time. The task of a filter is to capture patterns in the local receptive field, and smaller filters can capture patterns at a finer level of granularity. Even as we stack the layers, filters gradually capture patterns in larger areas of the

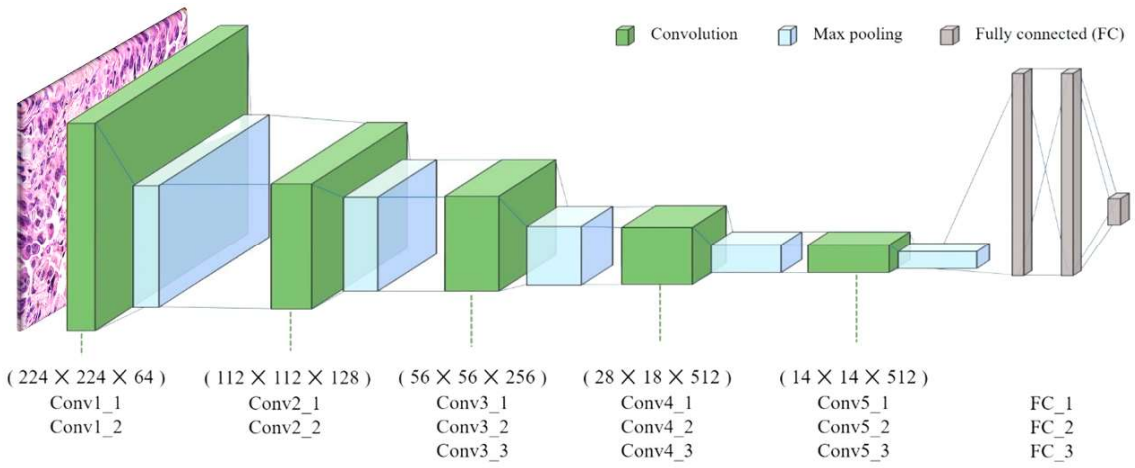


Figure 4.1: VGGNet-16 architecture

image at deeper levels. In the first convolutional layer, early CNNs like AlexNet [101] used large 11x11 filters. However, modern architectures such as VGGNet [115], use a 3x3 filter stack, which is a superior design compared to the single 11x11 filter because it uses a smaller number of parameters, incorporates more non-linearity and requires fewer computations.

Mathematically, we can illustrate how these models maintain the coarse granularity for covering the whole image in the form of small patches and different parameters at filters. Let h_i, w_i, d_i be the height, width and depth of volume in layer i and $h_{(i+1)}, w_{(i+1)}, d_{(i+1)}$ are the corresponding height, width and depth of volume in layer $i+1$. f_i is the size of filter, s_i is stride size going to layer $i \rightarrow i+1$ and p_i is the amount of zero padding. Therefore, the size of the volume in terms of width and height in the next layer i+1 can be calculated by the following formulae:

$$w_{(i+1)} = ((w_i - f_i + 2 \times p_i)/s_i) + 1 \tag{4.1}$$

$$h_{(i+1)} = ((h_i - f_i + 2 \times p_i)/s_i) + 1 \tag{4.2}$$

Total number of weight needed at i+1 layer is calculated as

$$(d_{(i+1)} \times (f_i \times f_i \times d_{(i+1)})) \; and \; d_{(i+1)} \; biases \tag{4.3}$$

Generally, pre-trained models use zero padding to maintain the size of the output activation map i.e., $w_{(i+1)} = w_i$, $h_{(i+1)} = h_i$ and $d_{(i+1)} = d_i$. For example, if we consider 3 stacked filter of size 3×3 with c number of channels, then total number of filter parameter is $3 \times c \times (3 \times 3 \times c + 1)$ that is $27 \times c^2 + 3 \times c$. This reduces the number of parameters by approximately 45% as compared to other single filters of size 7×7. However, pooling layers does not increase any parameter. Thus, it shows that deep CNN with a smaller filter can extract the information at a granular level of the

histopathological image without losing valuable information.

Since our CNN only needs to learn pixel-wide micro-texture and not the full tumour forms of cancer subtypes, we have examined several simplifications of the original VGGNet-16 architectures described later in section 4.4.1.

4.4 Proposed Methodology

4.4.1 VGGNet-16 as a feature extractor

Traditional machine learning methods are used to extract features from images using global feature descriptors such as LBP, HOG, etc., or Local descriptors such as ORB, SURF, SIFT, etc. These are hand-crafted features that require expertise in that domain. However, instead of using hand-crafted features, CNNs automatically learns these features from images in a hierarchical way. Lower layers learn low-level features like edges and corners, while middle layers learn to shape, colour, etc. Higher layers learn high-level features that represent the object in the image [165]. Instead of making a CNN model for classifying images, we use it as an extractor by considering the available activations map before the network's fully connected (FC) layer. This type of approach is well suited for image classification problems, where pre-trained CNNs are used for feature extracting instead of training time-consuming and tedious CNNs from scratch.

The CNN architecture VGGNet-16 is considered as a feature extractor in our proposed approach and has shown promising results in different image classification tasks. VGGNet-16 is a pre-trained model that has been trained using ImageNet data [166] consisting of millions of images showing animals, plants, vehicles, and other objects. Although the task of classifying these images is not closely related to our task of classifying histopathological cancer images, pre-trained architectures have been found suitable for transfer learning to various approaches as described in section 4.3.1. Yosinski et

