This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Partially-Independent Framework for Breast Cancer Histopathological Image Classification

Vibha Gupta, Arnav Bhavsar
School of Computing & Electrical Engineering, Indian Institute of Technology Mandi, India
gupta85vibha@gmail.com, arnav@iitmandi.ac.in

Abstract

The automated classification of histopathology images relives pathologists workload and, hence utilizing the resources to focus more on the most suspicious cases. More recently, inspired by the success of deep learning methods in computer vision application, such frameworks have also been applied in various medical image analysis applications. However, existing approaches showed less interest in exploring multi-layer features for improving the classification. We propose the integration of multi-layer features from a ResNet model for breast cancer histopathology image classification. Specifically, this work focuses on making a framework which considers both independent nature of layers as well as some partial dependency among them. Knowing that, not all the layers learn discriminative features, consideration of layers which learn to negative features will deteriorate the accuracy. Hence, we select the optimal subset of the layers based on an information-theoretic measure (ITS). Various experiments are performed on publicly available BreaKHis dataset, and demonstrate that the proposed multi-layer feature fusion yields better performance than the traditional way of using the highest layer features. This indicates that mid- and low-level features also carry useful discriminative information when explicitly considered. We also demonstrate improved performance, in most cases, over various state-of-the-art methods.

1. Introduction

Breast cancer (BC) has been the most common type of cancer detected in women and one of the most prevalent causes of women's death. Out of all diagnoses, 23% are identified to be breast cancer, making it one of the biggest cancer threats after lung cancer, with breast cancer accounting for 14% of all cancer deaths. Thus it's evident the amount of human population impacted by breast cancer is huge. Hence developing better diagnosing tools will, therefore, bring a positive change in lives of millions of people. A silver lining is that an early detection of breast cancer increases the range of available treatment options and enhances the probability of survival for patients. As in most cancers, and many other diseases, early detection can be a big advantage in treatment of breast cancer, helping reduce the mortality rate.

A number of tests are used in detection and diagnosis of breast cancer. Currently, the most common procedure is mammography followed by a biopsy. A biopsy is the most preferred examination, and a very certain way to know if it's cancer. While conducting a biopsy, the doctor uses a specialized needle device guided by X-ray or another imaging test to extract a sample of tissue from the tumorous area. Normally, a small metal marker is left at the site within the breast to facilitate identification during future imaging tests. While, different types of imaging technologies have been employed for diagnosis of BC, histopathology biopsy imaging has been a 'gold standard' in diagnosing breast cancer because it captures a comprehensive view of the effect of the disease on the tissues [1]. Examples of microscopy images of some breast samples (biopsies) are shown in Figure 1.

Image examination by pathologists can be tedious and can also be affected by observer variability. Computerized methods significantly improve the efficiency and objectivity, and have thereby attracted considerable attention in recent research. Computer-aided diagnosis (CADx) which is a concept established by taking into account equally the roles of physicians and computers, can relieve the pathologists workload such that attention can be focused on the most suspicious cases. Researchers face various challenges while designing CADx systems for histopathology images [2] such as: (1) Insufficient labeled images, (2) Inhomogeneous textured images, (3) Different magnification levels, and (4) Color variation.

The CADx system typically consists of steps such as preprocessing, feature extraction and classification framework. To develop the CADx system for the classification of breast histopathology images, the literature reported various feature definitions that include different color-texture representations along with appropriate classification frameworks. An important consideration in such classical approaches is the feature learning aspect that requires expertise in introspecting the target domain, thus, in turn, determining the effectiveness of target classification tasks. The deep learning (DL) techniques enable features to be learned directly from the training data, optimized for high discrimination. Thus, the DL paradigm absorbs the feature engineering step into a learning step [3]. More recently, inspired by the success of DL methods in computer vision applications, these frameworks have also been applied in various applications for medical image analysis.

