Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks

DALAL BARDOU[ID][1]**, KUN ZHANG**[1]**, AND SAYED MOHAMMAD AHMAD**[2]

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210000, China
[2]Lareb Technologies, New Delhi 11002, India

Corresponding author: Kun Zhang (zhangkun@njust.edu.cn)

ABSTRACT In recent years, the classification of breast cancer has been the topic of interest in the field of Healthcare informatics, because it is the second main cause of cancer-related deaths in women. Breast cancer can be identified using a biopsy where tissue is removed and studied under microscope. The diagnosis is based on the qualification of the histopathologist, who will look for abnormal cells. However, if the histopathologist is not well-trained, this may lead to wrong diagnosis. With the recent advances in image processing and machine learning, there is an interest in attempting to develop a reliable pattern recognition based systems to improve the quality of diagnosis. In this paper, we compare two machine learning approaches for the automatic classification of breast cancer histology images into benign and malignant and into benign and malignant sub-classes. The first approach is based on the extraction of a set of handcrafted features encoded by two coding models (bag of words and locality constrained linear coding) and trained by support vector machines, while the second approach is based on the design of convolutional neural networks. We have also experimentally tested dataset augmentation techniques to enhance the accuracy of the convolutional neural network as well as "handcrafted features + convolutional neural network" and "convolutional neural network features + classifier" configurations. The results show convolutional neural networks outperformed the handcrafted feature based classifier, where we achieved accuracy between 96.15% and 98.33% for the binary classification and 83.31% and 88.23% for the multi-class classification.

INDEX TERMS Histology images, convolutional neural networks, engineered features, bag of words, locality constrained linear coding.

I. INTRODUCTION

Breast cancer is the most common invasive cancer in women and the second main cause of cancer death in women, after lung cancer [1]. According to the International Agency for Research on Cancer (IARC), which is part of the World Health Organization (WHO) [2], the numbers of deaths caused by cancer in the year of 2012 alone come to around 8.2 million. The number of new cases is expected to increase to more than 27 million by 2030 [3].

Breast cancer can be diagnosed using medical images testing, like histology and radiology images [4]–[6]. The radiology images analysis can help to identify the areas where the abnormality is located. However, they cannot be used to determine whether the area is cancerous [5], [6]. The biopsy, where a tissue is taken and studied under a microscope to see if cancer is present, is the only sure way to identify if an area is cancerous [5]–[7]. After completing the biopsy, the diagnosis will be based on the qualification of the histopathologists, who will examine the tissue under a microscope, looking for abnormal or cancerous cells [5], [6]. The histology images allow us to distinguish the cell nuclei types and their architecture according to a specific pattern. Practically, histopathologists visually examine the regularities of cell shapes and tissue distributions and determine cancerous regions and malignancy degree [5], [6]. If the histopathologists are not well-trained, this may lead to an incorrect diagnosis. Also, there is a lack of specialists, which keeps the tissue sample on hold for up to two months, for example, this occurs often in Norway [8]. There is also the problem of reproducibility, as histopathology is a subjective science. This is true

especially between non-specialized pathologists, where we can receive a different diagnosis on the same sample. Therefore, there is an insistent demand for computer-assisted diagnosis.

With the onset of pattern recognition and machine learning, many handcrafted (engineered) features-based studies are proposed for classifying breast cancer histology images. Some studies have focused on nuclei segmentation such as in [9]–[11]. After selecting the region of interest, a set of features are extracted and fed into traditional classifiers to classify the breast histology images into either benign or malignant. These studies were conducted on small data sets comprising 500 and 92 images. Other studies have focused on extracting global or local features from the whole images such as in [12], where Zhang *et al.* combined local binary pattern, statistics from gray level co-occurrence matrix and curvelet transform, using a cascade random subspace ensembles scheme with reject options. In the first level of the cascade, the authors tried to solve the easy cases, while in the upcoming levels, the hard cases are sent to a more complex pattern classification system. The samples were classified into normal tissue, carcinoma in situ, and invasive carcinoma. In another study [13], the same authors evaluated a one-class kernel principal component analysis (KPCA) model ensemble using the same data (361 images) in [12]. The ensemble consists of one-class KPCA models trained with different features, a product-combining rule was then used to combine the KPCA models to get the scores. In [14], Spanhol *et al.* constructed a database of histology breast cancer images named "BreaKHis" to overcome the problem of small data sets that are the main obstacle leading to the lack of development of new analysis method as pointed by Veta *et al.* [15]. In addition to the database, the authors presented the performances of the baseline pattern recognition system, which was designed to distinguish between benign and malignant tumors.

The convolutional neural networks (CNN) are regarded as a variant of the standard neural networks. This variant introduces a new special network, which compromises so-called convolution and pooling layers [16], instead of fully connected hidden layers. They were first introduced for overcoming known problems of fully connected deep neural networks when handling high dimensionality structured inputs, such as images or speech [17]. CNNs have become state-of-the-art solutions for large-scale object classification [18], [19] and object detection tasks [19]–[21]. CNNs have been applied to address the task of breast cancer histology images classification such as in [22], where the authors divided the histology images into small batches and then used to train CNN. To get the final classification result, the patches results are combined for the whole image. In [23], Araújo *et al.* also used CNN to classify breast histology images into four classes - normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma. The authors also extracted a set of features from the CNN and fed them into support vector machines. In [24], Spanhol *et al.* used deep features extracted from a CNN.

The approach is based on the reuse of previously trained convolutional neural networks only, as a feature extractor. The extracted deep feature vectors are then used as an input to a traditional classifier. In [25], Bayramoglu *et al.* also used convolutional neural networks to classify breast histology images into benign and malignant. However, the approach was now magnification independent. The authors proposed two architectures, one named the "single task convolutional neural network" which was used to predict the malignancy and a "multi-task convolutional neural network" which was used to predict both the malignancy and the magnification factor. In [26], a structured deep learning model is proposed for addressing the breast cancer subordinate classes while in [27], local and frequency domains with convolutional neural networks are used.

With the recent advances in image processing and machine learning, the classification and diagnosis of breast cancer using histology images attract much interest, and developing pattern recognition-based diagnostic systems became a necessity to help experts optimize diagnostic quality. In this paper, we have proposed two approaches to classifying breast cancer histology images into benign and malignant, as well as their sub-classes. The first approach is based on the extraction of a set of handcrafted features encoded by two coding models (bag of words and locality constrained linear coding) and trained by support vector machines, while the second approach is based on the design of convolutional neural networks. The breast images consist of adenosis, fibroadenoma, phyllodes tumor, tabular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. The objectives of our study are: (a) to compare the handcrafted features-based classification methods and the convolutional neural networks, (b) to compare the performance of the same convolutional neural networks on binary and multi-class classification tasks., (c) to see how experimental dataset augmentations influence performances, (d), to compare between bag of words and locality constrained linear coding models, which are used to encode local descriptors.

The rest of the paper is organized as follows: in Section II, we give information about the data, while Section III is dedicated to the proposed convolutional neural network and the handcrafted features-based classification. In Section IV and Section V, the implementations settings, performances and experimental results comparisons are given and discussed.

II. MATERIALS

The data set used in this work is BreaKHis [14], which comprises microscopic images of biopsy for benign and malignant breast tumors with a total number of 7909 images. Each slide of breast tumors is stained with hematoxylin and eosin (HE). To obtain the digitized images from the slides, an Olympus BX-50 system microscope with a relay lens with magnification of $3.3\times$ coupled to a Samsung digital color camera SCC-131AN is used. The data set comprises eight types of benign and malignant tumors. The four benign tumors types are: Adenosis, fibro adenoma,

TABLE 1. The number of histology images samples per class.

