

Design of ensemble recurrent model with stacked fuzzy ARTMAP for breast cancer detection

Recurrent model for detecting breast cancer

Abhishek Das and Mihir Narayan Mohanty

*Department of Electronics and Communication Engineering,
Institute of Technical Education and Research,
Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India*

Received 21 March 2022

Revised 23 April 2022

20 May 2022

Accepted 3 June 2022

Abstract

Purpose – In time and accurate detection of cancer can save the life of the person affected. According to the World Health Organization (WHO), breast cancer occupies the most frequent incidence among all the cancers whereas breast cancer takes fifth place in the case of mortality numbers. Out of many image processing techniques, certain works have focused on convolutional neural networks (CNNs) for processing these images. However, deep learning models are to be explored well.

Design/methodology/approach – In this work, multivariate statistics-based kernel principal component analysis (KPCA) is used for essential features. KPCA is simultaneously helpful for denoising the data. These features are processed through a heterogeneous ensemble model that consists of three base models. The base models comprise recurrent neural network (RNN), long short-term memory (LSTM) and gated recurrent unit (GRU). The outcomes of these base learners are fed to fuzzy adaptive resonance theory mapping (ARTMAP) model for decision making as the nodes are added to the F₂ layer if the winning criteria are fulfilled that makes the ARTMAP model more robust.

Findings – The proposed model is verified using breast histopathology image dataset publicly available at Kaggle. The model provides 99.36% training accuracy and 98.72% validation accuracy. The proposed model utilizes data processing in all aspects, i.e. image denoising to reduce the data redundancy, training by ensemble learning to provide higher results than that of single models. The final classification by a fuzzy ARTMAP model that controls the number of nodes depending upon the performance makes robust accurate classification.

Research limitations/implications – Research in the field of medical applications is an ongoing method. More advanced algorithms are being developed for better classification. Still, the scope is there to design the models in terms of better performance, practicability and cost efficiency in the future. Also, the ensemble models may be chosen with different combinations and characteristics. Only signal instead of images may be verified for this proposed model. Experimental analysis shows the improved performance of the proposed model. This method needs to be verified using practical models. Also, the practical implementation will be carried out for its real-time performance and cost efficiency.

Originality/value – The proposed model is utilized for denoising and to reduce the data redundancy so that the feature selection is done using KPCA. Training and classification are performed using heterogeneous ensemble model designed using RNN, LSTM and GRU as base classifiers to provide higher results than that of single models. Use of adaptive fuzzy mapping model makes the final classification accurate. The effectiveness of combining these methods to a single model is analyzed in this work.

Keywords Breast cancer detection, Kernel-PCA, Ensemble learning, Recurrent neural network, Long short-term memory, Gated recurrent unit, Fuzzy ARTMAP

Paper type Research paper

© Abhishek Das and Mihir Narayan Mohanty. Published in *Applied Computing and Informatics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Conflict of interest: The authors declare that there is no conflict of interest for this work.



Applied Computing and Informatics
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-03-2022-0075

1. Introduction

Cancer is triggered by various genetic and environmental factors. In most cases, cancers originated as malignant tumors. Rapid growth in these tumors leads to cancer. Breast cancer comes under this category and is deadly. Non-invasive kinds of cancer are observed within the milk ducts with no growth and no spread to nearby tissues, whereas invasive ductal carcinoma (IDC) is just opposite to the previous type and appears in 80% of all cases of breast cancer. Breast cancer is causing the highest mortality rate among all other cancers as per the report shown in Ref. [1] till 2020. In time detection of breast cancer may lead to an increase in the life span of the affected person. The cancer detection technique involves various imaging methods to picture the location, shape and size of the affected tissue producing cancer. The physicians manually observe the scanning reports which may lead to a time-consuming process. Automatic and accurate detection of breast cancer from the images obtained through various means can be preferred in the era of artificial intelligence.

Different deep learning models are preferred in various fields of medical image processing, signal processing, speech enhancement, speech recognition, image generation and also in many other areas. Researchers are attracted to these methods due to their human-like training algorithms as well as for high performance. Initially, single models like convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM) and gated recurrent unit (GRU) have been designed for specific purposes. Then, the era of the concatenation of these models started to study the performance. Few concatenated models were designed by concatenating the LSTM layers with convolution layers [2], and even, the traditional machine learning model support vector machine (SVM) [3] has also been attached as the classifier with convolutional layers.

Ensemble learning emerged as the highly improved performing method as compared to single deep learning models. Ensemble learning models are broadly classified as homogeneous and heterogeneous depending upon the types of base learners chosen for initial stage processing. An ensemble model designed with the same base models is termed homogeneous, whereas different base models lead to a heterogeneous ensemble model [4]. The heterogeneous ensemble learning method has been mostly preferred over homogeneous ones. A deep ensemble learning-based approach for accurate detection of breast cancer is used. The main contributions of this work are summarized in the points as follows:

- (1) The proposed model is utilized for denoising and to reduce the data redundancy so that the feature selection is done using KPCA;
- (2) Training and classification are performed using a heterogeneous ensemble model designed using RNN, LSTM and GRU as base classifiers to provide higher results than that of single models and
- (3) The final classification by a fuzzy adaptive resonance theory mapping (ARTMAP) model that controls the number of nodes depending upon the performance makes it robust for accurate classification.

