

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins



Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer



Abhinav Kumar^{a,*}, Sanjay Kumar Singh^a, Sonal Saxena^{b,*}, K. Lakshmanan^a, Arun Kumar Sangaiah^c, Himanshu Chauhan^d, Sameer Shrivastava^b, Raj Kumar Singh^b

^a Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India

^b Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

^c School of Computing Science and Engineering, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India

^d Department of Mechanical Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India

ARTICLE INFO

Article history:

Received 4 January 2019

Revised 11 August 2019

Accepted 28 August 2019

Available online 30 August 2019

Keywords:

Canine mammary tumor (CMT)
Breast cancer
Deep learning
Histopathological classification

ABSTRACT

Canine mammary tumors (CMTs) have high incidences and mortality rates in dogs. They are also considered excellent models for human breast cancer studies. Diagnoses of both, human breast cancer and 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. The existence of heterogeneous and diverse types of CMTs and the paucity of skilled veterinary pathologists justify the need for automated diagnosis. Deep learning-based approaches have recently gained popularity for analyzing histopathological images of human breast cancer. However, so far, due to the lack of any publicly available CMT database, no studies have focused on the automated classification of CMTs. To the best of our knowledge, we have introduced for the first time a dataset of CMT histopathological images (CMTHis). Further, we have proposed a framework based on VGGNet-16, and evaluated the performance of the fused framework along with different classifiers on the CMT dataset (CMTHis) and human breast cancer dataset (BreakHis). We also explored the effect of data augmentation, stain normalization, and magnification on the performance of the proposed framework. The proposed framework, with support vector machines, resulted in mean accuracies of 97% and 93% for binary classification of human breast cancer and CMT respectively, which validates the efficacy of the proposed system.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Cancer continues to be one of the leading causes of human mortality and morbidity worldwide. In 2018 alone, 18.1 million new cases of cancer and 9.6 million cancer-related deaths were reported in humans [5]. Recent global cancer statistics

* Corresponding authors.

E-mail addresses: abhinav.rs.cse17@iitbhu.ac.in (A. Kumar), sks.cse@iitbhu.ac.in (S.K. Singh), sonalvet@gmail.com (S. Saxena), lakshmanank.cse@iitbhu.ac.in (K. Lakshmanan), arunkumarsangaiah@gmail.com (A.K. Sangaiah), himanshu.chauhan.mec16@iitbhu.ac.in (H. Chauhan), sameer_vet@rediffmail.com (S. Shrivastava), rks_virology@rediffmail.com (R.K. Singh).

have shown that breast cancer is still the most common type of cancer and the leading cause of mortality among women, accounting for 24.2% (2.1 million) new cases and 626,679 deaths per year [5]. Among animal species, pet animals, especially dogs, are more prone to cancer, often leading to poor prognosis and high mortality [15]. In the case of pet animals, mammary tumors have the highest incidences (16.8%) in females [34]. In unspayed female dogs, canine mammary tumor (CMT) is the most common malignancy with a thrice higher mortality rate as compared to human breast cancer [10]. Considering the increasing dog population and poor prognosis associated with CMTs, they present a major animal health issue, suggesting the need for newer diagnostic and therapeutic strategies for disease management. Spontaneously occurring CMTs share common characteristics between dogs and humans. CMTs are therefore considered excellent models for human breast cancer studies [20]. However, in humans, the majority of breast tumors are malignant, whereas, in canines, malignant and benign CMTs occur with similar frequencies [34]. The etiology of CMTs is still largely unknown; however, genetic, hormonal and nutritional risk factors are associated with CMT [4,39]. Owing to the influence of several hormonal, genetic and other associated factors, CMTs present diverse histological subtypes. Thus, the correct interpretation of CMTs is a major challenge for clinicians. Diagnosis of CMTs by routine cytology of biopsy, or by extirpated gland, is difficult and requires interpretation by trained veterinary pathologists. In humans, due to increased awareness about the disease, early diagnosis is possible with the help of routine self-check-ups and mammography followed by a biopsy. However, it is difficult to detect cancer at an early stage in pets because they are unable to convey warning signs and symptoms. Therefore, the diagnosis is made only when the tumor becomes visibly apparent to the animal owner. Thus, accurate diagnosis, as well as differentiation between benign and malignant neoplasms, is crucial for the successful outcome of treatment modalities, especially in canines.

Both in humans and canines, histopathological analysis remains the gold standard for cancer diagnosis. During the assessment, pathologists search for signs of cancer on microscopic portions of the Hematoxylin and Eosin (H&E) stained tissue sections by analyzing their histological properties, as well as changes in normal structures of breast parenchyma. It is worth mentioning that timely classification of breast samples into benign and malignant is essential for choosing the appropriate treatment regime. However, diagnosis using H&E stained biopsies is very time-consuming, costly, and laborious, requiring the intense efforts of specialized pathologists. Furthermore, diagnosis based upon manual analysis of slides suffers from inter-observer variability, with approximately 75% diagnostic concordance between specialists [11]. Hence, computer-aided approaches could be included in digital pathology in order to achieve rapid and reproducible results. They help in improving classification accuracy and reducing variability in interpretations, as errors made by machine learning methods have been reported to be less than those made by a single pathologist [41]. These techniques are also useful for assisting pathologists and reducing their labor in localizing and identifying abnormalities in the cancer tissue images. Therefore, researchers are trying to exploit the morphological criteria in the usual classification approach to develop computer-aided diagnostic (CAD) systems for improving the diagnostic efficacy and increasing the level of inter-observer agreement [44]. However, due to the complexity of the disease, it is a challenging task to develop a CAD system for cancer classification using histopathological images. Nevertheless, the latest advancements in machine learning approaches make this process more reliable and cost-effective than conventional methods. Deep learning has emerged as the leading machine learning tool for histopathological image analysis. Recently, Convolutional Neural Networks (ConvNets) based on deep learning architecture have been reported to be a powerful tool in the automated classification of human cancer histopathology images [22,27]. ConvNets automatically learn mid- and high-level abstractions obtained from RGB images, and generic descriptors extracted from ConvNets are extremely effective in object recognition and image segmentation or target localization in natural images [31]. ConvNets, along with multiple-instance learning, have accomplished high performance in the binary classification of human cancers and have evolved as a method of choice for analyzing histopathological images [43]. Despite the high performance of these systems for the binary classification of cancer, color variations in histopathology images are a concern for automated analysis. Color variations in images may occur because of several reasons, such as differences in the chemical reactivity of stains from different manufacturers, staining procedures, storage times, color responses of slide scanners or differences in slide thickness leading to variations in transmission of light. For a pathologist, these color variations may not hinder the analysis, but in automated image analysis, these variations can significantly affect image interpretation. Hence, stain normalization algorithms have been introduced recently to address this issue [1,21].

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 tumor image database for automated analysis. Therefore, in this study, we have introduced a canine mammary tumor image database (CMTHis) comprising images captured from 44 clinical cases of CMTs. The recent success of ConvNets for image classification has inspired us to use them for histopathology image classification. Accordingly, in this study, a framework based on VGGNet-16, a popular ConvNet, has been utilized for the generation of a robust and reliable feature set. Thereafter, different classifiers were applied to the model to enable learning of different patterns from these feature set. 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 human breast cancer dataset (BreakHis) and the canine mammary tumor dataset (CMTHis) introduced in this study. Besides this, the effects of data augmentation, stain normalization, and magnification on the performance of the proposed framework were also analyzed.

This paper is organized as follows. [Section 2](#) consists of the theoretical background for our proposed work; [Section 3](#) details the specifics of the proposed method and provides information on the datasets used to validate the framework. [Section 4](#) describes the experimental results and comparative analysis. Finally, [Section 5](#) discusses the conclusion and scope for future study.

2. Theoretical background

This section provides a brief overview of the techniques which are closely related to our work.