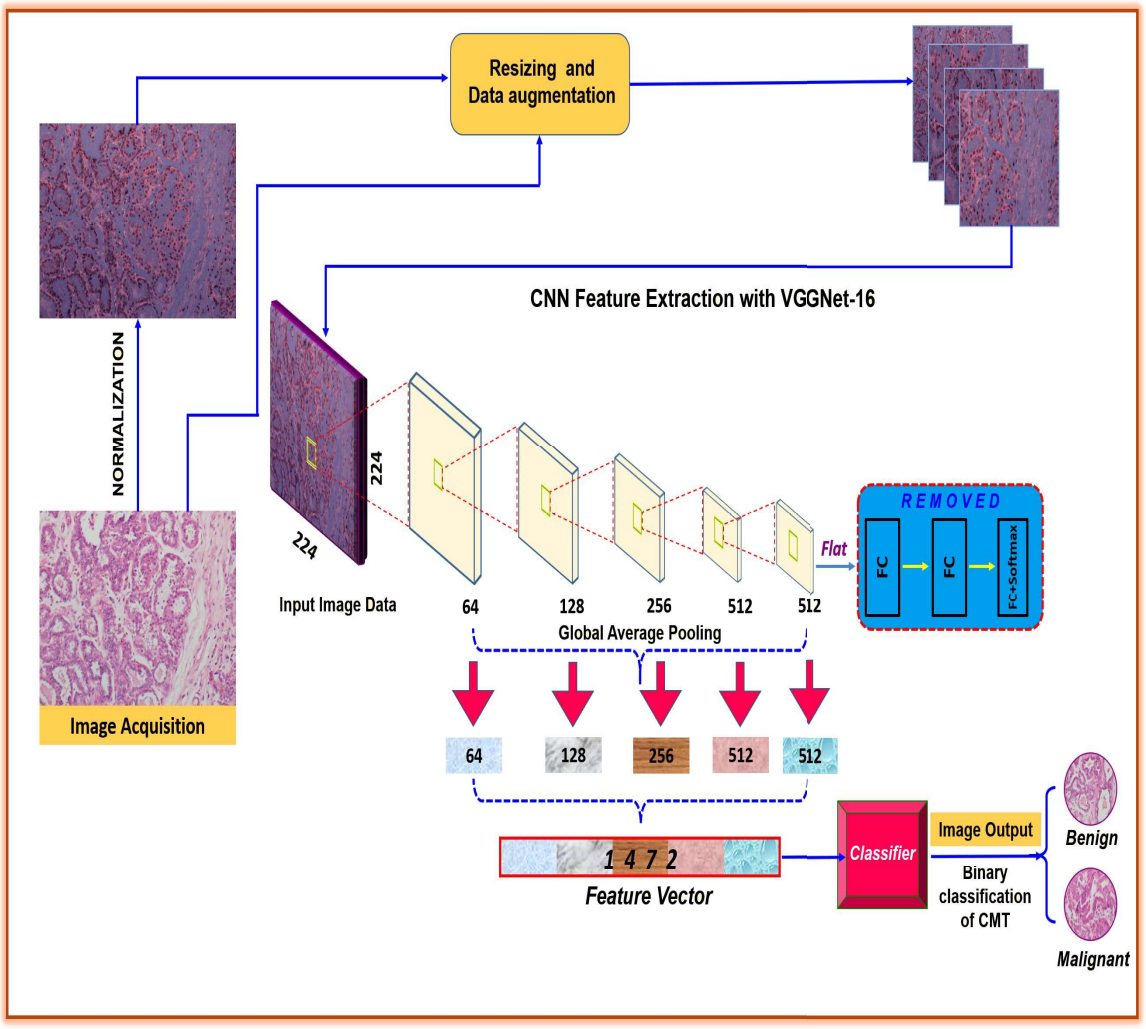


Figure 4.2: Proposed framework for histopathology image classification. H&E stained histopathological tumor images were initially subjected to data augmentation and stain normalisation. Finally, the images with and without stain normalisation were applied to the fused VGGNet-16 model for binary classification of cancer images, where all FC layers are removed.

al. [167] suggest that higher layers provide less-generic characteristics with a reduced target performance. The simple hierarchical structure found in most CNN architectures allows us to extract features in several prominent positions and explore their transferability further. Thus, we have removed FC layers from this model so that this network can consume arbitrary images, and each extraction layer is followed by a global average pooling (GAP) to limit the number of total features. Generally, a pooling layer is used

for dimension reduction, noise drop, and receptive field amplification. In the proposed method, GAP layers act as a regulariser that minimizes overfitting by performing a more extreme type of dimensionality reduction. As an illustration, if we assume a tensor with dimension $l \times b \times d$ is reduced to a size of $1 \times 1 \times d$, GAP layer reduces each $l \times b$ feature map to a single number by taking the average of all lb values. It reduces the response of each filter in a convolutional layer to just one feature, which is more robust and abstract. Therefore, global average pooling is applied to five external convolutional layers of all five blocks, respectively, in VGGNet-16 with channels 64, 128, 256, 512, 512. This results in a single vector of 1472 features after concatenation. This whole process is shown in Figure 4.2, and the visualization of feature maps from the five main blocks of the VGGNet-16 model is shown in Figure 4.8.

Mathematically, any CNN can be explained with the help of three layers: the convolution layer, defined as $\psi^1 = g(W^1 * \psi^0 + b^1)$, where $*$ is the convolution operator, ψ^0 represents input image and W^1 and b^1 are the filter and bias, respectively. The $\mathbf{g(.)}$ denotes activation function which is rectified linear unit (ReLU), and ψ^1 is output feature maps of convolution layer and input of pooling layer. GAP layer, written as $\psi^2 = g(h(\psi^1))$, where $\mathbf{h\ (.)}$ represents GAP function. FC layer is $\psi^{l+1} = g(W^l\psi^l + b^l)$, where l=2, 3, 4; ψ^{l+1} is the output of FC layer l, ψ^l, W^l, and b^l are the input, weights and bias for layer l, respectively.

4.4.2 Classifiers

Two different classifiers, SVM and RF, were used to access the aforementioned feature set. These algorithms were used to make a fair comparison to other state-of-the-art algorithms available in the literature, as these algorithms are widely used in histopathological image classification.

4.4.2.1 Support vector machine

SVM is a machine learning model with a strong mathematical background that can be used for classification and regression analysis. Originally, Vladimir Vapnik [168] designed and proposed this model to be able to efficiently perform linear classification and nonlinear classification by using a kernel trick. In cancer classification research, binary classification using SVM has often been adopted because of its ability to handle nonlinear classification and high-dimensional data. However, SVM itself cannot remove the noisy and irrelevant features. Therefore, we used VGGNet-16 to extract features before applying SVM.

4.4.2.2 Random forest

RF is one of the popular ensemble learning techniques that build a number of decision trees. The final class is decided by a majority vote on different decision trees [169]. To train this model, it requires the prior calculation of valid handcrafted features representing image characteristics. In histopathology images, the most commonly used features are pixel-based or object-based, such as image morphology and graph-based, which are assessed by RF for classification and detection. In our proposed model, the extracted feature from CNN is used by RF for classification. One advantage of RF is that it is quite fast and can handle large datasets and unbalanced data efficiently.

4.5 Experimental Results and Analysis

This section presents the experimental results and comparative analysis to demonstrate the efficacy of the proposed method over existing methods.

4.5.1 Experimental setup

The framework described in section 4.4 is implemented using the Keras library [170]. The SVM and RF classifiers are implemented using the scikit-learn library in python. The results are generated on a GPU with a 2.60 GHz Intel-Xeon E5-2660v3 processor, 128 GB DDR4 ECC RAM and a 12 GB NVIDIA Tesla K40C graphics engine. The proposed framework was applied to both datasets, BreakHis and CMTHis. The split protocol [14] was used to generate 5-folds, and the results were presented as an average of the five folds. To ensure that the classifier was generalized for unseen patients, the test set did not include patients from the training set. The dataset and split protocol are described in detail in Chapter 3.

