Computers in Biology and Medicine 150 (2022) 106155

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images

Soham Chattopadhyay [a], Arijit Dey [b], Pawan Kumar Singh [c,*], Diego Oliva [d], Erik Cuevas [d], Ram Sarkar [e]

[a] Department of Electrical Engineering, Jadavpur University, 188, Raja S.C. Mallick Road, Kolkata 700032, West Bengal, India
[b] Department of Computer Science and Engineering, Maulana Abul Kalam Azad University of Technology, Kolkata 700064, West Bengal, India
[c] Department of Information Technology, Jadavpur University, Jadavpur University Second Campus, Plot No. 8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata 700106, West Bengal, India
[d] División de Tecnologías para la Integración Ciber-Humana, Universidad de Guadalajara, CUCEI, Av. Revolución 1500, 44430, Guadalajara, Jal, Mexico
[e] Department of Computer Science and Engineering, Jadavpur University, 188, Raja S.C. Mallick Road, Kolkata 700032, West Bengal, India

ARTICLE INFO

Keywords:
Multi-scale features
Dual residual network
Recurrent connection
Breast cancer classification
Deep learning
Histopathological image

ABSTRACT

Histopathological image classification has become one of the most challenging tasks among researchers due to the fine-grained variability of the disease. However, the rapid development of deep learning-based models such as the Convolutional Neural Network (CNN) has propelled much attentiveness to the classification of complex biomedical images. In this work, we propose a novel end-to-end deep learning model, named Multi-scale Dual Residual Recurrent Network (MTRRE-Net), for breast cancer classification from histopathological images. This model introduces a contrasting approach of dual residual block combined with the recurrent network to overcome the vanishing gradient problem even if the network is significantly deep. The proposed model has been evaluated on a publicly available standard dataset, namely BreaKHis, and achieved impressive accuracy in overcoming state-of-the-art models on all the images considered at various magnification levels.

1. Introduction

Breast cancer is one of the most diagnosed diseases among women across the globe, almost one out of four cases are accounted for [1]. Day by day, it has been affecting women at a higher rate. According to a study, in the USA, breast cancer is the second cause of death among other types of cancer in women. However, almost one out of eight (almost 12%) women in the USA has developed an invasive breast over their entire life span. In the year 2020, over 2.2 million new cases of breast cancer were found which is almost 24.5% of cancer cases among women. Moreover, pet animals, principally dogs have also become victims of breast cancer. Since dogs are more vulnerable, studies found in Refs. [2,3] show that canine mammary tumor (CMT) happens to have more than three times higher mortality rates than that of humans. However, breast cancer has a higher rate of treatable malignancies if it is detected at an early stage. The most common way to diagnose cancerous cells is in the tissue biopsy, where a specialized needle is used as a screening device to extract the core of tissue from the suspicious area, which generates histopathological images at different magnification levels [4]. Proper analysis of those images can be very much helpful to detect cancerous cells [5]. However, the manual process of analyzing those images is a time-consuming process and accuracy may vary depending on the experience of the medical practitioners [6,7].

Therefore, it is worth mentioning that histopathological image analysis is considered to be one of the most standard tools for cancer detection for both humans and canines. In this process, experts try to find if there is any sign of cancer in the imperceptible regions of Hematoxylin & Eosin (H&E) tissues, which can be specifically analyzed by the histopathological characteristics. Additionally, if breast cancer develops, various changes occur in normal structures of the parenchyma, which are also taken into consideration during the extensive diagnosis process. However, the emergence of the Digital Image Scanner in the field of pathology has made the diagnosis and identification of cancer more effective. Given this, digital pathology has been explored by researchers and significant progress has been made owing to the low cost and fast diagnosis nature of whole slide image (WSI) scanners [8]. However, the H&E stained diagnosis happens to be very time-consuming, and extensive effort from skilled experts is required. In addition to

* Corresponding author.
E-mail addresses: chattopadhyaysoham99@gmail.com (S. Chattopadhyay), arijjitdey3413@gmail.com (A. Dey), pawansingh.ju@gmail.com (P.K. Singh), diego.oliva@cucei.udg.mx (D. Oliva), erik.cuevas@cucei.udg.mx (E. Cuevas), raamsarkar@gmail.com (R. Sarkar).

https://doi.org/10.1016/j.compbiomed.2022.106155
Received 7 June 2022; Received in revised form 31 August 2022; Accepted 24 September 2022
Available online 30 September 2022

this, Elmore et al. [9] show that the manual analysis-based diagnosis often suffers immensely from the frequently occurring changeability in inter-observers, which results in 75% diagnostic harmony between the experts.

To overcome the above limitation, researchers across the world have been trying hard to develop various computer-aided diagnosis-based systems, which reduce the prognosis time as well as achieve impressive results in a far more rapid way. Furthermore, it also effectively improves classification accuracy by achieving lesser inter-variability in the interpretations. The studies presented in [10] show that machine learning algorithms can achieve lesser errors than that of a single pathologist. Recently, it has been seen that computer-aided detection tools are used at a large scale to automate the diagnosis of cancerous cells from histopathological images. Besides, the rapid usage of artificial intelligence (AI) in the domain of biomedical image processing results in the development of many precise and efficient systems [11,12]. Due to the ever-growing importance of the automated healthcare system in our daily life, many researchers around the world have been striving to make such systems almost error-free with the application of either machine learning or a deep learning-based model. Therefore, various models, especially deep learning techniques can be found in the literature which accomplish such tasks in a far more accurate way. The advances and emergence of various such AI-based algorithms in the said field of research, along with their advantages and disadvantages, are discussed in the following section.

In this paper, we develop a novel deep learning framework, called Multi-scale Twofold Residual Recurrent Network (MTRRE-Net), for breast cancer detection from histopathology images. We have chosen the BreakHis dataset [10] to evaluate the model and report results by conducting various experiments. As the BreakHis dataset is well known for its fine-grained appearances, easy classification is not possible. To deal with this problem, it is necessary to emphasize the local features to learn the complex patterns, which boosts classification accuracy. To address this problem, we have proposed a novel deep learning model which utilizes the advantages of residual skip connections [13], feature learning from multiple scales of the images [14] and residual recurrent connections [15], thereby producing an efficient model. These components help to increase feature delivery and implementation of residual skip connection helps to merge shallow features with deep features with higher complexity. In a nutshell, the contributions of this paper are given as follows:

- A novel two-fold residual recurrent operation has been proposed, which efficiently deals with ***Degradation problem*** despite having noticeably deep architecture.
- The features are extracted by combining the two-fold residual recurrent operation with multi-scale feature learning.
- The proposed model has also shown better performance on significantly small datasets.
- Further, the obtained results are compared with different variants of our proposed MTRRE-Net model and also with some state-of-the-art models which report benchmark results on the BreakHis dataset. This is done to ensure that despite having such a simple architecture, the results of the proposed model are quite comparable to benchmark results.

The remainder sections are organized as follows: Section 2 presents the related work whereas Section 3 highlights the dataset used in the present work. In Section 4, we describe our proposed MTRRE-Net model. Section 5 presents the detailed evaluation of the proposed model by performing various experiments. Finally, Section 6 discusses some conclusions.

2. Related work

In the domain of computer vision, relevant pattern learning in an end-to-end manner has become one of the prime focuses of researchers in recent times. Research on histopathology images is not an exception as well. For example, Curz-Roa et al. [16] proposed a CNN model, which consists of three fully-convolution layers and evaluated on 100 × 100 patches of pixels at a magnification level of 2.5 and reported promising results. However, such a shallow network may work well on a smaller dataset but there is a high chance of failure of such a network while learning more complex patterns. In contrast, several deep networks have also been proposed in the field of medical image classification [17–20]. However, in many cases, their effectiveness is limited to the size of datasets, as it is also hard to develop deep nets that can perform promisingly well on small datasets. Moreover, having the class imbalance problem in medical image datasets is also a significant reason for the failure of several deep neural networks. To address both the aforementioned issues, in the papers [21–25] the authors used various types of generative adversarial networks (GANs) [26] to augment the data up to some desired level, and then classified with the aid of various deep learning-based classifiers. However, one problem associated with such approaches is the time required gets increased significantly for training such end-to-end models. Two deep networks (one is the GAN model [27] used to augment the data and the other one is the deep neural classifier used for the classification purpose) have to be trained separately for such cases, which increases the consumed time of training and the resource requirements as well. However, Wang et al. [28] incorporated a self adaptive mechanism into Extreme Learning Machine (ELM) to achieve faster learning speed and generalization problem during classification. Additionally, when we deal with some high-resolution histopathological images, the computational cost via deep learning gets higher. To reduce this problem XU et al. [29] developed a selective attention-based approach that identifies possible lesion regions from the original image instead of the whole image for the classification of breast cancer. Also, Yang et al. [30] introduced a guided soft attention network to predict the region of interest (ROI) for breast cancer classification. Zheng et al. [31] proposed a whole-slide histopathological image analysis technique that locates high-risk cancer regions by using probability mapping. Additionally, use of probabilistic algorithms [32,33] helped to reduce miss-prediction and improve the result of neural networks.