CLASS	SUB-CLASS	Magnification factor			
		40X	100X	200X	400X
BENIGN	ADENOSIS	114	113	111	106
	FIBROADENOMA	193	260	264	137
	PHYLLODES TUMORS	149	150	140	130
	TABULAR ADENOMA	109	121	108	115
MALIGNANT	DUCTAL CARCINOMA	864	903	896	788
	LOBULAR CARCINOMA	156	170	163	137
	MUCINOUS CARCINOMA	205	222	196	169
	PAPILLARY CARCINOMA	145	142	135	138

phyllodes tumors, and tabular adenoma. The four malignant tumors types are: Ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. The images are acquired using four magnification factors: 40X, 100X, 200X, and 400X. The number of samples for each magnification factor is provided in TABLE 1. FIGURE 1 shows a sample image from each type of breast benign and malignant breast seen in magnification factor 40X.
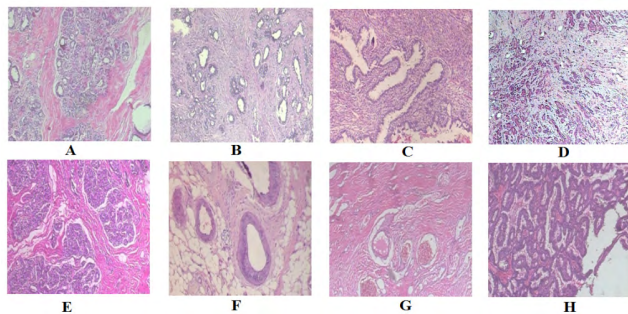
FIGURE 1. Samples from each type of breast tumors, (A): Adenosis, (B), Fibroadenoma, (C): phyllodes tumor, (D): Tabular adenoma, (E): ductal carcinoma, (F): lobular carcinoma, (G): mucinous carcinoma, (H): papillary carcinoma.

III. METHODOLOGY

The proposed methods comprising convolutional networks and the handcrafted features-based approach are presented in this section. The proposed CNN topology is given in Section A. A schematic representation of the handcrafted features approach is depicted in FIGURE 2. FIGURE 2 provides an overview of the proposed approach used for generating a feature representation from an image. A set of local features were extracted and a feature coding method was employed in order to aggregate the local features into an image representation. Spatial pyramid matching was used to capture spatial structure of the images. Support vector machines were then used to classify images. The dataset has been randomly split into 70% training and 30% testing for

each magnification factor. Twenty-five percent of the training data is retained and used for cross validation to select the best model's parameters. After selecting the best parameters, it has been added to the training data. The same testing dataset is used in all experiments. The dataset augmentation techniques are only applied to the CNN approach, where each image is flipped and rotated.

A. CNN TOPOLOGY

Convolutional neural network became a substantial trend in machine learning, and it had much success in many fields such as computer vision and speech recognition. The proposed CNN topology has been implemented using BVLC Caffe [28], which is a popular and robust framework of UC Berkeley. Caffe makes easy the design, the implementations of the CNNs, and comes up with simple interfaces for both MATLAB and Python. Caffe provides an expressive architecture that supports CPU and GPU training. The CNN model and the optimization parameters are specified in the configuration. Our proposed CNN topology is given below and it comprises five convolutional layers followed by two fully connected layers.

- 1st convolutional layer with filter size 3×3 and 64 feature maps.
- 2nd convolutional layer with filter size 3×3 and 96 feature maps.
- 3rd convolutional layer with filter size 3×3 and 128 feature maps.
- 4th convolutional layer with filter size 3×3 and 256 feature maps.
- 5th convolutional layer with filter size 3×3 and 256 feature maps.
- Fully connected layer with 2000 hidden units.
- Fully connected layer with number of hidden units equal to the number of classes.
- Softmax layer.

We have applied the RELU layer to all convolutional and fully connected layers to make fast the convergence learning and to introduce the non-linearity. The RELU layer changes all negative activation values to zero of the given input by applying the following function: $f(x) = \max(0, x)$. We have also used max pooling for the first two convolutional layers and the fifth layers and it has not applied for the third and the fourth layers. The max pooling layer is applied after RELU layer to reduce the spatial dimension with a filter of 3×3 and a stride of the length equal to two. The network's weights are initiated using the Gaussian distribution with low standard deviation equal to 0.01 for all the layers. Dropout layer was applied after the fully connected layer with keep probability of p = 0.5 which improved the performance of the network. We have also used weight decay as another common regularization technique through performing L2 regularization with a value of $\lambda = 10^{-3}$ to prevent the model from overfitting of the training data. We have also set the learning rate value to = 0.001 and it is decayed during training by using inverse decay policy after every five epochs. This parameter scales
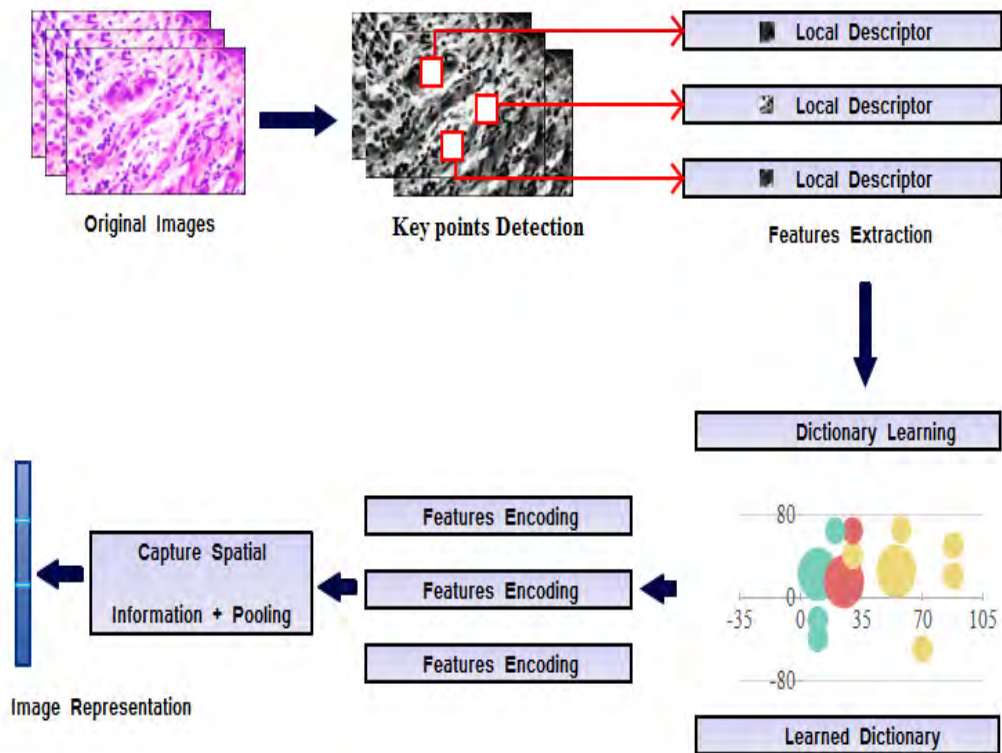
FIGURE 2. A schematic illustration of the proposed features-based approach's feature representation.

the size of the updates or the steps a neural network takes. The loss function was optimized using stochastic gradient descent with batch size equal to 32. The dataset was randomly shuffled to avoid any negative impact on the learning by using ordered training data. We have also set the value of the momentum factor to 0.9 which helps the loss function to move to the global minimum and do not stuck in one of the local minima.

The dataset augmentation techniques have been applied to increase the cardinality of the training set for all classes and also to overcome the problem of overfitting. We have used two augmentation techniques including rotation and horizontal flip.

- Rotation: Rotation by 90°, 180°, and 270° has been applied to the images, augmenting the size of the training dataset.
- Horizontal flip: Horizontal flip has been applied to the whole training images, also doubling the cardinality of the training dataset.

B. HANDCRAFTED FEATURES-BASED CLASSIFICATION

We have made a comparison between CNN and features-based classification. The feature-based approaches consist of the features extraction phase, features coding phase, and the classification phase. Two features-based approaches are proposed; the first approach focuses on extracting local descriptors and encoding them with the bag of words model, while

in the second approach, the local descriptors are encoded using locality constrained linear coding. Also, we have used spatial pyramid matching to capture the spatial structure of the images. The extracted features are trained using support vector machines. Finally, we compared the performance of the used feature coding models with the local descriptors.

