The Journal of Supercomputing (2021) 77:973–989
https://doi.org/10.1007/s11227-020-03321-y

A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model

Kadir Can Burçak[1] · **Ömer Kaan Baykan**[2] · **Harun Uğuz**[2]

Published online: 2 May 2020

Abstract
Deep learning algorithms have yielded remarkable results in medical diagnosis and image analysis, besides their contribution to improvements in a number of fields such as drug discovery, time-series modelling and optimisation methods. With regard to the analysis of histopathologic breast cancer images, the similarity of those images and the presence of healthy and tumourous tissues in different areas complicate the detection and classification of tumours on whole slide images. An accurate diagnosis in a short time is a need for full treatment in breast cancer. A successful classification on breast cancer histopathological images will overcome the burden on the pathologist and reduce the subjectivity of diagnosis. In this study, we propose a deep convolutional neural network model. The model uses various algorithms (i.e., stochastic gradient descent, Nesterov accelerated gradient, adaptive gradient, RMSprop, AdaDelta and Adam) to compute the initial weight of the network and update the model parameters for faster backpropagation learning. In order to train the model with less hardware in a short time, we used the parallel computing architecture with Cuda-enabled graphics processing unit. The results indicate that the deep convolutional neural network model is an effective classification model with a high performance up to 99.05% accuracy value.

Keywords Deep learning · Convolutional neural network · Breast cancer · Histopathology · Image classification

✉ Kadir Can Burçak
kcburcak@ahievran.edu.tr

Extended author information available on the last page of the article

1 Introduction

Deep learning methods are used in a variety of fields and have a significant computational performance on image classification, video analysis, speech recognition and natural language processing. The deep learning architectures comprise a combination of different layer types such as fully connected, convolutional and recurrent one. The multi-level deep neural networks (DNNs) are used to build systems that can detect features from big unlabelled training data [1]. Convolutional neural networks are the most successful model for image analysis up to now. This success results from the fact that computers extract the features best representing the data for the underlying problem [2]. Convolution and subsampling layers of DCNN do not require data-focused and properly decomposed cores, because they can extract image features automatically from a particular patch [3, 4]. Processing big data requires a set of powerful hardware and infrastructure [5]. Improvements in hardware processing units have facilitated the processing of huge amount of data and accelerated the research on DCNN [6]. Comparing to the thousands of computing cores and central processing unit (CPU), Cuda-enabled graphics processing unit (GPU) offers 10 to 100 times higher application performance [7]. The characteristic difficulties in classifying the medical images can be reduced to an optimal level by parallel computing [8, 9]. Therefore, specific DNN models using CPU and GPU in parallel are a need for machine learning.

Detection of breast cancer on digital histopathological images is a successful application area for deep learning algorithms. Different DCNN architectures have been proposed recently. Innovative architectural ideas and parameter optimisations show that DCNN performance can be boosted [6]. Medical image processing researchers have recently obtained promising results in this area [9, 10]. These results are generally based on specially cropped and labelled images, although the whole-slide histopathological images should be processed for a fully automatic computer-aided diagnosis. Nevertheless, processing the whole slide images in big sizes is difficult for a reasonable processing power and requires a long computational time. The presence of healthy or tumour tissues in different areas in different phases complicates the detection and classification of tumours on whole slide images.

In this study, we introduce a computer-aided system reducing the complexity of diagnosis. The system is based on a machine learning-based deep convolutional neural network (HCNN) model for an automated classification of breast cancer histopathological images as “benign” and “malignant”. The model consists of multi-convolution filters for the input layer. Using more than one feature of the multi-convolution filters improved the network performance. Multiple inputs were processed piece-by-piece using the mini-batch technique, and the batch normalisation was conducted for each mini-batch series. We compared the existing convolutional neural networks (CNNs) in the literature and evaluated the network performance. We analysed the behaviours of various algorithms to optimise the degradation of the model to improve the computational performance of HCNN. The model contributes to the literature with a high-performance deep learning network optimising the values of more than one features of the multi-convolution filters.

The remainder of this paper is organised as follows. A literature review positioning the research in the field is given in Sect. 2. The data set is described in Sect. 3.

Optimisation strategies of HCNN model are presented in Sect. 4. Experimental results are presented in Sect. 5. Finally, the main conclusions and future research directions are discussed in Sect. 6.