Another efficient and simpler way to achieve good classification results is by doing deep transfer learning [34,35], where the weights of a standard model, trained on very large datasets (like ImageNet) are transferred and fine-tuned to execute similar tasks on a relatively smaller dataset. Since transfer learning often brings promising results with fewer epochs of training, this has become very popular and adopted by many computer science researchers for developing various image classification problems including histopathological images [36–39]. Here we have briefed some of the recent models used for disease detection. Garud et al. [40] proposed a simple deep learning model considering pre-trained GoogleNet to classify benign and malignant cancers using high magnification multi-viewed samples. Litjens et al. [41] proposed a CNN-based model to increase the accuracy of pathological images using a patch extraction method. Further, Xie et al. [42] conducted experimentation including both supervised and unsupervised learning approaches with some pre-trained CNNs and come to the conclusion that in supervised learning pre-trained Inception ResNet V2 performs well in a comparison of the other transfer learning models they have considered. However, the fundamental problem that comes with transfer learning is the availability of a few such standard models and their variants, therefore the classification techniques become limited to some models only.

Apart from these traditional deep learning techniques, current computer vision researchers have shown a keen interest on developing several types of ensemble models by considering some standard deep neural networks. In such techniques, the probability distributions (confidence scores) of the test predictions of different trained deep learning classifiers are given as inputs to the ensemble algorithms to get the fine-tuned classification results. Some popular ensemble techniques are sum rule and product rule to mention some. Another type of

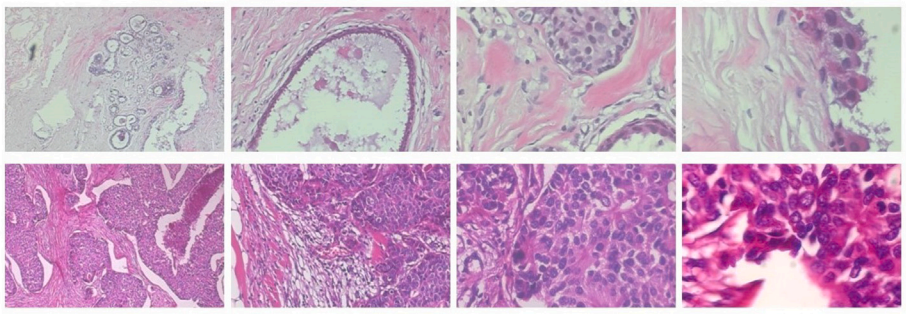


Fig. 1. Examples of breast cancer screening examinations. First row: both breasts with benign findings (40x, 100x, 200x, 400x respectively); second row: left breast with malignant findings (40x, 100x, 200x, 400x respectively).

Table 1
Image distribution by magnification factor and class of BreakHis dataset.

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

ensemble approach is extracting deep features from the pre-final layers of various finely tuned CNNs and concatenating them to form the final feature set, which is further fed to some machine learning classifiers like Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and so on for final predictions. Needless to say, the domain of medical imaging including histopathology images is also seen to benefit from the various ensemble techniques mentioned above. For example, S. H. Kassani et al. [43] extracted deep features from various CNNs: VGG16, MobileNet, and DenseNet, and thereafter the classification is done on the combined feature set by multi-layer perceptron (MLP) classifier. Z. Hameed et al. [44] performed an ensemble of VGG16 and VGG19 models by taking the average of the prediction probabilities of both of the models to generate the final prediction probability and the class having maximum confidence score is decided to be the predicted class.

3. Dataset description

BreaKHis is one of the popular public available datasets which consists of microscopic biopsy images of benign and malignant breast tumors [10]. The collection was done in the P&D Laboratory (Brazil), where all the patients were asked for the clinical detection of breast cancer. The study was approved by the institutional review board and all the participants had given their written consent. Every data sample is anonymized. Samples were generated from the breast tissue biopsy slides and stained with hematoxylin and eosin (HE). The samples were collected by surgical open biopsy (SOB), prepared for histological study, and labeled by pathologists in the lab. The diagnosis of each case was made by experienced pathologists and confirmed by complementary examinations such as immunohistochemistry analysis. Images were acquired in RGB color space, with a resolution of 752×582 using magnifying factors of 40×, 100×, 200×, and 400×. Fig. 1 shows these four magnifying factors of benign and malignant sample images. To date, the database is composed of 7909 images divided into benign and malignant tumors. Table 1 summarizes the image distribution.

It is a common practice to apply standard data augmentation techniques to increase the training images during the training phase of a CNN. Here in our case, we have used Gaussian blur with a certain kernel value, which randomly blurs images with Gaussian blur. By simply looking at the histopathological images of the dataset under consideration, we can say that they are not invariant to rotation. Therefore, we apply horizontal flip, vertical flip, and random rotation of the images at 45 degrees to augment the data.

It is to be noted that all the images are normalized before diving into any kind of experiments mentioned later. We have applied a contrast normalization step which can be considered as a pre-processing step followed in this paper.

4. Proposed MTRRE-Net

This article aims to develop a deep learning-based architecture that focuses on both learning and fusing prolific low-dimensional features obtained from the deeper section of the network while using less viable global features from the shallower end of a deep network in an efficient way. The more depth of the network, the more it learns complex and significant patterns of the input image. Therefore, the extraction of features from various depths of the network with different scales and different filter sizes, makes the feature set replete with various significant, complementary, and abundant information about the input image. However, there has always been a trade-off between depth and the successful learning of the network. A shallow network often fails to learn different complex patterns, and hence fails to extract salient features from the input images. On the other hand, for sufficiently deep networks, some other problems may exist such as vanishing gradient, overfitting on small datasets, loss of geometric information, etc. The weights and the biases of the initial layers of the network sometimes remain almost unchanged even after sufficient iterations of training, which is known as the vanishing gradient problem. Besides, if multi-scale feature extraction is performed from different layers, irrelevant and redundant features, mostly from the shallower section, might cause performance degradation of the network. However, if the network is kept shallower, then various significant features may not be retrieved properly which, in turn, might result in poor performance of the network. To address these problems, we introduce two-fold residual learning, which takes care of both the vanishing gradient problem as well as the recurrent connection which aids in maintaining the proficient information flow throughout the entire architecture. Due to such configuration, the proposed network can be trained with a comparatively smaller number of training samples, since the chances of occurring of the vanishing gradient and the overfitting problems are minimized even if the architecture is sufficiently deep (with more than 50 layers).

4.1. Network architecture

As shown in Fig. 2, our proposed MTRRE-Net model has three main parts: the root architecture, the multi-scale feature extractor, and the classifier. The root architecture consists of a shallow feature learner via convolution layers and a deep feature learner by two-fold residual recurrent operation. As Fig. 2 shows the basic building block of the MTRRE-Net model, which mainly consists of the multi-scale twofold residual recurrent groups (MTRRGs).

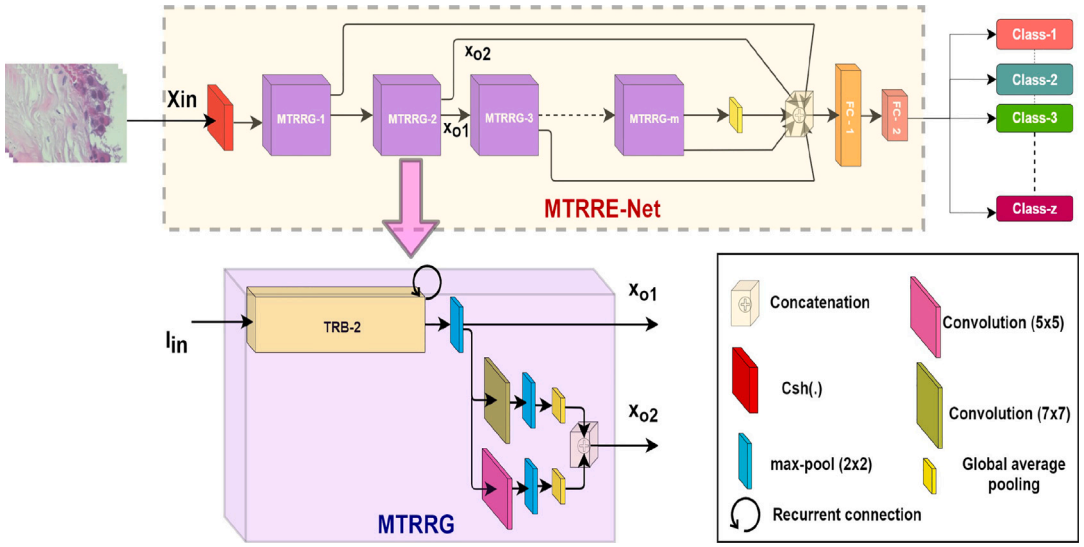


Fig. 2. Flowchart of the proposed model, called MTRRE-Net, used for breast cancer identification from histopathological images. The detailed description of the two-fold residual block (TRB) and the residual recurrent connection are illustrated in Fig. 3.

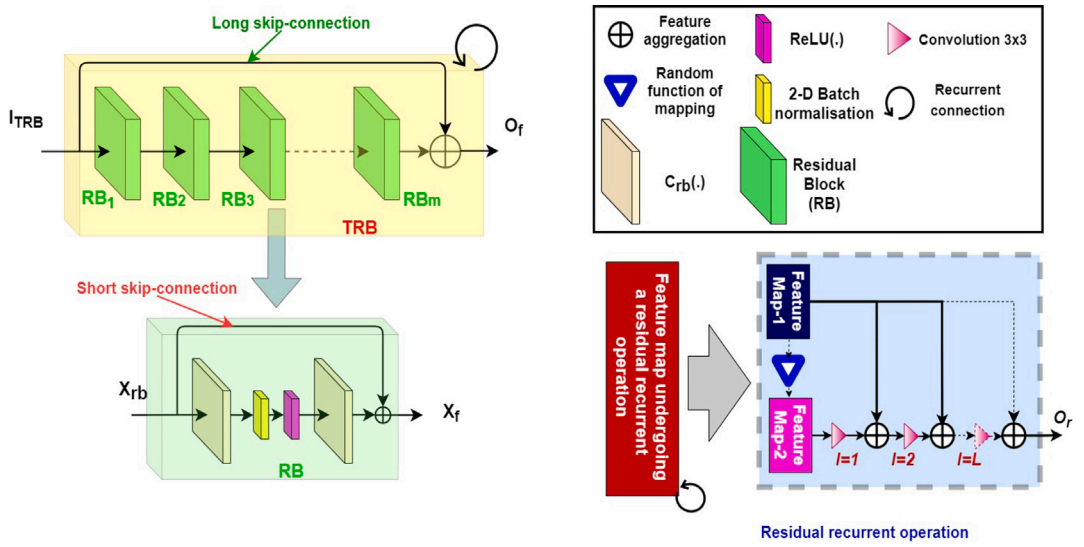


Fig. 3. Construction and specification of two-fold residual blocks (TRBs) and working principle of recurrent residual operation.