The rest part of the paper is arranged subsequently: The second section describes the various works proposed in the field of breast cancer detection. The third section provides a detailed description of the proposed technique designed for breast cancer detection. The fourth section delivers the outcomes found from the proposed model. In the fifth section, conclusions are drawn along with the future scope of this work followed by the references.

2. Literature survey

The deep learning algorithm has appeared as the most chosen method for the detection of breast cancer. Deep learning is performing with more accuracy on image and signal data in single [5, 6] and hybrid forms [7, 8]. Principal component analysis (PCA) was used to reduce

the data size followed by feature extraction by using a multilayer perceptron (MLP) model. The final classification was done by using SVM [9]. Semantic segmentation before detection using SVM and MLP has been used for breast cancer detection [10] from histopathology images. Color normalization followed by enhancement was carried out as preprocessing step. Refinement (R), correlation (C) and adaptive (A) algorithms were used for fine-tuning the features and classification by AdaBoost-oriented tree model [11].

Deep learning techniques are gaining the interest of researchers these days. Mitosis count-based cancer detection from breast histopathology images has been proposed using the Atrous Fully Connected Neural Network (A-FCNN) for segmentation and multi-scale and region-based CNN (MS-RCNN) model for detection [12]. A multi-instant pooling layer-based CNN (MI-CNN) model [13] has been suggested for breast cancer detection from histopathology images. Mislabeled patch correction and classification of histopathology images have been proposed for breast cancer detection [14]. The anomaly detection using generative adversarial network (AnoGAN) was used for anomaly detection, whereas classification was done using DenseNet-121. A deep learning framework has been proposed for mitotic cell count-based cancer detection from histopathology images using Recurrent Residual U-Net (R2U-Net) for segmentation and Inception Recurrent Residual CNN (IRRCNN) for classification [15]. An auto-encoder designed with the concept of the residual-CNN model [16] has been proposed for cancer detection from hematoxylin and Eosin-stained histopathology images. The use of CNN as a feature extractor and Extreme Machine Learning (ELM) as a classifier has been found in Ref. [17] for malignant versus benign classification. A combination of SVM as anomaly detector and FCNN as a classifier has been proposed in Ref. [18] for breast cancer detection from histopathology images. Labeling the data and then classification has been proposed for cancer detection from histopathology images of the BreastHis Dataset using a deep active learning model [19]. Mitotic cell detection for breast cancer classification has been proposed using two CNNs connected in parallel [20]. Application of empirical wavelet transform (EWT), as well as variational mode decomposition (VMD) as preprocessors followed by ensemble of three CNNs and MLP [21], has been proposed for breast cancer detection from histopathology images. That work also used gene data from breast cancer for image conversion using the DeepInsight framework [22]. For bioimage classification, an ensemble model containing CNNs as a base classifier and a sum rule for final decision-making has been proposed [23]. A modified residual neural network [24] has been proposed for breast cancer detection from histopathology images using modified ResNet34 and modified ResNet50 models. A Stochastic Dilated Residual Ghost (SDRG) method [25] has been proposed for cancer detection from breast histopathology images. In recent work, the combination of the attention technique and residual CNN model [26] has been utilized for breast cancer detection. A combination of the Xception model as feature extractor and radial basis function (RBF) kernel-based SVM as classifier [27] has been proposed for breast cancer detection from histopathology images. The authors have also studied the effect of magnification factors on performance. A combined deep learning model [28] designed using inception and residual blocks has been proposed for cancer detection from histopathology images. The magnification values are also considered for images as a preprocessing step.

From the literature analysis, it is found that ensemble models in this field have not yet been explored. In this work, we have considered the detection of breast cancer from breast histopathology images using data denoising by kernel principal component analysis (KPCA) followed by ensemble recurrent models stacked with fuzzy ARTMAP.

3. Proposed method

It has been studied in various literature that the heterogeneous form of ensemble learning provides better performance in comparison to homogeneous ensemble models and single models due to the use of different fine-tunes algorithms [29–31]. KPCA [32] is used to denoise the normalized raw signal and reduce the redundant information. The ensemble model is of

three recurrent models, i.e. RNN, LSTM and GRU. The outcomes are fed to the ARTMAP model for final detection. The workflow diagram of the proposed model is shown in Figure 1.