4.5.1.1 Data augmentation

Features extracted from pre-trained CNN are not guaranteed to be invariant in terms of the position or orientation of the tissue in an image or image patch. Also, CNN needs enough data to achieve impressive performance. Thus, to make our model robust for feature transformation, data augmentation was performed to increase the data size. This increases the chance that a subsequent classifier will rely on mostly invariant features or at least adjust to the variation within one feature. Several studies have examined the role of data augmentation in deep learning [171, 172], as this method often does not change image classes and allows for a high amount of data and the building of more generalized models. Various geometric transformations, like rotation by 90^o, 180^o, 270^o, positive scaling, and mirror projection such as left-right-top-bottom, have been applied to the original image. Besides, gaussian blur was also used to increase data on the original image. Thus, for all magnification factors, the total number of sample images is increased by around eight times.

4.5.1.2 Evaluation metrics

Traditionally, a malignant case is considered positive during cancer diagnosis, while a benign case is considered negative. Clinically, the sensitivity of a CAD system to positive cases (malignant) is more important. The accuracy is calculated by counting the number of correct predictions in both the negative and positive classes. Therefore, in addition to accuracy, we have used other evaluation metrics also such as F1 score and Area Under Curve (AUC), which are directly related to the CAD's sensitivity to positive cases. These are described as follows:

1. **Accuracy:** The accuracy is defined as the ratio of correctly classified samples to the total number of samples in the evaluation data set. This measure is one of the most widely utilized in machine learning applications. To measure the accuracy, we must compute the fraction of true positive and true negative results in all evaluated cases. This can be expressed mathematically as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.4}$$

where TP, FN, FP, and TN represent the number of true positives, false negatives, false positives, and true negatives, respectively.

2. **F1 Score:** The F1 score, also known as the F1 measure, is used to better emphasize a CAD system's sensitivity to (positive) malignant instances, which are of considerable interest in this type of medical diagnosis. The F1-score is the harmonic mean of recall (also known as sensitivity) and precision. It is defined as:

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{4.5}$$

where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$.

3. **Area Under Curve (AUC):** AUC measures how the sensitivity (recall) and

false-positive rate trade-off. More crucially, unlike standard accuracy, AUC is not a function of threshold. It allows the classifier to be evaluated across all possible threshold values. It indicates how well the model can distinguish between classes. The greater the AUC, the better the model predicts 0 classes as 0 and 1 classes as 1. By analogy, the higher the AUC, the better the model distinguishes between patients with the disease and no disease.

4.5.2 Performance evaluation of the proposed framework in magnification dependent model

The VGGNet-16 pre-trained model was evaluated on BreakHis and CMTHis datasets. Test accuracies for binary classification of the BreakHis dataset ranged from 86%-90%, whereas, in the CMTHis dataset, test accuracies ranging from 78%-82% were observed across all magnifications using VGGNet-16 architecture. Therefore, to further improve the classification performance, we tried a variant of VGGNet-16 in which the FC layers were removed and replaced with SVM and RF classifiers. Various kernels, such as linear, polynomial, and RBF, were tested for the SVM classifier. This framework was independently applied to each of the four magnifications available. To the best of our knowledge, we have evaluated for the first time the performance of a fused framework

Table 4.1: Mean test accuracy for different classifiers applied to proposed framework at different magnifications in the BreakHis dataset.

Proposed Models	**Magnification Factors**											
	40X			**100X**			**200X**			**400X**		
	Mean	**SD**	**F1**	**Mean**	**SD**	**F1**	**Mean**	**SD**	**F1**	**Mean**	**SD**	**F1**
FE-VGGNET16-RF	92.22	2.14	0.94	93.40	4.38	0.95	**95.23**	1.89	0.97	92.80	1.83	0.94
FE-VGGNET16-SVM(LIN)	93.82	1.45	0.94	94.98	1.13	0.95	**95.77**	1.02	0.97	92.40	0.62	0.95
FE-VGGNET16-SVM(RBF)	92.60	1.52	0.97	93.49	1.62	0.97	**95.13**	1.96	0.98	94.96	2.19	0.97
FE-VGGNET16-SVM(POLY)	94.11	1.83	0.96	95.12	1.10	0.97	**97.01**	1.14	0.98	93.40	1.01	0.96

VGGNet-16-SVM model on the CMTHis dataset (CMT image dataset) and BreakHis (HBC image dataset). The proposed fused framework using VGGNet-16 along with SVM, its variants and RF classifier resulted in mean testing accuracies ranging from 92.22% to 97.01% for all four classifiers at various magnifications used in the BreakHis dataset, as shown in Table 4.1. The differences between training and testing accuracies were very low, demonstrating that the model has the ability to avoid over-fitting. Performance of the proposed framework was also evaluated using a confusion matrix,