Now, if there are n MTRRGs and if each MTRRG consists of m residual blocks, then that particular version of MTRRE-Net is called MTRRE-Netmn. Each MTRRG produces two output feature maps, one is directly passed on to the next MTRRG via 2 × 2 max-pooling operation whereas the other one is flattened and concatenated to the final classification layer.

Now, let us suppose the input image X_{in} has three color channels. The shallow features are extracted from this input image via a convolution operation ($C_{sh}(.)$) with a kernel size of 3, and both stride and padding equal to 1. This convolution operation maps the input image to a higher dimensional feature (f_{hd}) space. The operation is given below as follows:

$$I_{sh} = C_{sh}(X_{in}) \quad | \quad C_{sh} \in \mathbb{R}^{3 \times f_{hd}} \tag{1}$$

where, I_{sh} is the shallow feature map which is to be fed to the following MTRRG.

Now, each MTRRG includes a max-pooling operation with the kernel of 2 × 2, the stride of 1, and the padding value of 0. This operation reduces the dimension of the feature map by a scale of 2 while keeping the number of channels of the output feature map constant to their previous values. Suppose the output feature map of an MTRRG has a dimension of H×W×C, then the dimension of the input feature map of the next MTRRG would be H/2×W/2×C after performing the max-pooling operation.

It is to be noted that to accomplish the two-fold residual (TR) operation, it is necessary to aggregate (pixel-wise addition) the input feature map of MTRRG to the output feature map of the MTRRG via longer skip connections. Therefore, the dimensions of both of the feature maps should be the same. So, the number of channels of the feature maps in the mainstream information flow in the root architecture is remained fixed to f_{hd}, whereas the height and the width of the feature maps are decreased by the scale of 2 as there is a max-pooling operation after each MTRRG.

From each MTRRG, two streams of information flow come as output, one of them is already discussed above, whereas the other stream is applied for learning more robust features with multiple scales of kernels. This stream is passed via convolution, max-pooling, and fully-connected layers. Finally, this gets concatenated with the final feature set coming out from the root architecture.

Suppose there are n MTRRGs which are $\{MTRRG_1, MTRRG_2, MTRRG_3, ..., MTRRG_n\}$. Now, the shallow feature map (I_{sh}) of the

initial convolution layer ($C_{sh}(.)$) with dimensions say $H_{hd} \times W_{hd} \times f_{hd}$, is to be fed to $MTRRG_1$. This produces two outputs, one is a deep feature map I_{d_1} with dimension $H_{hd}/2 \times W_{hd}/2 \times f_{hd}$ and the other one is a flattened deep feature set (f_{df_1}) with say k_1 features. The output I_{d_1} is going to be fed to $MTRRG_2$ and f_{df_1} is going to be concatenated to the final feature set of the root architecture. Similarly, $MTRRG_2$ produces I_{d_2} which has feature dimension of $H_{hd}/4 \times W_{hd}/4 \times f_{hd}$ and f_{df_2} with k_2 number of features and so on. Thus, in this way, the output I_{d_n} of $MTRRG_n$ has the feature dimension $H_{hd}/2^n \times W_{hd}/2^n \times f_{hd}$, and which is further passed through a global average pooling layer ($GP(.)$) that produces a feature map I_f having dimension of $1 \times 1 \times f_{hd}$. I_f is computed by using the following equation.

$$I_f = GP(I_{d_n}) \mid GP \in \mathbb{R}^{(H_{hd}/2^n \times W_{hd}/2^n \times f_{hd}) \rightarrow (1 \times 1 \times f_{hd})} \tag{2}$$

This is further flattened and concatenated to the features of previous MTRRGs. Therefore, the number of features present in this final feature set (f_0, see Eq. (3)) is $K = k_1 + k_2 + k_3 + \cdots ..k_n + f_{hd}$.

$$f_0 = flattened(I_f) + f_{df_n} + f_{df_{n-1}} + f_{df_{n-2}} + \cdots ... + f_{df_n} \tag{3}$$

Now to reduce redundancy as well as to extract a more prolific and low-dimensional feature set, another fully-connected ($FC_1(.)$) layer is embedded which maps K number of features to 1024 and this becomes the final feature set. Thereafter, another fully-connected layer ($FC_2(.)$) is embedded which maps 1024 features to the number of classes under consideration. Finally, Softmax ($\Omega(.)$) is applied for the classification as follows:

$$P = \Omega(FC_2(FC_1(f_0))) \mid FC_1 \in \mathbb{R}^{K \times 1024} \ \& \ FC_2 \in \mathbb{R}^{1024 \times c} \tag{4}$$

where, P is the probability distribution and c is the number of classes of the dataset under consideration.

4.2. Two-fold residual operation

In the proposed model, we introduce TR operation (see Fig. 3) for complex pattern learning and salient feature extraction. We know that relatively deeper CNN models learn complex features more effectively compared to shallower counterparts. But this associates a serious issue during training, that is the performance of the model decreases with an increase in depth of the model. This is known as the Degradation problem. Our proposed two-fold residual operation allows us to achieve a deep CNN architecture where the degradation problem is reduced significantly. This is done to address the central problem of having fewer data samples for training a deep CNN model. TR consists of two skip-connections: short skip-connection and long skip-connection.

As shown in Fig. 3, the basic residual block (RB), mentioned in [45], is the elementary building block of MTRRE-Net. Suppose a RB takes an input X_{rb}, having dimensions $H_{rb} \times W_{rb} \times Ch_{rb}$. The RB is basically consisted of two similar convolution layers ($C_{rb_1}(.)$ and $C_{rb_2}(.)$) of kernel size = 3, stride = 1 and padding = 1, and one ReLU(.) and batch normalization($bn(.)$) layers. Such a specific type of convolution operation keeps the height and width of the input feature map intact and produces the output feature map, which has the dimension same as that of the input feature map. Now in the RB, X_{rb} is first fed to $C_{rb_1}(.)$, which is followed by ReLU(.) and bn(.), and thereafter the resultant feature map is further fed to $C_{rb_2}(.)$. If the output of $C_{rb_2}(.)$ is I_{rb} then the final output feature map of RB would be as in the next equation.

$$X_f = X_{rb} + I_{rb} \tag{5}$$

Here, in the TR operation, a stack of several RBs is created, where the output of one RB is considered as the input to the following RB. This stack of RBs is named a twofold residual block (TRB). Now in each TRB, there is a single long skip connection. Let us suppose there are m RBs in a TRB, which are $\{RB_1, RB_2, \ldots ..., RB_m\}$. Then the input of TRB (I_{TRB}), which is also the input of RB_1, is aggregated with the output of RB_m (O_{RB_m}) via the long skip-connection, which forms the final output(O_f) of the TRB. The operation of TRB is mathematically expressed by Eqs. (6) and (7). The pseudo-code for TR operation is provided in **Algorithm 1**.

$$O_{RB_m} = RB_m(RB_{m-1}(RB_{m-2}(.....RB_1(I_{TRB})))) \tag{6}$$

$$O_f = I_{TRB} + O_{RB_m} \tag{7}$$

This is how via long and short skip-connections, TR operation is incorporated into the MTRRE-Net network.

Algorithm 1 Pseudo code of a TRB. The sign '+' denotes the aggregation of two different feature maps.

Input: $m \rightarrow$ Number of RBs in the TRB
$Ch_{rb} \rightarrow$ Input channels
$I_{TRB} \rightarrow$ Input feature map
Initiate RB = Residual Block (Ch_{rb})
$I = I_{TRB}$
for i = 1 : m **do**
  $I = RB(I)$
end for
$O_{RB_m} = I$
$O_f = O_{RB_m} + I_{TRB}$
Output: O_f

4.3. Multi-scale two-fold residual recurrent groups

The proposed model consists of several MTRRGs. The basic functionality of MTRRG is based on multi-scale feature extraction from the TRBs with recurrent connections (TRRB). The residual recurrent operation on a sample feature map is shown in Fig. 3. Say I_{fm} is a feature map undergoing any operation, which produces an output feature map of O_{fm}. Then the input feature map gets added to the output feature map recurrently for specifically L number of times via a convolution operation. The L should not be too high, otherwise, the impact of information extraction with that particular operation, which is undergoing recurrent operation too, will be lowered by a significant amount. Here, in our case, the operation is the two-fold residual operation. The TRRB is shown in detail in Fig. 3.

Now, the features from different levels of TRRBs have been extracted with different scales of convolution windows. The basic components of MTRRGs are TRB, maxpool layers, convolution with different kernel sizes for feature extraction, a fully-connected layer, and a global average pooling layer.

Algorithm 2 Pseudo code for implementation of MTRRG.