2 Literature review

In this section, we review the research papers on the detection of breast cancer on digital histopathology images in the field. Huh et al. [11] introduce a probabilistic model for microscopic images at two stages with the proposed method. The first stage entails the identification of spatio-temporal patch sequences, while the second one involves the localisation of a birth event. Albarqouni et al. [12] propose a network model (AggNet) from crowds for mitosis detection in breast cancer histology images. AggNet is a learning model that handles data aggregation directly as part of the learning process of the convolutional neural network. Bejnordi et al. [13] compared the metastases detected at the clinic with the results of deep learning-based systems. They report that 27.6% of the metastases are misdiagnosed by the pathologists, while the CNNs achieved better diagnostic performance. Filipczuk et al. [14] classify the feature vectors through a support vector machine with a set of 25 features, based on the analysis of cytological images of fine needle biopsies using the circular Hough transform. They achieve 98% effectiveness in classification of tumour cells. Xu et al. [15] present a Stacked Sparse Autoencoder (SSAE), an instance of a deep learning strategy, for efficient nuclei detection on high-resolution histopathological images of breast cancer. The SSAE model is based on just pixel intensities alone in order to identify distinguishing features of nuclei. The features obtained via the auto-encoder are divided into two, yielding an improved F-measure of 84%. Saha et al. [16] propose a supervised model to detect mitosis signature from breast histopathology WSI images, using deep learning architecture with handcrafted features. Their deep learning architecture mainly consists of five convolution layers. The model uses morphological, textural, intensity features and has 90% F-score. George et al. [17] proposed a diagnosis system for breast cancer based on the nuclei segmentation of cytological images, using different machine learning models such as neural networks and support vector machines. They report accuracy rates ranging from 76 to 94% on a data set of 92 images. Spanhol et al. [18] achieve a success rate of 90% at most by using BreakHis data set for histopathological images taken with different enlargement factors by classifying the cancer cells via deep learning methods. Han et al. [19] achieve an accuracy rate of 93.2% over BreakHis data set by using a different deep learning model recently proposed for grouping the breast cancer into multi-classes. Kausar et al. [20] proposed a binary and multiclass classification for breast histopathological images using a deep convolution neural network with Haar wavelet decomposed images. The proposed CNN model extracts and incorporates the deep features from 2-level Haar wavelet decomposed images. On BreakHis data set, this model showed an accuracy of 98.2% for both 4-class and 2-class recognition. Toğaçar et al. [21] proposed a new CNN model (BreastNet) for breast histopathological image classification using the BreakHis data set. The proposed model showed an accuracy of 98.8% for classification. Yang et al.

[22] proposed an ensemble of multi-scale convolutional neural networks (EMS-Net) to classify haematoxylin–eosin stained breast histopathological microscopy images into four categories, including normal tissue, benign lesion, in situ carcinoma, invasive carcinoma. They used the training patches cropped and augmented at each scale to fine-tune the pre-trained DenseNet-161, ResNet-152, and ResNet-101. They found that a combination of three fine-tuned models is more accurate than other combinations. The proposed EMS-Net model showed an accuracy of $91.75 \pm 2.32\%$ in the fivefold cross-validation using 400 training images. Budak et al. [23] introduced an end-to-end model (Bi-LSTM) based on fully convolutional network (FCN). They used the FCN as an encoder for feature extraction and turned the output of the FCN to a one-dimensional sequence. High-resolution images were thus used as direct input to the model. The proposed model achieved a performance of 98.10% through the fivefold cross-validation technique. Dabeer et al. [24] proposed a new CNN model for binary classification of breast histopathological images using the BreakHis data set. They obtained a prediction accuracy of up to 99.86%. Vo et al. [25] proposed an approach that utilises deep learning models with convolutional layers to extract the most useful visual features for breast cancer classification. This deep learning model is based on a novel boosting strategy to extract better features than handcrafted feature extraction approaches. The proposed model showed an accuracy of 99.5% for binary classification of breast histopathological images.

The literature review is summarised in Table 1 and the table clarify the literature contribution of this study.

Table 1 Literature summary on breast cancer detection approaches

Reviewed literature	Breast cancer detection approaches
Albarqouni et al. 2016	AggNet: deep learning from crowd
Bejnordi et al. 2017	Deep learning model
Han et al. 2017	Propose a breast cancer multi-classification method using a newly proposed deep learning model
Huh et al. 2011	CNN and handcrafted features
Malon et al. 2012	CNN and seeded blob features
Mitos-ATYPIA 2014	Deep CNN and max-pooling
Nahid et al. 2017	CNN and handcrafted features
Saha et al. 2018	Deep learning and handcrafted features
Spanhol et al. 2016	Parameter-free threshold adjacency statistics CNN
Srivastava et al. 2014	Convolutional neural network (CNN)
Xu et al. 2016	Stacked sparse autoencoder (SSAE)
Kausar et al. 2019	Decomposed image-based convolution neural network (HWDCNN)
Toğaçar et al. 2019	A novel convolutional neural network model (BreastNet)
Yang et al.2019	Ensemble of multiscale convolutional neural networks (EMS-Net)
Budak et al. 2019	Computer-aided diagnosis system combining FCN and Bi-LSTM model
Dabeer et al. 2019	CNN-based approach
Vo et al. 2019	Incremental boosting CNN
This research	Classification method using a newly proposed deep learning model

Accordingly, we propose a new convolutional neural network model (HCNN) that is to yield a higher accuracy rate by as well as to test and compare the gradient descent optimisation algorithms.

3 Data set

In this section, the experimental BreakHis data set is described [14]. BreakHis histopathologic database is made up of 9.109 microscopic images of breast tumour tissue enlarged in different rates (40×, 100×, 200× and 400×) collected from 82 patients. The distribution of 2480 benign and 5429 malign png format pictures with 3 channel RGB, 700×460 pixels in size and 8 bytes depth is shown in Table 2. This database is gathered incorporation with P & D Laboratory—Pathologic Anatomy and Sitopatology Parana Brasilia.

BreakHis data set divides tumours into two groups as benign tumours and malign tumours. Histopathologically benign tumours are lesions that do not overlap with any criteria of malignity. The benign tumours are comparatively innocent and slowly developing ones. Both tumour samples are presented in Fig. 1 as follows.