A variety of effective CNN architectures have been proposed in the past few years which have different salient aspects. Furthermore, in the deep learning frameworks, fea-

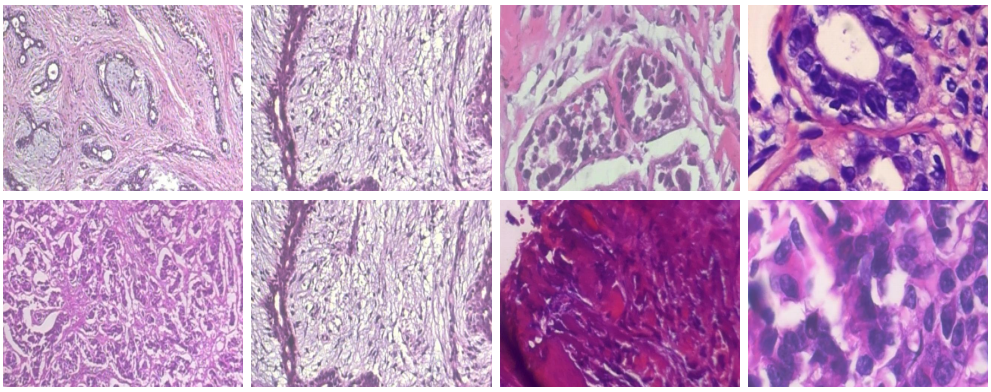

Figure 1. Heterogeneity in histopathological samples(first row: benign tumor, second row: malignant tumor) from BreakHis dataset at magnification factor of 40X-100X-200X-400X.

ture representations are learned at multiple levels of abstraction through multiple layers. Considering the appearance of histopathology images, inference in histopathology images may be based on low, mid and high level information. Thus, while important features for natural images can be observed in fully connected layers, for histopathology images, features learned by low, mid and high convolution layers may also play a role in classification. Noticing the prosperity of such deep multi-layered features in various domains including remote sensing scene classification [4], hyperspectral imagery [5] etc., there is still scope of improvement when employing them in an ensemble framework where mistakes made by each individual component can be compensated with the help of others. Thus, we envision that representations learned by different layers can potentially be used in an ensemble framework. There are various possible ways of designing an ensemble framework where such multi-layered features can be incorporated.

This study attempts to address the few of challenges discussed above, using ensembling over deep learning features. With regard to the first challenge that highlights the problem of data scarcity, we employ data augmentation to increase the data size. In addition, existing deep network with sufficient fine-tuning is exploited instead of learning the model from scratch. The second point is about the inhomogeneous textures, in a sense that it shows repetitive patterns of minimum components (usually cells) [2]. In this regard, deep textures learned by CNN would lead to the acquisition of invariance regarding cell position. The third challenge is the most informative magnification consensus [6]. Regarding to magnification related concerns, Spanhol at el. [7, 8], reported results on magnification-specific models. However, it would seem that one magnification model may not be able to handle images with other magnifications, and different frameworks are required at different magnifications, both in terms of features and classifiers. Bayramoglu at el. [9] proposed a magnification independent model which was trained with images of various magnifications (40x/100x/200x/400x) and can therefore handle the diversity of scale in microscopic images. The fourth challenge addresses the issue of color variation obtained due to a number of factors such as chemical reactivity from various manufacturers etc. [10].

In this work, we fine-tuned the ResNet [11] considering its popularity in recent contests, and utilize it for multi-layer feature extraction. Figure 3 shows the architecture of ResNet and residual module. The extracted deep features are employed in designed partially independent framework (more will be discussed in methodology). Knowing that, not all the layers learn discriminative features, we employ notion of layer selection based on information theory [12].

2. Related work

In this section, we discuss some previous works, including state-of-the-art methods, aiming to automate the diagnostic procedure in context of breast cancer histopathology.