2.1. Granularity in deep ConvNet

In computer science, granularity refers to a computation-to-communication ratio and also, in the classical sense, to the breakdown of larger holistic tasks into smaller, finer tasks. Explicit or implicit sparsity in the Deep Neural Network has been studied extensively [12,30]. Fine-grained sparsity and filter-wise sparsity are the two extreme cases that were widely studied [26]. As ConvNet 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 image at deeper levels. In the first convolutional layer, early ConvNets like AlexNet [23] used large 11x11 filters. However, modern architectures such as VGGNet [36], 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:

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

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

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

$$(d_{(i+1)} \times (f_i \times f_i \times d_{(i+1)})) \text{ and } d_{(i+1)} \text{ biases} \quad (3)$$

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 ConvNet with a smaller filter can extract the information at a granular level of the histopathological image without losing the valuable information.

2.2. VGGNet-16 ConvNet

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 [36]. VGGNet substantially simplified the design of ConvNet 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. 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 ConvNet architectures such as AlexNet, which uses larger filters to search for edges, macro-textures or other important object features as described earlier in the [Section 2.1](#). 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. Since our ConvNet only needs to learn pixel-wide micro-texture and not the full tumor forms of cancer subtypes, we have examined several simplifications of the original VGGNet-16 architectures described later in [Section 3.4](#).

2.3. Transfer learning along with fine tuning

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 modeling problems that use image data as input. With transfer learning, we basically try to exploit what has been learned in one task to improve generalization in another. It is difficult and time-consuming to collect very large medical imaging datasets of a specific domain. One of the problems associated with learning

the ConvNets parameter is that it requires a large amount of data and considerable computation power. When ConvNets are trained on image data, such as ImageNet, we have the luxury of having a database of millions of labeled images that allow us to train relatively large networks with hundreds of layers to a high degree of precision. Thus, when the data is insufficient, one way to mitigate this is to apply transfer learning in supervised ConvNet models, pre-trained from a natural image dataset or a different medical domain [14]. A pre-trained ConvNet is applied to an input image in one scheme, and the outputs are then extracted from the network layers. Fine-tuning is another approach used if medium-sized dataset exists for the task. It uses a pre-trained ConvNet 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 [50]. Moreover, in many computer vision tasks, deep ConvNet architectures containing millions of parameters have achieved state-of-the-art results with fine-tuning [2,8].

Training ConvNets from scratch, however, needs a large amount 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 [32].

3. Material and methods

This section explains the procedures used to develop the proposed method and provides information on the dataset used to validate our proposal.

3.1. Datasets

3.1.1. Canine mammary tumor dataset

In this paper, we introduce a database for canine mammary tumor, called CMTHis. The dataset comprises 352 images acquired from 44 clinical cases of canine mammary tumors (CMTs) that were presented to Referral Veterinary Polyclinics at ICAR-Indian Veterinary Research Institute (IVRI), Izatnagar. Tissue samples were fixed in 10% neutral buffered formalin, paraffin-embedded and, after cutting into 5 μm sections, were mounted on 3-aminopropyl-triethoxy-silane (APTES) coated slides. Sections were stained using H&E stain and covered with a glass coverslip and visualized microscopically for histopathological analysis. The histopathological classification of CMT tissues was done as per Goldschmidt et al. [13], and the tissues were classified as malignant or benign as described in [19]. Histopathological analysis of H&E stained CMT tissue sections was done by experienced veterinary pathologists and, wherever required, confirmation was done using complementary tests. Images were visualized on an Olympus BX-53 system and captured using an Olympus DP-73 Peltier cooled digital color camera with 17.28-megapixel resolution. The images were captured using objective lenses of 4 \times , 10 \times , 20 \times , and 40 \times corresponding to magnifying factors of 40 \times , 100 \times , 200 \times , and 400 \times , respectively. The captured images were of high quality and clarity with reduced noise because of advanced algorithms and fine detail processing provided by the DP73 CCD camera. Thus, 1600 \times 1200 pixel high-resolution RGB images with 24-bit color depth were captured, comprising in total 352 images from 20 benign and 24 malignant CMT cases. Details are given in Table 1, and CMT tissue processing and image acquisition setup are depicted in Fig. 1.

3.1.2. Breast cancer dataset

BreakHis dataset, containing a fairly large number of histopathology images of human breast cancer, was used to evaluate the performance of the proposed model. The dataset comprises 7909 breast cancer images collected from cancer tissue biopsies of 82 patients referred to the Pathological Anatomy and Cytopathology (P&D) Laboratory, Brazil from January–December 2014 [42]. The images represent microscopic images acquired using a system microscope Olympus BX-50 with a 3.3X magnification relay lens coupled with SCC-131AN, Samsung digital colored camera. The dataset consisted of 2480 benign and 5429 malignant images of 700 \times 460 pixel size from 82 patients each with different magnification of 40 \times , 100 \times , 200 \times , and 400 \times , as shown in Table 2.

3.2. Stain normalization

The images were pre-processed and normalized for variations in staining procedures based upon the methods given by Macenko et al.[28]. This method takes into account the staining technique used to prepare slides in histology. First, the

Table 1
CMTHis dataset image distribution in terms of class and magnification factor.

| Magnification | Benign ($n = 20$) | Malignant ($n = 24$) | Total ($n = 44$) |
|------------------------|---------------------|------------------------|--------------------|
| 40 \times | 40 | 48 | 88 |
| 100 \times | 40 | 48 | 88 |
| 200 \times | 40 | 48 | 88 |
| 400 \times | 40 | 48 | 88 |
| Total number of images | 160 | 192 | 352 |

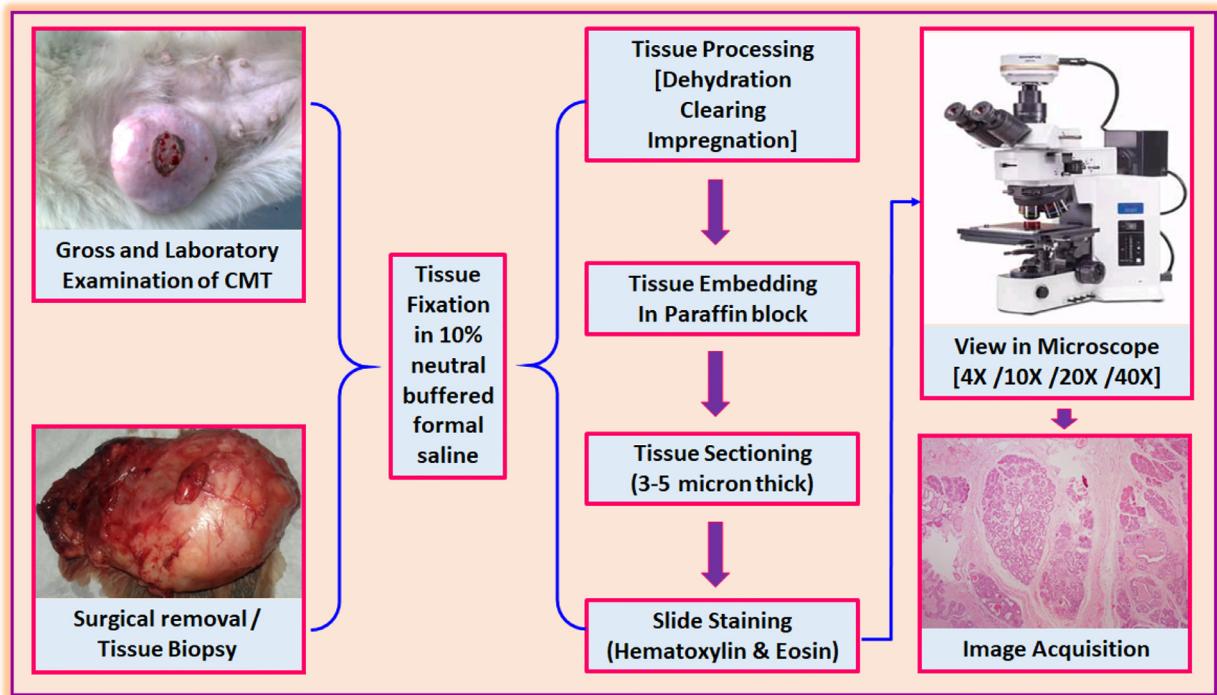


Fig. 1. Image acquisition framework.