Table 2 Image distribution by magnification factor and class

Magnification	Number of benign pictures (24 patients)	Number of malign pictures (58 patients)	Total (82 patients)
40×	625	1370	1995
100×	644	1437	2081
200×	623	1.390	2013
400×	588	1232	1820
Total	2480	5429	7909

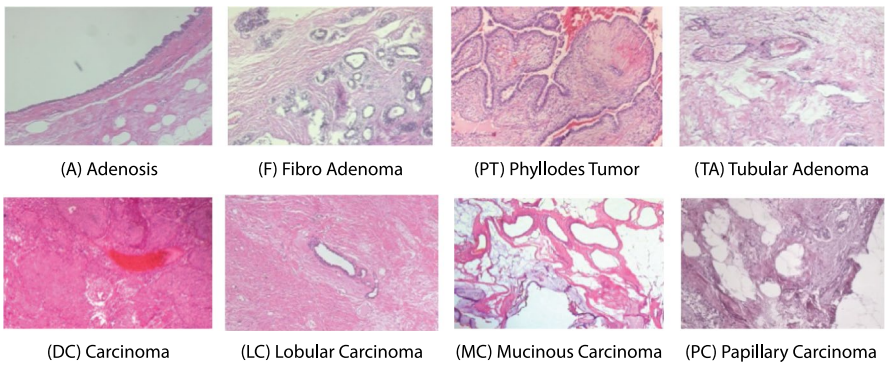

Fig. 1 BreakHis data set cluster images [14]—the first line for "normal" samples, the second line for "tumour" samples

4 Optimisation strategy of the deep convolutional neural networks

The inner parameters of the model play an important role to train the deep convolutional neural network (DCNN) models efficiently and to get correct results, as well as the parameters like layer number, neuron number, and dropout value. Parameters that change up to the model designer and according to the problem and data set are called hyperparameters. Therefore, we used various algorithms to compute and update the suitable and optimum model parameters that affect the training process and output of our model. These algorithms coordinate the forward extraction and feedbacks of a network with a solver and regulate the model optimisation for parameter updates, which improves the losses. Solver methods handle the loss minimisation problem of the general optimisation. In Eq. (1), D stands for the optimisation aim,$|D|$ for the average loss in all data conditions, and λ for controlling factor that makes trade-off between regularisation term (weight) and loss term [26, 27].

$$L(W) = \frac{1}{|D|} \sum_{i}^{|D|} fw\left(x^{(i)}\right) + \lambda r(w) \tag{1}$$

In Eq. (1), $fw(x^{(i)})$ is the loss in data sample $x^{(i)}$ and $r(w)$ is a regulation term in λ weight. $|D|$ may be very large (big), and hence, we practically use a stochastic zoom by drawing a mini-batch in Eq. (2) for the $N << |D|$ sample in each solver iteration [26].

$$L(W) \approx \frac{1}{N} \sum_{i}^{N} fw\left(x^{(i)}\right) + \lambda r(w) \tag{2}$$

The model computes fw in forward transition and gradient ∇fw in the backward transition. Parameter update is Δw, formed by the solver according to the error gradient ∇fw for each method, regulation gradient $\nabla r(W)$ and other details [26, 27].

4.1 Solvers used for the proposed model

Stochastic gradient descent (*Sgd*) is a simple but very effective approach to distinguish learning linear of classifiers under the convex loss functions such as logistic regression. Though *Sgd* has been used for a long time in learning software, lately it has been successfully applied to the large-scaled and dropout machine learning in which problems are frequently encountered as a part of large-scaled learning, text classification and language processing [27]. *Nesterov accelerated gradient* (*Nag*) is a technique to accelerate the standard momentum. Learning phase to optimise the deep neural networks is in direction to the reduction of the target all along the many gradient descent. *Nag* makes up a big jump towards the previous gradient. It measures the gradient descent and makes the correction after the update [26, 28]. *Adaptive Gradient* (*AdaGrad*) is an algorithm based on reduction. It applies the learning speed to the parameters. It is often used in the proximity of sparse data rather than the frequent one. One of the main advantages of adaptive gradient optimisation is

that it resolves the obligation of adjusting the learning rate. It minimises the learning speed in each dimension. It uses the default value (0.01) in many applications [26, 29]. *AdaDelta* is the extension of *AdaGrad*, which aims to decrease the learning rate in the gradients that are diminishing at the same level every time. *AdaDelta* limits the past gradients borders to specific dimensions. It doesn't keep the inefficient gradients. By taking the degradation average of all the previous gradients, it defines them as recursive [30]. *RMSprop* aims to decrease the learning rates that are diminishing in the same rate at the same time. It is generally suggested to adjust the momentum to 0.9 and learning speed to 0.01. For many times, this optimizer is a good choice for the recurrent neural network [29].

5 Experimental study

In this section, developing platforms and developing context are explained. The parameters of the chosen method suggested for each step are shown according to the experimental results. The results of the experimental studies are presented. The study has been conducted using one machine: Ubuntu 16.04 with Intel Core i7-6700 CPU @ 4.00 GHZ×8; NvidiaGeForce GTX 1080 8 GB; 32 GB DDR5 RAM and SSD hard disc on one platform. By installing the Python, OpenCV 3.0. and Pip with Caffe pane, CUDA 8.0 and cuDNN v6 drivers have been integrated to Caffe framework [31, 32]. By this way, the parallel training of the convolution neural network with GPU in Python context have been managed. The data set preparation for the convolution neural network has been done.