Zhang et al. [13] proposed two-stage cascade framework incorporating a rejection option, utilizing multiple image descriptors along with random subspace ensembles. In another work [14], same authors assessed an ensembles of one-class classifiers using same dataset. Linder et al. [15] extracted the local binary pattern combined with a contrast measure (LBP/C) and performance evaluated using support vector machine (SVM). However, we note that these methods use an independent dataset (not public). To take away the impediment of publicly available data set, Spanhol et al. [7] released the BreakHis dataset for beast histopathology. The detailed description of dataset is given in experimental section. Below, we discuss the methods that have been developed using BreakHis dataset. Spanhol et al. [7] performed various experiments that involved the state-of-art texture descriptors such as Local Binary Pattern (LBP), Threshold Adjancey Statistics (PFTAS) etc. and four traditional classifiers and reported accuracy at patient level. In [8], pre-trained CNN (AlexNet) is utilized for classifica-

Table 1. Related work (using Break-His dataset).

Approaches	Descriptors	Training-Testing Protocol
Traditional Descriptors		
Spanhol et al. [7]	CLBP,GLCM,LBP,LPQ,ORB,PFTAS	70%-30% (Patient-wise)
Gupta et al. [16]	NCSR,GCF,MCCR,OCLBP,CDWT,GCM	70%-30% (Patient-wise)
Dimitropoulos et al. [17]	VLAD encoding	70%-30% (Patient-wise)
CNN Based Descriptors		
Spanhol et al. [8]	Variant based on AlexNet	70%-30% (Patient-wise)
Song et al. [18]	FV encoding of features extracted from CNN	70%-30% (Patient-wise)
Han et al. [19]	Class-structure based Deep Convolution Neural Network	75%-25% (Patient-wise)
Bardou et al. [20]	BOW and CNN	70%-30% (Not clear)
Nahid et al. [21]	Combination of LSTM and CNN	Not provided
Bayramolgu et al. [9]	Single and Multi-task CNN	70%-30% (Patient-wise)

tion. In [16] authors used various color-texture features in ensemble framework which contained different classifiers. Song et al. [18] proposed a classification model by combining convolution neural network with supervised intra-embedding of Fisher vectors. In [23], authors computed feature vector by fisher vector encoding of features extracted from pre-trained VGG-VD model and used with linear support vector machine. Han et al. [19], trained a deep model form scratch with GoogleNet architecture to identify subordinate classes of breast cancer (8) as well as for main classes (2). Dimitropoulos et al. [17] proposed a model for the grading of invasive breast carcinoma through the encoding of histological images as VLAD (Vector of Locally Aggregated Descriptors) representations on the Grassmann manifold. Nahid et al. [21] proposed CNN, Long-short-Term-Memory (LSTM) and, combination of CNN and LSTM for classification of breast cancer histopathology images. Bardou et al. [20] compared the performance of convolution neural network (designed) with handcrafted features encoded by bag of words and locally constrained linear coding. Bayramoglu at el. [9] proposed multi-tasking network utilizing deep learning that predicts magnification factor and malignancy (benign/malignant) simultaneously. Table 1 provides a summary of frameworks which have been proposed in the last five years; utilizing the Break-His dataset for breast cancer image analysis.

The main contributions are listed as follows: (1) A framework which uses the multi-layered deep features in a partially-independent manner for classification of breast cancer histopathology images. (2) Information-theoretic layer selection. (3) Various experiments are performed and related comparisons are provided to gauge importance of such a proposed study.

3. Proposed approach

Here we will discuss the architecture chosen for the proposed study: the learning of features involving extraction of features and their dimensionality reduction, layer selection based on information theortic based measure (ITS), designed multi-layered framework, and classification framework.

3.1. ResNet architecture

In neural network, network depth is of crucial importance, but deeper networks are more difficult to train. The residual learning framework eases the training of these networks, and, hence enabling them to be substantially deeper, leading to improved performance in both visual and non-visual tasks. These residual networks are much deeper than their 'plain' counterparts, yet they require a similar number of parameters. The main power of deep residual networks is in residual blocks.

Residual blocks: The layers are copied from the learned shallower model, and the added layers are identity mapping. The existence of this constructed solution indicates that a deeper model should produce no higher training error than its shallower counterpart. Fig.2 shows the residual block (right) which makes it different from other deep networks. Fig.3 shows the overall architecture of ResNet and plain net.