Input: $m \rightarrow$ Number of RBs in the TRB
$Ch_{in} \rightarrow$ Input channels
$I_{in} \rightarrow$ Input feature map
Initiate TRB = Dual residual block (m, Ch_{in})
Initiate $C_{m,5}$ = 2-D Convolution(Ch_{in}, Ch_{in}, kernel size = 5×5, stride = 2, padding = 2)
Initiate $C_{m,7}$ = 2-D Convolution(Ch_{in}, Ch_{in}, kernel size = 7×7, stride = 2, padding = 2)
Initiate MP = 2-D Max-pool((2×2))
Initiate GP = 2-D Global average pooling((1×1))
$x_i = TRB(I_{in})$
$x_{o_1} = MP(x_i)$
$x_1 = GP(MP(C_{m,5}(x_{o_1})))$
$x_2 = GP(MP(C_{m,7}(x_{o_1})))$
$x_{o_2} = Flatten(Channelwiseconcatenate(x_1, x_2))$
Output: x_{o_1}, x_{o_2}

It is to be noted that the input of MTRRG (I_{in}) is also the input of its TRRB. Suppose the dimension of I_{in} is $H_{in} \times W_{in} \times Ch_{in}$, therefore the output of the TRB (x_i) also has the dimension of $H_{in} \times W_{in} \times Ch_{in}$. x_i

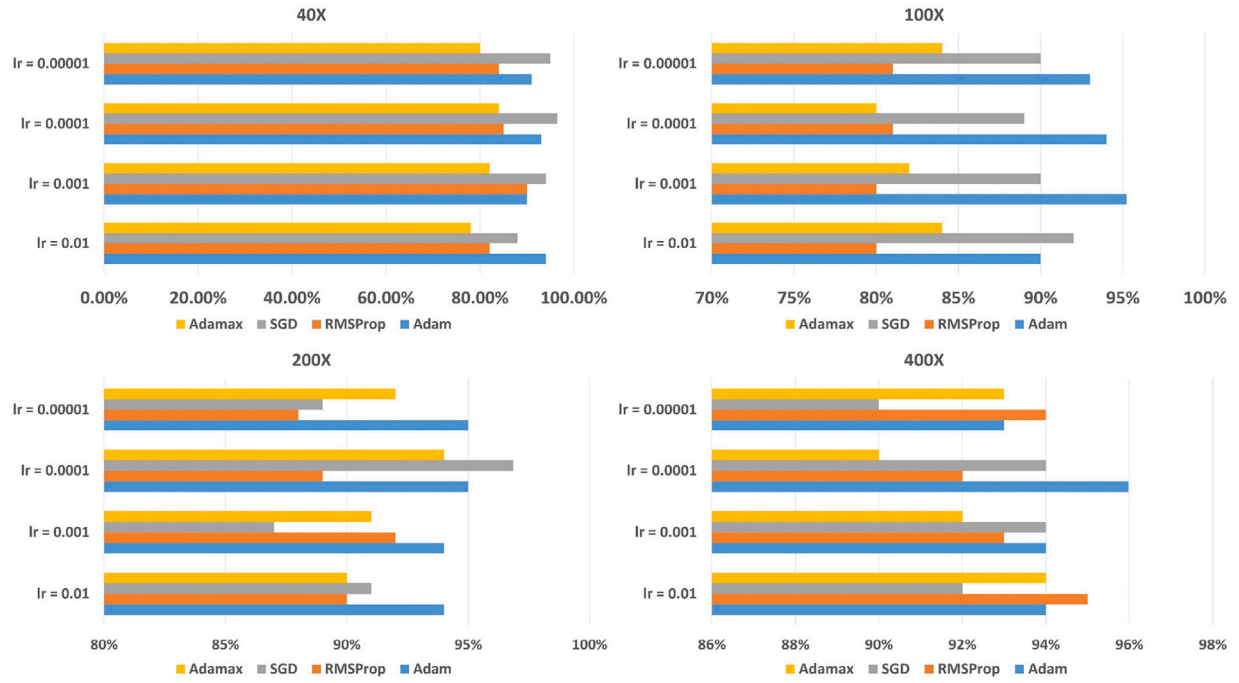


Fig. 4. The accuracy of the proposed model is obtained with various combinations of optimizers and corresponding learning rates.

is further passed through a max-pooling layer with a pooling kernel of 2×2, stride $= 1$, and padding $= 0$ for dimensionality reduction of the feature map. The output feature map of the max-pooling layer (x_{o_1}) has the feature map of $H_{in}/2 \times W_{in}/2 \times Ch_{in}$. This x_{o_1} is one of three outputs of the MTRRG which is further fed to the next MTRRG for more complex pattern learning. From the feature map x_{o_1}, another two streamlines of information flow are extracted with two different convolution operations with different kernel sizes from x_{o_1}. One is passed through a convolution operation($C_{m,5}(.) \mid C_{m,5} \in \mathbb{R}^{Ch_{in} \times 2 * Ch_{in}}$) of kernel size of 5×5 and the other($C_{m,7}(.) \mid C_{m,7} \in \mathbb{R}^{Ch_{in} \times 2 * Ch_{in}}$) of kernel size 7×7. The output of both convolution operations is further fed through the max-pooling operation, which is followed by a global average pooling layer. Both of the outputs of the global average pooling layers are concatenated and flattened, and passed through a fully-connected layer, which is to extract deep features (see **Algorithm 2**). Thereafter, the deep features from that particular level are further concatenated with the final feature vector of the network.

5. Experiments and analysis

To estimate the performance of the MTRRE-Net model with several experimentation, we have relied on some primary measures of evaluation. These elementary measures include accuracy, precision, recall, and F1 score. We have considered F1 score as it enhances the reliability of the results when the dataset is class imbalanced which is present in the dataset under consideration. Moreover, to deal with any kind of hidden biases of the dataset, we perform a 5-fold cross-validation for evaluating our proposed model. Here, the entire dataset is divided into 5 different subsets of almost equal dimensions in such a manner that there is no intersection of data samples between any two of the subsets. Thereafter, 5 different mutually exclusive and independent experiments are performed by considering four datasets for training and the remaining one for testing. It is to be noted that the test dataset is different for each of these five cases of experimentation, thereby resulting in no overlapping between the two experiments. The final results reported are the linear average of the results obtained by performing the same experiments five times for each of the five folds.

5.1. Implementation details

It has been mentioned in Section 4.1 that our proposed model is flexible in terms of its architecture and based on the numbers of RBs and TRBs used, the nomenclature of the network is done. For breast cancer detection from histopathological images, several experiments are performed considering different combinations of RBs and TRBs. Table 2 shows the results obtained with different versions of the MTRRE-Net model at all magnification levels of the BreakHis dataset. A detailed discussion of these versions is already explained in Section 5.2. It is observed that the best results are obtained using 4 TRBs and 7 RBs in each TRB while keeping the number of recurrent connections (L) to 1. In addition to this, the numbers of features from different levels of the architecture are also fixed to the values as shown below:

$$k_1 = 512\ ,\ k_2 = 1024\ ,\ k_3 = 1024\ ,\ k_4 = 2048 \tag{8}$$

and the high dimensional feature (f_{hd}) has been fixed to 64. Therefore, the final feature set f_0 consists of $k = (512 + 1024 + 1024 + 2048 + 64) = 4672$ features. The results obtained by the MTRRE-Net74 model at four different magnification levels are reported in Table 3.

The network is trained for 300 epochs with four different optimizers, namely Adam, RMSProp, SGD, and Adamax. The learning rates of the optimizers range from $1e^{-2}$ to $1e^{-5}$. The overall results obtained by the mentioned combinations with different magnification levels of the images of the BreakHis dataset are given in Fig. 4. Though the reported results are much comparable, in some cases, significant differences are also observed. In most of the cases mentioned in Fig. 4, it can be found that Adam optimizer with a learning rate ranging between 0.001 to 0.0001, reports optimal performance with maximum accuracies. For the images with a magnification level of 200X, the SGD optimizer with a learning rate of 0.0001 achieves the best among others. Since the results of any deep learning model with different optimizers and learning rates may vary from one dataset to another and although the Adam optimizer produces good results with the proposed framework, it is quite difficult to generalize the results completely.

5.2. Performance with different configurations of the network

It has been mentioned in Section 4.1 that based on the requirement, the configuration of the proposed network can be customized. Here, we

Table 2
Comparison among different configurations of proposed MTRRE-Net model. In the table, f_0 is the number of features in the final feature set and k is the number of features extracted from different levels of the network. A, P, R, and F respectively denote the accuracy, precision, recall, and F1 score. All the results are mentioned in percentage format and the maximum result of each column is highlighted in bold and underlined text format.

Version	k	f_0	Different magnification levels of the images of BreakHis dataset															
			40X				100X				200X				400X			
			A	P	R	F	A	P	R	F	A	P	R	F	A	P	R	F
MTRRE-Net53	k_1 = 1024, k_2 = 1024, k_3 = 1024	3136	91.10	90	92	91	92.29	91	91	94	93.49	92	94	94	91.19	90	90	91
MTRRE-Net34	k_1 = 1024, k_2 = 1024, k_3 = 1024, k_4 = 2048	5184	86.00	82	92	80	91.19	88	92	90	91.10	93	90	89	92.4	90	92	92
MTRRE-Net35	k_1 = 512, k_2 = 1024, k_3 = 1024, k_4 = 1024, k_5 = 2048	5696	92.21	92	92	94	91.19	87	91	90	90.09	88	90	91	91.48	92	90	93
MTRRE-Net73	k_1 = 1024, k_2 = 1024, k_3 = 1024	3136	91.17	91	94	91	93.10	90	**95**	93	94.49	95	94	**97**	95.24	**97**	**96**	**95**
MTRRE-Net93	k_1 = 1024, k_2 = 1024, k_3 = 1024,	3136	93.39	94	91	94	93.33	**93**	92	93	94.77	**96**	93	95	94.68	95	**96**	91
MTRRE-Net74	k_1 = 512, k_2 = 1024, k_3 = 1024, k_4 = 2048	4672	**97.12**	**97**	**96**	**96**	**95.22**	92	**95**	**97**	**96.85**	**96**	**98**	94	**97.81**	96	94	**95**

Table 3
Results obtained by MTRRE-Net74 model using four different magnification levels of BreakHis dataset.