1) LOCAL FEATURES EXTRACTION
We have used two types of local descriptors named Dense Scale Invariant Feature Transform (DSIFT) features and Speeded-Up Robust Features (SURF).

a: DENSE SIFT
Dense SIFT [29]–[31] is a variant of SIFT and, as opposed to SIFT, dense SIFT does not identify nor compute interest points. In other words, dense SIFT does not consist of both detection and description; it only uses the descriptor on the regular dense grid (uniform way). DSIFT uses a histogram of the oriented gradient method. Thus, it produces a 128-dimensional feature vector. DSIFT outputs a very large set of local descriptors, which provide more information when compared to SIFT.

b: SURF
SURF [29]–[31] is a fast and robust version of SIFT; SURF computation also consists of three main steps, like the SIFT descriptor. The steps are: (a): create a scale-space

representation of a given input image, (b): detect SURF interest points' in the scale and assignment of the orientation to keypoints, and (c) finally, construct the SURF descriptor for each keypoint. The main difference between SIFT and SURF in the detection phase is that SIFT uses the Gaussian derivatives while SURF uses a simple 2D box filters. It uses a scale-invariant blob detector based on the determinant of the Hessian matrix for both scale selection and locations [34]. After computing the interest points, the next step is to calculate their descriptions. A square region centered on the interesting point is constructed and then the sum of the Haar wavelet response is applied around the point of interest. The extracted descriptor is a 64-dimensional feature vector.

2) FEATURES CODING

Features coding is an important step in the classification process when working with local features (SURF and DSIFT descriptors …etc.). Most classifiers only accept a fixed-length vector as the input for image recognition. Thus, providing the descriptors as inputs is not possible because there are a variable number of feature vectors per image [35]. We have used two types of feature coding, named ''bag of words'' and ''locality constrained linear coding'' which are used to aggregate the local features into an image representation.

a: BAG OF WORDS

Bag of Words (BoW) is widely applied as a feature coding method for medical [36] and natural image classification [37], [38]. To construct the bag of words model, we first need to build a vocabulary of visual words (a codebook) from the extracted local descriptors by applying clustering. The sets of descriptors extracted from the training set are clustered into K-clusters using K-means. Afterward, the centroids of these clusters will be considered as visual words and will be used to encode the local descriptors. Each descriptor activates a visual word and generates a code vector. The length of the code vector is equal to the number of clusters. After computing the code vector of each descriptor, the output of this step is a coding multi-dimensional vector for the whole image. The final step is final image representation, also called pooling, where a frequency histogram of visual words is built of length K for the entire image.

b: LOCALITY CONSTRAINED LINEAR CODING

Locality constrained linear coding is an efficient variant of sparse coding, which utilizes the local linear property of manifolds to project each descriptor into its local coordinate system, and the projected coordinates are integrated by max pooling to generate the final representation [39]. It has shown improved performance over Bag of words model for image classification [40], [41]. Also, it is very a representative model with fast coding speed. Let $X \in R^{D \times N}$ is a set of D-dimensional local descriptors in an image, i.e. $X = [x_1, x_2 \ldots x_N] \in R^{D \times N}$. Given a visual codebook with M entries $B = [b_1, b_2 \ldots b_M] \in R^{D \times M}$, each descriptor is converted into M-dimensional code. LLC coding uses the

following criteria to compute the codes $C = [c_1, c_2 \ldots c_N] \in R^{D \times N}$

$$\arg\min_c = \sum_{i=1}^{N} \|X_i - Bc_i\|^2 + \gamma \|d_i \otimes c_i\|^2$$
$$\text{s.t. } 1^T c_{ij} = 1 \tag{1}$$

Where: $\otimes$ denotes the element wise multiplication and

$$d_i = \exp(\frac{dist(x_i, B)}{\sigma}) \tag{2}$$

And

$$dist(x_i, B) = [dist(x_1, b_1), dist(x_2, b_2) \ldots dist(x_i, b_M)] \tag{3}$$

Where the distance $dist(x_i, b_j)$ is the Euclidean distance between x_i and b_j. σ is decay parameter which used to adjust the weight decay speed for the locality adaptor d_{ij}.

A fast approximation to LLC was proposed in [39] to speed up the encoding process by ignoring the second term in equation (1), where a k nearest basis descriptors of x_i are directly used to minimize the first term. The encoding process is simplified by just solving a much smaller linear system [38] given below:

$$\arg\min_c = \sum_{i=1}^{N} \|X_i - Bc_i\|^2$$
$$\text{s.t. } 1^T c_{ij} = 1, \quad \forall i \tag{4}$$

Thus, the coefficients of the selected K basis vectors are given while the others are set to zero. The image representation z_i of an image I_i is then obtained by pooling the sparse codes associated with the local descriptor.

Preserving the spatial relationship between the code vectors may enhance the performance of the classification. Thus, we decided to apply spatial pyramid matching [41], [42], which is an efficient extension of pyramid matching [43]. The main idea is to divide the image into levels. Each level divides the images into sub-regions, building a histogram of codes in each block. Level 0 will consist of a single histogram, level 2 will comprise of level 0, comprising a single histogram, level 1, comprising of 4 histograms [44]…etc. The constructed histograms are finally pooled to achieve the final presentation of the image.

IV. RESULTS

The results in Section A are related to the CNN classification performance applied to the original dataset and the dataset with augmentation. The results of each test are given in terms of accuracy (% of correctly classified instances). The learning rate curves for testing and training losses are given too. We have also tested 'CNN+SVM' configuration by replacing the fully-connected layer with linear support vector machines, 'CNN Features+Classifiers' configuration through extracting CNN features at the fully connected layer with 2,000 outputs and classifying them by the K-nearest neighbor (KNN), radial basis support vector machines

(RBF SVM), linear SVM and random forests, and we also tested 'Handcrafted Features+CNN' configuration. In Section C, we gave the performances of the handcrafted features-based while in Section D, we provided the results of "Handcrafted Features+CNN" configurations for both BOW and LLC models. To evaluate the proposed models, we used accuracy, precision, recall, F1-score and confusion matrix as evaluation metrics.

$$PRECISION = \frac{TP}{TP + FP}, \quad RECALL = \frac{TP}{TP + FN}$$

TP: True positive, FP: false positive, FN: false negative. F1-score is a measure of test accuracy, and it uses both precision and recall to compute the scores. F1-score is calculated using the following formula:

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall}$$

A. CNN RESULTS

1) BINARY CLASSIFICATION

a: BASELINE CONFIGURATIONS WITH DATASET AUGMENTATION

Following the proposed CNN topology in Section III.A, the images were trained for 20,000 iterations. The obtained accuracies are 94.65%, 94.07%, 94.54%, and 93.77% for images with magnification factors 40X, 100X, 200X, and 400X, respectively. After augmenting the training dataset by applying flipping and rotation and by training the model for 20,000 iterations, we reached an accuracy of 96.82%, 96.96%, 96.36% and 95.97. Applying image augmentation techniques has improved the accuracy for all images with different magnification factors. The data set was trained on Tesla K40m GPU. The training log and test accuracy curves regarding the number of iterations are shown in FIGURE3 (A), FIGURE3 (B) for the original data set and the augmented one respectively applied to the images with 40X magnification factor. The confusion matrixes related to the augmented dataset are given in FIGURE4. All accuracy results are given in TABLE 2. The accuracy performances per class are given in TABLE 3. In the FIGURE5, we gave the visualization of the convolution layers' activations for an example image.

TABLE 2. The accuracy performance of CNN configuration applied to the original and augmented, and "CNN+ Classifier" configuration as well as ensemble model.

