Optimised CNN in conjunction with efficient pooling strategy for the multi-classification of breast cancer

Shallu Sharma1 | Rajesh Mehra1 | Sumit Kumar2,3

1 Department of Electronics and Communication Engineering, National Institute of Technical Teachers’ Training & Research, Chandigarh, India
2 School of Electronics and Electrical Engineering, Division of Communication Systems, Lovely Professional University, Jalandhar, Punjab, India
3 Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, India

Abstract

Tissue analysis using histopathological images is the most prevailing as well as a challenging task in the treatment of cancer. The clinical assessment of tissues becomes very tough as high variability in the magnification levels makes the situation worst for any pathologist to deal with the benign and malignant stages of cancer. One of the possible ways to address such a pathetic situation could be an advanced machine learning approach. Hence, a convolutional neural network (CNN) architecture is proposed to create an automated system for magnification independent multi-classification of breast cancer histopathological images. This automated system offers high productivity and consistency in diagnosing the eight different classes of breast cancer from a balanced BreakHis dataset. The system utilises an efficient training methodology to learn the discerning features from images of different magnification levels. Data augmentation techniques are also employed to overcome the problem of overfitting. Additionally, the performance of CNN architecture has been improved in a significant manner by adopting an appropriate pooling strategy and optimisation technique. Based on that, we have achieved an accuracy of 80.76%, 76.58%, 79.90%, and 74.21% at the magnification 40X, 100X, 200X, and 400X, respectively. The proposed model outperforms the handcrafted approaches with an average accuracy of 80.47% at 40X magnification level.

1 INTRODUCTION

Multi-classification of cancer from histopathology slides have a prominent place in pathology for clinical practices to provide a reliable diagnosis. It is very challenging to manually classify the histopathological images in multiple classes of cancer including benign, malignant, and baffling categories. Confounder tissue patterns and similar clinical expressions of the classes are some factors due to which classes often emulate and share common attributes [1]. As a matter of fact, it becomes difficult to identify a class correctly. Besides, a worldwide dearth of experienced and skilled pathologists has burdened the available pathologist due to the increasing incidence of new cancer cases [2]. In the manual analysis of biopsy slides, the chances of misdiagnosis increase due to over-fatigue of pathologists [3]. To overcome this situation, there is an urgent need to develop an automated and effective multi-classification system that not only assists pathologists but also helps in reducing their workload with improved efficacy of the diagnosis process as well.

However, automated multi-classification systems also confront critical obstacles.

Handcrafted feature engineering is the primary obstacle in the development of an efficient automated system. Pre-processing steps play a key role in handcrafted feature engineering but require remarkable domain expertise in extracting task-specific features from the images [3, 4]. In most of the previous studies, texture-based features have been extracted from histopathology images in order to differentiate between classes. To accomplish this, various feature descriptors such as scale-invariant feature transform (SIFT), speeded up robust features (SURF), oriented fast and rotated brief (ORB), histogram of gradient (HOG), graph run length matrix (GRLM), grey level co-occurrence matrix (GLCM), local binary patterns (LBP), local phase quantization (LPQ), and wavelet-based descriptors are already in practice by a large research community [5–12]. The handcrafted feature engineering is time-consuming and demands high computational resources. In general, segmentation is a prerequisite to derive the most relevant features from...
the segmented nuclei and glands in histopathological images, and it would be possible only when an area of interest is segmented properly from the image [13]. Therefore, the overall performance of a system highly depends on the success of the preprocessing step. Intra-class variability, inhomogeneous colour distribution, and high coherency in cancerous cells are other serious obstacles in the analysis of histopathological images. The pictorial representation of these obstacles is given in Figure 1, which represents different classes of breast cancer histopathological images obtained from the BreakHis dataset [14]. The magnification level of the histological image is another challenge for the automated multi-classification system because the graininess of an image becomes worse at higher magnification. With a significant rise in the level of magnification, more coloured dots appeared in the images and make them noisy [15]. Therefore, capturing a digital image from biopsy slides at a different level of magnification introduces complexity in the background and makes it tedious to extract distinct features for differential diagnosis [16]. To reduce variation in the background due to different levels of magnification, Peikari et al. [17] and Loukas et al. [18] utilise images of the same magnification for model training. The experiment for each magnification factor is conducted independently. In this context, Spanhol et al. [14] and Keskin et al. [8] utilised multiple classification models in which each model is associated with a specific magnification level of images. But the prior knowledge of image magnification factor and inadaptability to new magnification put serious constraints on the performance of the systems designed by Spanhol et al. and Keskin et al. Thus, the automated system should be efficient in handling different magnification level as well as adaptable to images with new magnification level.