Table 2
BreakHis dataset image distribution in terms of class and magnification factor.

| Magnification | Benign (<i>n</i> = 24) | Malignant (<i>n</i> = 58) | Total (<i>n</i> = 82) |
|------------------------|-------------------------|----------------------------|------------------------|
| 40 × | 625 | 1370 | 1995 |
| 100 × | 644 | 1437 | 2081 |
| 200 × | 623 | 1390 | 2013 |
| 400 × | 588 | 1232 | 1820 |
| Total number of images | 2480 | 5429 | 7909 |

image colors are converted to an optical density(OD) using the logarithmic transformation shown in the Eq. (4).

$$OD = -\log_{10}(I), \quad \text{where } I = (i_r, i_g, i_b) \quad (4)$$

Here, I is an RGB color vector with each component normalized to [0,1] and representing the OD converted RGB color space by following matrix A of size $[m,n]$, where m represents number of stains and n is number of color channel. In our case n is 3.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Here, rows represent specific stains and columns represent the optical density detected for each stain by the red, green, and blue channels.

Further, to acquire independent information for each stain, color deconvolution as described in [28] was used. Here the color values were transformed using the ortho-normal transformation of the RGB information using the equation:

$$A = VS \Rightarrow S = V^{-1}A \quad (5)$$

In this equation, A is the observed optical density, V and S are the matrices of the stain vectors and the saturation of each of the stains, respectively.

To find 2D projections with higher variance, singular value decomposition (SVD) algorithm is applied to the OD tuples. The resulting transformation of color space applies to the original image. Finally, the image histogram is extended to cover the dynamic range of the lower 90% of the data. This process is described in Algorithm 1 and 2.

Algorithm 1 SVD-geodesic method.

Input: RGB Slide $\{(X_i, y_i)\}_{i=1}^N$
Output: Optimal stain vectors

- 1: $X \leftarrow$ Normalized input to $[0, 1]$
- 2: $A \leftarrow \text{OPTICAL_DENSITY}(X)$
- 3: $A' \leftarrow \text{UPDATE_OPERATION}(A)$
- 4: $A'' \leftarrow \text{SVD}(A')$
- 5: $(\alpha_1, \alpha_2) \leftarrow$ Two maximum values of A''
- 6: $P \leftarrow \text{CONSTRUCT_PLANE}(\alpha_1, \alpha_2)$
- 7: Project A'' on P and normalize to unit length
- 8: $\delta \leftarrow$ Compute angle of A'' using first SVD direction
- 9: Identify the robust extremes from δ
- 10: Transform extreme value back to OD space

return : Optimal stain vectors

Algorithm 2 Update_Operation.

Input: A, β
Output: A

- 1: **for** $\forall a_{ij} \in A$ **do**
- 2: **if** $a_{ij} < \beta$ **then**
- 3: $a_{ij} \leftarrow 0$

return : A

3.3. Data augmentation

Features extracted from pre-trained ConvNet are not guaranteed to be invariant in terms of the position or orientation of the tissue in an image or image patch. Also, ConvNet 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 [35,46], 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 $90^\circ, 180^\circ, 270^\circ$, 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 12 times.

3.4. ConvNets as a feature extractor

Traditional machine learning methods are used to extract features from images using global feature descriptors such as Local Binary Patterns, Histogram of Oriented Gradients, etc., or Local descriptors such as ORB, SURF, SIFT, etc. These are hand-crafted features which require expertise at that domain. However, instead of using hand-crafted features, ConvNets 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, color, etc. Higher layers learn high-level features that represent the object in the image [49]. Instead of making a ConvNet 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 ConvNets are used for feature extraction instead of training time-consuming and tedious ConvNets from scratch.

The ConvNet 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 [9] 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 2.3. Yosinski et al. [47] suggest that higher layers provide less-generic characteristics with a reduced target performance. The simple hierarchical structure found in most ConvNet 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 regularizer 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 con-

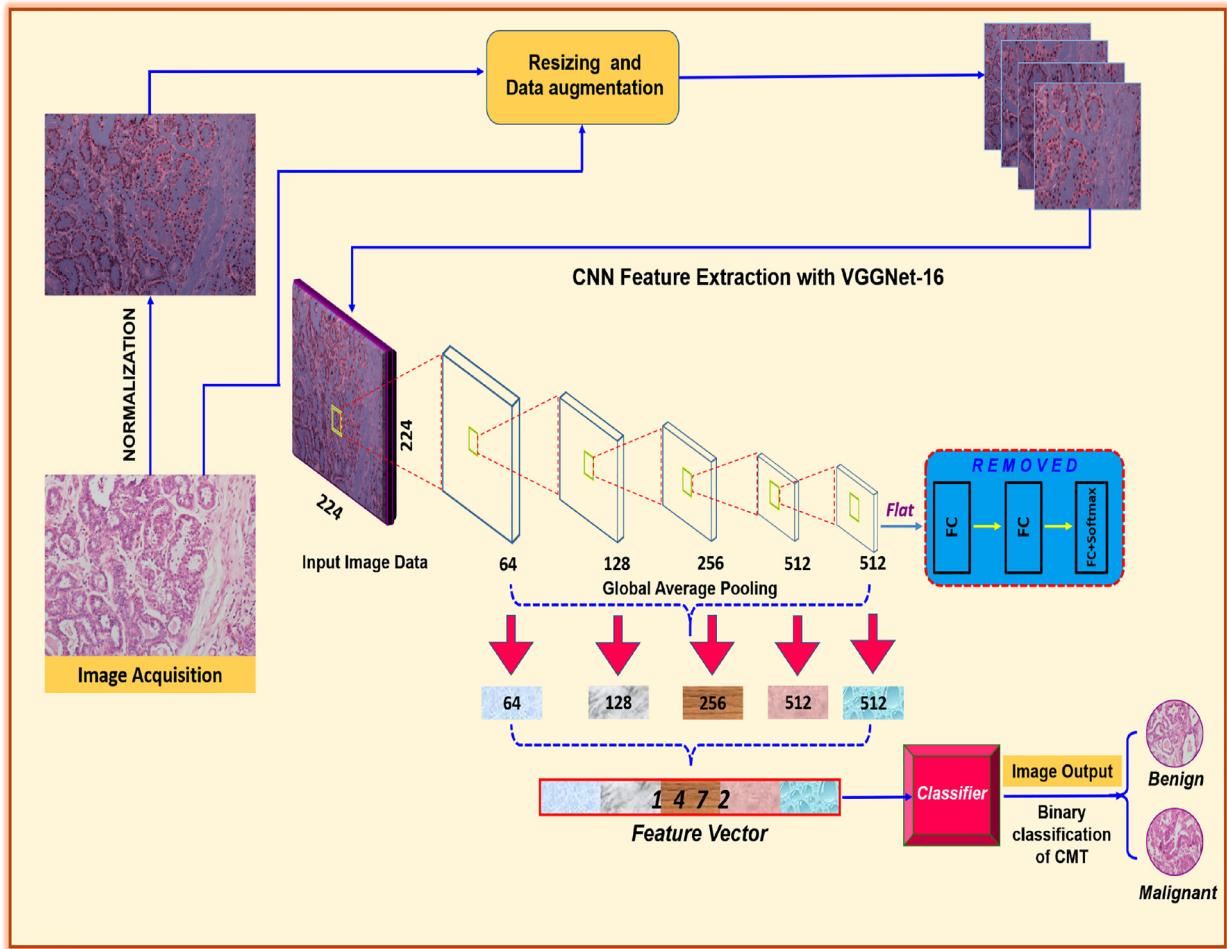


Fig. 2. Proposed framework for histopathology image classification. H&E stained histopathological tumor images were initially subjected to data augmentation and stain normalization. Finally, the images with and without stain normalization were applied to the fused VGGNet-16 model for binary classification of cancer images, where all FC layers are removed.

volutional 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 Fig. 2.