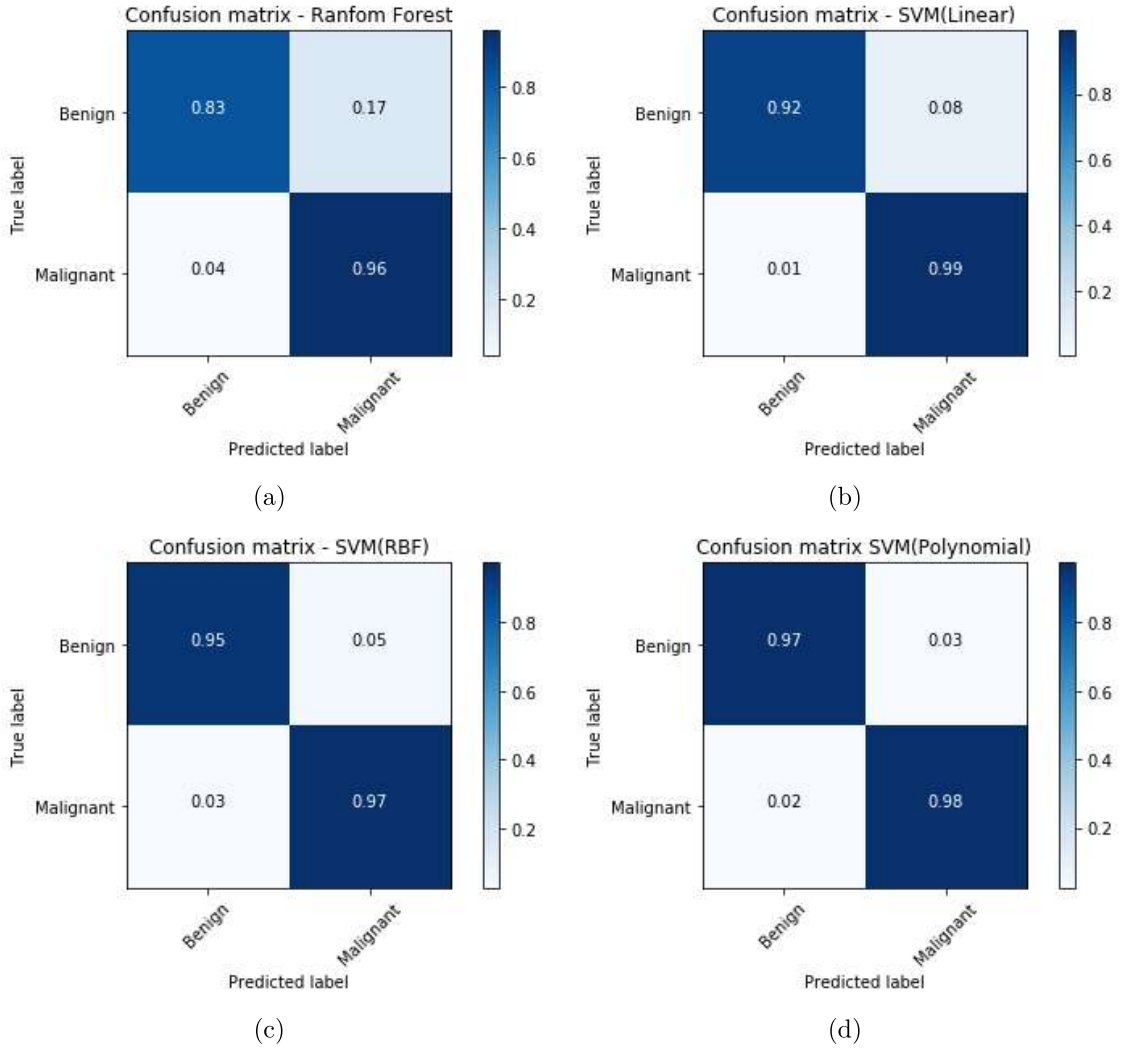


Figure 4.3: Confusion matrix plots of BreakHis dataset at 200× magnification for (a) RF, (b) SVM(LIN), (c) SVM(RBF), and (d) SVM(POLY).

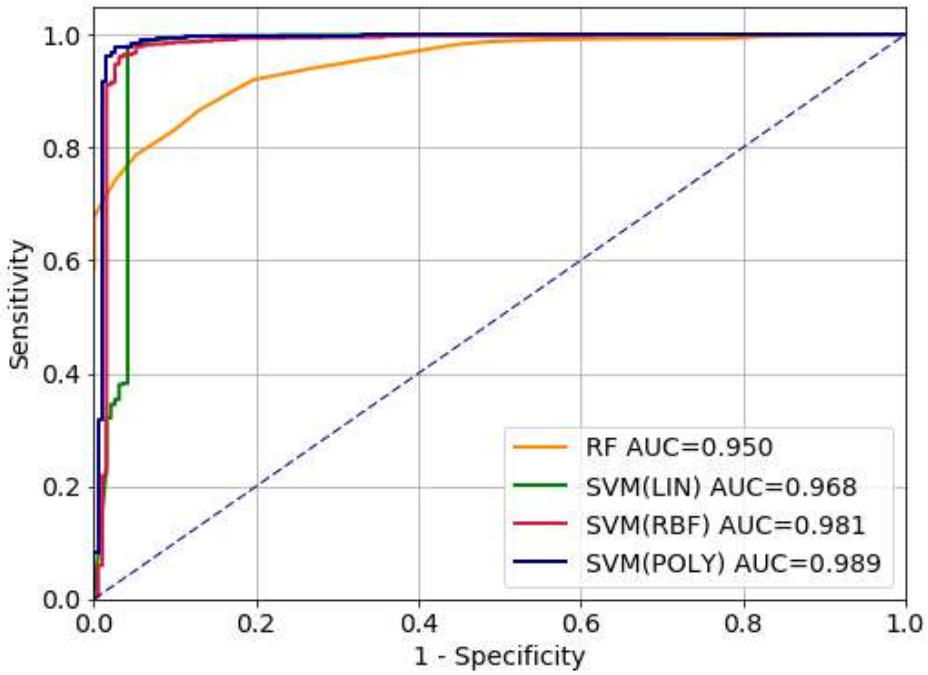
as shown in Figure 4.3. The ROC curve analysis, shown in Figure 4.4(a), revealed that AUCs of the proposed framework with various classifiers ranged from 0.950 to 0.989, indicating the high ability of the proposed framework to distinguish between benign and malignant breast cancers. Once the high performance of our proposed framework was validated on a standard BreakHis dataset, the framework was applied to the CMTHis dataset. The proposed framework was able to successfully classify malignant and benign CMTs with accuracy ranging from 81.15% to 92.75% at various magnifications across all four classifiers used in the study, as shown in Table 4.2. ROC curve analysis revealed AUCs ranging from 0.890 to 0.969 for different classifiers applied to CMTHis dataset at 200× magnification factors as shown in Figure 4.4(b), demonstrating the efficacy of the framework to classify CMTs.

Table 4.2: Mean test accuracy for different classifiers applied to proposed framework at different magnifications in the CMTHis dataset.

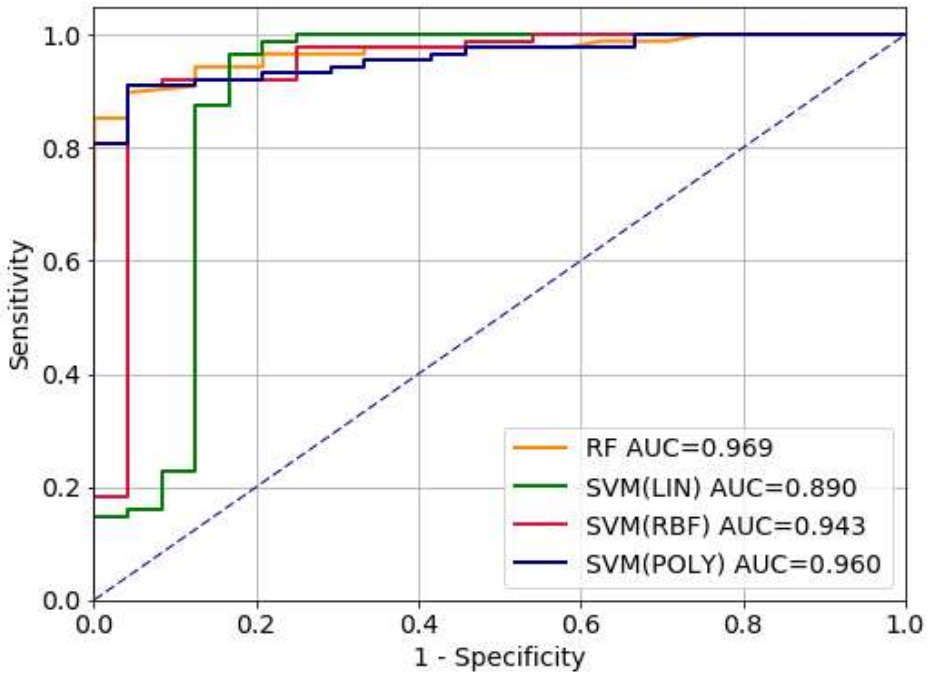
Proposed Models	**Magnification Factors**											
	40X			**100X**			**200X**			**400X**		
	Mean	**SD**	**F1**	**Mean**	**SD**	**F1**	**Mean**	**SD**	**F1**	**Mean**	**SD**	**F1**
FE-VGGNET16-RF	91.27	2.93	0.94	**91.40**	3.00	0.94	86.13	1.31	0.89	81.63	1.56	0.86
FE-VGGNET16-SVM(LIN)	91.31	1.69	0.94	**91.35**	4.50	0.94	89.07	3.74	0.93	83.35	6.07	0.88
FE-VGGNET16-SVM(RBF)	89.09	4.10	0.93	**89.15**	3.40	0.93	85.44	2.52	0.90	81.87	0.69	0.87
FE-VGGNET16-SVM(POLY)	91.95	2.59	0.95	**92.75**	4.47	0.95	89.12	2.95	0.93	81.15	1.69	0.86