Most of the recent literature based on the classification of histopathological images has not considered the effect of magnification factor on the performance of the classifier. Murtaza et al. [19] developed a classification model named BMIC_Net to classify the BreakHis dataset into multiple classes. They achieved a significant performance in the subclassification of benign and malignant classes but did not consider the effect of the magnification factor in their study. Similarly, Dhahri et al. [20] proposed an automated cancer diagnosis system that is further optimised using the Tabu search algorithm for feature selection but no attention has been paid to the magnification factor of the images. Adaboost and logistic regression (LR) are found as the high-performing classifier when trained on these selected features. Kassani et. al [21] proposed an ensemble model of three pre-trained convolutional neural networks (CNNs; VGG19, MobileNet, and DenseNet), and a variety of tuning techniques (stain normalisation, data augmentation, hyperparameter tuning, and fine-tuning) are also applied. However, no consideration has been given to the magnification factor while performing the classification. The magnification factor in histopathological images is one of the root causes that completely alters the behaviour of a classifier for solving the same problem over similar settings of the classifier parameters. In this context, we are trying to bridge the gap of literature by conducting this study to determine the effect of the magnification factor under different settings of input data.

In order to provide an effectual, reliable, and robust automated system for multi-classification, we have developed a magnification independent multi-classification model for the classification of breast cancer histopathological images. This study is an extension of our previous study [22], where we designed a CNN for magnification independent binary classification. The proposed CNN performed efficiently on the entire magnification levels of the BreakHis dataset as well as outperformed the research contribution given by Spanhol et al. in [23]. Handcrafted feature engineering is not required in CNN as it is leveraged with the hierarchical representation of learning the semantic and discriminate features from low to high level [24]. In view of the foregoing, the key contributions of the present study are listed below:

1. We have designed a less complex and efficient CNN architecture for classifying breast cancer histopathological images. The classification process has enough capability of handling images of unknown magnification and also demonstrates the efficacy of the proposed methodology in multi-classification of breast cancer under eight different classes.
2. To further improve the effectiveness of the proposed CNN architecture, we have incorporated the pooling technique along with a suitable optimiser in an appropriate manner so that the overall performance of the network could be enhanced.
3. The present study demonstrates the capability of the network in learning the discerning features from the images.
Additionally, we have validated the influence of training iterations in the features learning process through visual analysis.

4. The current study also provides a comparative analysis between the proposed CNN and the handcrafted approach for the multi-classification of magnification independent histology images. The obtained results are compared that are based on various performance metrics such as accuracy, precision, recall, F1-score, receiver operating characteristic (ROC), and area under the curve (AUC).

5. It is worth mentioning that, for the first time, we have reported the magnification independent multi-classification for the detection of breast cancer from histopathological images to the best of authors' knowledge concerned.

6. Moreover, the study helps in determining the most confusing and complicated sub-classes of breast cancer to felicitate the proposed network for defining the benign and malignant classes in the histology dataset.

2 MATERIALS AND METHODS

In this section, the detail of the database used for the execution of experimental work along with the description of the employed methodology is provided.

2.1 Database

To examine the viability of the proposed CNN for the multi-classification of cancer using histological images, the BreakHis database is utilised. BreakHis database is large enough to make statistical analysis because it consists of a total of 7909 histological images related to eight classes of breast cancer at a magnification level of 40, 100, 200, and 400 X (Figures 2 and 3). Image samples are acquired from the hematoxylin and eosin-stained biopsy slides of 82 anonymous patients at Pathology Anatomy and Cytopathology Lab, Brazil [14]. The size of each image is 700 × 460 pixels and available in portable network graphics format (see [25]). A detailed distribution of the BreakHis dataset is tabularised in Table 1.