The parameters which are set before training to obtain better performance and accuracy are termed hyper-parameters. In this work, each image of the dataset has 2500 numbers of pixels. Such enormous data will cause higher processing time. The presence of noise in the image data also increases the data complexity. To lower this effect of data complexity, KPCA is used to select the 500 highest eigenvalues. The data are mapped into higher space with the use of KPCA. These features are used to train each base model for further processing and feature extraction. To receive the 500 features provided by the KPCA, each neural network model has 500 nodes in the input layer. Each base model has 2 hidden layers with 32 and 64 numbers of nodes for deep feature extraction. Each model provides two predictions on the same data with two nodes in the output layer. The prediction values are the corresponding probabilities generated by the base models along with the labels of data. These parameters are concatenated to form six values (two values from three base classifiers) as metadata used to train the meta classifier fuzzy ARTMAP for final classification.

3.1 Dataset

The proposed model is experienced with the dataset [33]. The dataset contains 198,738 non-IDC images and 787860 images of the IDC category. 0 as class value represents non-IDC and IDC is considered as 1. Sample images from each category are shown in Figure 2. As 80% of the total dataset is used for training, the rest 20% is used for testing. However, the training set

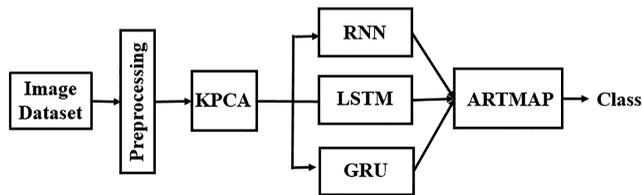


Figure 1.
Workflow diagram of the proposed model

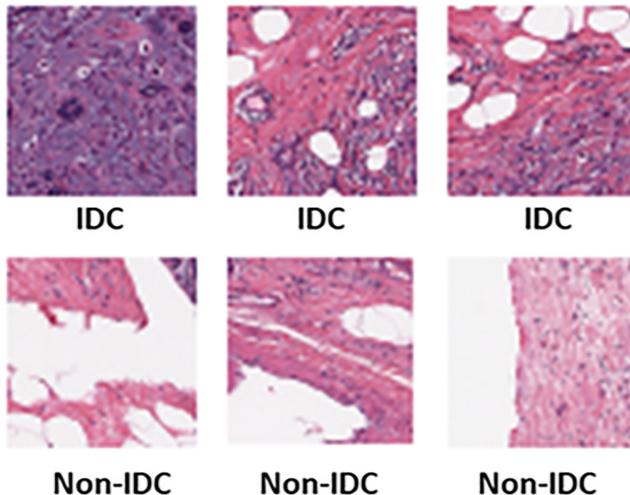


Figure 2.
Samples of breast histopathology images

is again divided into train and validation subsets using 80:20 ratios. Validation data are used for training, whereas testing data are neither used for training nor validation. Test data are separate and meant for testing only. The accuracy along with the robustness is explained in the result section.

3.2 Preprocessing

The color images of the breast histopathology are converted to grayscale images to reduce the computational requirements. The height and width values of each image remain the same as it is. Only the depth of each image was reduced from 3 to 1, i.e. the $50 \times 50 \times 3$ images are converted to $50 \times 50 \times 1$ sized images. Figure 3 shows the sample image after conversion from a color image to a gray image.

The grayscale images with attributes (d) are then normalized using the mean normalization technique as given in Eq. (1).

$$d' = \frac{f - \text{average}(d)}{\max(d) - \min(d)} \quad (1)$$

3.3 Image denoising and feature selection

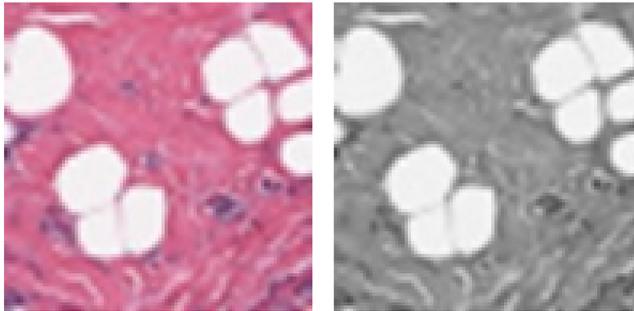
RNN model may pass through vanishing gradient problems due to the large size of data, whereas the LSTM and GRU models are free from this problem [34]. Each image of the dataset has 2,500 numbers of pixels. Such enormous data will cause higher processing time. The presence of noise in the image data also increases the data complexity. To lower this effect of data complexity, KPCA is used to select the 500 highest eigenvalues. Cosine kernel is considered as it is a metric that measures similar documents irrespective of size.