MAGNIFICATION	40X	100X	200X	400X
ORIGIN	94.65	94.07	94.54	93.77
AUGMENTED	96.82	96.96	96.36	95.97
SVM	92.71	93.75	92.72	92.12
ENSEMBLE MODEL	98.33	97.12	97.85	96.15

As an assessment way of the deep learned features, we have visualized the first layer weights. The visualization of the weight is an efficient way that indicates if the network is well trained or not. Well trained networks display smooth and satisfactory filters without noisy patterns. Having noisy filters
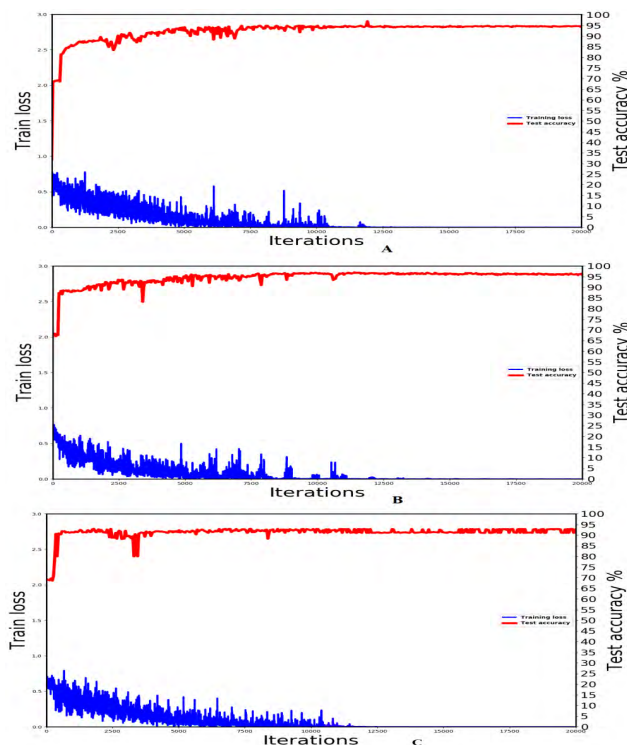

FIGURE 3. CNN results of images with 40X magnification factor: Log loss and test accuracy of the tested dataset: (A). Accuracy results of the original baseline dataset, (B). Accuracy results of the augmented dataset, (C). Accuracy results of the "CNN+ Classifier" configuration.

TABLE 3. The accuracy performance per class for the four magnification factors applied to the augmented data set.

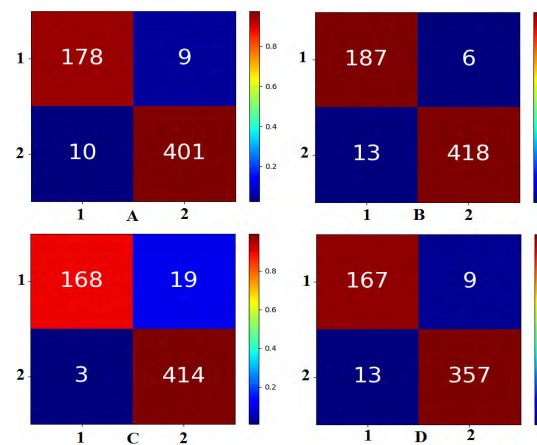| ACCURACY PER CLASS	40X	100X	200X	400X
BENIGN	95.19	96.89	89.84	94.89
MALIGNANT	97.57	96.98	99.28	96.49

FIGURE 4. Confusion matrix: (A). Confusion matrix of CNN models applied to the augmented dataset with 40X magnification factor, (B). Confusion matrix of 100X magnification factor, (C). Confusion matrix of 200X magnification factor, (D). Confusion matrix of 400X magnification factor. 1: Benign, 2: malignant.

may indicate that the network has not been trained for long enough. FIGURE6 shows a visualization of the weights of the first layer (with 64 filters) which look smooth. The weights
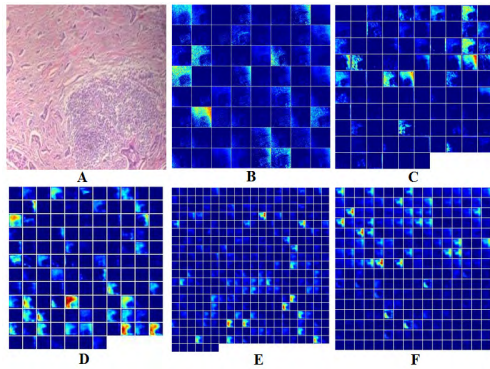
FIGURE 5. Visualization of the convolutional layers' activations, (A): the input images, (B): The first convolutional layer's activation, (C): The second convolutional layer's activation, (D): The third convolutional layer's activation, (E): The fourth convolutional layer's activation, (F): The last convolutional layer's activation.
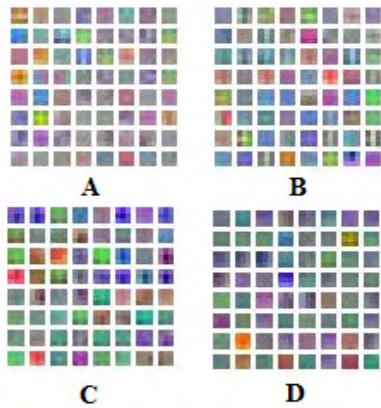

FIGURE 6. Illustration of the learned filters on the first layer of the proposed CNN, (A): The learned filters for images with 40X magnification factor, (B): The learned filters for images with 100X magnification factor, (C): The learned filters for images with 200X magnification factor, (D): The learned filters for images with 400X magnification factor.

were visualized for four examples from each magnification factor.

To enhance the accuracy, we thought of using an ensemble model. CNNs have a final fully connected layer with softmax activation, which provides a vector of probabilities for each sample. At different iterations, the accuracy may be different and also the probabilities for a given sample. Therefore, we have used 10 predictive models that provide the highest accuracies using the augmented data at different iterations to get the probabilities for the testing data. Then, we have summed them (the 10 vectors) up. Finally, the max value among them is taken and its class is given as output. This simple ensembling method was efficient and it helped to increase the performance for the four magnification factor to 98.33%, 97.12%, 97.85% and 96.15 respectively. The results are given in TABLE 2.

TABLE 4 provides the evaluation metrics including precision, recall and F1-score values computed from the CNN models applied to the augmented data. TABLE 5 provides the

TABLE 4. Precision, recall and F1-score values computed from the CNN models applied to the augmented data set.

MAGNIFICATION	40X	100X	200X	400X
PRECISION	97.80	98.58	95.61	97.54
RECALL	97.57	96.98	99.28	96.49
F1-SCORE	97.68	97.77	97.41	97.01

TABLE 5. Training time of CNN applied to original and augmented data sets with different magnification factors.

MAGNIFICATION	40X	100X	200X	400X
ORIGIN DATA	46 MINUTES, 17 SECONDS	47 MINUTES, 14 SECONDS	42 MINUTES, 29 SECONDS	46 MINUTES, 47 SECONDS
AUGMENTED DATA	59 MINUTES, 05 SECONDS	1 HOUR, 22 SECONDS	59 MINUTES, 47SECONDS	1 HOUR 7 MINUTES, 30 SECONDS

training time of CNN applied to the original and augmented datasets with different magnification factors.

b: USE OF LINEAR SVM ON THE TOP OF CNN

We investigated a hybrid approach 'CNN+ Classifier'. Following the same used in [21], we put a small fully-connected layer downstream of the last convolutional layer, with neurons equal to two, which is the number of classes, and without using activation layer (non-linearities), feeding a hinge loss layer. As mentioned in [21], learning this architecture is equivalent to work with a linear SVM acting on features learned by the CNN. The accuracies obtained were 92.71%, 93.75%, 92.72%, and 92.12%, respectively, which is less than the accuracy reached by using CNN in Section 1.1. The decreasing result is due to the linearity of support vector machines.

2) BENIGN AND MALIGNANT CANCER SUB-CLASSES CLASSIFICATION

a: BASELINE CONFIGURATIONS WITH DATASET AUGMENTATION

Following the same CNN topology used for the binary classification, the images were trained for 20,000 iterations. The obtained accuracies are 86.34%, 84.00%, 79.83%, and 79.74% for images with magnification factors 40X, 100X, 200X, and 400X, respectively.