2.2 Methodology

A new CNN topology has been proposed that utilises a defined training protocol and is elaborated in Section 2.2.2. A stratified training set is utilised in which image samples are equal for each class. Image distribution in the BreakHis dataset is not uniform at the entire magnification level, therefore balancing has been

| TABLE 1 Distribution of images in the BreakHis dataset |
|-------------------------------------------|-----|-----|-----|-----|-----|
| BC                         | 40 X | 100 X | 200 X | 400 X | Total | No. of Patients |
| Adenosis (A)               | 114  | 113   | 111   | 106   | 444   | 4               |
| Fibro-Adenoma (F)          | 253  | 260   | 264   | 237   | 1014  | 10              |
| Tublar Adenoma (TA)        | 109  | 121   | 108   | 115   | 453   | 3               |
| Phyllods Tumor (PT)        | 149  | 150   | 140   | 130   | 569   | 7               |
| Ductal Carcinoma (DC)      | 864  | 903   | 896   | 788   | 3451  | 38              |
| Lobular Carcinoma (LC)     | 156  | 170   | 163   | 137   | 169   | 5               |
| Mucinous Carcinoma (MC)    | 205  | 222   | 196   | 169   | 792   | 9               |
| Papillary Carcinoma (PC)   | 145  | 142   | 135   | 138   | 560   | 58              |
done before the training process. Class with least image samples is kept as it is, while the remaining classes are randomly downsampled to make them equivalent to the class with the least samples. Down-sampling is applied to the classes at each magnification level. All the images of the dataset are put together in a common folder without considering their magnification to design a magnification independent multi-classification system. But testing has been performed independently to determine the ability of the model in classifying the data based on the magnification. A schematic representation of the proposed methodology is shown in Figure 4. The dataset has been split into 90% training set and 10% testing set. The training set is further divided into 75% training and 25% validation set for each experiment. The data augmentation has been applied to both balanced training and validation sets, in which each image is rotated to an angle of 90, 180, and 270 degrees to enlarge the dataset [22].

### 2.2.1 CNN topology

CNN is a special type of neural network that identifies visual representations directly from images [26–29]. In the present study, the proposed CNN has been implemented using the Tensorflow framework. CNN architecture begins with two convolutional layers (conv1 and conv2), where the input image of size $224 \times 224$ mapped to $56 \times 56$ feature maps (Table 2). In both convolutional layers, the numbers of filters are 32 with a kernel of size $3 \times 3$. One more convolutional layer (conv3) is appended in the network with 64 filters of size $3 \times 3$. The architecture then proceeds with one pooling and flatten layer. Max-pooling operation is performed using a kernel of size $2 \times 2$ with a stride value 2. To maintain the size of an input image, zero paddings are applied at the second and third convolutional layers because a convolutional operation always resulted in an image with reduced size. The flattened layer of the architecture

| Layer   | Type       | Activation | Regularisation | No. of filters | Filtre size | Pooling size/stride | Zero padding |
|---------|------------|------------|----------------|----------------|-------------|---------------------|--------------|
| Conv1   | Convolutional | ReLU       | –              | 32             | $3 \times 3$ | –                   | –            |
| Conv2   | Convolutional | ReLU       | –              | 32             | $3 \times 3$ | –                   | 2 × 2        |
| Conv3   | Convolutional | ReLU       | –              | 64             | $3 \times 3$ | –                   | 2 × 2        |
| Max-Pool| Pooling    | –          | –              | –              | –           | $2 \times 2/2$      | –            |
| FC1     | Fully connected | ReLU       | Dropout        | 128 nodes     | $1 \times 1$ | –                   | –            |
| FC2     | Fully connected | ReLU       | Dropout        | 128 nodes     | $1 \times 1$ | –                   | –            |
provides a set of 50,176 features, which is further followed by two fully connected layers (FC1 and FC2). Each fully connected layer in the network is composed of 128 nodes. Eventually, a softmax layer is added at the end to produce the desired output. In addition to this, ReLU activation is used at the entire convolutional and fully connected layer to introduce non-linearity in the network.

### 2.2.2 Training protocol

Training is a very crucial step in a multi-classification system because a model works efficiently only when it is trained properly. An effectual training of a network requires adequate tweaking of the network parameters that can be done by determining the error between the desired and predicted output. Various types of loss function are computed in classification to measure this error but the cross-entropy is the most commonly used loss function for the classification problem, which can be represented as

\[
\text{Cross Entropy Loss (CEL)} = -\sum_{i=1}^{N} y_i \log(p_i)
\]

where \(y_i\) represents correct classification and \(p_i\) represents predicted probability observation of class \(y\), and \(N\) is the number of classes. Cross entropy loss (CEL) is a positive continuous function that becomes zero if the predicted output exactly equals the desired output. The backpropagation algorithm is used to minimise this loss function. Adam optimiser is then applied to optimise this loss function. But to control the optimisation, a single-scalar value is required as an output of CEL function. In this context, CEL is computed for all the images during the process of training and averaged. Later, this averaged CEL is considered for further processing in the classification.