4.5.2.1 Influence of magnification on test performance

After establishing the validity and performance of our classification approach, we studied the effect of magnification on the proposed framework and other state-of-the-art classifiers. The VGGNet-16 architecture, along with traditional classifiers, i.e., RF, SVM(Linear), SVM(Poly) & SVM(RBF), were tested on each of the four magnifications independently to assess the effect of magnification on the method performance.



(a)



(b)

Figure 4.4: AUC for different classifiers at 200× magnification for (a) BreakHis dataset and (b) CMTHis dataset.

The results obtained by each of the four methods varied with the magnification factor. It is well known that magnification influences the interpretation and clinical diagnosis by a pathologist; therefore, a pathologist first analyses the H&E stained tissue sections on lower magnification and then switches to higher magnifications with areas of interest. Higher magnification helps a pathologist in the fine-tuning of results. However, using the proposed framework, strikingly higher accuracies were achieved at mid-range magnifications (100× and 200×) in comparison to 400× magnifications.

For all the methods applied, the best results were obtained with 200× magnification, with test accuracies ranging from 95.13% to 97.01% for the BreakHis dataset. As shown in Table 4.1, accuracies with all the algorithms were influenced by magnification, with the highest accuracies at 200× magnification, followed by 100×. Several studies have shown the importance of magnification in automated breast cancer image binary classification using various algorithms [6, 173, 120, 130]. Various researchers [6, 130, 174] have shown that lower (100× and 200×) magnification yields higher accuracies as compared to 400× magnification. This may be due to the fact that at 100× and 200×, a larger region of interest (ROI) is captured as compared to 400×. Thus, 100× and 200× cover a larger ROI and at the same time provide enough resolution to extract the feature details, which could be the reason behind higher accuracies as compared to 400× magnification. However, in this study, the 40× specific model gave lower performance than 400× magnification for the BreakHis dataset, which might be due to the large variation in patterns on which the model was trained and tested.

In the CMTHis dataset, accuracy was found to be highest at 100× magnification, with test accuracies ranging from 89.15% to 92.75% for various classifiers, as shown in Table 4.2. There were only marginal differences between accuracies at 40× and 100×. Thus, in the CMTHis dataset, the highest accuracies were observed at the lowest magnifications (40× and 100×), followed by 200×, and lowest accuracies were observed at 400× magnification. Spanhol et al. [6] also showed that the accuracy of

CNN (ImageNet) decreases with an increase in magnification.

4.5.3 Performance comparison with state-of-the-art CNN architectures

We compared the average performance of the proposed framework with the state-of-the-art CNN architectures and other multi-layered framework approaches. The comparative results shown in Table 4.3 demonstrate that the proposed framework outperforms most state-of-the-art approaches. Spanhol et al. in [6] tested LeNet [177], a traditional CNN consisting of 2 convolutional layers and 3 fully-connected layers for binary classification of breast cancers with 72% accuracy. In the same study, Spanhol et al. used AlexNet [101] architecture and achieved a maximum average accuracy of 84.4% in the binary classification of histopathological images of HBC. In yet another study, Spanhol et al. [111] used pre-trained BVLC CaffeNet architecture along with a logistic regression classifier resulting, accuracy in the range of 81.6%-84.8%. Bayramoglu et al. [120], designed two networks: (1) a single task CNN for prediction of malignancy with an accuracy of 83.25%, (2) a multi-task CNN for predicting malignancy, as well as, magnification factor with 82.13% accuracy for binary classification. Very recently, a structure-based deep convolutional neural network (CSDCNN) has been proposed for HBC image classification with 94.9%- 96.9% accuracy for binary classification [174]. A non-linear representation learning model CSDCNN eliminates feature extraction steps into feature learning and bypasses feature engineering requiring a hand-designed approach. The VGGNet-16 based classification framework proposed by us outperforms most of the state-of-the-art strategies. Furthermore, the proposed framework is quite simple and straightforward compared to the existing frameworks.

4.5.4 Effect of contemporary classifiers in magnification independent model

The proposed framework was also applied in the magnification independent model, and mean accuracies at different magnifications were compared between all contemporary

Table 4.3: Comparison of our framework with state-of-the-art CNNs at different magnifications.

	Methods	Dataset	Test accuracies(%) at different magnifications			
			40×	100×	200×	400×
Existing framework	Bayramoglu et al.[120]	BreakHis	83.08±2.08	83.17±3.51	84.63±2.72	82.10±4.42
	Spanhol et al. [6]	BreakHis	89.60±6.50	85.00±4.80	82.80±2.10	80.02±3.40
	Spanhol et al. [111]	BreakHis	84.60±2.90	84.80±4.20	84.20±1.70	81.60±3.70
	Song et al. [175]	BreakHis	90.02±3.20	88.90±5.00	86.90±5.20	86.30±7.00
	Song et al. [176]	BreakHis	90.02±3.20	91.20±4.40	87.80±5.30	87.40±7.20
	Han et al. [174]	BreakHis	95.80±3.10	96.90±1.90	96.70±2.00	94.90±2.80
	Gupta and Bhavsar [130]	BreakHis	86.74±2.37	88.56±2.70	90.31±3.76	88.31±3.01
	Gupta and Bhavsar [116]	BreakHis	84.72	89.44	95.65	82.65
	Gupta and Bhavsar [116]	BreakHis	91.90	93.64	95.84	90.15
Proposed Approach	FE-VGGNET16-RF	BreakHis	92.22 ± 2.14	93.40 ± 4.38	95.23 ± 1.89	92.80 ± 1.83
	FE-VGGNET16-SVM(LIN)	BreakHis	93.82 ± 1.45	94.98 ± 1.13	95.77 ± 1.02	92.40 ± 0.62
	FE-VGGNET16-SVM(RBF)	BreakHis	92.60 ± 1.52	93.49 ± 1.62	95.13 ± 1.96	**94.96 ± 2.19**
	FE-VGGNET16-SVM(POLY)	BreakHis	**94.11 ± 1.83**	**95.12 ± 1.10**	**97.01 ± 1.14**	93.40 ± 1.01
	FE-VGGNET16-RF	CMTHiS	91.27 ± 2.93	91.4 ± 3.00	86.13 ± 1.31	81.63 ± 1.56
	FE-VGGNET16-SVM(LIN)	CMTHiS	91.31 ± 1.69	91.35 ± 4.5	89.07 ± 3.74	**83.35 ± 6.07**
	FE-VGGNET16-SVM(RBF)	CMTHiS	89.09 ± 4.1	89.15 ± 3.4	85.44 ± 2.52	81.87 ± 0.69
	FE-VGGNET16-SVM(POLY)	CMTHiS	**91.95 ± 2.59**	**92.75 ± 4.47**	**89.12 ± 2.95**	81.15 ± 1.69