Let the data $\{d_i\} \in \mathbb{R}^D \forall_i = 1 \dots n$, of the high dimension D to higher dimension space $\phi(d_i)$. $\phi(d_i)$ is the kernelized version of the input data space and the ability to capture the reduced form of the data. It is common to consider the feature space to have zero mean, which is given by Eq. (2).

$$\frac{1}{N} \sum_{i=1}^N \phi(d_i) = 0 \quad (2)$$

The covariance matrix is then computed using Eq. (3).

$$\Sigma = \frac{1}{N} \sum_{I=1}^N \phi(d_i)\phi(d_i)^T \quad (3)$$



(a)

(b)

Figure 3.
(a) Dataset image of size $50 \times 50 \times 3$ and (b) images of size 50×50 after conversion

ACI

The corresponding eigenvalues and eigenvectors are given by Eq. (4).

$$\sum u_i = \lambda_k u_i \quad \forall_i = 1, \dots, k \quad (4)$$

Substituting Eq. (3) in Eq. (4), we get

$$\frac{1}{N} \sum_{l=1}^N \phi(d_l) \{ \phi(d_l)^T u_k \} = \lambda_k u_k \quad (5)$$

$$\text{Hence, } u_k = \sum_{i=1}^N a_{ki} \phi(d_i) \quad (6)$$

$$\text{Where } a_{ki} = \frac{1}{\lambda_k N} \phi(d_i)^T u_k \quad (7)$$

According to Mercer's theory, the kernel κ is represented by

$$\kappa(d_i, d_j) = \phi(d_i)^T \phi(d_j) \quad (8)$$

Substituting Eq. (6) into Eq. (5) and multiplying $\phi(d_i)$. On both the sides of equations, we get

$$\frac{1}{N} \sum_{i=1}^N \phi(d_i) \phi(d_i) \sum_{j=1}^N a_{kj} \phi(d_j) \phi(d_j) = \lambda_k a_{ki} \phi(d_i) \phi(d_i) \quad (9)$$

Rewriting Eq. (9) in terms of the kernel, yields

$$\frac{1}{N} \kappa(d_i, d_i) \sum_{i=1}^N a_{kj} \kappa(d_i, d_j) = \lambda_k \sum_{i=1}^N \kappa(d_i, d_i) \quad (10)$$

$$\mathbf{K}^2 a_k = \lambda_k N a_k \quad (11)$$

$$\text{Where } \mathbf{K} = \kappa(d_l, d_i) \quad (12)$$

and a_k represents the N-dimensional column eigenvectors of a_{ki} .

The denoised form obtained using KPCA is given by

$$\hat{d} = \phi(d)^T u_k = \sum_{i=1}^N a_{ki} \kappa(d, d_i) \quad (13)$$

After computing the zero mean of the kernel, we obtained

$$\hat{\mathbf{K}} = \mathbf{K} - \frac{1}{N} \mathbf{K} - \mathbf{K} \frac{1}{N} + \frac{1}{N} \mathbf{K} \frac{1}{N} \quad (14)$$

where $\mathbf{K} = \mathbf{K}_{ij}$. $\hat{\mathbf{K}}$ is known as the Gram matrix.

In this work, we have used a cosine kernel in the KPCA. The cosine kernel is given by Eq. (15).

$$\kappa(d_i, d_j) = \frac{d_i d_j^T}{\|d_i\| \|d_j\|} \quad (15)$$

3.4 Training

The denoised data are used to train the three base learners connected in parallel to each other. All the base models are designed with four layers. The output layer has two nodes activated

with the SoftMax activation function as two classes are there for classification. These deep learning models are evaluated using the binary cross-entropy loss (BCE) that is mathematically represented by Eq. (16) as follows:

$$BCE(y_i, \hat{y}_i) = - \sum_{l=1}^2 y_i \log \hat{y}_i \quad (16)$$

where y_i represents the actual prediction given as target and \hat{y}_i represents the experimental prediction done by the model.

From Figure 1, it can be visualized that the base classifiers are indirectly creating the features for ARTMAP. Majority voting, MLP [35] and Fuzzy Min-Max [36] models have been used as the meta classifier in various works. In those models, the number of nodes in each hidden layer is fixed to a certain value before training and cannot be updated according to the training requirements. But in fuzzy ARTMAP, the nodes are added to the F_2^a layer if the winning criteria are fulfilled. That makes the ARTMAP model more robust in comparison to earlier ones. ARTMAP is designed with the structure as shown in Figure 4.

Fuzzy ARTMAP is designed with two ART models, i.e. ART_a and ART_b, connected by a map field [37] that is capable of developing maps between the clusters generated in the input domain ART_a and output domain ART_b. In Figure 4, F_0 , F_1 , and F_2 represent the normalization layer, input layer and recognition layer, respectively, for both the domains.

The ART_a module receives the concatenated outputs a of M dimension from ensemble recurrent models and forms a $2M$ -dimensional complement-coded vector A as given in Eq. (17).

$$A = (a, a^c) = (a_1, \dots, a_M, 1 - a_1, \dots, 1 - a_M) \quad (17)$$

The category nodes in the network are selected using the category choice function (CCF) given by Eq. (18).