Magnification level	Accuracy	Precision	Recall	F1 Score
40x	97.12%	97%	96%	96%
100x	95.22%	94%	95%	97%
200x	96.85%	96%	98%	94%
400x	97.81%	96%	94%	95%

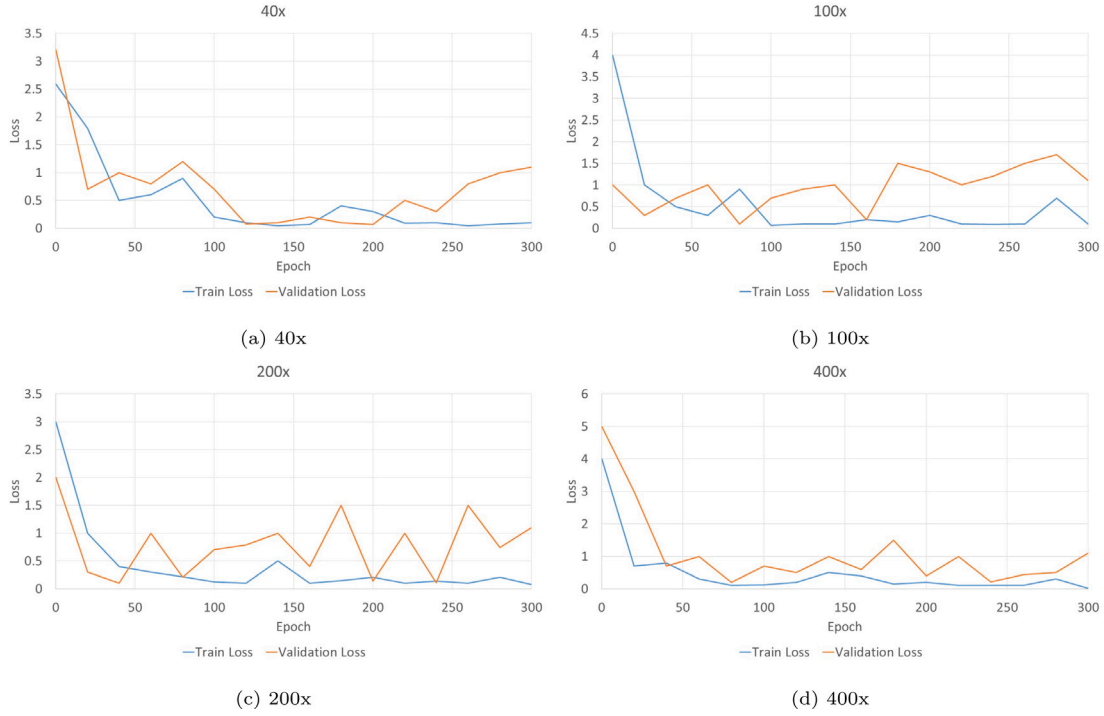
have performed some experimentation using varied configurations of MTRRE-Net on the BreakHis dataset.

The results obtained by different configurations with all magnification levels of the images along with the explicit parameters of the configurations are shown in Table 2. It can be observed that larger networks like MTRRE-Net73, MTRRE-Net93, and MTRRE-Net74 report better results than the shallower networks such as MTRRE-Net34, MTRRE-Net53, and MTRRE-Net35. Since the larger networks learn more complex patterns that generate more relevant features in the final feature set, they outperform the shallower networks in terms of accuracy. Though the dataset used here for experimentation is not having a large number of samples, then also deeper networks achieve quite good results which justifies the fact we have claimed in Section 4, i.e., the twofold residual skip-connections can overcome vanishing gradient problem, and the recurrent connections maintain the consistency of relevant information flow throughout the whole network. Since it is observed that the MTRRE-Net74 model performs better than other configurations, the rest of the experimentation are performed using the best performing MTRRE-Net74 model only.

5.3. Performance analysis

Therefore, from the above comparison, we can see that the MTRRE-Net74 has been considered the best version of its other kinds. Here in this section, we analyze the performance of MTRRE-Net74 by reporting some classical but effective results on all four magnification levels. These experimental results include the convergence loss plots and accuracy plots over the iterations, the confusion matrix, ROC plots, and their corresponding AUC.

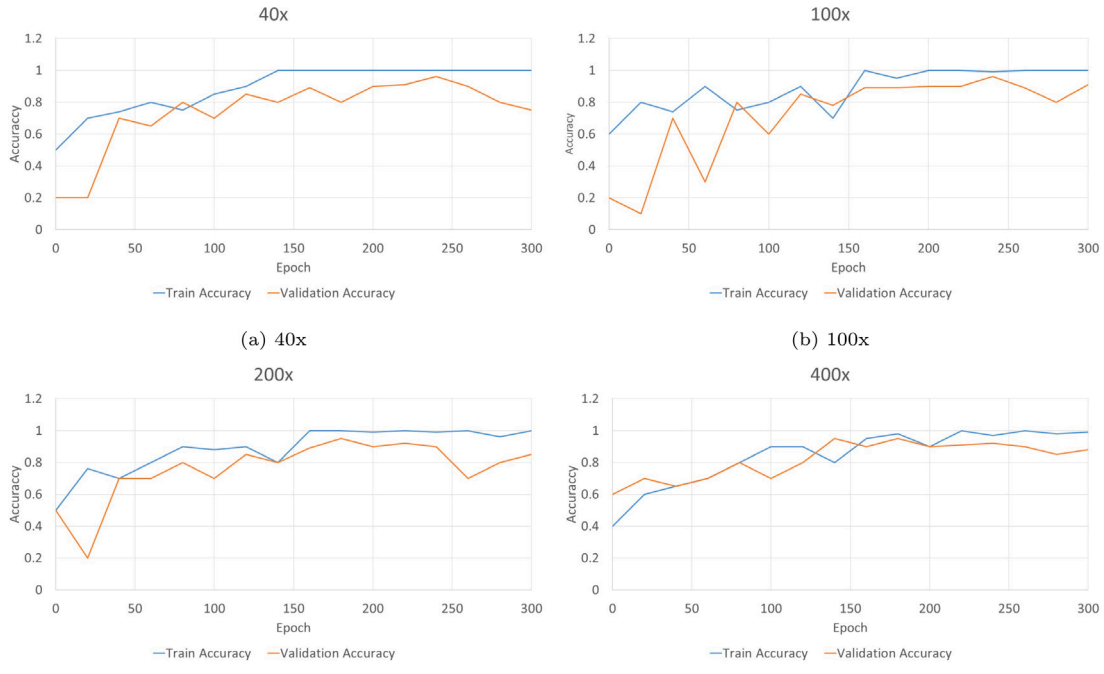
The convergence loss plots depict the variations of train and validation losses over the iterations of training. The loss plot corresponding to the training of a deep learning model carries out significant information regarding the learning of the model. Ideally, the train and validation losses decrease simultaneously concerning the training epochs. But in reality, this property cannot be followed rigorously or rigidly. Instead, after certain epochs of training the train and validation losses get saturated and start oscillating under certain boundaries. Such a nature of losses shows the positive learning of the models. In some cases, after certain epochs, the validation losses keep increasing and the deviation of the validation plot from the train plot also starts increasing over the iterations. This indicates the overfitting of the model in the dataset. For some cases the validation loss varies indefinitely during the entire training process, this also indicates the poor learning of the model. Ideally, both the train and validation losses converge, and the variation of the validation loss follows the variation of the training loss noticeably. The converging loss plot of each of the magnification levels of the BreakHis dataset used in this paper is depicted in Fig. 5. It can be observed from all of the loss plots that the validation loss follows the training loss and converges well over the iterations. This specifies that the model is learning and is not overfitting in such small datasets of different levels of BreakHis. Similar to loss plots, the accuracy also shows converging characteristics when the disparity is plotted against epochs. Unlike loss plots, instead of decreasing, the accuracies increase over the iterations. In this case, also, the validation accuracies follow the train accuracies and reach saturation after certain iterations. If some indefinite oscillations in the validation accuracy plot are observed which cause significant deviation from the train accuracy plot, we can say that the model is not learning at all. For ideal cases, the validation accuracy plot strictly follows the train accuracy plot. Fig. 6 contains the



(a) 40x (b) 100x

(c) 200x (d) 400x

Fig. 5. Graphs showing the train versus validation loss plots at multiple magnification levels of BreakHis dataset.



(a) 40x (b) 100x

(c) 200x (d) 400x

Fig. 6. Graphs showing the variation of train and validation accuracies over iterations on multiple magnification levels of BreakHis dataset.

accuracy plots of different magnification levels of the histopathological dataset. Now, by observing the reported accuracy plots, we also arrive at the same conclusion that the model is learning significantly over the iterations during the training process.

The confusion matrix obtained on all four levels is given in Fig. 7. Here the class Benign is considered as the Positive class and the Malignant class is considered as the Negative class. A thorough pattern of getting greater True Positive values than the True Negative values for higher levels of magnification can be observed in the figure. Therefore, it infers that, for lower magnification levels (40x), the model MTRRE-Net74 learns to identify the malignant cells better than the benign cells, whereas for the higher magnification levels(>= 100x) the reverse of the above statement holds.