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

3.5. Classifiers

Two different classifiers were used to access the aforementioned feature set:

3.5.1. Support vector machine

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 [45] 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.

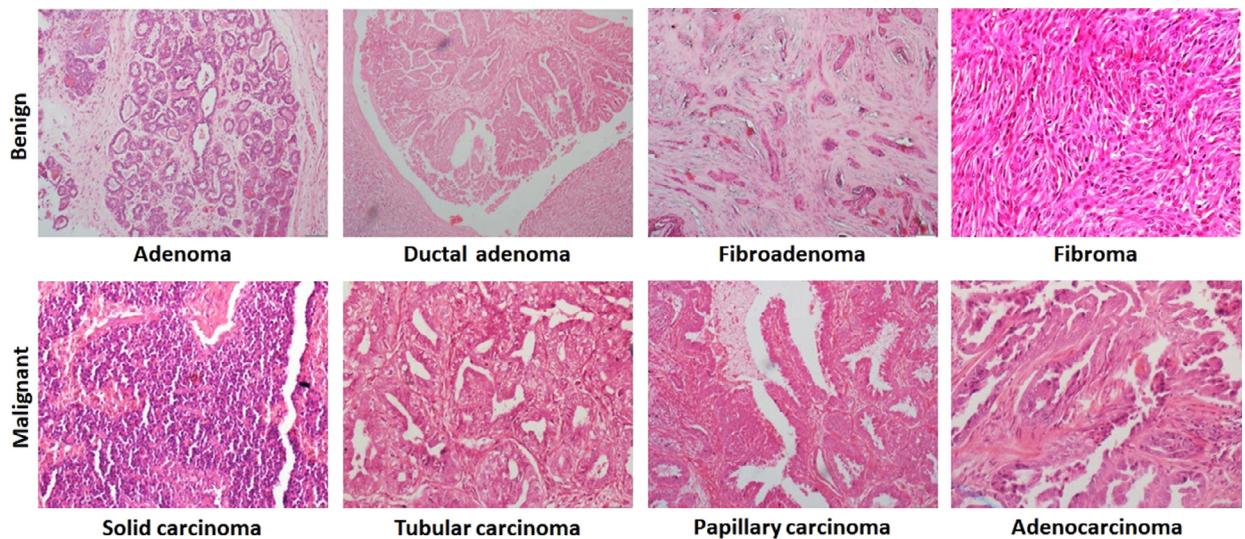


Fig. 3. Representative H&E stained images from CMTHis dataset.

3.5.2. Random forest

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 [6]. 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 random forest for classification and detection. In our proposed model, the extracted feature from ConvNet is used by random forest for classification. One advantage of RF is that it is quite fast and can handle large databases and unbalanced data efficiently.

4. Experimental results and analysis

This section presents the experimental results and comparative analysis to demonstrate the efficacy of the proposed method over existing methods. The framework described in [Section 3](#) is implemented using the Keras library [7]. 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.

4.1. Experimental data

Two datasets, one of the canine mammary tumor (CMTHis) and the other of human breast cancer histopathological images (BreakHis), were used in this study. Hereby, we have introduced a dataset of canine mammary tumor histopathology images. This dataset currently contains four histopathologically distinct types of benign tumors, namely, adenoma, ductal adenoma, fibroadenoma, and fibroma, and four malignant tumors, namely, adenocarcinoma, solid carcinoma, tubular carcinoma, and papillary carcinoma. The representative H&E stained images from the CMTHis dataset showing typical benign and malignant CMTs are illustrated in [Fig. 3](#). The proposed framework was first evaluated on a standard and challenging BreakHis dataset comprising of 7909 images from 82 human breast cancer patients. The dataset included four distinct types of benign breast tumors (fibroadenoma, adenosis, tubular adenoma, and phyllodes tumor) and four types of malignant tumor (lobular carcinoma, ductal carcinoma, papillary carcinoma, and mucinous carcinoma). Representative histopathological images from the BreakHis dataset are shown in [Fig. 4](#).

4.2. Experimental protocol

The framework proposed in [Section 3.4](#) was applied to both datasets (BreakHis and CMTHis), which were randomly divided into a training set (70%) and test set (30%). The split protocol [42] was used to generate 5-folds and results were presented by taking an average of 5 folds. To ensure that the classifier was generalized for unseen patients, it was ensured that the test set did not include patients used to create the training set. This was done to ensure that the efficiency of the test is assessed on a dataset that is not used to train the classifier. Since the CMTHis dataset is introduced for the first time in this study, the accuracy and performances of the proposed algorithms were first validated on a BreakHis dataset to ensure that the results were tested on the standard dataset with a large number of images. In our experiments with CMTHis, we have randomly chosen 31 patients (70%) for training and the remaining 13 for testing (30%). The results of each test are given in terms of accuracy (percentage of correctly classified instances).

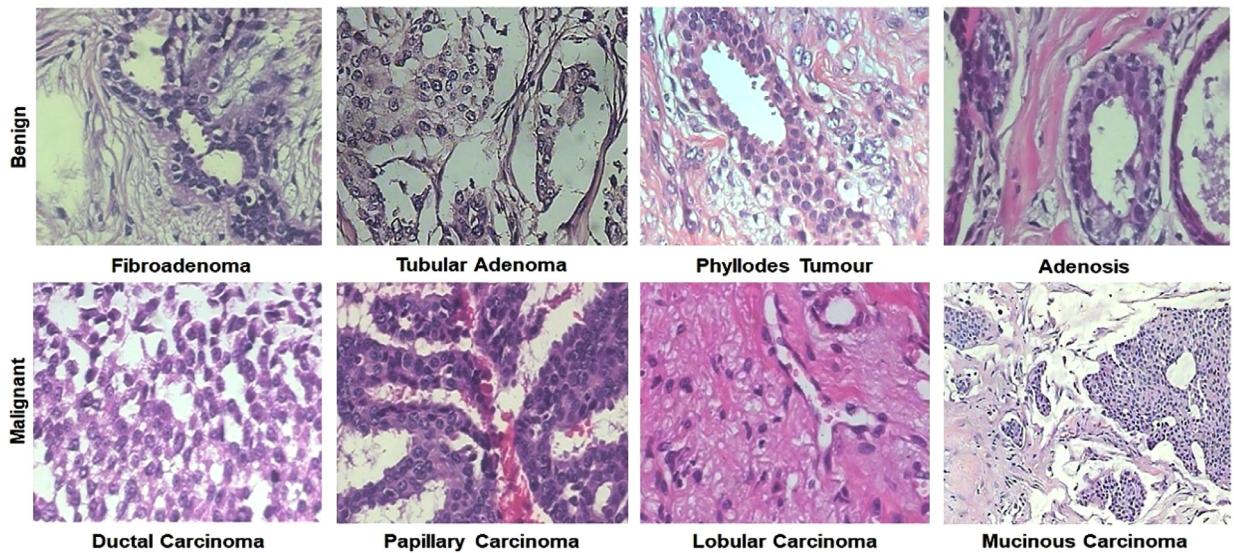


Fig. 4. Representative H&E stained images from BreakHis dataset.