After augmenting the training dataset by applying flipping and rotation and by training the model for 20,000 iterations, we reached an accuracy of 83.79%, 84.48%, 80.83%, and 81.03%. Applying image augmentation techniques has improved the accuracy for all images with different magnification factors, except for images with magnification factor 40X, where the accuracy is decreased from 86.34% to 83.79%. The confusion matrixes related to the original dataset are given in FIGURE7. All accuracy results are given in TABLE 6. The accuracy performances per class are
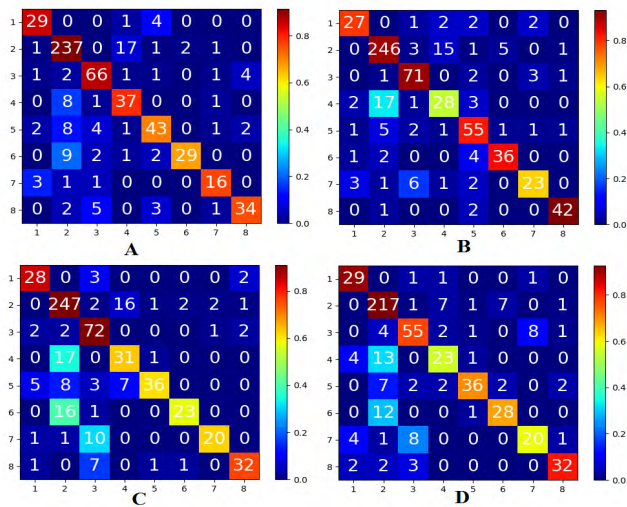
FIGURE 7. Confusion matrixes, (A): Confusion matrix of CNN model applied to the augmented data set with 40X magnification factor, (B): Confusion 100X magnification factor, (C): 200X magnification factor, (D): 400X magnification factor. 1: Adenosis, 2: ductal carcinoma, 3: fibroadenoma, 4: lobular carcinoma, 5: mucinous carcinoma, 6: papillary carcinoma, 7: phyllodes tumor, 8: tabular adenoma.

TABLE 6. The accuracy performance (%) of CNN configuration applied to the original and augmented, and "CNN+ Classifier" configuration as well as ensemble model.

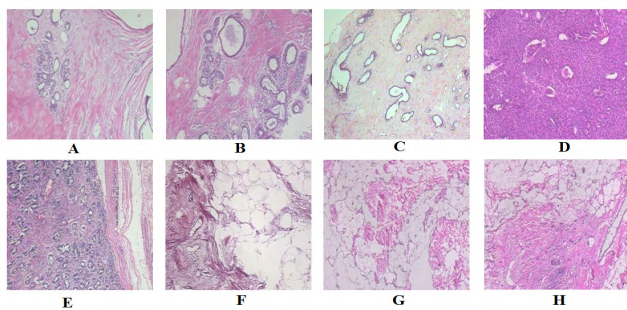| MAGNIFICATION	40X	100X	200X	400X
ORIGIN	86.34	84.00	79.83	79.74
AUGMENTED	83.79	84.48	80.83	81.03
SVM	82.89	80.94	79.44	77.94
ENSEMBLE MODEL	88.23	84.64	83.31	83.98

FIGURE 8. Classification results examples, (A): Well-classified adenosis, (B): Adenosis classified as phyllodes tumor, (C): Tabular adenoma classified as fibroadenoma, (D): Well-classified ductal carcinoma, (E): Well-classified tabular adenoma, (F): Well-classified papillary carcinoma, (G): Well-classified lobular carcinoma, (H): Lobular carcinoma classified as ductal carcinoma.

given in TABLE 7. The data set was also trained on Tesla K40m GPU. TABLE 8 provides the evaluation metrics including precision, recall and F1-score values computed from the CNN models applied to the augmented data. TABLE 9 provides the training time of CNN applied to the original and augmented datasets with different magnification factors. In FIGURE8, we gave classification results for some

TABLE 7. The accuracy performance (%) per class for the four magnification factors applied to the augmented data set.

ACCURACY PER CLASS	40X	100X	200X	400X
ADENO	85.29	79.41	84.85	90.63
DUCTAL	91.51	90.77	91.14	92.74
FIBRO	86.84	91.03	91.14	77.46
LOBULAR	78.72	54.90	63.27	56.10
MUCINOUS	70.49	82.09	61.02	70.59
PAPILLARY	67.44	83.72	57.50	68.29
PHYLLODES	76.19	63.89	62.50	58.82
TABULAR	75.56	93.33	76.19	82.05

TABLE 8. Precision, recall and F1-score values computed from the CNN models applied to the augmented data set.

MAGNIFICATION	40X	100X	200X	400X
PRECISION	84.27	84.29	81.85	80.84
RECALL	83.79	84.48	80.83	81.03
F1-SCORE	83.74	84.31	80.48	80.63

TABLE 9. Training time of CNN applied to original and augmented data sets with different magnification factors.

MAGNIFICATION	40X	100X	200X	400X
ORIGIN DATA	37 MINUTES, 29 SECONDS	39 MINUTES, 54 SECONDS	40 MINUTES, 30 SECONDS	46 MINUTES, 47 SECONDS
AUGMENTED DATA	53 MINUTES, 54 SECONDS	53 MINUTES, 01 SECOND	1 HOUR, 9 MINUTES	1 HOUR 43 MINUTES

TABLE 10. The softmax probabilities output of images given in figure 8.

IMAGE	A	B	C	D	E	F	G	H
ADENO	99.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DUCT	21.39	0.0	0.57	0.0	0.0	0.0	78.04	0.0
FIBRO	0.0	0.0	99.49	0.0	0.12	0.0	0.0	0.39
LOBU	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0
MUCI	0.0	0.0	0.41	0.0	0.0	0.28	0.0	99.31
PAPI	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0
PHY	0.0	0.16	0.0	99.84	0.0	0.0	0.0	0.0
TABU	0.0	94.84	0.0	4.63	0.54	0.01	0.0	0.0

right-classified and misclassified instances, and we also provided their Softmax probabilities output in TABLE 10. Similarly to the binary classification, the model ensembling method has been applied for the multi-class classification and the results are given in TABLE 6.

b: USE OF LINEAR SVM ON THE TOP OF CNN
We also investigated the hybrid approach 'CNN+ Classifier' for benign and malignant cancer sub-classes classification. Following the same used in [21], we put a small fully-connected layer downstream of the last convolutional layer, with several neurons equal to eight, which is the number of

classes, and without using activation layer (non-linearities), feeding a hinge loss layer. The accuracies obtained were 82.89%, 80.94%, 79.44%, and 77.94%, respectively, which is less than the accuracy reached by using CNN in Section 2.1. The decreasing result is due to the linearity of support vector machines.

B. 'CNN FEATURES + CLASSIFIER'CONFIGURATION

Convolutional neural networks' layers produce activations for input images. A set of features can be extracted from these layers and used for classification with traditional classifiers. We opted for 'CNN Features + Classifier' configuration, where we have extracted features from the fully connected layer with 2000 outputs and then gave them as the input to K-nearest neighbors, support vector machines, and random forests. For K-nearest neighbors, the number of neighbors is set to 3. Support vector machines were tested with two kernels, linear and radial basis kernels. Radial basis kernels were used with the following parameters, which provided the best accuracies: C = 5 and gamma = (1/number of features). The construction of the trees for random forests was carried out using Gini index which used as a split criterion for the attributes and the number of trees was set to 50. The features are extracted from the convolutional neural network applied to the augmented dataset for all magnification factors. The performance of the classifiers is given in TABLE 11 for both binary and multi-class classification. By comparing the performances of the classifiers for the binary classification, we find that linear SVM provided the highest accuracies for the four magnification factor. This is due to a decision boundary that separates the training data (the data distribution is linearly separable). For the multi-class classification, the RBF kernel worked better on images with 40X and 100X, while KNN worked better on images with a 200X magnification factor and linear SVM worked better on images with a 400X magnification factor. Thus, the performance of the classifiers greatly depends on the characteristics of the data and the ability to discriminate between the classes.