classifiers applied to VGGNet-16 architecture. The results, shown in Figure 4.5, clearly demonstrate that the mean accuracies across all classifiers were almost similar. Further, the magnification independent model of the framework proposed in this study was also compared with other reported magnification independent models [173, 120, 130] for binary classification of breast cancers, and the results are compiled in Table 4.4. It is clearly evident from Table 4.4 that the proposed magnification independent model

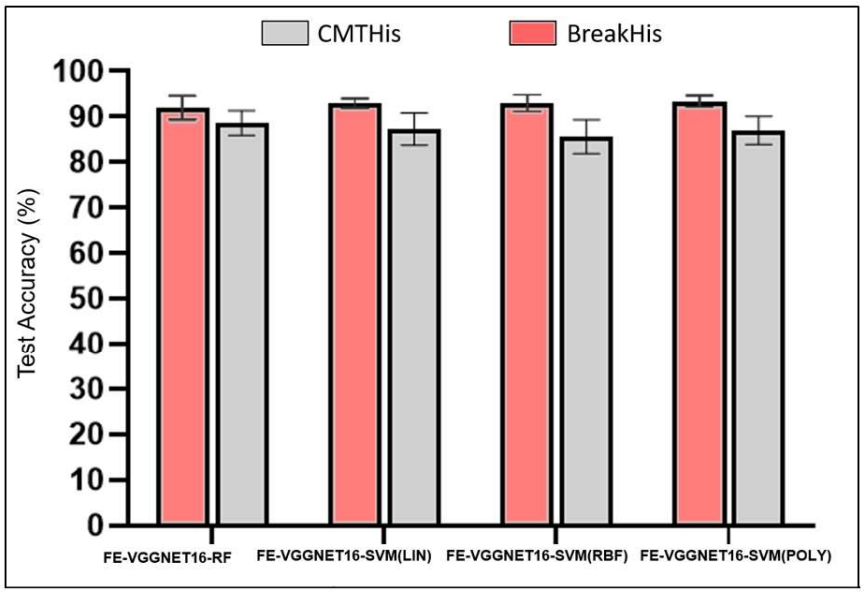


Figure 4.5: Mean test accuracy for different classifiers applied to proposed framework in magnification independent model.

Table 4.4: Comparison of the magnification independent model of proposed framework with other reported magnification independent models.

	Method	**Mean accuracy (%)**
Existing	Bayramoglu [120]	83.24
	Gupta and Bhavsar [130]	87.53
	Sharma and Mehra [173]	85.30
Proposed	FE-VGGNET16-RF	**92.05**
	FE-VGGNET16-SVM(LIN)	**92.30**
	FE-VGGNET16-SVM(RBF)	**93.00**
	FE-VGGNET16-SVM(POLY)	**93.05**

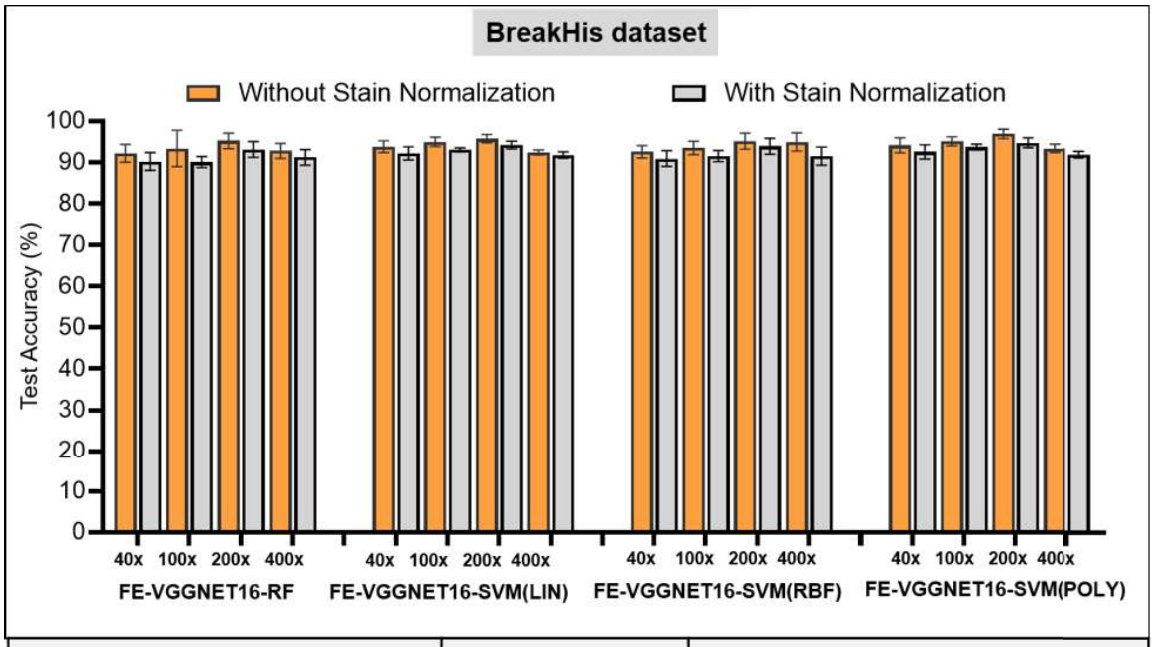
outperforms other such models reported so far. The magnification independent models have the advantage of simplicity as there is only a single CNN for all magnification factors, thus reducing the training time and complexity.