In this study, three different convolutional neural network architectures are trained and compared with each other. All of the trained CNN networks contain convolution and maximum pooling layers. Each model is explained in the sub-categories, and the suggested model's architecture is given as follows;

5.1 Input layer

The three models share the same input structure in 700×460×3 size in RGB format of BreakHis data set which are put into the network by turning them into randomly cropped in 256×256×3 size, mean filtered images. Training, verification and test sets were formed in order to test the performance of the models. The training set is randomly selected from the data set groups with 40×, 100×, 200×, 400×zoom rates (Fig. 2). The prepared training set is separately trained in the network according to the zoom rates with 20% verification and in 15% data set rates. The sample required for our network is prepared from the selected pictures with 40×zoom, 1398 training set, 399 verification and 198 tests with the total samples of 1995.

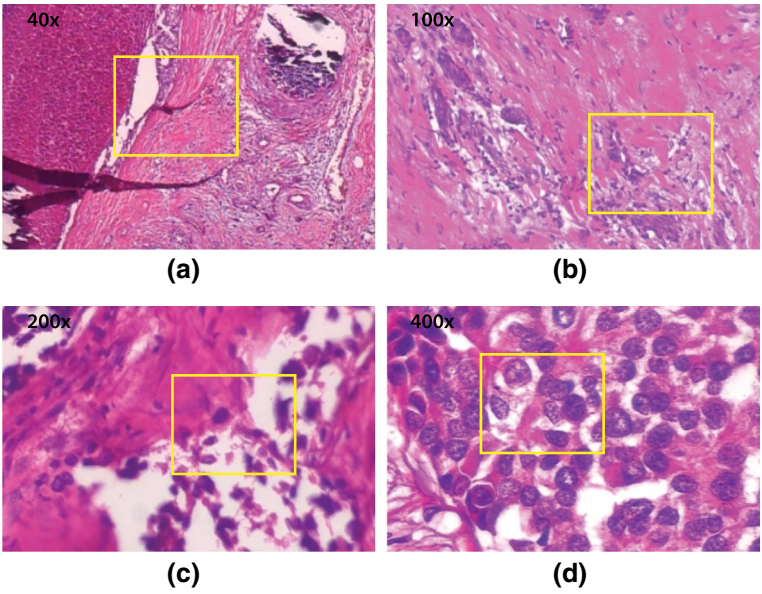

Fig. 2 An example of breast malign tumour seen in different enlargement factors of the same image **a** 40×, **b** 100×, **c** 200×, and **d** 400×

5.2 Convolution layer

The block diagram of the model is shown in Fig. 3. The output of the first convolution layer is max pooled, normalised and conveyed to the second convolution layer as input. The output of the second layer is first normalised and then max pooled. The input image with a 3×3 filter size and zero padding passes through the first convolution layer.

Following the first convolution layer, the pooling diagram is applied by making the input pass through the second convolution layer with a filter size of 3×3

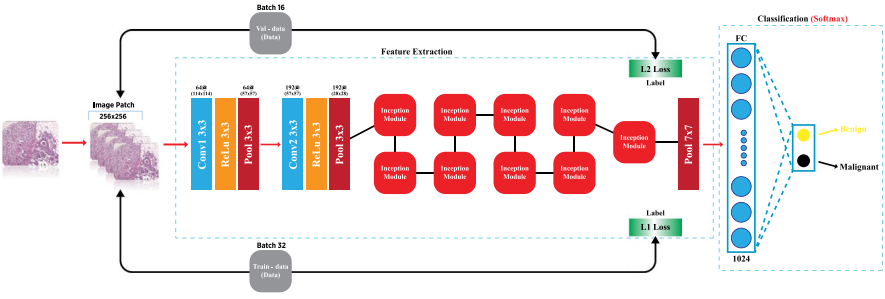

Fig. 3 Block diagram of the deep convolutional neural network model for an automated classification of breast cancer histopathological images

and at step 2. This process reduces the data size by 4 in width and length. The output of the first pool layer passes through the second convolution layer and the maximum pool layer once more. In order to catch the features in different scales, they use the start modules which hold various convolutional kernels in different sizes and gather the outputs along the depth dimension. L1 and L2 in Fig. 3 represent the regularisation of normalised softmax probabilities for classification. These regularisations induce using all, rather than some, of the network inputs. Table 3 summarises the parameters of the HCNN model in layers.

Module basically works as multi-convolution filters applied to the same input of pooling. Results are compiled to let utilise the multi-feature acquisition feature. As seen in Fig. 4, the passing of the network after the convolution layers have been given in detail.

For example, general (5×5) filter and local (1×1) filter acquire the features at the same time. Here, the optimisation of the network performance is provided by using more than one feature of the multi-filters. Before the beginning, all architectures have convolution together in spatial and on channel area. By applying convolution with (1×1) filter to the entry picture, entry block's spatial dimensions are disregarded. This is followed by the cross-spatial and cross-channel correlation composed with (3×3) and (5×5) filter. Hence, a feature vector is acquired in every phase. In order to create a probability distribution in the last phase of the model, Softmax is used to classify the input images.