TABLE 11. The performance of different classifiers with deep features extracted from the fully connected layer with 2000 output.

	MAGNIFICATION FACTOR	KNN	RBF SVM	Linear SVM	Random Forest
BINARY	40X	88.80	84.45	90.64	86.70
	100X	85.42	86.21	89.58	86.05
	200X	88.58	87.09	90.23	88.25
	400X	75.16	73.88	75.96	75.32
MULTI-CLASS	40X	70.48	75.43	72.35	66.38
	100X	68.00	71.20	67.68	65.12
	200X	70.08	67.27	66.45	69.80
	400X	66.38	65.12	64.95	67.96

C. HANDCRAFTED FEATURES-BASED APPROACH RESULTS

In this section, we will provide the results of DSIFT and SURF features encoded with bag of words and

locality constrained linear coding. Firstly, we will give the implementation setting which chosen carefully to provide the best performances. The size of codebook used for both coding models is 1500 visual words. The public library, vlfeat [45], was used for vocabulary building for bag of words model. After extracting both DSFIT and SURF features, first, a vocabulary of 1500 visual is built. The 1500 visual words are selected using K-means clustering where the extracted features are divided into 1500 clusters and the centroid of each cluster is taken as a visual word. The max number of iterations for K-means is set to 100. In order to encode a descriptor and assign it to one of the centroids, we need to use K-nearest neighbors to measure the distance. We have set the number of neighbors to 20. For LLC, we used the implementation of LLC from [39] with also 20 nearest neighbors (K = 20) and max pooling to pool the encoded descriptors and to provide the final image presentation. K-means is also used for clustering for LLC to build the dictionary (Codebook). Spatial pyramid matching was also used in order to enhance the performance of LLC with different levels. Libsvm [46] library is used to train and build the SVM. For DSIFT features, we used the following settings: Binsize = 8 and the step = 4. The results of BoW model are given in Section 1, and the results of LLC are given in Sections 2.

1) DSIFT/SURF FEATURES ENCODED WITH BOW MODEL

In this sub-section, we provided the performance of the set of features extracted from the images and encoded with the bag of words model. The BoW model's codebook comprises 1500 visual words. TABLE 12 provides the accuracies for each magnification factor for the binary and the multi-class classification. The performances of the binary classification using SURF features were satisfying compared to the one of DSIFT features (The difference is between 18% and 23%). For the multi-class classification, all the accuracies were less than 50% due to the number of handled classes. Also, bag of words model ignores the high-level semantics and spatial information which is another reason of the low performances. Much semantic information is lost during the generation of the codebook, because the visual vocabulary is constructed by directly clustering the low-level visual feature vectors [47], [48]. The extracted features were trained using support vector machines with radial basis kernel, which offered the best results compared to linear SVM.

TABLE 12. The performance of the features encoded with bow and classified with svm for the binary and the multi-class classification tasks.

	FEATURES	40X	100X	200X	400X
BINARY	DSIFT	52.68	56.22	50.12	50.00
	SURF	79.95	74.80	70.96	72.02
MULTI-CLASS	DSIFT	18.77	17.28	20.16	17.49
	SURF	49.65	47.00	38.84	29.50

TABLE 13. The performance of the features encoded with LLC model, classified with svm for the binary and the multi-class classification tasks. SPL: spatial pyramid level.

Task	SPL	40X		100X		200X		400X	
		DSFIT	SURF	DSIFT	SURF	DSIFT	SURF	DSIFT	SURF
Binary	0	69.40	84.78	72.11	81.57	70.36	83.77	70.32	82.05
	1	69.73	85.27	70.67	82.85	70.53	84.93	71.61	82.60
	2	70.06	85.79	71.80	81.73	71.19	79.64	67.77	83.15
Multi-class	0	48.46	55.80	49.44	54.24	43.97	40.83	32.60	37.20
	1	47.44	54.61	44.32	53.92	44.46	48.10	32.86	38.12
	2	44.54	53.75	51.68	44.30	44.30	45.30	35.54	40.88

TABLE 14. The performance of DSIFT/SURF features encoded with BOW/ LLC models and classified with CNN FOR the binary and the multi-class classification tasks.

Model	Task	Feature	40X	100X	200X	400X
BoW	Binary Classification	DSIFT	66.72	69.06	62.42	52.75
		SURF	85.45	79.77	78.97	78.57
	Multi-class Classification	DSIFT	41.80	38.56	49.75	38.67
		SURF	53.07	60.80	70.00	51.01
LLC	Binary Classification	DSIFT	72.74	78.04	78.97	75.00
		SURF	87.00	82.50	84.00	87.91
	Multi-class Classification	DSIFT	60.58	57.44	70.00	46.96
		SURF	80.37	63.84	74.54	54.70

2) DSIFT/SURF FEATURES ENCODED WITH LLC MODEL

This section focuses on the performance of locality constrained linear coding applied to the DSIFT and SURF features, with and without adding spatial information. The set of features vectors are also trained with radial basis kernel SVM. The results are given in TABLE 13 for the binary classification and the multi-class classification. Similarly to the configuration in Section 1, the results of SURF features were higher than the one of DSIFT. Also, the results of the binary classification were much better than the one on the multi-class classification. Incorporating the spatial information of the local descriptors allowed increasing the performance of the models, in most of the cases.

D. 'HANDCRAFTED FEATURES+ CNN' RESULTS

In order to test an alternative 'Handcrafted Features+CNN' configuration, and instead of using traditional classifiers (such as SVM and K-nearest neighbors) to classify the handcrafted features extracted from the images, we replaced the traditional classifiers with fully-connected layers and we gave the set of features as an input to CNN. The features were stored in the HDF5 format. Fully-connected layers are simply convolutional layers. The only difference is that fully connected layers use filters with the same size as the input. Also, for each unit in this layer, it has full connections with units of the previous layer. The same CNN topology is applied to the binary and the multi-class classification. The used CNN for features encoded with BoW and LLC coding comprises three fully-connected layers. The first fully-connected layer has 2,000 outputs followed by a RELU function with a dropout rate of 50%. The second fully-connected layer has 100 outputs followed by a RELU function with a dropout

rate of 50%. The third fully-connected layer has (two/ eight) outputs and the number of classes without a dropout or RELU function. The network weights are initiated using Gaussian distribution. The regularizing parameter weight decay is set to 0.1. The number of iterations is set to 20000 iterations. The learning rate is initiated to 0.001, and the used policy is inverse decay. The batch size is set to 32. The standard gradient descent is used in training. This topology has given higher accuracy compared to the traditional approach. For example, this topology has given an accuracy of 70% for SURF features extracted from images with 200X magnification factor for the multi-class classification, while the traditional classifier (SVM) gave an accuracy of 38.84%. All the results are given in TABLE 14.