Table 3

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

| BreakHis dataset | | | |
|------------------------------|---------------|---------------------|----------|
| Proposed framework | Magnification | Test accuracy(%) | F1 score |
| FE-VGGNET16-RF | 40 × | 92.22 ± 2.14 | 0.94 |
| | 100 × | 93.40 ± 4.38 | 0.95 |
| | 200 × | 95.23 ± 1.89 | 0.97 |
| | 400 × | 92.80 ± 1.83 | 0.94 |
| FE-VGGNET16-SVM(LIN) | 40 × | 93.82 ± 1.45 | 0.94 |
| | 100 × | 94.98 ± 1.13 | 0.95 |
| | 200 × | 95.77 ± 1.02 | 0.97 |
| | 400 × | 92.40 ± 0.62 | 0.95 |
| FE-VGGNET16-SVM(RBF) | 40 × | 92.60 ± 1.52 | 0.97 |
| | 100 × | 93.49 ± 1.62 | 0.97 |
| | 200 × | 95.13 ± 1.96 | 0.98 |
| | 400 × | 94.96 ± 2.19 | 0.97 |
| FE-VGGNET16-SVM(POLY) | 40 × | 94.11 ± 1.83 | 0.96 |
| | 100 × | 95.12 ± 1.10 | 0.97 |
| | 200 × | 97.01 ± 1.14 | 0.98 |
| | 400 × | 93.40 ± 1.01 | 0.96 |

4.3. 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 BreakHis dataset ranged from 86 to 90%, whereas, in CMTHis dataset, test accuracies ranging from 78 to 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 VGGNet-16-SVM model on the CMTHis dataset (CMT image dataset) and BreakHis human breast cancer 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 3](#). 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, as shown in [Fig. 5](#). The ROC curve analysis, shown in [Fig. 6\(a\)](#), 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

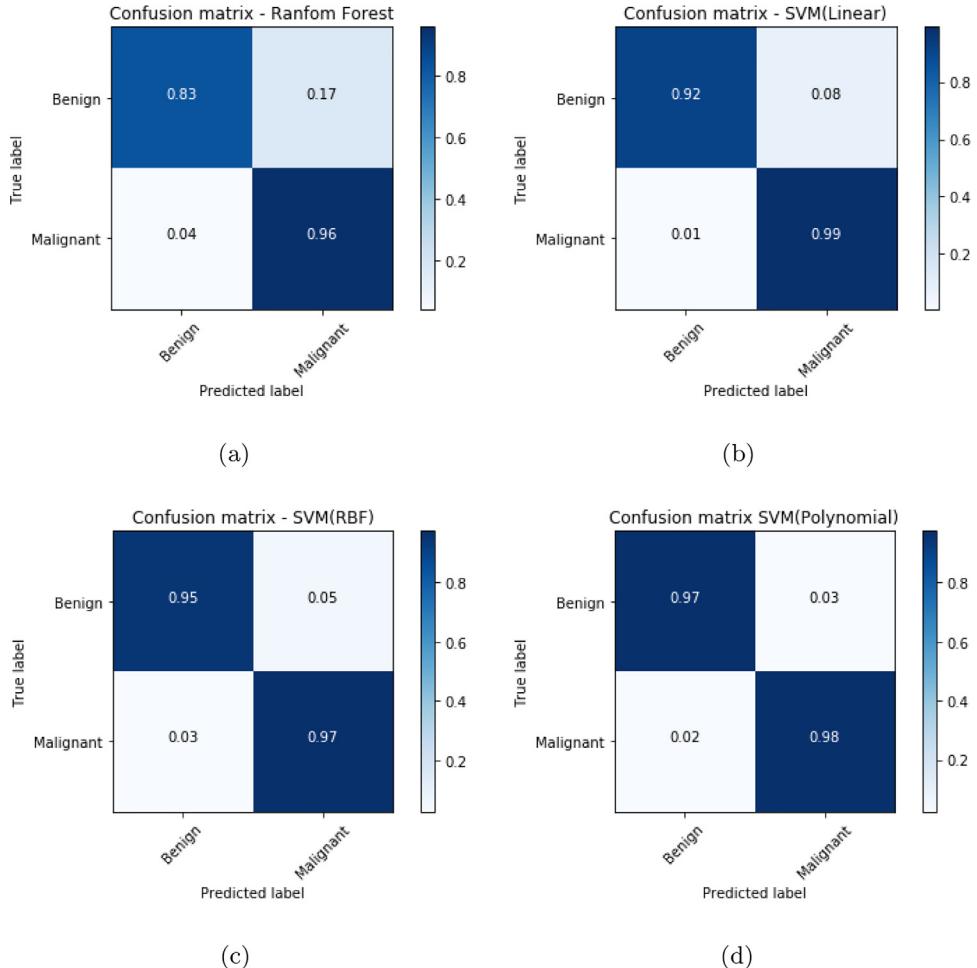


Fig. 5. Confusion matrix plots of BreakHis dataset at $200 \times$ for (a) Random Forest(RF), (b) SVM(LIN), (c) SVM(RBF) and (d) SVM(POLY).

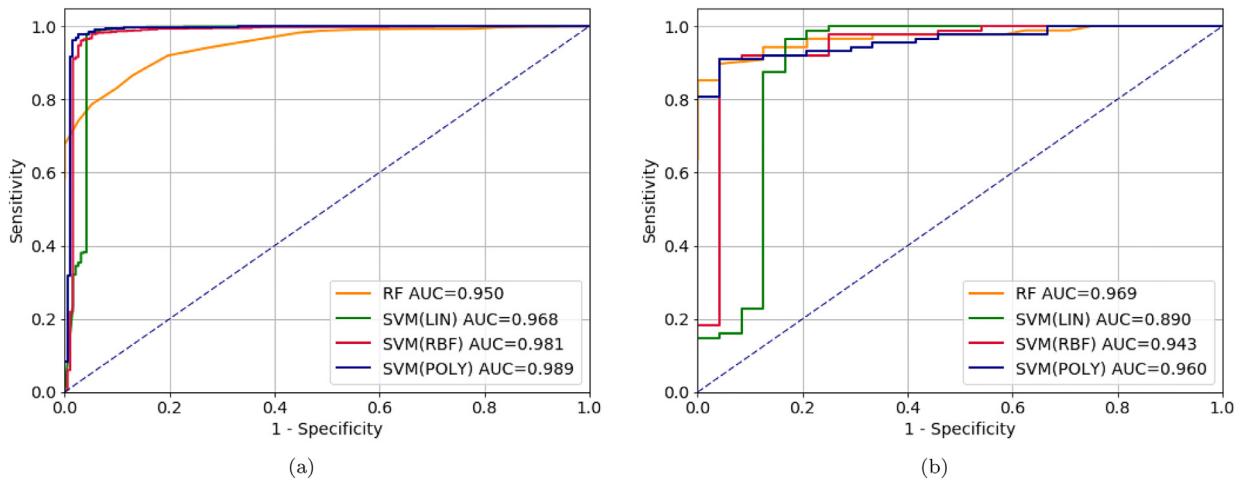


Fig. 6. AUC for different classifiers at $200 \times$ for (a) BreakHis dataset and (b) CMTHis dataset.

Table 4

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

| CMTHis | | | |
|------------------------------|---------------|---------------------|----------|
| Proposed framework | Magnification | Test accuracy(%) | F1 score |
| FE-VGGNET16-RF | 40 × | 91.27 ± 2.93 | 0.94 |
| | 100 × | 91.40 ± 3.00 | 0.94 |
| | 200 × | 86.13 ± 1.31 | 0.89 |
| | 400 × | 81.63 ± 1.56 | 0.86 |
| FE-VGGNET16-SVM(LIN) | 40 × | 91.31 ± 1.69 | 0.94 |
| | 100 × | 91.35 ± 4.50 | 0.94 |
| | 200 × | 89.07 ± 3.74 | 0.93 |
| | 400 × | 83.35 ± 6.07 | 0.88 |
| FE-VGGNET16-SVM(RBF) | 40 × | 89.09 ± 4.10 | 0.93 |
| | 100 × | 89.15 ± 3.40 | 0.93 |
| | 200 × | 85.44 ± 2.52 | 0.90 |
| | 400 × | 81.87 ± 0.69 | 0.87 |
| FE-VGGNET16-SVM(POLY) | 40 × | 91.95 ± 2.59 | 0.95 |
| | 100 × | 92.75 ± 4.47 | 0.95 |
| | 200 × | 89.12 ± 2.95 | 0.93 |
| | 400 × | 81.15 ± 1.69 | 0.86 |

able to successfully classify malignant and benign canine mammary tumors 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](#). 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 [Fig. 6\(b\)](#), demonstrating the efficacy of the framework to classify CMTs.