Table 3 Detailed parameters of each layer

Layer name	Filter size	Padding	Output size	1×1	3×3	5×5
Conv1	3×3	3	[64 114 114]			
Pooling1	3×3	2	[64 57 57]			
Conv2	3×3	1	[192 57 57]		192	
Pooling2	3×3	2	[192 28 28]			
Inception1			[256 28 28]	64	128	32
Inception2			[480 28 28]	128	192	96
Inception3			[512 14 14]			
Inception4			[512 14 14]	192	208	48
Inception5			[512 14 14]	160	224	64
Inception6			[528 14 14]	128	256	64
Inception7			[832 14 14]	112	288	64
Pooling3	3×3	2	[832 7 7]	256	320	128
Inception8			[832 7 7]	256	320	128
Inception9			[1024 7 7]	384	384	128
Pooling4	7×7	1	[1024 1 1]			
Softmax			[1024]			

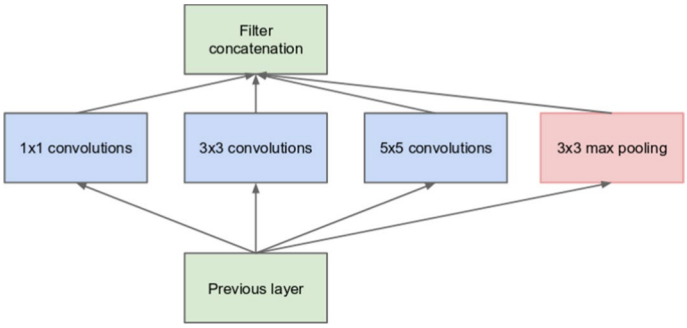

Fig. 4 Inception module

5.3 Experimental results and analysis

The HCNN architecture is trained via six different optimisation algorithms. The same data sets were trained and compared with Alex Krizhevsky's AlexNet model, the winner of ImageNet Large-Scale Visual Recognition Challenge 2012 (lLS-VRCI2) [33]. GoogLeNet became the winner of ILSVRC 2014. The top-5 error was 6.67%. The network utilised a CNN architecture inspired from LeNet. A new inception module was applied in that architecture. With tiny convolutions, this model significantly reduced the number of parameters [34]. Figure 5 demonstrates the evaluation scheme for the performance of the HCNN model.

Our evaluation criteria include precision-recall and confusion matrix. We tested 4 magnification factors for each model and 24 complexity matrices for 6 optimisation algorithms. In total, 72 complexity matrices were generated for 3 different DCNN models. As seen in Table 4, we computed true positive (TP), false positive (FP), true negative (TN), false negative (FN), as well as sensitivity, precision, specificity, and F1 score of the model with confusion matrix. The confusion matrices in Tables 4, 5, 6, and 7 denote the percentile representations of per-class accuracy of the other successful network models we trained. The results regarding the best performance our model attained are reported in detail in Table 8.

In this study, we classify the breast cancer histopathological images into two groups as benign and malign cancer cells. We train the HCNN, AlexNet and

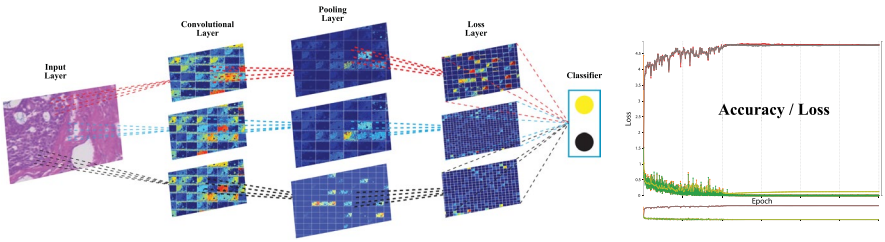

Fig. 5 Test-stage assessment scheme and the performance of the deep convolutional neural network model for an automated classification of breast cancer histopathological images

Table 4 Confusion matrix for test data according to HCNN-Nag and 200×enlargement factor

Confusion matrix	Benign	Malign	Per-class accuracy
Benign	89	4	95.7%
Malign	1	207	99.52%
Recall = sensitive	= TP/(TP + FN)	= 207/211	= 0.981
Precision	= TP/(TP + FP)	= 207/208	= 0.9951
Specificity	= TN/(TN + FP)	= 89/90	= 0.988
F1-score	$=2\times\frac{\text{precision}*\text{recall}}{\text{precision}+\text{recall}}$	= 2 (0.9761/1.9761)	= 0.9879

Table 5 Confusion matrix for test data according to HCNN-Sgd and 40×enlargement factor

Confusion matrix	Benign	Malign	Per-class accuracy (%)
Benign	88	6	93.62
Malign	2	203	99.02

Table 6 Confusion matrix for test data according to HCNN-Sgd and 100×enlargement factor

Confusion matrix	Benign	Malign	Per-class accuracy (%)
Benign	85	11	88.54
Malign	3	213	98.61

Table 7 Confusion matrix for test data according to HCNN-Nag and 400×enlargement factor

Confusion matrix	Benign	Malign	Per-class accuracy (%)
Benign	83	5	88.54
Malign	4	181	98.61