A ResNet is composed of multiple computational blocks, and each block commonly consists of three layers: 1) a convolutional layer; 2) an activation layer; and 3) a pooling layer.

Convolutional layer. The main building block used to construct a CNN architecture is the convolutional layer. A hierarchy of features can be formed by stacking up multiple layers. Each layer learns some weights, also called as kernels, and represents a set of features which it tries to track

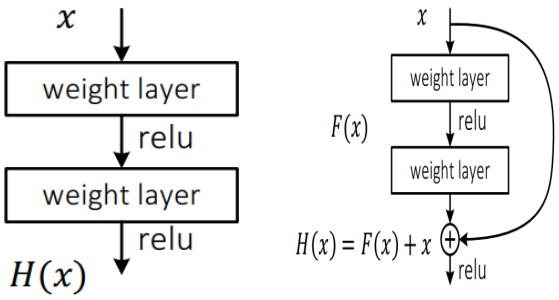

Figure 2. (a)Overall architecture of ResNet [11] and plain net.

down in whole of the input image. Each filter in a CNN is replicated across the entire visual field, which uses the same bias and weight parameters, also called as the shared weights. The weights learned by the initial layers generally represent low-level features, for example, horizontal or vertical edges. As the number of hidden layers in an architecture increases, the network is beginning to learn features of a high level.

Pooling layer Pooling layers are usually used immediately after convolutional layers. A pooling layer takes each feature map output from the convolutional layer and prepares a condensed feature map.

ReLU activation function ReLU is the state of the art non-linearity used in deep neural networks. The rectifier activation function is defined as, $f(x) = max(0, x)$, where x is the input to the neuron. Rectified linear units allow for faster effective training of deep neural architectures as compared to sigmoid or similar activation functions, with further advantages of sparsity and reduced likelihood of the gradient to vanish.

Softmax layer It is a generalization of the logistic function that squashes a K-dimensional vector of arbitrary real values to a K-dimensional vector of real values in the range of 0 to 1 which add upto 1. It highlights the largest values and suppresses the values which are significantly below the maximum value.

Dropout Dropout is a regularization technique in which filters are randomly turned off during training [24]. Hence, the convolutional network learns to not depend too heavily only on one filter and reduces the chances of over-fitting, especially in our case where we have relatively less training data.

Batch Normalization Batch normalization is a solution to the covariate shift problem that works by using per batch statistics to convert the output of each convolutional network into one that has zero mean and unit variance [25]. Batch normalization also functions as a regularization method [26].

For the proposed study, we fine tune the resnet by freezing starting 100 layer and learning remaining layers. We also add two fully connected layer at last. This fine-tune ResNet is used as generic feature extractor. We extract features from all convolution layers (total of 143).

3.2. Dimensionality reduction using XGBoost [27]

The feature vector at the output of deep network is often not utilized directly with other frameworks due to their high dimensionality. Hence, some methods [4] applied principal component analysis (PCA) for such purpose. In proposed study, XGboost [27] is utilized due to its various advantages. XGBoost is short form of "Extreme Gradient Boosting". It is an additive tree classification model where new tree is added in such a fashion that it compliments the already built Trees. The final output is the weighted sum of all decision given by each individual tree. It is very famous due to its high efficiency and performance. We consider the scores assigned by XGboost to each feature, to reduce the dimension. Features correspond to top scores (<3000) are chosen for next stage of work. Default values of parameters are used (max depth:2, number of trees: 250). The ranking of features (based on scores given by XGboost), is utilized to extract top features.

3.3. Layer selection (Based on ITS measure [12])

knowing that not all layers contribute to decision making, as many of them learn similar representation (small variance). Hence, consideration of layers which learn to negative features will deteriorate the accuracy. This work explores an information theoretic measure [12] for selection of an optimal set of layers. Different from traditional methods which only consider the diversity measure, it includes both accuracy and diversity. The summary of ITS measure is given below. More details can be found in [12].