In addition to this, the Receiver Operating Characteristic or ROC curve also carries significant information about the classification performance of the MTRRE-Net74. The ROC curve as well as the Area Under the Curve (AUC) of each ROC, of different magnification levels of BreakHis dataset, is depicted in Fig. 8. The ROC curve is the variation of

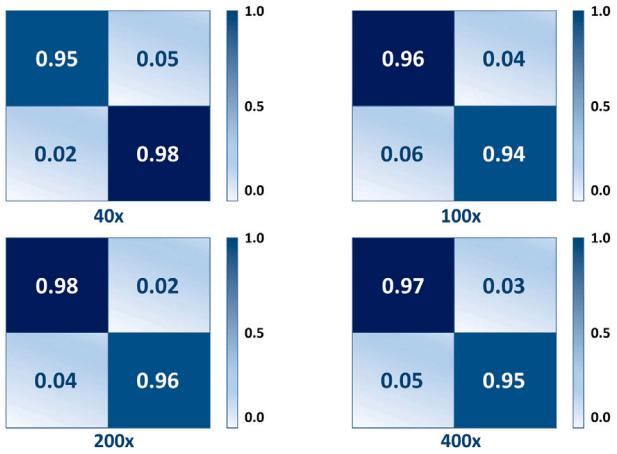


Fig. 7. The binary-class confusion matrix on different magnification levels of BreakHis dataset.

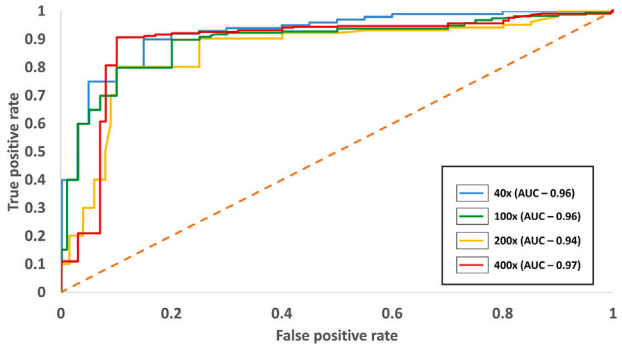


Fig. 8. The ROC plots obtained by the MTRRE-Net74 on all four magnification levels of BreakHis dataset.

the True Positive Rate (TPR) concerning the False Positive Rate (FPR), which eventually helps to determine the right threshold value for a problem by considering the variations of TPR and FPR. Ideally, for the ROC plot, the TPR at 0 FPR should be 1, therefore it can be observed from Fig. 8 that the ROC obtained by MTRRE-Net74 on all four levels of magnification is very much close to the ideal ROC. Moreover, the high value of AUC depicts the good performance of the model. In our case, the obtained AUC values are very high and are almost equal to 1. Therefore the ROC and corresponding AUC constructively support the selection of the model MTRRE-Net74 in this particular task.

5.4. Comparison with different deep learning networks

To confirm the superiority of the MTRRE-Net74, we report the accuracies obtained by some state-of-the-art methods along with the accuracies of MTRRE-Net74 with all four magnification levels of the BreakHis dataset. It is to be noted that traditional transfer learning-based models have been used for classifications using pre-trained CNN. These CNNs include GoogLeNet, 2 different versions of DenseNet, 4 versions of ResNet, and 3 versions of VGG and SqueezeNet. Two versions of a certain model differ by the number of layers or the depths of models only, and root architectures always remain the same. Obtained accuracies are reported in Table 4.

The neural networks with the pre-trained weights of the ImageNet dataset are fine-tuned by 50 epochs of training using the BreakHis dataset with the Adam optimizer and the learning rate of 0.001. The other parameters are fixed to their standard values. It can be observed that the results reported by VGG16, ResNet18, and GoogLeNet are competent, but the proposed model MTRRE-Net74 outperforms the state-of-the-art models with a good margin.

Table 4
Results obtained by several standard CNNs on BreakHis dataset. Experimentation has been performed on all four magnification levels of the dataset and corresponding accuracies are reported.

CNN model	Obtained accuracies on different magnification scales			
	40x	100x	200x	400x
GoogLeNet [46]	89.97%	93.45%	91.20%	94.12%
DenseNet169 [47]	92.17%	91.19%	92.44%	91.99%
DenseNet161 [36]	91.91%	90.00%	92.44%	88.7%
ResNet18 [48]	92.12%	93.22%	92.29%	91.01%
ResNet50 [48]	92.12%	92.29%	91.47%	90.91%
ResNet121 [48]	93.23%	92.11%	90.00%	92.44%
ResNet152 [48]	90.19%	94.01%	91.06%	92.01%
VGG11 [49]	92.54%	91.17%	93.10%	92.20%
VGG16 [49]	93.22%	95.19%	94.00%	95.22%
VGG19 [49]	92.11%	92.00%	92.00%	93.45%
SqueezeNet [50]	94.42%	91.2%	93.12%	91.19%
MTRRE-Net74	**97.12%**	**95.22%**	**96.85%**	**97.81%**

5.5. Performance comparison with state-of-the-art models

To further prove the effectiveness of the MTRRE-Net, we have compared its performance with some recently proposed techniques, which have reported benchmark accuracies so far on the BreakHis dataset. The results and approaches of these state-of-the-art methods are reported in Table 5. It can be observed that several models have been developed for the breast cancer classification task, and most of them are based on pure deep learning models only. A. Kumar et al. [51] extracted deep features from VGG16 and classified them using popular machine learning classifiers — Support Vector Machine (SVM) and Random Forest. The results reported by A. Kumar are quite good. In another work, S. Sharma et al. [52] developed a pure transfer learning model using pre-trained AlexNet, however, they failed to achieve good results. Apart from that M. Toğaçar et al. [53], U. Budak et al. [54] and M. Saini et al. [21] achieved benchmark accuracies on this particular dataset. Overall, it can be seen from Table 5 that the proposed MTRRE-Net74 outperforms most of the techniques reported hereby. Though the reported results by V. Gupta et al. [55] are found to be better than ours however, their reported results are produced by performing the average of three different data splits (80:10:10, 70:10:20 and 60:10:30). On the other hand, the results, reported in our proposed work, are the average of 5-fold cross-validation outcomes. Since data samples were chosen for training, validation, and testing impacts the final results significantly, we cannot simply compare these two different sets of results.

5.6. Statistical significance test

To validate the comparison reported above, we have performed a ranking-based statistical significance test called the Friedman test. The Friedman test performs a statistical analysis, which is used to determine whether 3 or more experiments (algorithms in our case) from the same kind of subjects are different from each other on an ordinal variable of interest. The variable considered should be continuous, and have a similar spread across all the groups. Another advantage of the Friedman test over other popular statistical tests is that it is a non-parametric test.

However, similar to all other statistical tests [58,59], the Friedman test also has Null and Alternative hypotheses. The Null-hypothesis of Friedman's test states that the median value of accuracy considering both datasets for each contender algorithm is the same. Let there are N contender algorithms among which M_n is the median of accuracy values of k^th algorithm, achieved on multiple datasets where $n \in [1, N]$. In contrast, the Alternative Hypothesis states all median values would not be the same rather there would be significant differences between the performances of the competitor algorithms. Mathematically, the algorithms can be expressed as follows:

H_0 : $M_1 = M_2 = M_3 = \cdots ..... = M_{N-1} = M_N$

H_1 : M_n values for $k \in [1, N]$ can be same and cannot be

Table 5
Comparison of our proposed MTRRE-Net74 model with some recent state-of-the-art models on the BreakHis dataset. Results of the best models are underlined and that of our model is made bold for easier identification.

Work ref.	Method description	Achieved accuracy			
		40x	100x	200x	400x
C. Zhu et al. [56]	A hybrid CNN with inception block and Squeeze–Excitation–Pruning (SEP)	85.6%	83.9%	85.4%	81.2%
X. Li et al. [57]	Interleaved DenseNet(IDSNet)	89.1%	85.0%	87.00%	84.5%
U. Budak et al. [54]	A combination of FCN and BiLSTM	95.69%	93.6%	96.3%	94.29%
A Kumar et al. [51]	Deep features obtained from VGG16 are fed to SVM and Random Forest	94.11%	95.1%	97.0%	94.96%
M. Saini et al. [21]	Augmentation using DCGAN and classification using modified VGG16	96.4%	94.0%	95.5%	93.00%
S. Sharma et al. [52]	Transfer learning using pre-trained AlexNet	89.31%	85.7%	83.9%	84.33%
V. Gupta et al. [55]	Modified Residual Networks	99.11%	99.52%	98.74%	–
Proposed model	**Deep learning model — MTRRE-Net74**	**97.12%**	**95.2%**	**96.8%**	**97.81%**

Table 6
Results of the Friedman test on all four magnification levels of BreakHis dataset. Here, r_n^{av} denotes the average rank in all four magnification levels of BreaKHis dataset.