GoogLeNet models via original data sets and directly compared the alternate networks. In order to train the three models, we used six different optimisation algorithms which produced different results as reported in Table 9. The *RMSprop* optimisation algorithm was not able to ensure that the gradient descent was lost in all trainings, thus causing overlearning, in other words, the network memorisation. It resulted from the fact that the learning coefficient was determined in accordance with the network's gradient descent value. The *AdaGrad* optimisation algorithm is similar to the *RMSprop* optimisation algorithm; the only difference is the calculation of the total gradient descent in the former. Based on this, we can say that the methods for calculating the gradient descents cause considerable changes in the network performance. The *Nag* optimisation algorithm has been successful in all the trained networks. We can conclude that this success positively affects the network

Table 8 Sensitivity, specificity, precision, and F1-score performance for the HCNN model four enlargement factors

	Method	HCNN-Sgd	HCNN-Nag	HCNN-AdaGrad
40×	*Sensitivity*	**0.971**	**0.935**	0.926
	Specificity	**0.977**	**0.975**	0.951
	Precision	**0.990**	**0.990**	0.980
	F1	**0.980**	**0.962**	0.952
100×	*Sensitivity*	**0.950**	0.945	0.949
	Specificity	**0.965**	0.933	0.914
	Precision	**0.986**	0.972	0.963
	F1	0.968	**0.958**	**0.956**
200×	*Sensitivity*	0.980	**0.981**	**0.971**
	Specificity	0.936	**0.988**	**0.966**
	Precision	0.971	**0.995**	**0.985**
	F1	0.975	0.988	0.978
400×	*Sensitivity*	**0.988**	0.973	0.946
	Specificity	**0.905**	0.954	0.917
	Precision	**0.951**	0.978	0.962
	F1	**0.969**	0.975	0.954

Bold values signify the highest values of HCNN

Table 9 Per-class accuracy performance for the HCNN, AlexNet and GoogLeNet model of four enlargement factors

Methods	Accuracy of the enlargement factors			
	40×	100×	200×	400×
Hcnn + Rms + Softmax	0.74	0.76	0.87	0.68
Hcnn + Adam + Softmax	0.69	0.69	0.69	0.92
Hcnn + Adadelta + Softmax	0.92	0.91	0.91	0.92
Hcnn + Adagrad + Softmax	0.94	0.95	**0.97**	0.95
Hcnn + Nag + Softmax	0.96	**0.99**	**0.97**	0.96
Hcnn + Sgd + Softmax	**0.97**	**0.97**	**0.96**	**0.96**
AlexNet + Rms + Softmax	0.70	0.69	0.69	0.69
AlexNet + Adam + Softmax	0.70	0.69	0.69	0.69
AlexNet + Adadelta + Softmax	0.86	0.85	0.91	0.88
AlexNet + Adagrad + Softmax	0.82	0.81	0.88	0.88
AlexNet + Nag + Softmax	0.94	0.92	**0.96**	0.94
AlexNet + Sgd + Softmax	0.94	0.93	**0.96**	0.95
GoogLeNet + Rms + Softmax	0.76	0.76	0.87	0.69
GoogLeNet + Adam + Softmax	0.69	0.69	0.69	0.96
GoogLeNet + Adadelta + Softmax	0.92	0.90	0.90	0.91
GoogLeNet + Adagrad + Softmax	0.94	0.96	0.96	0.95
GoogLeNet + Nag + Softmax	0.96	**0.97**	0.96	0.96
GoogLeNet + Sgd + Softmax	**0.97**	0.96	**0.98**	**0.97**

Bold values signify the highest values of HCNN, AlexNet and GoogLeNet

performance by preventing excessive diminution while reducing the error function. Although *AdaDelta* is similar to *RMSprop*, it shows higher performance by receiving a certain number, rather than all, of the priority gradient descents. The self-determination of learning coefficient may positively affect its performance as well. *Adam* optimisation algorithm is not as successful in all trainings as *RMSprop*. Based on the common features of these two optimisation algorithms, taking the exponentially weighted averages of the squares of the past gradient descents at each iteration negatively affected the performance of the networks we tested. We reached this result in consideration of the fact that other *AdaGrad* optimisation algorithms similar to these optimisation algorithms showed high performance by calculating the total past gradient descents.

The HCNN model we propose was trained via six different optimisation algorithms with data sets separately created according to the magnification factors of 40×, 100×, 200×, 400×. The data set created according to the 40× magnification factor achieved a performance of 97% on the HCNN with *Sgd* optimisation algorithm. The data set created according to the 100× magnification factor achieved a performance of 99% on the HCNN with *Nag* optimisation algorithm. The data set created according to the 200× magnification factor achieved the same performance (97%) on the HCNN with *Nag* and AdaGrad optimisation algorithms. The data set created according to the 400× magnification factor achieved the same performance (96%) on the HCNN with *Nag* and *Sgd* optimisation algorithms. *RMSprop* and *Adam* optimisation algorithms stuck around a certain performance value in the network training. Regarding the training of other models, optimisation algorithms were tested on separate networks and their effects were observed as seen in Table 9. *RMSprop* ve *Adam* optimisation algorithms showed lower performance in the training of these models as well. Per-class accuracy performances for the HCNN and AlexNet architectures according to four enlargement factors are reported in Fig. 6.