$$\frac{|A \wedge W_j^a|}{|W_j^a| + \alpha_a} = Z_j(A) \quad (18)$$

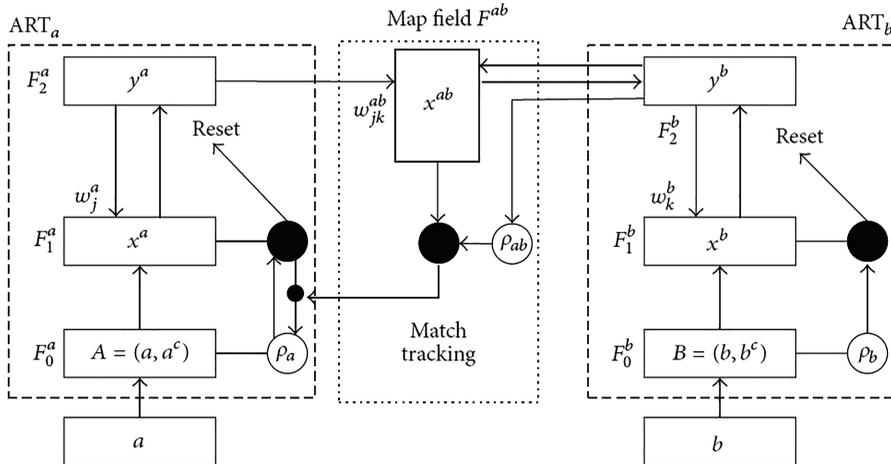


Figure 4.
Structure of the
ARTMAP

ACI

where Λ represents the min operator, W_j^a are the weights learned by the j th node and α_a represents the ART_a choice function.

Once the winning node is selected, a similarity check vigilance test is carried out against a vigilance parameter ρ_a , where

$$\frac{|A \wedge W_j^a|}{|A|} \geq \rho_a \quad (19)$$

Resonance occurs when the above winning criterion is fulfilled and the weights are updated using the weight update equation given in Eq. (20).

$$w_j^{new} = \beta (A \wedge w_j^{old}) + (1 - \beta)w_j^{old} \quad (20)$$

where β is the learning rate.

A new node is generated when the winning criterion is fulfilled in the F_2^a layer. The following algorithm is utilized in training the proposed model. The symbols used in the algorithm are as follows:

d_i : image data;

y_i : target;

n : dataset size and

B : base classifier

Algorithm 1: The complete proposed method

Input: Dataset $D = \{d_i, y_i\}_{i=1}^n$

Output: Class

Step 1. Preprocessing

- a. Resize the images to 50x50
- b. Denoise the data using kPCA
- c. $\hat{d} = kPCA(d)$
- d. $D_{new} = \{\hat{d}_i, y_i\}_{i=1}^n$

Step 2. Train the Base classifiers

- a. for $t = 1$ to 3
- b. Train B_t with D_{new}
- c. End for

Step 3: Form the input for Meta Learner

- a. for $t = 1$ to 3
- b. $D_{Meta} = \{a_t, b_t\}$, where $a_t = B_1(\hat{d}_i), B_2(\hat{d}_i), B_3(\hat{d}_i)$
Where b_t are the predicted labels of each image
- c. end for

Step 4: Train Meta Learner

- a. Class = ARTMAP(D_{meta})
 - b. Return Class
-

4. Results and discussion

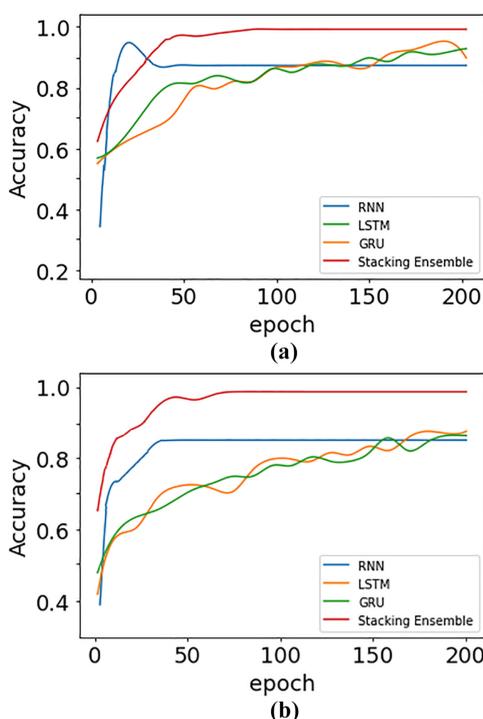
4.1 Denoised features for training and testing

The size of each image of the dataset is denoised and the 2,500 numbers of pixel values are converted to high priority 500 values using KPCA. It is maintained because a low number of features degrade the performance of deep learning models.

4.2 Training, validation and classification results

The three recurrent models are trained with the denoised data and the well-fitted models are stored for stacking with each other. The performance of each model is represented in terms of accuracy in Figure 5. The overall performance of the stacked ensemble model is also shown in the same figure to show the improvement in the performance.

The details of accuracies are provided in Table 1. Figure 6 provides the confusion matrix for each base model and the whole proposed model for validation data.