V. DISCUSSION

In this work, we have made a direct comparison of convolutional neural networks against the handcrafted features-based classification in classifying breast cancer histology images into benign and malignant (binary classification) and also into benign and malignant sub-classes (multi-class classification). The same CNN topology was applied to both classification tasks. For the binary classification, we reached an accuracy ranging from 93.77 % to 94.65 % using the original data. By applying data augmentation, we succeeded to increase it for all the magnification factors (ranging from 95.97% to 96.96). Although the CNN topology provided very promising results for the binary classification, the performance for the multi-class classification was a bit less (ranging from 79.74 % to 86.34%). The reasons behind that are: (1) the number of classes is large (eight) compared to the binary classification, (2) the difficulty of discriminating
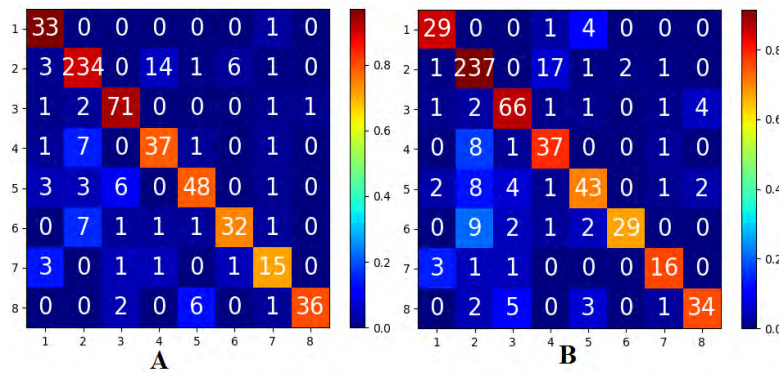
FIGURE 9. Confusion matrixes, (A): Confusion matrix of CNN models applied to the original data set with 40X magnification factor, (B): Confusion matrix of CNN models applied to the augmented data set with 40X magnification factor, 1: adeno, 2: ductal, 3: Fibro, 4: lobular, 5: mucinous, 6: papillary, 7: phyllodes, 8: tabular.

between the classes and especially between ductal carcinoma and lobular carcinoma and also between tabular adenoma and fibroadenoma, as shown in the confusion matrix depicted in FIGURE9, (3). Another reason could be the small number of training samples present per class, even after adopting data augmentation techniques. Data augmentation techniques allowed us to extend the data with a set of transformations (rotation and flipping) of the original data. Applying data augmentation helped to increase the accuracy of images with the 100X, 200X, and 400X magnification factors.

TABLE 15. Accuracy per class for images with 40X magnification factor with and without applying data augmentation.

ACCURACY PER CLASS	ORIGINAL DATA SET	Augmented
ADENO	97.06	85.29
DUCTAL	90.35	91.51
FIBRO	93.42	86.84
LOBULAR	78.72	78.72
MUCINOUS	78.69	70.49
PAPILLARY	74.42	67.44
PHYLLODES	71.43	76.19
TABULAR	73.33	75.56

For images with the 40X magnification factor, the accuracy dropped from 86.39% to 83.79%. To analyze the reason for this, we computed the accuracy per class for the original and the augmented data set given in TABLE 15 and also depicted the confusion matrixes in FIGURE 9. As we can see in TABLE 15, data augmentation worked well for some classes (ductal carcinoma, lobular carcinoma, phyllodes tumor and tabular adenoma) and for some (adenosis, fibroadenoma, mucinous and papillary carcinoma), it led to a drop in the performance. The reason could be that the augmented data for the classes where the performance decreased does not represent the intended output, thus creating confusion for the CNN model with other classes' outputs. For example, in the confusion matrix FIGURE 9 (A) and by using the original data set, the number of correctly classified

samples for fibroadenoma was 71. However, after applying data augmentation, this number dropped to 66 samples. Five samples are misclassified as tabular adenoma (3 samples), lobular carcinoma (1 sample) and mucinous carcinoma (1 sample).

The design of the right convolutional neural networks is a crucial step and it affects the performance of the CNN. However, there are no firm rules about the exact number of layers that should be used in order to attain the highest performances. It is usually based on experimentation, background knowledge, and the type of problem at hand. In this work, we have used a topology comprising five convolutional layers which have been chosen using cross-validation. The same topology is used for the binary and the multi-class classification tasks. This topology brought the model to have the flexibility to capture high-level features and non-linearities in the data, while also providing promising results. We have used small filters of 3×3 for all the layers in order to capture fine details in the images. Now, we will compare our topology and the achieved results with the recent works where CNN and the handcrafted features-based approaches are used. The comparison results are given in TABLE 16.

In [22], Spanhol *et al.* tested out LeNet [17] topology (which consists of two convolutional layers and three fully-connected layers) to classify breast cancer tumors into benign and malignant. However, they only reported an accuracy of 72%. Thus, they decided to use a complex topology which is a variant of AlexNet [18] networks. This topology consists of three convolutional layers with a filter size of 5×5 and two fully-connected layers. To train the network, the images were divided into small patches using sliding window and random strategy with different sizes and to get the final classification result, the patches results are combined for the whole image. The authors reported an accuracy ranging from 80.8% to 85.6% where they used sum fusion rules on the different trained models with different patch sizes. On the other hand, in [24], Spanhol *et al.* reused the weights of the pre-trained BVLC CaffeNet architecture to extract a set of deep

TABLE 16. Comparative analysis OF CNN'S results with existing CNN in the literature.

AUTHORS	DATA SET	Classes	Accuracy	Precision	Recall	F1-score
SPANHOL ET AL [22]	**BreaKHis** a. 40X b. 100X c. 200X D. 400X	BENIGN/ MALIGNANT	89.6±6.5 85.0±4.8 84.0 ±3.2 80.8 ±3.1	---	---	92.9 88.9 88.7 85.9
ARAÚJO ET AL [23]	269 IMAGES	CARCINOMA/NON-CARCINOMA	90.00	---	---	---
		NORMAL/BENIGN/IN SITU/ INVASIVE	85.00	---	---	---
SPANHOL ET AL [24]	**BreaKHis** a. 40X b. 100X c. 200X D. 400X	BENIGN/ MALIGNANT	84.6 ±2.9 84.8 ± 4.2 84.2 ±1.7 81.6 ± 3.7	---	---	88.0 88.8 88.7 86.7
BAYRAMOGLU ET AL [25]	**BreaKHis** a. 40X b. 100X c. 200X D. 400X	BENIGN/ MALIGNANT	83.0±3.0 83.1±3.5 84.6±2.7 82.1±4.4	---	---	---
NAHID ET AL [26]	**BreaKHis** a. 40X b. 100X c. 200X D. 400X	BENIGN/ MALIGNANT	94.40 95.93 97.19 96.00	94.00 98.00 98.00 95.00	96.00 96.36 98.20 97.79	95.00 97.00 98.00 96.00
HAN ET AL [27]	**BreaKHis** a. 40X b. 100X c. 200X D. 400X	BENIGN/ MALIGNANT	95.8 ±3.1 96.9±1.9 96.7±2.0 94.9±2.8	---	---	---
		MULTI-CLASS	92.8±2.1 93.9±1.9 93.7±2.2 92.9±1.8	---	---	---

features and input them into logistic regression classifier. The reported accuracy was ranging from 83.6% to 84.8%. In [25], Bayramoglu *et al.* proposed an approach that is magnification independent. The authors proposed two architectures, one named ''single task convolutional neural network'' which is used to predict the malignancy, and a ''multi-task convolutional neural network'' which is used to predict both the malignancy and the magnification factor. For the single task, the average accuracy (based on patient score) was 83.25% while for the multi-task; the average accuracy was 82.13% for the benign/malignant classification task. In [26], raw images, handcrafted features, and frequency-domain information were fed into a CNN to classify the breast histology images into benign and malignant. The reported performance was between 94.40% and 97.19%. In [27], the use of a structured deep learning model allowed reaching an accuracy of between 94.9% and 96.9% for the binary classification and accuracy between 92.8% and 93.9% for the multiclass classification. When we compare our CNN network results for the binary classification, we find that our topology outperforms the proposed approaches in [22] and [24]–[27]

in terms of accuracy where we achieved the state-of-the-art accuracy between 96.15% and 98.33 %. By comparing the accuracies and the F1-scores, we find that our CNN topology (without using ensemble model) provided the highest accuracies and F1-scores for images with 40X and 100X magnification factors compared to the study in [26] and it provided the highest accuracies and F1-scores for images with 40X, 100X and 400X magnification factors compared to the study in [27]. The results of [27] for the multi-class classification were better (between 92.8% and 93.9%).

In [23], Araújo *et al.* classified breast histology images into four classes, normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma. The proposed CNN comprises three convolutional layers and three fully-connected layers. The authors also extracted deep features from the CNN and fed them into support vector machines. The highest accuracy the authors reported was 85% for four classes' problem and 90% for carcinoma/ non-carcinoma classification. While the authors only reported 85% for the four-class classification; with our CNN, and for eight-class classification problem, we achieved 88.23% for images with a 40X magnification

factor as the highest accuracy. Also, for the binary classification, we achieved higher results.