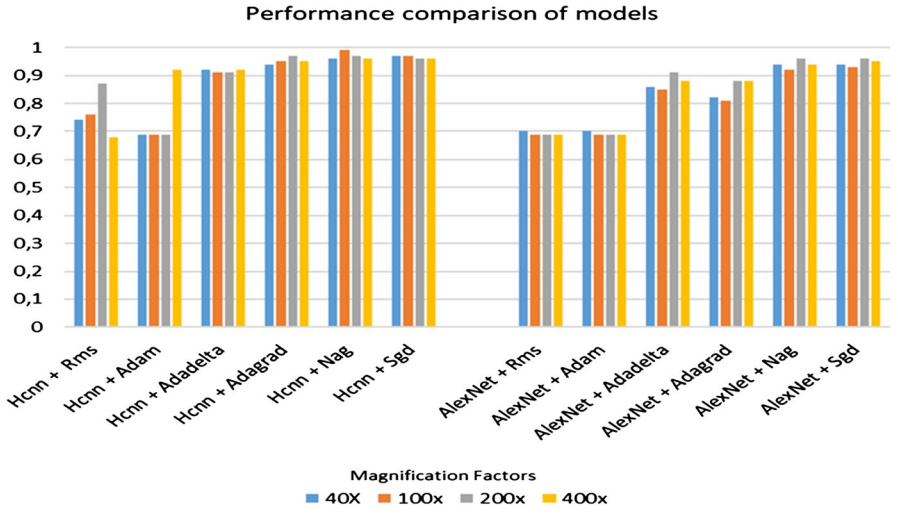

Fig. 6 Comparison of model performances

Table 10 Comparative analysis of the results in the related literature

Category	Breast cancer detection approaches	Data set	Accuracy
Texture based and colour based	Filipczuk et al. [14]	FNB-ZGBH	0.975
	Huh et al. [11]	–	0.89
Deep learning	Saha et al. [16]	MITOS-ATYPIA-14	0.92
	Bejnordi et al. [13]	CAMELYON16	0.96
	Spanhol et al. [18]	BREAKHIS	0.90
	Malon and Cosatto [36]	ICPR 2012	0.76
	Albarqouni et al. [12]	MICCAI-AMIDA13	0.86
	Han et al. [19]	BREAKHIS	0.932
	Xu et al. [15]	–	0.84
	Nahid et al. [35]	BREAKHIS	0.96
	Yang et al. [22]	BREAKHIS	91.75 ± 2.32
	Budak et al. [23]	BREAKHIS	0.981
	Vo et al. [25]	BREAKHIS	0.995
	Dabeer et al. [24]	BREAKHIS	0.9986
	Toğaçar et al. [21]	BREAKHIS	0.988
	Kausar et al. [20]	BREAKHIS	0.982

Our HCNN model yields promising results for the binary classification. In the related studies as seen in Table 10, Filipczuk et al. [14] obtained a performance of 97.5–98% in classifying tumour cells with 25-dimension feature vector with a different data set. However, our model has an advantage, as feature extraction via fine needle biopsy is a costly method in medical imaging. Han et al. [19] propose a new model for grouping breast cancer into multi-classes with BreakHis data set. They succeed by 93% and 96%, respectively, for multi-class classification and binary classification. With BreakHis data set again, Nahid et al. [35] achieve an accuracy rate of 94.40–97.19%, a precision of 98% according to $100\times$ and $200\times$ enlargement factors, a maximum recall value of 98.2%, and an F1 score of 98% with $200\times$ enlargement factor. Our model increases the accuracy rate of pathologic diagnosis to shorten the decision-making process, as well as to help minimise unnoticed cancer cells and get a quicker diagnosis.

6 Conclusion

In this paper, we introduced a DCNN-based model for automatic detection and classification of cancerous regions on histopathological images taken from breast cancer patients, using BreakHis data set. The large size of images and the large amount of training data, which are among the major problems in medical image analysis, complicate learning in terms of building the systems with the ability of feature detection. In order to solve this problem, we divided the data set into two small groups and carried out the process of learning on the mini batches. Besides, we conducted the

batch normalisation for each mini-batch series. We used a Cuda-enabled GPU to train the model in a shorter time with less hardware. Another problem in designing deep learning models is the selection of network parameters and their congruence with the network. Solving problems via deep learning is equivalent to designing the multi-layered network structure in an optimal way. In the proposed model, we tested the different optimisation algorithms that could accelerate the learning process and perform better in terms of cost function. We proved that the optimisation algorithms produce different results in accordance with different problems and data clusters in the DCNN architectures. It is possible to conclude that the HCNN model achieves the best performance (% 99.05).

Compliance with ethical standards

Conflict of interest There is no conflict of interest in this study.