4.5.5 Effect of stain normalization

The images were pre-processed and normalized for variations in staining procedures based upon the methods given by Macenko et al.[161]. The details of the stain normalization algorithm are explained in Chapter 3. As discussed by several researchers, [178, 179], major factors limiting the large-scale application of automated histopathological image classification are the inter-laboratory variations in H&E staining, which significantly affect the results, requiring the application of stain normalization algorithms to neutralize these. Thus, having established the performance and effect of magnification on the proposed framework, the effect of stain normalization was evaluated on BreakHis and CMTHis datasets. A stain normalization algorithm, as discussed in section 4.4, was applied to our proposed framework, and the results before and after stain normalization were compared. H&E stained histopathological images from the CMTHis dataset, with and without stain normalization are shown earlier in Chapter 3. To our surprise, for the method proposed by us, the stain normalization does not significantly affect either the training or the test accuracy, as shown in Figure 4.6 and 4.7. The performances of our proposed framework without stain normalization were slightly better than with stain normalization, though the differences were not significant. Thus, the proposed algorithm gives equally good results without stain normalization, and thus the method avoids the unnecessary time and labour in the application of stain normalization algorithms. Therefore, using the proposed framework, histopathological slides with wide colour variations may not require stain normalization.

Similar results have been demonstrated by Reinhard [180], who also showed that the application of a stain normalization algorithm on CNNs slightly reduces the perfor-

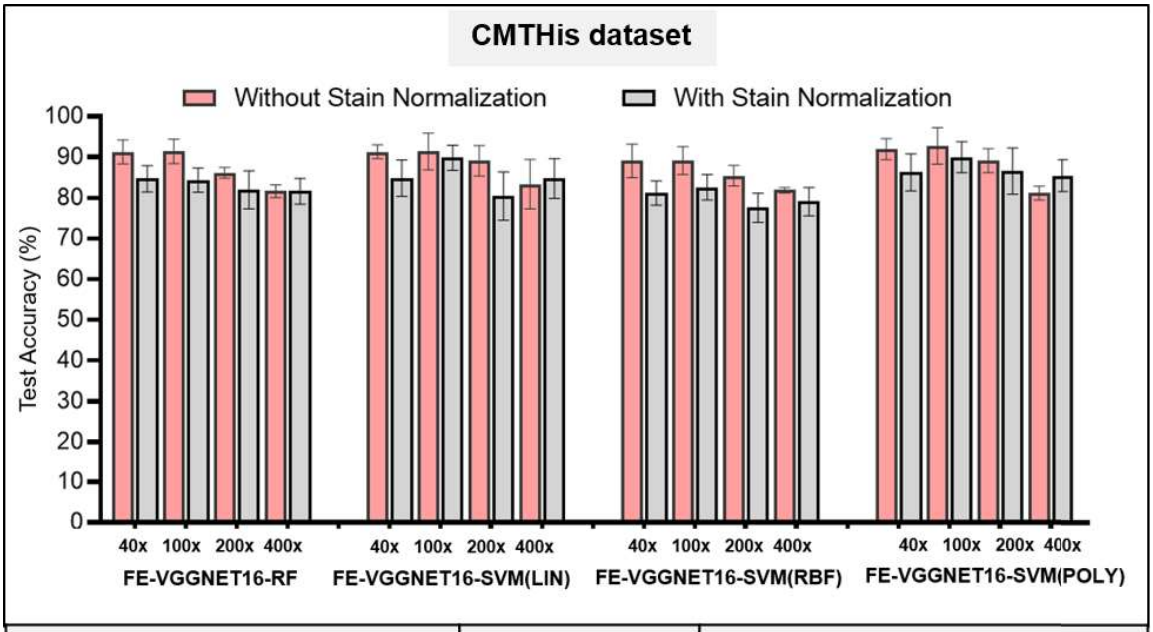
mance. Studies have previously found the usefulness of colour normalization algorithms when using handcrafted features based computer algorithms. Results from this study



Classifier	Magnification	Test Accuracy	
		Without stain normalization	With stain normalization
FE-VGGNET16-RF	40x	92.22 ± 2.14	90.22 ± 2.14
	100x	93.40 ± 4.38	90.08 ± 1.35
	200x	**95.23 ± 1.89**	**93.14 ± 1.92**
	400x	92.80 ±1.83	91.22 ± 1.9
FE-VGGNET16-SVM(LIN)	40x	93.82 ± 1.45	92.18 ± 1.57
	100x	94.98 ± 1.13	93.05 ± 0.46
	200x	**95.77 ± 1.02**	**94.23 ± 0.98**
	400x	92.40± 0.62	91.77 ± 0.77
FE-VGGNET16-SVM(RBF)	40x	92.60 ± 1.52	90.91 + 1.89
	100x	93.49 ± 1.62	91.52 ± 1.33
	200x	**95.13 ± 1.96**	**93.91 ± 1.92**
	400x	94.96 ± 2.19	91.49 ± 2.21
FE-VGGNET16-SVM(POLY)	40x	94.11 ± 1.83	92.53 ± 1.73
	100x	95.12 ± 1.10	93.73 ± 0.7
	200x	**97.01 ± 1.14**	**94.95 ± 1.18**
	400x	93.40 ± 1.01	91.77 ± 0.8

Figure 4.6: Mean test accuracy for different classifiers of proposed framework with and without stain normalisation on BreakHis dataset.

and other [2, 180] suggest that CNNs are effective in learning the task in the presence of colour variation. Another reason may be less colour variability in the BreakHis dataset collected from a single laboratory with a single microscopic system. Existing colour



Classifier	Magnification	Test Accuracy	
		Without stain normalization	With stain normalization
FE-VGGNET16-RF	40x	91.27± 2.93	84.68 ± 3.25
	100x	**91.41 ± 3.00**	**84.37 ± 2.97**
	200x	86.13 ± 1.31	81.96 ± 4.68
	400x	81.63 ± 1.56	81.62 ± 3.16
FE-VGGNET16-SVM(LIN)	40x	91.31 ± 1.69	84.84 ± 4.48
	100x	**91.35 ± 4.5**	**89.84 ± 3.06**
	200x	89.07 ± 3.74	80.44 ± 5.93
	400x	83.35 ± 6.07	84.73 ± 4.9
FE-VGGNET16-SVM(RBF)	40x	89.09 ± 4.1	81.15 ± 2.97
	100x	**89.15 ± 3.4**	**82.58 ± 3.13**
	200x	85.44 ± 2.52	77.58 ± 3.57
	400x	81.87 ± 0.69	79.07 ± 3.47
FE-VGGNET16-SVM(POLY)	40x	91.95 ± 2.59	86.26 ± 4.54
	100x	**92.75 ± 4.47**	**89.95 ± 3.8**
	200x	89.12 ± 2.95	86.55 ± 5.69
	400x	81.15 ± 1.69	85.44 ± 3.95

Figure 4.7: Mean test accuracy for different classifiers of proposed frameworks with and without stain normalisation on CMTHis dataset.

normalization methods were mostly developed for machine learning schemes employing handcrafted features (as opposed to deep learning/CNNs). These schemes may or may not perform other pre-processing processes on the input image. CNNs, on the other hand, typically perform (channel-wise) mean subtraction on the images. This may have an interaction with a specific colour normalization method.