Note(s): (a) Training and (b) validation

Figure 5.
Accuracy
comparison plot

Model	Accuracy (%)	F1-score	Training				Validation accuracy (%)
			Recall	Precision	Sensitivity	Specificity	
RNN	88.46	0.9115	0.9035	0.9196	0.9035	0.7857	87.18
LSTM	93.59	0.9333	0.9210	0.9459	0.9211	0.8571	90.38
GRU	90.03	0.9292	0.9210	0.9375	0.9211	0.8333	89.74
Proposed stacked ensemble	99.36	0.9912	0.9912	0.9912	0.9912	0.9762	98.72

Table 1.
Performance
comparison of base
models and
proposed model

ACI

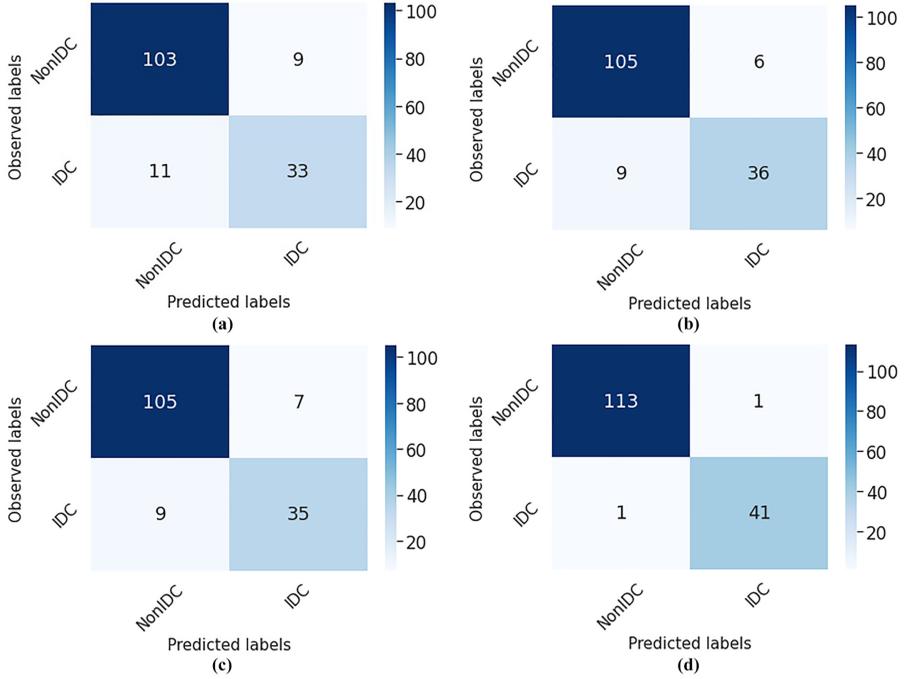


Figure 6.
Confusion matrix

Note(s): (a) RNN, (b) LSTM, (c) GRU and (d) proposed stacked ensemble model

The performance of the whole model is increasing in comparison to base models due to the second stage training by the ARTMAP which helps in decision making.

The evaluation parameters such as accuracy, F1-score, recall, precision, sensitivity and specificity are calculated using the mathematical expressions given in Eq. (21) to Eq. (26).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (21)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (23)$$

$$F1 \text{ Score} = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (24)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (25)$$

$$Specificity = \frac{TN}{TN + FP} \quad (26)$$

The proposed model is also compared to a few *state-of-the-art* models and is provided in Table 2. The results in italics show the highest result.

The training and validation loss graphs of the proposed stacked ensemble model are shown in Figure 7.

Work	Method used Preprocessing	Feature extraction	Classification	Accuracy (%)	F1- score
Chiu <i>et al.</i> [9]	Data reduction using PCA	MLP	SVM	86.97	–
Rashmi <i>et al.</i> [10]	Color normalization	Texture Feature	SVM + MLP	–	0.83
Li <i>et al.</i> [11]	–	RCA	AdaBoost	93.30	–
Kausar <i>et al.</i> [12]	Segmentation using A-FCNN	MS-RCNN	–	–	0.902
Das <i>et al.</i> [13]	–	MI-CNN	–	96.63	–
Alom <i>et al.</i> [15]	Segmentation using R2U-Net	IRRCNN	–	–	0.878
Brancati <i>et al.</i> [16]	–	Residual-CNN	–	89.57	–
Wang <i>et al.</i> [17]	–	CNN	ELM	86.50	–
Li <i>et al.</i> [18]	–	SVM	CNN	76	–
Qi <i>et al.</i> [19]	–	Pre-trained AlexNet	–	92.84	–
Sebai <i>et al.</i> [20]	–	Ensemble of 2 CNNs	–	–	0.698
Das <i>et al.</i> [21]	EWT and VMD	Ensemble of 3 CNNs	MLP	98.08	0.98
Sharma <i>et al.</i> [27]	Magnification	Xception	SVM R	96.25	0.96
<i>Proposed Work</i>	Data denoising using kPCA	Ensemble of RNN, LSTM and GRU	Fuzzy ARTMAP	98.72	0.985

Table 2.
Performance
comparison of the
proposed model with
related works

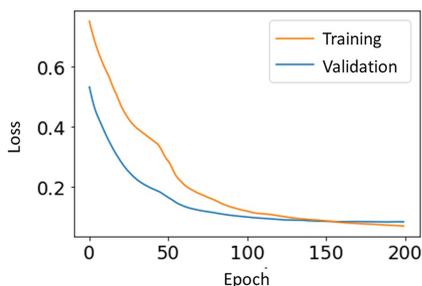


Figure 7.
Loss plots for training
and validation

Very low loss values, i.e. 0.0192 and 0.0331 in training and validation, respectively, represent the efficiency of the proposed model in classifying breast histopathology images classification.