- Best performing layer is chosen first:

$$\hat{i_1} = \underset{i=1,2,...,M}{\arg\max}\ I(C_i, C) \tag{1}$$

  Where M is total number of layers. C and C_i are true class labels and predicted class labels assigned by layer i respectively. I refers to mutual information.

- Now, the layer for which ITS score [12] between $\hat{i_n}$ and the previously chosen layer is maximized, is added to optimal subset.

$$\hat{i_{n+1}} = \underset{i=1,2,...,M-n}{\arg\max}\ I(C_{\hat{i_n}}, C_i); \quad n = 1.2...M-1 \tag{2}$$

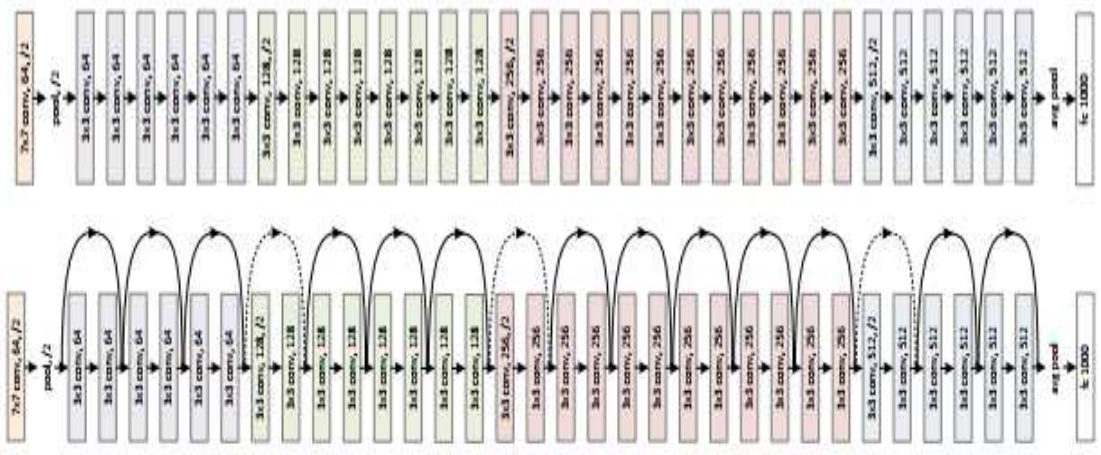

Figure 3. (a) plain Net (b) Residual block (ResNet)

- We will iterate step 2 until the improvements are small enough by adding member. The set of selected layers is considered to be an optimal subset.

3.4. Multi-layered deep framework: Partially Independent

Figure 4 shows the proposed multi-layered model where some integration among multi-layered features is considered (for illustration purpose, we have only shown 20 layers in figure). In figure, each color shows the feature (top 500) of different layers. At each level features (top 500) from all previous layers are concatenated to features of current layer. The top ranked features (based on ranking given by XGboost) are used in the concatenation. These new feature vector, is used to learn the classification framework, and to make decision. The motivation behind considering such a framework is to see how performance is being effected when feature integration considered. As we will note in results section, a comparable performance can be achieved with the independent system with features from much fewer layers . We believe that the effectiveness of incorporating such feature integration stems from the possibility that the dimensionality reduction that is applied earlier on the features, may compromise some dependency inherent in CNN. This might be improved by such a feature concatenation.

3.5. Baseline ResNet classification framework

We also calculate the baseline accuracy which is produced by pre-trained ResNet. In baseline network, features of last fully connected layers is used on the trained neural network. We calculable it for the comparison with the multi-layered framework which comprises the low-mid-high level features.