Algorithm	Datasets				r_n^{av}	χ_F^2	Status
	40x	100x	200x	400x			
MTRRE-Net7	1	1	2	1	1.25	30.667	**Rejected**
C. Zhu et al. [56]	7	7	6	7	6.75		
X. Li et al. [57]	6	6	5	5	5.5		
U. Budak et al. [54]	3	4	3	3	3.25		
A Kumar et al. [51]	4	2	1	2	2.25		
M. Saini et al. [21]	2	3	4	4	4.66		
S. Sharma et al. [52]	5	5	7	6	5.75		

To carry out the Friedman test, let us consider r_n^m is the rank of n^th algorithm on m^th dataset based on classification accuracies obtained by all contender algorithms such that the algorithm which has achieved the best accuracy is considered to have rank 1 and the algorithm having worst performance is ranked as N. Now if there are M datasets in total then the average ranking r_n^av can be defined as

$$r_n^{av} = \frac{1}{M}\sum_{m=1}^{M} r_n^m \tag{9}$$

In the Friedman test, the performance evaluating criterion is χ_F^2 which is calculated by the following equation:

$$\chi_F^2 = \frac{12M}{N(N+1)}\left[\sum_{n=1}^{N}(r_n^{av})^2 - \frac{N(N+1)^2}{4}\right] \tag{10}$$

The individual ranks assigned by applying the Friedman test at all four magnification levels are depicted in Table 6. It is to be noted that since In this case, there are 7 competitor algorithms (N=7) and 4 datasets (M=4), therefore from the Friedman distribution chart, it is known that the value of $\chi^2_{N-1=6,M=4,\alpha=0.5}$ is 7.84 which is eventually significantly less than the obtained χ_F^2 which is 30.667. Therefore, we can conclude that the null hypothesis gets rejected and the superiority of the proposed MTRRE-Net74 model is established.

6. Conclusion

In this paper, we have proposed a deep learning model, called MTRRE-Net, for the screening of breast cancer from histopathological images. The proposed deep learning model is trained in an end-to-end fashion to produce meticulous results. Our experiments show that the different variants of the MTRRE-Net may be used to develop a higher-level computer-aided system to classify different diseases from medical images. Our idea of the twofold residual recurrent operation works well on small datasets. Besides, the multi-scaling operation helps to learn significant features and to focus on generating prolific features, which improves the classification accuracy of the model as well. While evaluated on the challenging BreaKHis dataset, our model achieves accuracy values of 97.12%, 95.22%, 96.85%, 97.81% on the images of 40x, 100x, 200x, and 400x magnificent levels respectively to classify benign and malignant cells. We have compared the obtained results with some pre-trained models that include GoogLeNet, DenseNet169, DenseNet161, ResNet18, VGG16, etc., and we have found that it outperforms all these models, thereby confirming the superiority of the model. Moreover, we have compared our model with some state-of-the-art models which reported the benchmark results, and the comparison validates the effectiveness of the proposed model. As a future work scope, we have a plan to make this MTRRE-Net a clinical diagnostics tool, by examining its performance on different medical image datasets such as lung cancer, prostate cancer, renal cancer, etc. Moreover, we have also proposed different competent variants of the MTRRE-Net, which can also be evaluated on different classification problems as well. Some further modifications such as taking the two-fold residual operation to a new level like three-fold or multi-folded residual operation depending upon the dimension of the datasets under consideration can also be done. Additionally, we can consider the use of computational intelligence algorithms on deep features extracted from this MTRRE-Net to improve the result further. Some standard and newly developed optimization algorithms like Monarch butterfly optimization (MBO) [60], Earthworm optimization algorithm (EWA) [61] Elephant herding optimization (EHO) [62], Moth search (MS) algorithm [63], Runge–Kutta Optimization [64], Manta Ray Foraging Optimizer (MRFO) [65] can be used to reduce the feature dimension generated by the deep learning model which will make the model more time and memory efficient.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L.J. Van't Veer, H. Dai, M.J. Van De Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. Van Der Kooy, M.J. Marton, A.T. Witteveen, et al., Gene expression profiling predicts clinical outcome of breast cancer, Nature 415 (6871) (2002) 530–536.

[2] A. Egenvall, B.N. Bonnett, P. Öhagen, P. Olson, A. Hedhammar, H. von Euler, Incidence of and survival after mammary tumors in a population of over 80,000 insured female dogs in Sweden from 1995 to 2002, Prevent. Vet. Med. 69 (1–2) (2005) 109–127.

[3] Y. Salas, A. Márquez, D. Diaz, L. Romero, Epidemiological study of mammary tumors in female dogs diagnosed during the period 2002–2012: A growing animal health problem, PLoS One 10 (5) (2015) e0127381.

[4] F. Viola, Biopsy system, 2004, Google Patents, US Patent 6, 712, 773.

[5] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, IEEE Trans. Med. Imaging 35 (1) (2015) 119–130.

[6] A. Belsare, M. Mushrif, Histopathological image analysis using image processing techniques: An overview, Signal Image Process. 3 (4) (2012) 23.

[7] M.H. Nadimi-Shahraki, S. Taghian, S. Mirjalili, L. Abualigah, Binary aquila optimizer for selecting effective features from medical data: A COVID-19 case study, Mathematics 10 (11) (2022) 1929.

[8] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases, J. Pathol. Inform. 7 (2016).

[9] J.G. Elmore, G.M. Longton, P.A. Carney, B.M. Geller, T. Onega, A.N. Tosteson, H.D. Nelson, M.S. Pepe, K.H. Allison, S.J. Schnitt, F.P. O'Malley, D.L. Weaver, Diagnostic concordance among pathologists interpreting breast biopsy specimens, JAMA 313 (11) (2015) 1122–1132.

[10] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: 2016 International Joint Conference on Neural Networks, IJCNN, IEEE, 2016, pp. 2560–2567.

[11] M.H. Nadimi-Shahraki, H. Zamani, S. Mirjalili, Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study, Comput. Biol. Med. (2022) 105858.

[12] M. Fasihi, M.H. Nadimi-Shahraki, A. Jannesari, A shallow 1-D convolution neural network for fetal state assessment based on cardiotocogram, SN Comput. Sci. 2 (4) (2021) 1–9.

[13] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, 2015, arXiv preprint arXiv:1506.01497.

[14] C. Guo, B. Fan, Q. Zhang, S. Xiang, C. Pan, AugFPN: Improving multi-scale feature learning for object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12595–12604.

[15] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation, 2018, arXiv preprint arXiv:1802.06955.

[16] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N.N. Shih, J. Tomaszewski, F.A. González, A. Madabhushi, Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent, Sci. Rep. 7 (1) (2017) 1–14.

[17] S.D. Roy, S. Das, D. Kar, F. Schwenker, R. Sarkar, Computer aided breast cancer detection using ensembling of texture and statistical image features, Sensors 21 (11) (2021) 3628.

[18] R. Sanyal, D. Kar, R. Sarkar, Carcinoma type classification from high-resolution breast microscopy images using a hybrid ensemble of deep convolutional features and gradient boosting trees classifiers, IEEE/ACM Trans. Comput. Biol. Bioinform. (2021).

[19] P. Bhowal, S. Sen, J.D. Velasquez, R. Sarkar, Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification, Expert Syst. Appl. 190 (2022) 116167.

[20] S. Chattopadhyay, A. Dey, P.K. Singh, R. Sarkar, DRDA-Net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images, Comput. Biol. Med. 145 (2022) 105437.

[21] M. Saini, S. Susan, Deep transfer with minority data augmentation for imbalanced breast cancer dataset, Appl. Soft Comput. 97 (2020) 106759.

[22] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using GAN for improved liver lesion classification, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, IEEE, 2018, pp. 289–293.

[23] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing 321 (2018) 321–331.

[24] T. Iqbal, H. Ali, Generative adversarial network for medical images (MI-GAN), J. Med. Syst. 42 (11) (2018) 1–11.

[25] D. Mukherkjee, P. Saha, D. Kaplun, A. Sinitca, R. Sarkar, Brain tumor image generation using an aggregation of GAN models with style transfer, Sci. Rep. 12 (1) (2022) 1–16.

[26] Y. Chen, X.-H. Yang, Z. Wei, A.A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, Q. Guan, Generative adversarial networks in medical image augmentation: A review, Comput. Biol. Med. (2022) 105382.

[27] Q. Guan, Y. Chen, Z. Wei, A.A. Heidari, H. Hu, X.-H. Yang, J. Zheng, Q. Zhou, H. Chen, F. Chen, Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN, Comput. Biol. Med. 145 (2022) 105444.

[28] G.-G. Wang, M. Lu, Y.-Q. Dong, X.-J. Zhao, Self-adaptive extreme learning machine, Neural Comput. Appl. 27 (2) (2016) 291–303.

[29] B. Xu, J. Liu, X. Hou, B. Liu, J. Garibaldi, I.O. Ellis, A. Green, L. Shen, G. Qiu, Attention by selection: A deep selective attention approach to breast cancer classification, IEEE Trans. Med. Imaging 39 (6) (2019) 1930–1941.

[30] H. Yang, J.-Y. Kim, H. Kim, S.P. Adhikari, Guided soft attention network for classification of breast cancer histopathology images, IEEE Trans. Med. Imaging 39 (5) (2019) 1306–1315.

[31] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, Y. Zhao, Histopathological whole slide image analysis using context-based CBIR, IEEE Trans. Med. Imaging 37 (7) (2018) 1641–1652.

[32] J.-H. Yi, J. Wang, G.-G. Wang, Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem, Adv. Mech. Eng. 8 (1) (2016) 1687814015624832.

[33] Z. Cui, F. Xue, X. Cai, Y. Cao, G.-g. Wang, J. Chen, Detection of malicious code variants based on deep learning, IEEE Trans. Ind. Inf. 14 (7) (2018) 3187–3196.

[34] M. Talo, U.B. Baloglu, Ö. Yıldırım, U.R. Acharya, Application of deep transfer learning for automated brain abnormality classification using MR images, Cogn. Syst. Res. 54 (2019) 176–188.

[35] O. Yildirim, M. Talo, B. Ay, U.B. Baloglu, G. Aydin, U.R. Acharya, Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals, Comput. Biol. Med. 113 (2019) 103387.