4.3 Discussion

The number of base learners is chosen as three for certain reasons. Two or four base models may affect the performance of the meta classifier if half of them will provide opposite results to that of another half. Three base learners are chosen so that the training of the meta classifier will not be adversely affected and the computation time will be less in comparison to five base classifiers.

Figure 5 shows the accuracy of base models with the proposed stacked ensemble model. As it is observed, the accuracy of each individual starts from 40% and increased gradually. However, the stacked model provides 99.36% accuracy, which is given in Table 1. Also, this is proved in terms of the confusion matrix given in Figure 6. A total of 156 images were taken from the test data including both categories. In total, 114 images were from non-IDC images, whereas 42 number images belong to IDC. From this figure, it is observed that the three base learners have the results of a maximum of 105 images as true positive (TP), whereas the

stacked ensemble models detected 113 images as TP. Similarly, the proposed model predicted 41 images as true negative (TN) out of 42 IDC images, whereas this count is 33, 36 and 35, respectively, for RNN, LSTM and GRU models. Very low numbers of data are classified as false positive (FP) and false negative (FN) by the proposed model. Data overfitting is analyzed from the validation loss plot shown in Figure 7. The proposed model is free from data overfitting which is verified through convergence. Within 200 epochs, the convergence is found for both training and validation that proves the efficacy. Table 1 and Table 2 are showing the improvement in performance due to the proposed stacked ensemble method in comparison to both base models and *state-of-the-art* methods, respectively. The proper utilization of data preprocessing to avoid data redundancy and a suitable number of base classifiers of the proposed ensemble model combined together improved the performance.

5. Conclusion

In this work, we have utilized the data denoising by KPCA to reduce the data complexity burden from the classification model for faster processing. Features are extracted using the best form of deep learning, i.e. the deep ensemble learning developed using three recurrent models. The final classification is done with the stacked fuzzy ARTMAP. The proposed model is also free from data overfitting by considering a suitable number of iterations. It provides that the histopathology images are efficiently classified into IDC and non-IDC with 99.36% training and 98.72% validation accuracy.

Research in the field of medical applications is an ongoing method. More advanced algorithms are being developed for better classification. Still, the scope is there to design the models in terms of better performance, practicability and cost efficiency in the future. Also, the ensemble models may be chosen with different combinations and characteristics. Only signals instead of images may be verified for this proposed model. Experimental analysis shows the improved performance of the proposed model. This method needs to be verified using practical models and is kept for future work. Also, the practical implementation will be carried out for its real-time performance and cost efficiency.

References

1. Available from: <http://gco.iarc.fr/> (accessed 15 February, 2022).
2. Wang, J, Yu, LC, Lai, KR, Zhang, X (2016). Dimensional sentiment analysis using a regional CNN-LSTM model. in Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers), 225-30.
3. Xue, DX, Zhang, R, Feng, H, Wang, YL (2016). CNN-SVM for microvascular morphological type recognition with data augmentation. J Med Biol Eng, 36(6), 755-64.
4. Nguyen, TT, Nguyen, MP, Pham, XC, Liew, AWC (2018). Heterogeneous classifier ensemble with fuzzy rule-based meta learner. Inf Sci, 422, 144-60.
5. Mambou, SJ, Maresova, P, Krejcar, O, Selamat, A, Kuca, K (2018). Breast cancer detection using infrared thermal imaging and a deep learning model. Sensors, 18(9), 2799.
6. Shen, L, Margolies, LR, Rothstein, JH, Fluder, E, McBride, R, Sieh, W (2019). Deep learning to improve breast cancer detection on screening mammography. Scientific Rep, 9(1), 1-12.
7. Ragab, DA, Sharkas, M, Marshall, S, Ren, J (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ, 7, e6201.
8. Wang, D, Khosla, A, Gargeya, R, Irshad, H, Beck, AH (2016). Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718.
9. Chiu, H-J, Li, T-HS, Kuo, P-H (2020). Breast cancer-detection system using PCA, multilayer perceptron, transfer learning, and support vector machine. IEEE Access, 8, 204309-24.