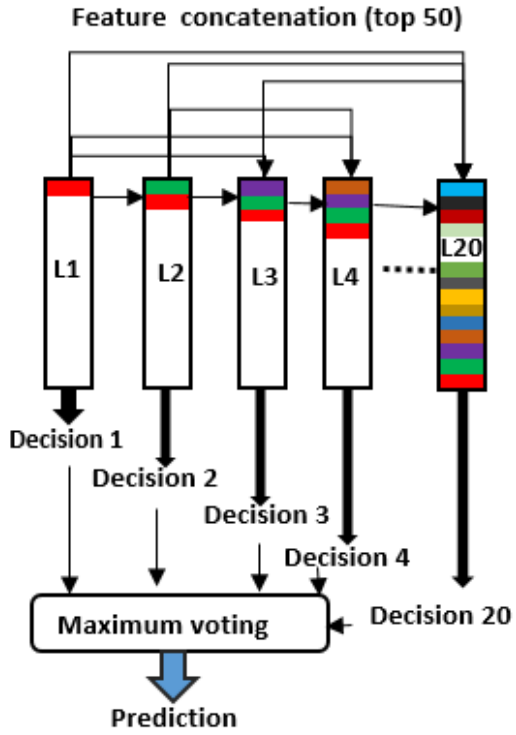

Figure 4. Partially independent framework

3.6. Classification framework

Due to ease, for classification with the above features, we use the popular quadratic SVM classifier with the order two polynomial kernel and consider the output probability scores for further processing. The definition of SVM is given below:

Support Vector Machine (SVM) [28]: It learns a hyperplane that separates a set of positive examples from a set

of negative examples with maximum margin. The hyperplanes can be learnt in higher dimensional space using kernels. Based on the kernel and their parameters, a variety of SVMs can be defined.

4. Results & discussion

In this section we discuss the dataset, training & testing protocol, evaluation metric, designed experimentation and their discussion along with various comparisons.

4.1. Dataset description

To evaluate the effectiveness of the proposed study, we utilize break-His [7] dataset. This dataset contains total 7909 images of begin and malignant tumors, taken from random 82 patients. The images are captured at four different magnification to capture more comprehensive view of disease. Figure 1 shows the samples images of benign and malignant tumor captured at various magnifications. Table 2 illustrates the distribution of data.

Table 2. Detailed description of BreaKHis dataset [7].

	Magnifications				**Total**	**Patient**
	40x	**100x**	**200x**	**400x**		
Benign	625	644	623	588	2480	24
Malignant	1370	1437	1390	1232	5429	58
Total	**1995**	**2081**	**2013**	**1820**	**7909**	**82**

4.2. Training & Testing Protocol

For the fair comparison with exiting approaches, in our experimentation, we use images of 58 (70%) patients chosen randomly for training/validation and remaining 25 for testing (30%). We report the accuracy averaged over three trails, where each trial provides random training-testing patients. We also do the data argumentation to remove the scarcity of labeled data. We adopt rotation, flip, height shift, sift width and translation to increase. We have six times the original training data after the augmentation.

4.3. Evaluation metrics

Patient recognition rate (PRR) and Image recognition rate (IRR) are the two metrics which mainly utilized by the existing methods to report the performance. PRR is a ratio of total correctly classified images (N_{rec}) to total images of cancer images (N_P). The details of patient recognition rate is given as follows:

$$PRR = \frac{\sum_{i=1}^{N} PS_i}{N} \qquad PS = \frac{N_{rec}}{N_P}; \tag{3}$$

In the equation, N is the available testing patients, and PS is a patient score. *IRR* is simply a ratio of total classified images to total number of images (N_{all}).

$$IRR = \frac{N_{re}}{N_{all}} \tag{4}$$

4.4. Partially independent framework

Table 3 & 4 represents the results for multi-layered partially independent framework. In Table 3, results are reported for the layers which are chosen randomly. Various experiments are performed to see the effect on accuracy with chosen number of layers. In table , TF denotes total features that each previous layer contributes to current layer while considering dependency, and TL denotes the total number of layers considered in the model. In all the experiments, the value of these parameters are chosen experimentally. To see the effect of number of layers, we have experimented with three variations TL=10, 20, 30, and TF=50.

Table 3. Performance of partially independent framework with randomly selected layers.