4.5.6 Ablation study

To study the importance of different blocks of VGGNet16 and classifiers in the proposed FE-VGGNET16, we perform an ablation study on CMTHis dataset by choosing different combinations of blocks to extract features. We perform three different ablation experiments, which are presented in Table 4.5. It is observed that each block has an important role in histopathological image classification. Removing features from either of the blocks deteriorates the performance. The VGG16 network used in our experiments was pre-trained on natural images using the ImageNet dataset, which explains why features from multiple blocks were concatenated. The variability in natural images is very different from that of pathology images. However, because VGG16 is layered, each layer is responsible for extracting different types of features from an image; thus, as we go deeper into the network, the extracted features become more suitable for nat-

Table 4.5: Ablation study on reducing the number of block in VGG16 and observing its effect on CMTHis dataset

	Classifier	**Magnification Factors**			
		40X	**100X**	**200X**	**400X**
		Mean	Mean	Mean	Mean
All 5 blocks	RF	**91.27**	**91.41**	**86.13**	**81.63**
	SVM(POLY)	**91.95**	**92.75**	**89.12**	**81.15**
Last 4 blocks	RF	90.32	90.18	85.44	80.82
	SVM(POLY)	90.45	91.10	87.92	79.80
Last 3 blocks	RF	88.98	89.12	84.02	79.30
	SVM(POLY)	89.20	89.50	86.80	78.22

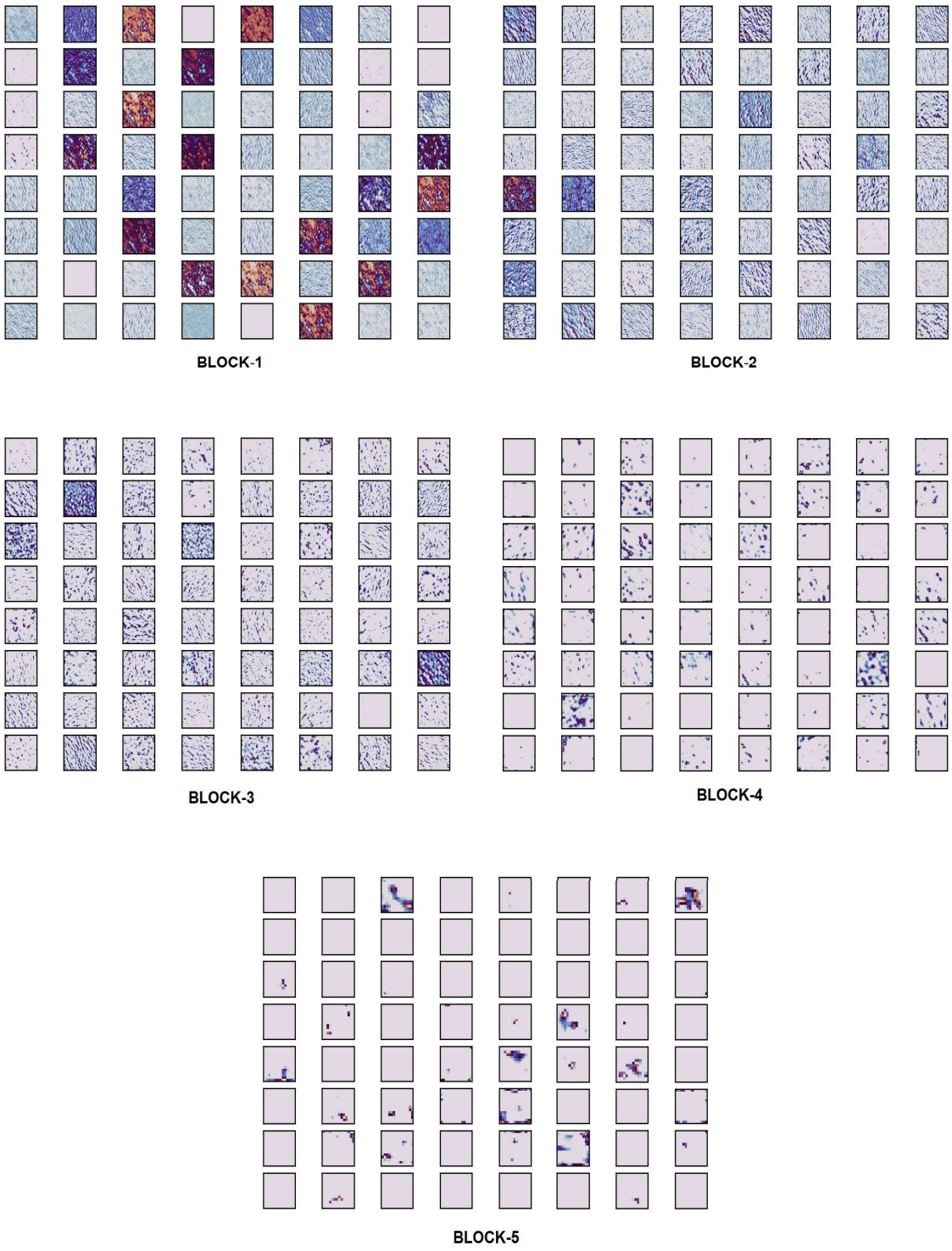


Figure 4.8: Visualization of the feature maps of one of the images extracted from Block-1 to Block-5 in the VGGNet-16 Model.

ural images and lose their generality for other types of images such as histopathology images, which can be seen from the Figure 4.8.

4.6 Summary

In this chapter, we proposed a framework in which features are extracted at various levels by removing the FC layer from VGGNet-16 and experimented with various classifiers. The model used in this study achieved high accuracy ($\approx 97\%$) and outperformed most state-of-the-art approaches used so far for binary classification of HBC. The high performance of our framework on a challenging BreakHis breast cancer dataset proved that it is capable of learning higher-level discriminating features. Therefore, the framework was applied to the CMT dataset introduced in this study, and the model achieved reasonably high ($\approx 93\%$) accuracy in the binary classification of CMT histopathological images. The reason behind the lower accuracy for the CMTHis dataset, as compared to the BreakHis dataset, might be the small size of the CMTHis dataset.

Chapter 5

Secure Deep Feature Classification using Convolutional Neural Network

The previous chapter discussed a deep learning-based feature extractor for breast cancer histopathology images. However, data privacy and security of the predictive model is a major concern when sharing sensitive healthcare data over the cloud. Therefore, this chapter focused on the privacy and confidentiality of medical data, as well as model security for a cloud-based approach.

5.1 Introduction

The advancements in the Internet of Things (IoT) and cloud services have enabled the availability of smart e-healthcare services in a distant and distributed environment. The integration of IoT and powerful infrastructure like the cloud into the health sector has provided new opportunities for designing new frameworks for different applications, with a view to limited storage space and fast computing resources. However, this also raised major privacy and efficiency concerns that need to be addressed. While sharing