4.3.1. 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., Random Forest, SVM(Linear), SVM(Poly) and SVM(RBF), were tested on each of the four magnifications independently to assess the effect of magnification on the method performance. 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 × magnification.

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 3](#), 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 [3,16,29,41]. Various researchers [16,18,41] 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](#). 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. [41] also showed that the accuracy of ConvNet (ImageNet) decreases with increase in magnification.

4.4. Performance comparison with state-of-the-art ConvNet architectures

We compared the average performance of the proposed framework with the state-of-the-art ConvNet architectures and other multi-layered framework approaches. The comparative results shown in [Table 5](#) demonstrate that the proposed framework outperforms most state-of-the-art approaches. Spanhol et al. in [41] tested LeNet [24], a traditional ConvNet 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 [23] architecture and achieved a maximum average accuracy of 84.4% in the binary classification of histopathological images of human breast cancer. In yet another study, Spanhol et al. [40] used

Table 5

Comparison of our framework with state-of-the-art ConvNets at various magnifications.

| | Methods | Dataset | Test accuracies(%) at different magnification factors | | | |
|---------------------------|------------------------|----------|---|---------------------|---------------------|---------------------|
| | | | 40 × | 100 × | 200 × | 400 × |
| Existing framework | Bayramoglu et al. [3] | BreakHis | 83.08 ± 2.08 | 83.17 ± 3.51 | 84.63 ± 2.72 | 82.10 ± 4.42 |
| | Spanhol et al. [41] | BreakHis | 89.60 ± 6.50 | 85.00 ± 4.80 | 82.80 ± 2.10 | 80.02 ± 3.40 |
| | Spanhol et al. [40] | BreakHis | 84.60 ± 2.90 | 84.80 ± 4.20 | 84.20 ± 1.70 | 81.60 ± 3.70 |
| | Song et al. [38] | BreakHis | 90.02 ± 3.20 | 88.90 ± 5.00 | 86.90 ± 5.20 | 86.30 ± 7.00 |
| | Song et al. [37] | BreakHis | 90.02 ± 3.20 | 91.20 ± 4.40 | 87.80 ± 5.30 | 87.40 ± 7.20 |
| | Han et al. [18] | BreakHis | 95.80 ± 3.10 | 96.90 ± 1.90 | 96.70 ± 2.00 | 94.90 ± 2.80 |
| | Gupta and Bhavsar [16] | BreakHis | 86.74 ± 2.37 | 88.56 ± 2.70 | 90.31 ± 3.76 | 88.31 ± 3.01 |
| | Gupta and Bhavsar [17] | BreakHis | 84.72 | 89.44 | 95.65 | 82.65 |
| | Gupta and Bhavsar [17] | BreakHis | 91.90 | 93.64 | 95.84 | 90.15 |
| | FE-VGGNET16-RF | BreakHis | 92.22 ± 2.14 | 93.40 ± 4.38 | 95.23 ± 1.89 | 92.80 ± 1.83 |
| Proposed Approach | 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 ± 4.50 | 89.07 ± 3.74 | 83.35 ± 6.07 |
| | FE-VGGNET16-SVM(RBF) | CMTHiS | 89.09 ± 4.10 | 89.15 ± 3.40 | 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 |

pre-trained BVLC CaffeNet architecture along with a logistic regression classifier resulting in accuracy in the range of 81.6–84.8%. Bayramoglu et al. [3], 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, structure-based deep convolutional neural network (CSDCNN) has been proposed for human breast cancer image classification with 94.9–96.9% accuracy for binary classification [18]. 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 straight-forward compared to the existing frameworks.

4.5. Effect of contemporary classifiers in magnification independent model

The proposed framework was also applied in magnification independent model and mean accuracies at different magnifications were compared between all contemporary classifiers applied to VGGNet-16 architecture. The results, shown in Fig. 7,

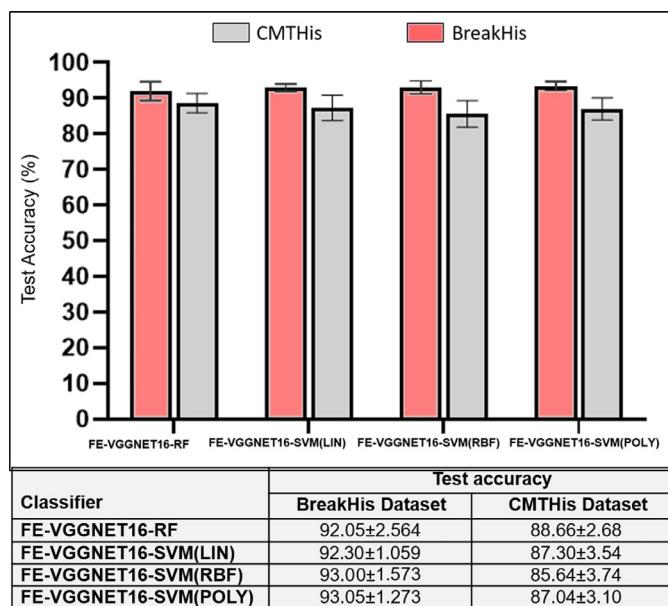


Fig. 7. Mean test accuracy for different classifiers applied to proposed framework in magnification independent model.

Table 6

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

| Method | Mean accuracy (%) |
|-----------------|------------------------------|
| Existing | Bayramoglu [3] 83.24 |
| | Gupta and Bhavsar [16] 87.53 |
| | Sharma and Mehra [29] 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 |

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 [3,16,29] for binary classification of breast cancers, and the results are compiled in **Table 6**. It is clearly evident from the **Table 6** that the proposed magnification independent model 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.

4.6. Effect of stain normalization

As discussed by several researchers [25,48], 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 3**, was applied to our proposed framework and the results before and after stain normalization were compared. **Fig. 8** shows H&E stained histopathological images from the CMTHis dataset, with and without stain normalization. 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 **Figs. 9** and **10**. The performances of our proposed framework without stain normalization were slightly better than with the stain normalization, though the differences were not significant. Thus, the proposed algorithm gives equally good results without stain normalization, and thus the method avoids unnecessary time and labor in the application of stain normalization algorithms. Therefore, using the proposed framework, histopathological slides with wide color variations may not require stain normalization.

Similar results have been demonstrated by Reinhard [33], who also showed that the application of a stain normalization algorithm on ConvNets slightly reduces the performance. Studies have previously found the usefulness of color normalization algorithms when using handcrafted features based computer algorithms. Results from this study and other [15,33] suggest that ConvNets are effective in learning the task in the presence of color variation. Another reason may be less color variability in the BreaKHis dataset collected from a single laboratory with a single microscopic system. Existing color normalization methods were mostly developed for machine learning schemes employing handcrafted features (as opposed to deep learn-

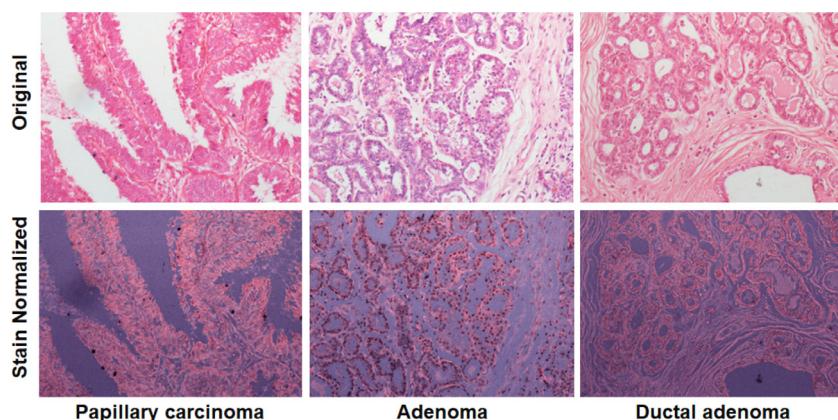


Fig. 8. H&E stained images from CMTHis dataset, before and after stain normalization.