Mag.	**Magnification-specific (%)**		
	TF50, TL:10	**TF:50, TL:20**	**TF:50, TL:30**
40x	87.88	91.99	93.99
100x	90.79	92.93	95.00
200x	88.90	90.79	91.75
400x	82.72	86.60	87.70

Table 4. Performance comparison with variants of designed framework.

Mag.	**Magnification-specific (%)**		
	Partially independent	**Independent**	**Baseline**
40x	97.0	96.81	88.37
100x	96.10	95.26	90.29
200x	94.69	93.78	90.54
400x	90.85	90.76	86.11

We repeat the same experiment with the more number of layers (depends on ITS measure: 70, 45, 66, 40) which are chosen based on ITS measure. And to see the effect with more feature fraction, we consider value of TF is equal to 500. The results corresponding to this experiment is shown in Table 4. In table, first column shows the results of proposed framework, and remaining two show the results corresponding to independent modeling of selected layers, and baseline accuracy. We observed that as the performance improves with inclusion of features from more layers, and the performance while considering feature integration, is in most cases, higher than when features are considered independently. Improving accuracy over baseline signifies the role of low-mid-level features in multi- layered framework together with high-level features. Finally, as indicated earlier, when considering independent multi-layered features, the proposed framework also outperforms the case.

Table 5. Performance Comparison of magnification specific system (in %). For the proposed method, the numbers in bracket provide its rank based on the performance among all approaches.

	Methods	Magnifications (values in percentage (%))			
		40x	100x	200x	400x
Existing works	Spanhol et al. [7]	83.8±4.1	82.1±4.9	85.1±3.1	82.3±3.8
	Spanhol et al. [8]	90.0±6.7	88.4±4.8	84.6±4.2	86.10±6.2
	Bayramoglu et al. [9]	83.08±2.08	83.17±3.51	84.63±2.72	82.10±4.42
	Gupta et al. [16]	86.74±2.37	88.56±2.73	90.31±3.76	88.31±3.01
	Song et al. [18]	90.02±3.2	88.9±5.0	86.9±5.2	86.3±7.0
	Song et al. [23]	90.02±3.2	91.2±4.4	87.8±5.3	87.4±7.2
	Han et al. [19]	97.1±1.5	95.7±2.8	96.5±2.1	95.7±2.2
Proposed	Baseline-accuracy (ResNet)	88.37	90.29	90.54	86.11
	Independent framework	96.81	95.26	93.78	90.76
	Partially independent model	97.0±1.18	96.10±1.0	94.69±1.19	90.85±2.12

4.5. Performance comparison with state-of-art

In Table 6, we compare the performance of magnification specific study with state-of art methods, and depicts the proposed method outperforms all contemporary methods except [19], in some cases. With regard to comparison with [19], we note that this work [19] was designed in a learning based and data-driven manner. They employed class structure-based deep convolution neural network (CSDCNN) embedded with an additional distance constraint for multi-classification. Their work, non-trivally extends upon a baseline existing deep learning network. Our purpose work is mainly to understand the role of deep multi-layered features from an existing architecture. The proposed framework outperforms the various state-of- the-art approaches even when fine-tuning an existing model, used as it is, rather than formulating a new deep architecture. In comparison to [19], proposed model is fairly simple as it does not add much to existing network. Thus, expectedly, the multi-layered partial independent modeling yields better results than the others. We don't compare our work with [29, 30] as some details about training-testing protocol (imagewise or patinetwise) are not clear. In addition to that, only image level accuracy has been calculated (IRR from eq. 4) which ignores the patient information.

5. Conclusion

In this paper we focus on better exploring the potential of fine-tuned pre-trained CNN models for classifying the images of breast cancer histopathology. In this regard, we present a model which combine the decision of layers selected through ITS, in partially independent framework. We conclude that, by integrating dependency in terms of feature fraction among selected layers, will perform better than all layer model. Hence, the utility of low-mid-high level features have shown through such improved performance over baseline. The proposed approach is also shown to outperform most state-of-the-art classification methods.