References

1. Sudharshan P, Petitjean C, Spanhol F, Oliveira LE, Heutte L, Honeine P (2019) Multiple instance learning for histopathological breast cancer image classification. Expert Syst Appl 117:103–111
2. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
3. Wahab N, Khan A, Lee YS (2017) Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. Comput Biol Med 85:86–97
4. Wahab N, Khan A, Lee YS (2019) Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images. Microscopy 68(3):216–233
5. Qin C, Yao D, Shi Y, Song Z (2018) Computer-aided detection in chest radiography based on artificial intelligence: a survey. Biomed Eng Online 17(1):113. https://doi.org/10.1186/s12938-018-0544-y
6. Khan A, Sohail A, Zahoora U, Qureshi AS (2019) A survey of the recent architectures of deep convolutional neural networks. arXiv preprint arXiv:190106032
7. Tan Y, Sim K, Ting F (2017) Breast cancer detection using convolutional neural networks for mammogram imaging system. In: 2017 International Conference on Robotics, Automation and Sciences (ICORAS). IEEE, pp 1–5. https://doi.org/10.1109/ICORAS.2017.8308076
8. Upasani N, Om H (2019) A modified neuro-fuzzy classifier and its parallel implementation on modern GPUs for real time intrusion detection. Appl Soft Comput 82:105595
9. Zhou SK, Greenspan H, Shen D (2017) Deep learning for medical image analysis. Academic Press, New York
10. Kleesiek J, Biller A, Urban G, Kothe U, Bendszus M, Hamprecht F (2014) Ilastik for multi-modal brain tumor segmentation. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp 12–17
11. Huh S, Ker DF, Bise R, Chen M, Kanade T (2011) Automated mitosis detection of stem cell populations in phase-contrast microscopy images. IEEE Trans Med Imaging 30(3):586–596. https://doi.org/10.1109/TMI.2010.2089384
12. Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N (2016) AggNet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans Med Imaging 35(5):1313–1321. https://doi.org/10.1109/TMI.2016.2528120
13. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JA, Hermsen M, Manson QF, Balkenhol M, Geessink O (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318(22):2199–2210. https://doi.org/10.1001/jama.2017.14585
14. Filipczuk P, Fevens T, Krzyzak A, Monczak R (2013) Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. IEEE Trans Med Imaging 32(12):2169–2178. https://doi.org/10.1109/TMI.2013.2275151

15. Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J, Madabhushi A (2016) Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Trans Med Imaging 35(1):119–130. https://doi.org/10.1109/TMI.2015.2458702
16. Saha M, Chakraborty C, Racoceanu D (2018) Efficient deep learning model for mitosis detection using breast histopathology images. Comput Med Imaging Graph 64:29–40. https://doi.org/10.1016/j.compmedimag.2017.12.001
17. George YM, Zayed HH, Roushdy MI, Elbagoury BM (2014) Remote computer-aided breast cancer detection and diagnosis system based on cytological images. IEEE Syst J 8(3):949–964. https://doi.org/10.1109/Jsyst.2013.2279415
18. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) A dataset for breast cancer histopathological image classification. IEEE Trans Biomed Eng 63(7):1455–1462. https://doi.org/10.1109/TBME.2015.2496264
19. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S (2017) Breast cancer multi-classification from histopathological images with structured deep learning model. Sci Rep 7(1):4172. https://doi.org/10.1038/s41598-017-04075-z
20. Kausar T, Wang M, Idrees M, Lu Y (2019) HWDCNN: multi-class recognition in breast histopathology with Haar wavelet decomposed image based convolution neural network. Biocybern Biomed Eng 39(4):967–982
21. Toğaçar M, Özkurt KB, Ergen B, Cömert Z (2020) BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. Phys A Stat Mech Appl 545:123592
22. Yang Z, Ran L, Zhang S, Xia Y, Zhang Y (2019) EMS-Net: ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. Neurocomputing 366:46–53
23. Budak Ü, Cömert Z, Rashid ZN, Şengür A, Çıbuk M (2019) Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images. Appl Soft Comput 85:105765
24. Dabeer S, Khan MM, Islam S (2019) Cancer diagnosis in histopathological image: CNN based approach. Inform Med Unlocked 16:100231
25. Vo DM, Nguyen N-Q, Lee S-W (2019) Classification of breast cancer histology images using incremental boosting convolution networks. Inf Sci 482:123–138
26. Yangqing Jia ES (2018) Berkeley artificial intelligence research. https://caffe.berkeleyvision.org/tutorial/solver.html
27. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791
28. Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning. In: International Conference on Machine Learning, pp 1139–1147
29. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv preprint arXiv:160904747
30. Dozat T (2016) Incorporating Nesterov momentum into Adam. ICLR Workshop (1):2013–2016
31. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. ACM, pp 675–678
32. Jia Y, Darrell T (2011) Heavy-tailed distances for gradient based image descriptors. In: Advances in Neural Information Processing Systems, pp 397–405
33. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp 1097–1105
34. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9
35. Nahid A-A, Kong Y (2017) Local and global feature utilization for breast image classification by convolutional neural network. In: 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, pp 1–6
36. Malon CD, Cosatto E (2013) Classification of mitotic figures with convolutional neural networks and seeded blob features. J Pathol Inform 4:9. https://doi.org/10.4103/2153-3539.112694

Affiliations

Kadir Can Burçak[1] · Ömer Kaan Baykan[2] · Harun Uğuz[2]

Ömer Kaan Baykan
okbaykan@ktun.edu.tr

Harun Uğuz
huguz@ktun.edu.tr

[1] Computer Science Research and Application Center, Ahi Evran University, Kırşehir, Turkey

[2] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Konya Technical University, Konya, Turkey