The network parameters used in training of network are determined through an extensive set of trial and error experiments. From the analysis, the value of learning rate equal to 0.0001 is found as a reasonable choice because the use of a lower learning rate delayed the process of convergence, whereas a higher learning rate resulting in convergence failure. In this network, random initialisation of the weights has been carried out with a standard deviation of 0.05, and the network is trained from scratch on the dataset. To avoid the problem of a computer crash and run out of memory, a mini-batch of size 32 is considered. At last, the training process is ended only when the maximum accuracy is attained on the validation set.

### 2.3 Performance metrics

All experiments in the present study are executed on an Intel system with configuration Core (TM) i7-7500U CPU @ 2.90 GHz, and 8 GB onboard memory. The performance of the developed model in the classification of breast cancer is evaluated on the basis of the image recognition rate for each magnification independently to determine the robustness of the model. Image recognition rate can be defined as the ratio of correctly classified images of cancer to the total number of images in the test set [14, 30]. Using the protocol devised in Section 2.2.2, the training set for each experiment is composed of total 2400 images (400 images per class) which include images from all the magnifications, while the size of the validation set and testing set is taken as 800. The performance of the model is evaluated using confusion matrix and activations of convolutional layers to provide a visual analysis for further computation and visualisation purpose in a respective manner.

### 3 RESULTS AND DISCUSSION

The potential of the developed classification system is qualitatively described and analysed through an experimental study whose implementation is explained in this section. The process of hyperparameter selection is described in Section 3.1, a detailed analysis of system performance is presented in Section 3.2, and a visual analysis to illustrate the network capacity in extracting discerning features is explained in Section 3.3.

#### 3.1 Hyper-parameter tuning

Here, the effect of the most essential selections for the training procedures and system performance is described. In this context, Table 3 demonstrates the effects of different optimisers on classification performance and convergence time in the training of CNN. The parameters of the optimiser such as learning rate, momentum, and decay rate have also been tuned on the validation set accordingly. A learning rate of 0.0001 is used for Adam, AdaGrad, and gradient descent. On the other hand, a learning rate of 0.001 provided the best performance for RMSProp optimiser.

It has been observed from Table 3 that Adam optimiser yielded the best performance by minimising the cross-entropy error function in 700 iterations. Adam optimiser is further followed by an RMSProp optimiser with about 8% lower performance, while gradient descent and AdaGrad optimisers provide a significant drop in the performance of the network around 24% and 37%, respectively. The convergence curves in terms of validation accuracy and iterations for four different optimisers are shown in Figure 5. It has been analysed from the figure that the AdaGrad optimiser failed to converge and provides the poorest performance on the histological data. AdaGrad optimiser scales the learning rate for each parameter as per the

| Optimiser          | Error function | Acc. (%) | Iteration |
|--------------------|----------------|----------|-----------|
| Gradient descent   | Cross-entropy  | 69.4     | 426       |
| AdaGrad            | Cross-entropy  | 46.8     | 426       |
| Adam               | Cross-entropy  | 83.67    | 700       |
| RMSProp            | Cross-entropy  | 76.9     | 146       |

**TABLE 3** Performance analysis of designed CNN during training with different optimiser
gradient calculated in the previous step for the same parameter. The learning rate is reduced if the computed gradient for the parameter is large and vice-versa. A quick reduction in the learning rate may be a reason behind their poor performance due to which the optimiser stuck to the local minima instead of the global minima.

On the other hand, gradient descent optimiser performed better in terms of convergence as compared to AdaGrad for the same number of iterations, but still, the performance is not up to the mark. RMSProp is an adaptive optimiser where the learning rate for the parameters is adapted on the basis of the average of gradient’s magnitude (the first moment). However, Adam optimiser utilises the uncentred variance of the gradient (the second moment) in conjunction with the first moment to determine the learning rate of the parameters due to which Adam often performs better in comparison to RMSProp. Therefore, despite the convergence of the model in a small number of iterations with RMSProp, we select Adam because it provides better classification accuracy (Table 3), which is of main concern in the classification of cancer disease based on histopathological images.