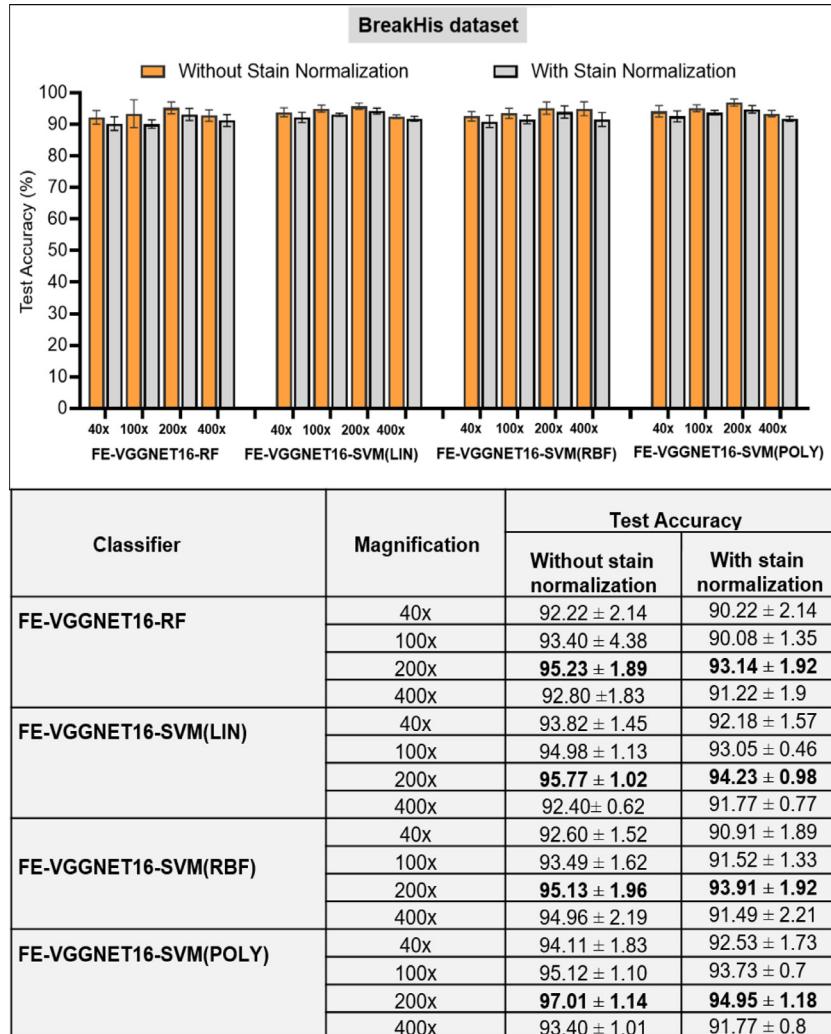


Fig. 9. Mean test accuracy for different classifiers of proposed framework with and without stain normalization on BreakHis dataset.

ing/ConvNets). These schemes may or may not perform other pre-processing processes on the input image. ConvNets, on the other hand, typically perform (channel-wise) mean subtraction on the images. This may have an interaction with a specific color normalization method.

5. Conclusion and future scope

In this study, for the first time, we have introduced a dataset of canine mammary tumor histopathological images, and we have presented a preliminary study on automated binary classification of canine mammary tumors (CMTs). We have also proposed a variant of VGGNet-16, wherein a FC layer was removed and experimented with various classifiers. The model used in this study achieved high accuracy ($\approx 97\%$) and outperforms most state-of-the-art approaches used so far for binary classification of human breast cancer. 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 database, as compared to BreakHis database, might be the small size of the CMTHis database. Therefore, in the future, further studies with a large number of CMT patients and histopathological images are required to prove the efficacy of the proposed framework for binary classification of canine mammary tumors.

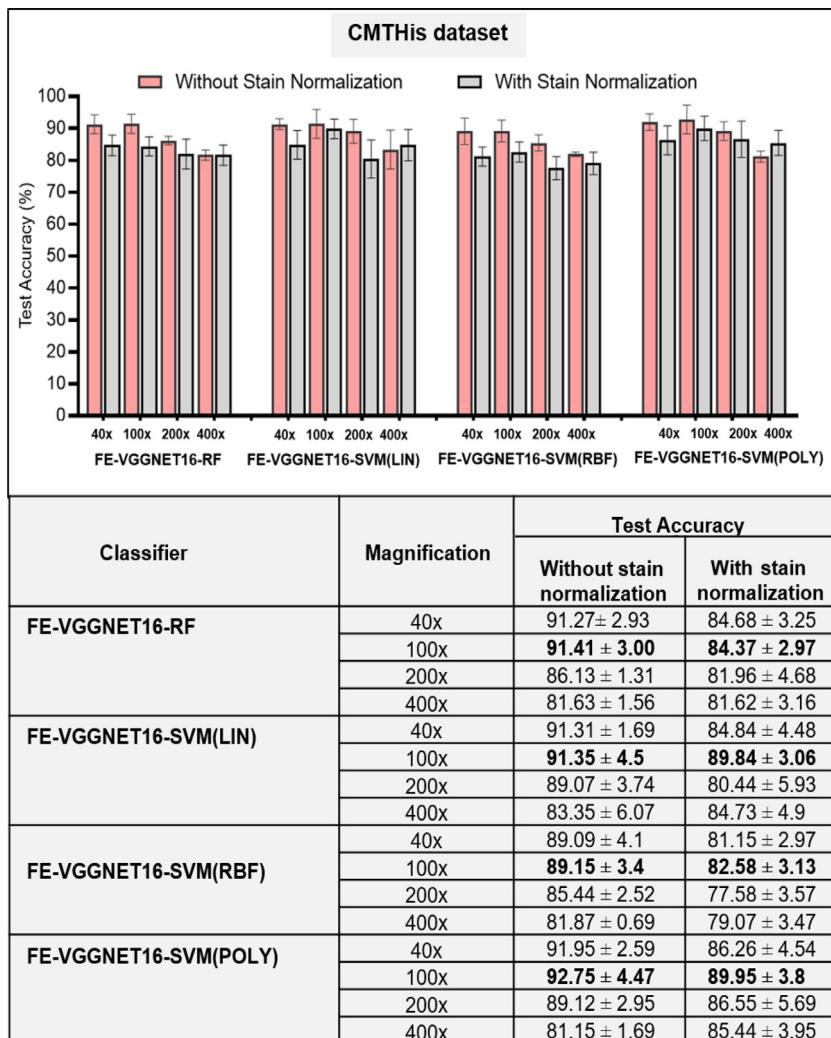


Fig. 10. Mean test accuracy for different classifiers of proposed frameworks with and without stain normalization on CMTHis dataset.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgement

Authors are thankful to Director IIT(BHU) and Director ICAR-IVRI for providing necessary infrastructure facilities. This study was supported by the grants received from, Department of Biotechnology (DBT), Government of India.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ins.2019.08.072](https://doi.org/10.1016/j.ins.2019.08.072).

References

- [1] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, N.M. Rajpoot, Stain deconvolution using statistical analysis of multi-resolution stain colour representation, *PloS One* 12 (1) (2017) e0169875.
- [2] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan, Chest pathology detection using deep learning with non-medical training, in: Proceedings of the International Symposium on Biomedical Imaging (ISBI), Citeseer, 2015, pp. 294–297.
- [3] N. Bayramoglu, J. Kannala, J. Heikkilä, Deep learning for magnification independent breast cancer histopathology image classification, in: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 2440–2445.