[36] M. Talo, Automated classification of histopathology images using transfer learning, Artif. Intell. Med. 101 (2019) 101743.

[37] N. Bayramoglu, J. Heikkilä, Transfer learning for cell nuclei classification in histopathology images, in: European Conference on Computer Vision, Springer, 2016, pp. 532–539.

[38] S. Khan, N. Islam, Z. Jan, I.U. Din, J.J.C. Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, Pattern Recognit. Lett. 125 (2019) 1–6.

[39] A. Kensert, P.J. Harrison, O. Spjuth, Transfer learning with deep convolutional neural networks for classifying cellular morphological changes, SLAS Discov.: Adv. Life Sci. R&D 24 (4) (2019) 466–475.

[40] H. Garud, S.P.K. Karri, D. Sheet, J. Chatterjee, M. Mahadevappa, A.K. Ray, A. Ghosh, A.K. Maity, High-magnification multi-views based classification of breast fine needle aspiration cytology cell samples using fusion of decisions from deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2017, pp. 76–81.

[41] G. Litjens, C.I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-Van De Kaa, P. Bult, B. Van Ginneken, J. Van Der Laak, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, Sci. Rep. 6 (1) (2016) 1–11.

[42] J. Xie, R. Liu, J. Luttrell IV, C. Zhang, Deep learning based analysis of histopathological images of breast cancer, Front. Genet. 10 (2019) 80.

[43] S.H. Kassani, P.H. Kassani, M.J. Wesolowski, K.A. Schneider, R. Deters, Classification of histopathological biopsy images using ensemble of deep learning networks, 2019, arXiv preprint arXiv:1909.11870.

[44] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, A. María Vanegas, Breast cancer histopathology image classification using an ensemble of deep learning models, Sensors 20 (16) (2020) 4373.

[45] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[46] M. Al-Qizwini, I. Barjasteh, H. Al-Qassab, H. Radha, Deep learning algorithm for autonomous driving using Googlenet, in: 2017 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2017, pp. 89–96.

[47] R.A. Aral, Ş.R. Keskin, M. Kaya, M. Hacıömeroğlu, Classification of trashnet dataset based on deep learning models, in: 2018 IEEE International Conference on Big Data, Big Data, IEEE, 2018, pp. 2058–2062.

[48] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3147–3155.

[49] N.S. Rani, A. Subramani, A. Kumar, B. Pushpa, Deep learning network architecture based kannada handwritten character recognition, in: 2020 Second International Conference on Inventive Research in Computing Applications, ICIRCA, IEEE, 2020, pp. 213–220.

[50] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size, 2016, arXiv preprint arXiv:1602.07360.

[51] A. Kumar, S.K. Singh, S. Saxena, K. Lakshmanan, A.K. Sangaiah, H. Chauhan, S. Shrivastava, R.K. Singh, Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer, Inform. Sci. 508 (2020) 405–421.

[52] S. Sharma, R. Mehra, Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image, Vis. Comput. 36 (9) (2020) 1755–1769.

[53] M. Toğaçar, K.B. Özkurt, B. Ergen, Z. Cömert, BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer, Physica A 545 (2020) 123592.

[54] Ü. Budak, Z. Cömert, Z.N. Rashid, A. Şengür, M. Çıbuk, Computer-aided diagnosis system combining FCN and bi-LSTM model for efficient breast cancer detection from histopathological images, Appl. Soft Comput. 85 (2019) 105765.

[55] V. Gupta, M. Vasudev, A. Doegar, N. Sambyal, Breast cancer detection from histopathology images using modified residual neural networks, Biocybern. Biomed. Eng. 41 (4) (2021) 1272–1287.

[56] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, J. Liu, Breast cancer histopathology image classification through assembling multiple compact CNNs, BMC Med. Inform. Decis. Mak. 19 (1) (2019) 198.

[57] X. Li, X. Shen, Y. Zhou, X. Wang, T.-Q. Li, Classification of breast cancer histopathological images using interleaved DenseNet with senet (IDSNet), PLoS One 15 (5) (2020) e0232127.
[58] P.K. Singh, R. Sarkar, M. Nasipuri, Statistical validation of multiple classifiers over multiple datasets in the field of pattern recognition, Int. J. Appl. Pattern Recogn. 2 (1) (2015) 1–23.
[59] P.K. Singh, R. Sarkar, M. Nasipuri, Significance of non-parametric statistical tests for comparison of classifiers over multiple datasets, Int. J. Comput. Sci. Math. 7 (5) (2016) 410–442.
[60] G.-G. Wang, S. Deb, Z. Cui, Monarch butterfly optimization, Neural Comput. Appl. 31 (7) (2019) 1995–2014.
[61] G.-G. Wang, S. Deb, L.D.S. Coelho, Earthworm optimisation algorithm: A bio-inspired metaheuristic algorithm for global optimisation problems, Int. J. Bio-Inspired Comput. 12 (1) (2018) 1–22.
[62] G.-G. Wang, S. Deb, L.d.S. Coelho, Elephant herding optimization, in: 2015 3rd International Symposium on Computational and Business Intelligence, ISCBI, IEEE, 2015, pp. 1–5.
[63] G.-G. Wang, Moth search algorithm: A bio-inspired metaheuristic algorithm for global optimization problems, Memetic Comput. 10 (2) (2018) 151–164.
[64] H. Chen, I. Ahmadianfar, G. Liang, H. Bakhsizadeh, B. Azad, X. Chu, A successful candidate strategy with runge-kutta optimization for multi-hydropower reservoir optimization, Expert Syst. Appl. (2022) 118383.
[65] W. Zhao, Z. Zhang, L. Wang, Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications, Eng. Appl. Artif. Intell. 87 (2020) 103300.

Soham Chattopadhyay completed his Bachelor's degree in Electrical Engineering in Jadavpur University, Kolkata, India. His research interests include machine learning, deep learning and optimization algorithms.

Arijit Dey completed his Bachelor's degree in Computer Science and Engineering in Maulana Abul Kalalm Azad University of Technology, Kolkata, India. His research interests include machine learning, deep learning and optimization algorithms.

Pawan Kumar Singh received his B.Tech degree in Information Technology from West Bengal University of Technology in 2010. He received his M.Tech in Computer Science and Engineering and Ph.D. (Engineering) degrees from Jadavpur University (J.U.) in 2013 and 2018 respectively. He also received the RUSA 2.0 fellowship for pursuing his post-doctoral research in J.U. in 2019. He is currently working as an Assistant Professor in the Department of Information Technology in J.U. He also worked as a visiting postdoctoral researcher in Dongguk University (South Korea), Sejong University (South Korea) and Nottingham Trent University (UK). He has authored or coauthored more than 100 research papers in peer-reviewed journals and international conferences. He also serves as Editorial Board Member, Reviewer, and Technical Program Committee Member for a number of IEEE and Springer journals and conferences. His areas of current research interest are Computer Vision, Pattern Recognition, Handwritten Document Analysis, Image & Video Processing, Feature Optimization, Machine Learning, Deep Learning and Artificial Intelligence. He is a senior member of the IEEE (U.S.A.), member of The Institution of Engineers (India) and Association for Computing Machinery (ACM) as well as a life member of the Indian Society for Technical Education (ISTE, New Delhi) and Computer Society of India (CSI).

Diego Oliva received the B.S. degree in Electronics and Computer Engineering from the Industrial Technical Education Center (CETI) of Guadalajara, Mexico in 2007, the M.Sc. degree in Electronic Engineering and Computer Sciences from the University of Guadalajara, Mexico in 2010. He obtained the Ph.D. in Informatics in 2015 from the Universidad Complutense de Madrid. Currently he is an Associate Professor at the University of Guadalajara in Mexico and has been visiting professor at the Tomsk Polytechnic University in Russia. He has the distinction of National Researcher Rank 2 by the Mexican Council of Science and Technology. Since 2022 he is Senior Member of the IEEE. His research interest includes Evolutionary and swarm algorithms, Hybridization of evolutionary and swarm algorithms, and Computational intelligence. He has been a guest editor for the journal of Mathematical Biosciences and Engineering of AIMS press, the journal of Remote Sensing and Entropy both of MDPI, Currently he is associate editor of IEEE Access, IEEE Latin America Transaction and Engineering Applications of Artificial Intelligence to mention some.

Erik Cuevas received the B.S. degree with distinction in Electronics and Communications Engineering from the University of Guadalajara, Mexico in 1995, the M.Sc. degree in Industrial Electronics from ITESO, Mexico in 2000, and the Ph.D. degree from Freie Universität Berlin, Germany in 2005. From 2001 he was awarded a scholarship from the German Service for Academic Interchange (DAAD) as fulltime researcher. Since 2007 he has been with University of Guadalajara, where he is currently a fulltime Professor in the Department of Electronics. From 2008, he is a member of the Mexican National Research System (SNI). His current research interest includes computer vision and artificial intelligence.

Ram Sarkar received his B. Tech degree in Computer Science and Engineering from University of Calcutta, India in 2003. He received his M.E. degree in Computer Science and Engineering and Ph.D. (Engineering) degree from Jadavpur University, India in 2005 and 2012 respectively. He joined in the department of Computer Science and Engineering of Jadavpur University as an Assistant Professor in 2008, where he is now working as a Full Professor. He received Fulbright-Nehru Fellowship (USIEF) for post-doctoral research in University of Maryland, College Park, USA in 2014–15. His areas of current research interest are Image Processing, Computer Vision, Machine Learning and Deep Learning. He is a senior member of the IEEE, and member of ACM.