Further, the effect of the pooling strategy is also studied by changing the pooling method from max-pooling to average pooling. This is evident from the experiment that the change in pooling strategy from max to average pooling resulted in a significant drop of 17.57% in the accuracy of the classification system (Table 4). Thus, a max-pooling strategy is found to be more effective for the problem of histopathological image classification. Staining in histopathology is performed to highlight important and discerning features of the tissue by enhancing the contrast in tissues. Loss of the information in terms of contrast may be one of the major reasons behind the failure of average pooling in histopathological image classification because tissue contrast is an important attribute in histopathological image analysis [29].

### Table 4: Performance analysis of designed CNN during training with different pooling strategy

| Pooling strategy | Optimiser | Accuracy | Iterations |
|------------------|-----------|----------|------------|
| Max-pooling      | Adam      | 83.67    | 700        |
| Average pooling  | Adam      | 66.1     | 700        |

### Table 5: Image recognition rate for multi-classification of Breast Cancer Histological Images at different magnifications

| Magnification | Classification | 40 X | 100 X | 200 X | 400 X | Ref. |
|---------------|----------------|------|-------|-------|-------|------|
| Multi-classification | 80.76 | 76.58 | 79.90 | 74.21 | Present study |
| Binary classification | 90.4  | 86.3  | 83.1  | 81.3  | [22]  |

#### 3.2 System performance analysis

Following the methodology discussed in Section 2.2, the proposed CNN is trained for 1000 iterations, initially. The model was not trained well, so it is terminated after achieving the highest accuracy of 45.9% on the validation set. Thus, the number of iterations has been increased to 9000 to train the network efficiently. A significant improvement in performance has been noticed by attaining accuracy of 67.2%. We have continuously increased the iterations to determine the point of convergence for training-validation loss and accuracy. It has been found that the model converged for approximately 50,000 iterations with 83.67% accuracy on the validation set. However, the problem of overfitting has occurred on the further increment of iterations.

The performance of the model has been analysed on four testing sets in which each set comprised of an equal number of images and have the same magnification specifications from 40, 100, 200, and 400 X. The idea behind this selection is to examine the performance of proposed CNN on unseen data to make an efficient magnification independent classification system to diagnose the class of breast cancer disease from each testing set. The obtained image recognition rates for multi-classification and binary classification using the same architecture are tabulated in Table 5. The network performance is not so impressive for multi-classification as for binary classification. In multi-classification, the numbers of classes are more than binary and the same data is distributed into many classes. Due to this fact, the number of image samples per class has also been reduced which leads to a drop in performance of the network.

Further, it has been observed from Table 5 that the performance deteriorates with a rise in magnification factor and least obtained with 400 X magnification. It has happened because histological images become more complicated due to the noise and creation of some artefacts at higher magnification. Consequently, only a few pixels provide the information required for...
FIGURE 6  Confusion matrices for the testing set at (a) 40 X, (b) 100 X, (c) 200 X, and (d) 400 X, where the classes from 0 to 7 represent adenosis, ductal carcinoma, fibroadenoma, lobular carcinoma, mucinous carcinoma, papillary carcinoma, phyllodes tumor, and tubular adenoma, respectively.

FIGURE 7  Few illustrations of misclassified images produced by the proposed convolutional neural network (CNN).

come are shown in Figure 7. It has been noticed that adenosis shows conflicts with the phyllodes tumor and fibroadenoma during the classification and nearly all the sub-classes of the malignant group create a little bit of confusion in the process.
of class identification. Eventually, the lobular carcinoma, mucinous carcinoma, phyllodes tumor and fibroadenoma cause higher misclassification and require a precise description of the texture, contrast and intensity values from the images.

### 3.3 Visualisations

The capacity of the network in learning discriminative features from the images plays a key role in classification. However, there is some probability to lose relevant details when a pooling operation is performed on the input image. Therefore, the selection of pooling operations and their position should be in a manner that preserves task-specific information and discards irrelevant information. From the extensive set of experimentation, we found that the application of pooling operation only after the last convolutional layer (conv3) in the proposed network provides the best performance due to less information loss. The effectiveness of the proposed model in the context of feature-level has been analysed in which feature maps (activations) from all the convolutional layers are extracted and visualised for a test image (Figure 8). Figures 8(b)–(d) represent the weights of convolutional filters in conv1, conv2, and conv3, respectively, in which red and blue colours represent the positive and negative weights of convolutional filters. The corresponding activations obtained from the convolutional layers 1, 2, and 3 for the considered test image are given in Figures 9(a)–(c). These feature maps represent the low-, mid-, and high-level features for the test image. From the visual analysis of the feature maps, one can easily determine whether the network is trained properly or not. A well-trained network always displays smooth activations that are free from noisy patterns (Figure 9), while a poorly trained network provides irregular and missed activations. From Figure 10, it has been observed that the activations obtained from the first convolutional layer are not smooth because the network is trained only for 1000 iterations.