In [14], Spanhol *et al.* have compared the performances of texture features including local binary pattern (LBP), completed local binary pattern (CLBP) and Gray Level Co-Occurrence Matrices (GLCM)) and also Oriented FAST and Rotated BRIEF (ORB) which is a developed alternative for SIFT and SURF for the classification of histology images into benign and malignant categories. By comparing the achieved results with ours, where we used encoded local descriptors, we found that the encoded SURF features with LLC coding and classified with support vector machines outperformed both texture features and also ORB features.

VI. CONCLUSION

In this work, we compared the performance of convolutional neural networks with various configurations for the classification of breast cancer histology images into benign and malignant, and also into benign and malignant sub-classes too. The designed CNN topology worked well on both binary and multi-class classification tasks. However, the performance of the multi-class classification was lower when compared to the one of the binary classification due to the number of handled classes and also due to the similarities between the sub-classes. The performance of the handcrafted features-based approach where we used coding models to encode the local descriptors to build image representation was too low compared to the CNN. Convolutional neural networks are also used to replace traditional classifiers with fully-connected layers to train the handcrafted features (DSIFT and SURF), which helped increase the performance of the handcrafted features. Convolutional neural networks have become state-of-the-art, demonstrating an ability to solve challenging classification tasks. Our proposed CNN topology has beaten the previous ones for the binary classification task, where we reached a performance of between 96.15% and 98.33%.

ACKNOWLEDGMENT

REFERENCES

[1] M. C. Chun. (2018). *Breast Cancer: Symptoms, Risk Factors, and Treatment*, Medical News Today. Accessed: Mar. 10, 2018. [Online]. Available: https://www.medicalnewstoday.com/articles/37136.php

[2] (2018). *World Health Organization.* Accessed: Mar. 10, 2018. [Online]. Available: http://www.who.int/en/

[3] P. Boyle and B. Levin. (2008). *World Cancer Report.* [Online]. Available: http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr2008.pdf

[4] J. Joy, E. Penhoet, and D. Petitti, *Saving Women's Lives.* Washington, DC, USA: National Academies Press, 2005.

[5] L. He, L. R. Long, S. Antani, and G. Thoma, "Computer assisted diagnosis in histopathology," *Sequence Genome Anal., Methods Appl.*, 2010, pp. 271–287.

[6] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," *Comput. Methods Programs Biomed.*, vol. 107, no. 3, pp. 538–556, 2012.

[7] Nbcf. *Biopsy: The National Breast Cancer Foundation.* Accessed: Mar. 10, 2018. [Online]. Available: http://www.nationalbreastcancer.org/breast-cancer-biopsy

[8] *Machines Will be Able to Diagnose Cancer.* Accessed: Mar. 10, 2018. [Online]. Available: http://sciencenordic.com/machines-will-be-able-diagnose-cancer

[9] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1563–1572, 2013.

[10] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2169–2178, Dec. 2013.

[11] Y. M. George, H. H. Zayed, M. I. Roushdy, and B. M. Elbagoury, "Remote computer-aided breast cancer detection and diagnosis system based on cytological images," *IEEE Syst. J.*, vol. 8, no. 3, pp. 949–964, Sep. 2014.

[12] Y. Zhang, B. Zhang, F. Coenen, and W. Lu, "Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles," *Mach. Vis. Appl.*, vol. 24, no. 7, pp. 1405–1420, 2012.

[13] Y. Zhang, B. Zhang, F. Coenen, J. Xiao, and W. Lu, "One-class kernel subspace ensemble for medical image classification," *EURASIP J. Adv. Signal Process.*, vol. 2014, p. 17, Dec. 2014.

[14] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[15] M. Veta, J. Pluim, P. van Diest, and M. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.

[16] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio, Speech Language Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2015.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.

[19] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[21] A. Ferrari, S. Lombardi, and A. Signoroni, "Bacterial colony counting with convolutional neural networks in digital microbiology imaging," *Pattern Recognit.*, vol. 61, pp. 629–640, Jan. 2017.

[22] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2560–2567.

[23] T. Araújo *et al.*, "Classification of breast cancer histology images using convolutional neural networks," *PLoS ONE*, vol. 12, no. 6, p. e0177544, 2017.

[24] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 1868–1873.

[25] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2440–2445.

[26] A.-A. Nahid and Y. Kong, "Histopathological breast-image classification using local and frequency domains by convolutional neural network," *Information*, vol. 9, no. 1, p. 19, 2018.

[27] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.*, vol. 7, Jun. 2017, Art. no. 4172.

[28] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

[29] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 524–531.

[30] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification via pLSA," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 517–530.

[31] A. Bosch, A. Zisserman, and X. Muñoz, "Image classification using random forests and ferns," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[32] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[33] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[34] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, "Image features detection, description and matching," in *Image Feature Detectors and Descriptors* (Studies in Computational Intelligence). Cham, Switzerland: Springer, 2016, pp. 11–45.

[35] D. A. Lisin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, "Combining local and global image features for object class recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops (CVPR)*, Sep. 2005, p. 47.

[36] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, and B. C. Lovell, "Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 95–102.

[37] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2006, p. 13.

[38] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1794–1801.

[39] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained Linear Coding for image classification," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 3360–3367.

[40] S. Manivannan, W. Li, S. Akbar, R. Wang, J. Zhang, and S. J. Mckenna, "An automated pattern recognition system for classifying indirect immunofluorescence images of HEp-2 cells and specimens," *Pattern Recognit.*, vol. 51, pp. 12–26, Mar. 2016.

[41] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 2169–2178.

[42] S. Lazebnik, C. Schmid, and J. Ponce, "Spatial pyramid matching," in *Object Categorization: Computer and Human Vision Perspectives*. Cambridge, U.K.: Cambridge Univ. Press, 2009, pp. 401–415.

[43] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Oct. 2005, pp. 1458–1465.

[44] K. Vyas, Y. Vora, and R. Vastani, "Using bag of visual words and spatial pyramid matching for object classification along with applications for RIS," *Proc. Comput. Sci.*, vol. 89, pp. 457–464, Jan. 2016.

[45] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. 18th Int. Conf. Multimedia*, 2010, pp. 1469–1472.

[46] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.

[47] Z. Lu, L. Wang, and J.-R. Wen, "Image classification by visual bag-of-words refinement and reduction," *Neurocomputing*, vol. 173, pp. 373–384, Jan. 2016.

[48] L. Wu, S. C. H. Hoi, and N. Yu, "Semantics-preserving bag-of-words models and applications," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1908–1920, Jul. 2010.

DALAL BARDOU received the B.S. degree in computer science and the M.S. degree in fundamental and applied computer from Abbes Laghror University, Khenchela, Algeria. She is currently pursuing the Ph.D. degree with the Nanjing University of Science and Technology, Nanjing, China. Her research interests are focused on artificial intelligence, deep learning, machine learning, image processing, and medical engineering.

KUN ZHANG received the B.S., M.S., and Ph.D. from the Nanjing University and Science and Technology. She is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology. Her research interests are focused on big data, pattern recognition, medical engineering, networks, and trusted computing.

SAYED MOHAMMAD AHMAD received the M.Tech. degree (Hons.) in language technology from the Centre for Development of Advanced Computing, India. He has also received several awards and accolades nationally and internationally. He has worked in capacity of a Senior Manager, the Head-Technical Delivery in multinational companies and Researcher with King Saud University, Saudi Arabia. He is currently a Chief Technology Officer and the Head of the research and development with Lareb Technologies. His current research interests and work are designing system architecture, artificial intelligence, medical engineering, image processing, language technology, cloud computing, security solutions, software development, data mining, parallel computing and conversion of outcome of research into commercial products. He is also a Reviewer of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

• • •