References

[1] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009. 1

[2] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42, 2018. 1, 2

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 1

[4] Erzhu Li, Junshi Xia, Peijun Du, Cong Lin, and Alim Samat. Integrating multilayer features of convolutional neural networks for remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2017. 2, 4

[5] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016. 2

[6] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016. 2

[7] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, 2016. 2, 3, 6, 7

[8] Fabio Alexandre Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *Neural Networks (IJCNN), 2016 International Joint Conference on*, pages 2560–2567. IEEE, 2016. 2, 3, 7

[9] Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä. Deep learning for magnification independent breast cancer histopathology image classification. 2, 3, 7

[10] Vibha Gupta, Apurva Singh, Kartikeya Sharma, and Arnav Bhavsar. Automated classification for breast cancer histopathology images: Is stain normalization important? In *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures*, pages 160–169. Springer, 2017. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 4

[12] Julien Meynet and Jean-Philippe Thiran. Information theoretic combination of pattern classifiers. *Pattern Recognition*, 43(10):3412–3421, 2010. 2, 4

[13] Yungang Zhang, Bailing Zhang, Frans Coenen, and Wenjin Lu. Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles. *Machine vision and applications*, 24(7):1405–1420, 2013. 2

[14] Yungang Zhang, Bailing Zhang, Frans Coenen, Jimin Xiao, and Wenjin Lu. One-class kernel subspace ensemble for medical image classification. *EURASIP Journal on Advances in Signal Processing*, 2014(1):17, 2014. 2

[15] Nina Linder, Juho Konsti, Riku Turkki, Esa Rahtu, Mikael Lundin, Stig Nordling, Caj Haglund, Timo Ahonen, Matti Pietikäinen, and Johan Lundin. Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagnostic pathology*, 7(1):22, 2012. 2

[16] Vibha Gupta and Arnav Bhavsar. Breast cancer histopathological image classification: Is magnification important? In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 769–776. IEEE, 2017. 3, 7

[17] Kosmas Dimitropoulos, Panagiotis Barmpoutis, Christina Zioga, Athanasios Kamas, Kalliopi Patsiaoura, and Nikos Grammalidis. Grading of invasive breast carcinoma through grassmannian vlad encoding. *PloS one*, 12(9):e0185110, 2017. 3

[18] Yang Song, Ju Jia Zou, Hang Chang, and Weidong Cai. Adapting fisher vectors for histopathology image classification. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 600–603. IEEE, 2017. 3, 7

[19] Zhongyi Han, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li, and Shuo Li. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*, 7, 2017. 3, 7

[20] Dalal Bardou, Kun Zhang, and Sayed Mohammad Ahmad. Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access*, 6:24680–24693, 2018. 3

[21] Abdullah-Al Nahid and Yinan Kong. Histopathological breast-image classification using local and frequency domains by convolutional neural network. *Information*, 9(1):19, 2018. 3

[22] Md Zahangir Alom, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *arXiv preprint arXiv:1811.04241*, 2018.

[23] Yang Song, Hang Chang, Heng Huang, and Weidong Cai. Supervised intra-embedding of fisher vectors for histopathology image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 99–106. Springer, 2017. 3, 7

[24] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1):1929–1958, 2014. 4

[25] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015. 4

[26] David A Van Valen, Takamasa Kudo, Keara M Lane, Derek N Macklin, Nicolas T Quach, Mialy M DeFelice, Inbal Maayan, Yu Tanouchi, Euan A Ashley, and Markus W Covert. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology*, 12(11):e1005177, 2016. 4

[27] Tianqi Chen and Carlos Guestrin. Xgboost: reliable large-scale tree boosting system. arxiv. *2016a. ISSN*, pages 0146–4833, 2016. 4

[28] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011. 5

[29] Abdullah-Al Nahid and Yinan Kong. Histopathological breast-image classification using local and frequency domains by convolutional neural network. *Information*, 9(1):19, 2018. 7

[30] Dalal Bardou, Kun Zhang, and Sayed Mohammad Ahmad. Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access*, 6:24680–24693, 2018. 7