Therefore, numbers of activations are rough and blanked in the resultant feature maps which illustrate that the network has not been trained for enough time. However, the network provides uniform activations when trained for 50,000 iterations and becomes a piece of evidence for its proper training. Also, we noticed that the conv2 and conv3 layers learn more task-specific features from the image which provides the ability to learn more discerning features (Figure 9).

### 4 STATE-OF-THE-ART COMPARISON

The prime objective of this study is to reveal the scope of the classification network in multiple classifications of breast cancer histopathological images. Therefore, it is essential to validate the devised framework over other methods of classification. In view of this, many descriptors (such as Hu Moment, Haralick texture, colour histogram) have been applied to extract the features from images to evaluate the potential of CNN in contrast to handcrafted approaches for the task of image
classical classification. Concurrently, several conventional machine-learning methods have also been examined for performance comparison.

A box-plot analysis has been done to evaluate the performance of the employed handcrafted approaches (Figure 11). Here, several conventional machine-learning algorithms have been applied in conjunction with the handcrafted feature descriptors, and the achieved accuracy is tabularised in Table 6. It has been observed from Figure 11 that the proposed CNN outperforms the handcrafted approaches in the multi-classification of the breast cancer images. The classification accuracy achieved by CNN for the 40 X magnification level is the highest where the value of mean ($\mu$) is 80.47 and the standard deviation ($\sigma$) is 1.42. However, the classification performance of CNN is comparable to the performance which is obtained by the random forest (RF) classifier ($\mu = 79.47$ and $\sigma = 1.38$). This is because the RF is composed of several decision trees. Every decision tree in RF is built over a random extraction of features and considers a different set of features. Consequently, the trees remain uncorrelated with each other which makes them less prone to the problem of overfitting. Moreover, support vector machine (SVM) achieves a significant fall in the classification accuracy when implemented using radial kernel and polynomial kernel i.e. 25.49% and 18.58% respectively. On the other hand, the implementation of SVM with a linear kernel improves the accuracy of the model significantly (77.92%) which confirms the linearity of extracted features from the histopathological images. It is also observed that the implementation of classification networks with LR, linear discriminant analysis, Classification and Regression Tree (CART), and naïve bays (NB) classifiers drop the network performance considerably.

Additionally, no state-of-the-art method has been reported yet for the multi-classification of the breast cancer histopathological images that is independent of the magnification factor with an ability to handle the images of unknown magnification. Hence, in the scarcity and unavailability of the state-of-the-art method, we are unable to compare the results obtained from the present study. Despite that, we have compared the results with a study on magnification independent classification of the breast cancer images which was carried out by Bayramoglu et al. [16]. However, the conducted classification is a binary classification instead of a multi-classification. The authors proposed two different architectures to predict malignancy and the magnification level of the images. The average accuracy obtained with the proposed architectures are 81.87%, 83.39%, 82.56% and 80.69% at 40, 100, 200 and 400 X, respectively. It can be easily analysed from the results mentioned in Table 7 that the classification performance of our designed CNN is considerably better than the performance shown by the proposed architectures of Bayramoglu et al. for binary classification. Since the task of multi-classification is more tedious as compared to binary

### TABLE 6 Comparison of proposed CNN and the handcrafted approaches in multi-classification of BreakHis dataset

| Model                        | Approach     | Accuracy ± Standard deviation (40 X) |
|------------------------------|--------------|--------------------------------------|
| Logistic regression          | Handcrafted  | $71.02 \pm 2.52$                     |
| Linear discriminant analysis| Handcrafted  | $70.61 \pm 2.04$                     |
| Classification and regression tree | Handcrafted | $70.87 \pm 1.68$                     |
| Naïve bays                   | Handcrafted  | $45.25 \pm 2.05$                     |
| SVM                          | Handcrafted  | $77.92 \pm 3.04$                     |
| Random forest                | Handcrafted  | $79.47 \pm 1.38$                     |
| CNN                          | Deep learning| $80.47 \pm 1.42$                     |
classification, it is concluded from the comparison that our approach is superior to that devised by Bayramoglu et al. in [16].