- [4] W. Beauvais, J. Cardwell, D. Brodbelt, The effect of neutering on the risk of mammary tumours in dogs—a systematic review, *J. Small Anim. Pract.* 53 (6) (2012) 314–322.
- [5] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: Cancer J. Clin. 68 (6) (2018) 394–424.
- [6] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [7] F. Chollet, 2015.
- [8] F. Ciompi, B. de Hoop, S.J. van Riel, K. Chung, E.T. Scholten, M. Oudkerk, P.A. de Jong, M. Prokop, B. van Ginneken, Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box, *Med. Image Anal.* 26 (1) (2015) 195–202.
- [9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, IEEE, 2009, pp. 248–255.
- [10] A. Egenvall, B.N. Bonnett, P. Öhagen, P. Olson, Å. Hedhammar, H. von Euler, Incidence of and survival after mammary tumors in a population of over 80,000 insured female dogs in sweden from 1995 to 2002, *Prev. Vet. Med.* 69 (1–2) (2005) 109–127.
- [11] J.G. Elmore, G.M. Longton, P.A. Carney, B.M. Geller, T. Onega, A.N. Tosteson, H.D. Nelson, M.S. Pepe, K.H. Allison, S.J. Schnitt, et al., Diagnostic concordance among pathologists interpreting breast biopsy specimens, *JAMA* 313 (11) (2015) 1122–1132.
- [12] C.L. Giles, C.W. Omlin, Pruning recurrent neural networks for improved generalization performance, *IEEE Trans. Neural Netw.* 5 (5) (1994) 848–851.
- [13] M. Goldschmidt, L. Peña, R. Rasotto, V. Zappulli, Classification and grading of canine mammary tumors, *Vet. Pathol.* 48 (1) (2011) 117–131.
- [14] H. Greenspan, B. Van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1153–1159.
- [15] K. Grünzig, R. Graf, M. Häsig, M. Welle, D. Meier, G. Lott, D. Erni, N. Schenker, F. Gusetti, G. Boo, et al., The swiss canine cancer registry: a retrospective study on the occurrence of tumours in dogs in switzerland from 1955 to 2008, *J. Comp. Pathol.* 152 (2–3) (2015) 161–171.
- [16] V. Gupta, A. Bhavsar, Breast cancer histopathological image classification: is magnification important? in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 2440–2445.
- [17] V. Gupta, A. Bhavsar, Sequential modeling of deep features for breast cancer histopathological image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2254–2261.
- [18] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, S. Li, Breast cancer multi-classification from histopathological images with structured deep learning model, *Sci. Rep.* 7 (1) (2017) 1–10.
- [19] S. Hussain, S. Saxena, S. Shrivastava, R. Arora, R.J. Singh, S.C. Jena, N. Kumar, A.K. Sharma, M. Sahoo, A.K. Tiwari, et al., Multiplexed autoantibody signature for serological detection of canine mammary tumours, *Sci. Rep.* 8 (1) (2018) 15785.
- [20] S. Hussain, S. Saxena, S. Shrivastava, A.K. Mohanty, S. Kumar, R.J. Singh, A. Kumar, S.A. Wani, R.K. Gandham, N. Kumar, et al., Gene expression profiling of spontaneously occurring canine mammary tumours: insight into gene networks and pathways linked to cancer pathogenesis, *PLoS One* 13 (12) (2018) e0208656.
- [21] A.M. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, *IEEE Trans. Biomed. Eng.* 61 (6) (2014) 1729–1738.
- [22] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, I. Hajirasouliha, Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images, *EBioMedicine* 27 (2018) 317–328.
- [23] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
- [24] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, et al., Comparison of learning algorithms for handwritten digit recognition, in: Proceedings of the International Conference on Artificial Neural Networks, 60, Perth, Australia, 1995, pp. 53–60.
- [25] G. Lee, M. Bajer, K. Clark, Deep learning and color variability in breast cancer histopathological images: a preliminary study, in: Proceedings of the Conference Series Society of Photo-Optical Instrumentation Engineers (SPIE), 10718, 2018.
- [26] H. Li, A. Kadav, I. Durdanovic, H. Samet, H.P. Graf, Pruning filters for efficient convnets, (2016) arXiv:1608.08710.
- [27] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88.
- [28] M. Macenko, M. Niethammer, J.S. Marron, D. Borland, J.T. Woosley, X. Guan, C. Schmitt, N.E. Thomas, A method for normalizing histology slides for quantitative analysis, in: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI, IEEE, 2009, pp. 1107–1110.
- [29] R. Mehra, et al., Automatic magnification independent classification of breast cancer tissue in histological images using deep convolutional neural network, in: International Conference on Advanced Informatics for Computing Research, Springer, 2018, pp. 772–781.
- [30] D. Mehta, K.J. Kim, C. Theobalt, On implicit filter level sparsity in convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 520–528.
- [31] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1717–1724.
- [32] S.J. Pan, Q. Yang, et al., A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345–1359.
- [33] E. Reinhard, M. Adhikmin, B. Gooch, P. Shirley, Color transfer between images, *IEEE Comput. Gr. Appl.* 21 (5) (2001) 34–41.
- [34] Y. Salas, A. Márquez, D. Diaz, L. Romero, Epidemiological study of mammary tumors in female dogs diagnosed during the period 2002–2012: a growing animal health problem, *PLoS One* 10 (5) (2015) e0127381.
- [35] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. Van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sánchez, B. van Ginneken, Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1160–1169.
- [36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, (2014) arXiv:1409.1556.
- [37] Y. Song, H. Chang, H. Huang, W. Cai, Supervised intra-embedding of fisher vectors for histopathology image classification, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 99–106.
- [38] Y. Song, J.J. Zou, H. Chang, W. Cai, Adapting fisher vectors for histopathology image classification, in: Proceedings of the IEEE 14th International Symposium on Biomedical Imaging (ISBI), IEEE, 2017, pp. 600–603.
- [39] K. Sorenmo, Canine mammary gland tumors, *Vet. Clin. Small Anim. Pract.* 33 (3) (2003) 573–596.
- [40] F.A. Spanhol, L.S. Oliveira, P.R. Cavalin, C. Petitjean, L. Heutte, Deep features for breast cancer histopathological image classification, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2017, pp. 1868–1873.
- [41] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2016, pp. 2560–2567.
- [42] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, *IEEE Trans. Biomed. Eng.* 63 (7) (2016) 1455–1462.
- [43] P. Sudharshan, C. Petitjean, F. Spanhol, L.E. Oliveira, L. Heutte, P. Honeine, Multiple instance learning for histopathological breast cancer image classification, *Expert Syst. Appl.* 117 (2019) 103–111.
- [44] J. Tang, R.M. Rangayyan, J. Xu, I. El Naqa, Y. Yang, Computer-aided detection and diagnosis of breast cancer with mammography: recent advances, *IEEE Trans. Inf. Technol. Biomed.* 13 (2) (2009) 236–251.
- [45] V. Vapnik, The Nature of Statistical Learning Theory, Springer Science & Business Media, 2013.
- [46] R. Wu, S. Yan, Y. Shan, Q. Dang, G. Sun, Deep image: scaling up image recognition, (2015) arXiv:1501.02876.

- [47] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? in: *Advances in Neural Information Processing Systems*, 2014, pp. 3320–3328.
- [48] F.G. Zanjani, S. Zinger, B.E. Bejnordi, J.A. van der Laak, P.H. de With, Stain normalization of histopathology images using generative adversarial networks, in: *Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2018, pp. 573–577.
- [49] M.D. Zeiler, G.W. Taylor, R. Fergus, et al., Adaptive deconvolutional networks for mid and high level feature learning, in: *Proceedings of the ICCV*, 1, 2011, p. 6.
- [50] S.K. Zhou, H. Greenspan, D. Shen, *Deep Learning for Medical Image Analysis*, Academic Press, 2017.