5 | CONCLUSION

In this study, a general framework based on CNN is proposed to extract the features from the breast cancer histological images independent of their magnification. The proposed framework is more robust and faster due to the involvement of a single training step whereas other classification approaches utilise multiple training steps to recognise the images of specific magnification. Scalability is another important feature of the magnification independent model and provides an ability to handle the image with a new magnification. This is because images with new magnifications are utilised in the training process by just placing the images in the training set. Hence, it becomes very easy to generalise the proposed framework for the new images with different magnifications in contrast to previously trained data. Moreover, Adam optimiser is found as the most suitable optimisation technique as compared to AdaGrad, RMSProp, and gradient descent due to its better tuning capability. It is also noticed that the max-pooling operation works better than the average pooling and helps in enhancing the overall performance of the model by extracting adequate information related to contrast. Based on these facts, we can conclude that the proposed system is capable enough to classify the lobular carcinoma and mucinous carcinoma from the malignant class, and phyllodes tumor and fibroadenoma from the benign class which are the most complicated classes in BreakHis dataset with a maximum efficiency of 80.76% in case of 40 X magnification. The situation arises because of the high coherency among the tissues appearances which creates more confusion in these classes and leads to misclassification. Overall, the experimental results of the proposed CNN demonstrate the ability of the model in the efficient classification of eight classes of breast cancer.

To further enhance the classification accuracy, one can utilise a variety of weight initialisation techniques (Xavier, Lecun, and He) and advanced pooling regimes (stochastic, rank-based pooling, S3pool etc.) which could pave a leading step in improving the overall performance of computer-aided diagnosis system. Additionally, a significant increment in the training dataset could be achieved through different data-augmentation techniques which would open another aspect of this study.
20. Dhahri, H., et al.: Tabu search and machine-learning classification of benign and malignant proliferative breast lesions. Biomed. Res. Int. 2020(2), 1–10 (2020)
21. Kassani, S.H., et al.: Classification of histopathological biopsy images using ensemble of deep learning networks. In: CASCON ’19: Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, IBM Corp, Riverton, New Jersey, 92–99 (2019). arXiv:1909.11870
22. Sharma, S., Mehra, R.: Automatic magnification independent classification of breast cancer tissue in histological images using deep convolutional neural network. In: International Conference on Advanced Informatics for Computing Research, Shillong, India, 772–781 (2018). https://doi.org/10.1007/978-981-13-3140-4_69
23. Sharma, S., Mehra, R.: Breast cancer histology images classification: Training from scratch or transfer learning?. ICT Express 4(4), 247–254 (2018)
24. Fang, Y., et al.: Image classification toward breast cancer using deeply-learned quality features. J. Visual Commun. Image Represent. 64, 102609 (2019)
25. BreakHis Database. https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/
26. Sharma, S., Mehra, R.: Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image. Visual Comput, 36, 1755–1769 (2020)
27. Spanhol, F.A., et al.: Deep features for breast cancer histopathological image classification. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, Canada, 1868–1873 (2017). https://doi.org/10.1109/SMC.2017.8122889
28. Sharma, S., Mehra, R.: Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. J. Digital Imaging, 33(3), 632–654 (2020)
29. Sharma, S., Mehra, R.: Implications of pooling strategies in convolutional neural networks: A deep insight. Found. Comput. Decis. Sci. 44(3), 303–330 (2019)
30. Spanhol, F.A., et al.: Breast cancer histopathological image classification using convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada, 2560–2567 (2016). https://doi.org/10.1109/IJCNN.2016.772519
31. Yu, J., Bhargava, R., Dabbs, D.J.: Invasive lobular carcinoma with extracellular mucin production and HER-2 overexpression: A case report and further case studies. Diagn. Pathol. 5(1), 36 (2010)
32. Chao, T.C., et al.: Sonographic features of phyllodes tumors of the breast. Ultrasound Obstetrics Gynecol. 20(1), 64–71 (2002)

How to cite this article: Sharma S, Mehra R, Kumar S. Optimised CNN in conjunction with efficient pooling strategy for the multi-classification of breast cancer. IET Image Process 2021;15:936–946.
https://doi.org/10.1049/ipr2.12074