10. Rashmi, R, Prasad, K, Udupa, CBK, Shwetha, V (2020). A comparative evaluation of texture features for semantic segmentation of breast histopathological images. *IEEE Access*, 8, 64331-46.
11. Li G, Yuan T, Li C, Zhuo J, Jiang Z, Wu J, Ji D, Zhang H. Effective breast cancer recognition based on fine-grained feature selection. *IEEE Access*. 2020; 8: 227538-55.
12. Kausar, T, Wang, M, Ashraf, MA, Kausar, A (2020). SmallMitosis: small size mitotic cells detection in breast histopathology images. *IEEE Access*, 9, 905-22.
13. Das, K, Conjeti, S, Chatterjee, J, Sheet, D (2020). Detection of breast cancer from whole slide histopathological images using deep multiple instance CNN. *IEEE Access*, 8, 213502-11.
14. Man, R, Yang, P, Xu, B (2020). Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks. *IEEE Access*, 8, 155362-77.
15. Alom, MZ, Aspiras, T, Taha, TM, Bowen, T, Asari, VK (2020). MitosisNet: end-to-end mitotic cell detection by multi-task learning. *IEEE Access*, 8, 68695-710.
16. Brancati, N, De Pietro, G, Frucci, M, Riccio, D (2019). A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. *IEEE Access*, 7, 44709-20.
17. Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H, Xin J. Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*. 2019; 7: 105146-58.
18. Li, X, Radulovic, M, Kanjer, K, Plataniotis, KN (2019). Discriminative pattern mining for breast cancer histopathology image classification via fully convolutional autoencoder. *IEEE Access*, 7, 36433-45.
19. Qi Q, Li Y, Wang J, Zheng H, Huang Y, Ding X, Rohde GK. Label-efficient breast cancer histopathological image classification. *IEEE J Biomed Health Inform*. 2018; 23(5): 2108-16.
20. Sebai, M, Wang, T, Al-Fadhli, SA (2020). PartMitosis: a partially supervised deep learning framework for mitosis detection in breast cancer histopathology images. *IEEE Access*, 8, 45133-47.
21. Das, A, Mohanty, MN, Mallick, PK, Tiwari, P, Muhammad, K, Zhu, H (2021). Breast cancer detection using an ensemble deep learning method. *Biomed Signal Process Control*, 70, 103009.
22. Sharma, A, Vans, E, Shigemizu, D, Boroevich, KA, Tsunoda, T (2019). DeepInsight: a methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific Rep*, 9(1), 1-7.
23. Nanni L, Ghidoni S, Brahnam S. Ensemble of convolutional neural networks for bioimage classification. *Appl Comput Inform*. 2021; 17(1): 19-35.
24. Gupta, V, Vasudev, M, Doegar, A, Sambyal, N (2021). Breast cancer detection from histopathology images using modified residual neural networks. *Biocybernetics Biomed Eng*, 41(4), 1272-87.
25. Kashyap, R (2022). Breast cancer histopathological image classification using stochastic dilated residual Ghost model. *Int J Inf Retrieval Res (IJIRR)*, 12(1), 1-24.
26. Zou, Y, Zhang, J, Huang, S, Liu, B (2022). Breast cancer histopathological image classification using attention high-order deep network. *Int J Imaging Syst Technol*, 32(1), 266-79.
27. Sharma, S, Kumar, S (2022). The Xception model: a potential feature extractor in breast cancer histology images classification. *ICT Express*, 8(1), 101-8.
28. Singh, S, Kumar, R (2022). Breast cancer detection from histopathology images with deep inception and residual blocks. *Multimedia Tools Appl*, 81(4), 5849-65.
29. Chali, Y, Hasan, SA, Mojahid, M (2014). Complex question answering: homogeneous or heterogeneous, which ensemble is better? in *International Conference on Applications of Natural Language to Data Bases/Information Systems* (pp. 160-3). Springer.
30. Alshdaifat, E.A., Al-hassan, M., Aloqaily, A (2021). Effective heterogeneous ensemble classification: an alternative approach for selecting base classifiers. *ICT Express*, 7(3), 342-9.
31. Zhong Y, Chen W, Wang Z, Chen Y, Wang K, Li Y, Yin X, Shi X, Yang J, Li K. HELAD: a novel network anomaly detection model based on heterogeneous ensemble learning. *Computer Networks*. 2020; 169: 107049.

-
32. Schölkopf, B, Smola, A, Müller, K-R (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput*, 10(5), 1299-319.
 33. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform*. 2016; 7(1): 29.
 34. Das A, Patra G R, Mohanty M N. A comparison study of recurrent neural networks in recognition of handwritten Odia numerals. *Advances in electronics, communication and computing*. Singapore: Springer; 2021. 251-260.
 35. Das A, Mohanty MN. Classification of magnetic resonance images of brain using concatenated deep neural network. *Int J Model, Identification and Control, Inderscience*. In press.
 36. Das, A, Mohapatra, SK, Mohanty, MN (2022). Design of deep ensemble classifier with fuzzy decision method for biomedical image classification. *Appl Soft Comput*, 115, 108178.
 37. Carpenter, GA, Grossberg, S, Markuzon, N, Reynolds, JH, Rosen, DB (1992). Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans Neural netw*, 3(5), 698-713.

Corresponding author

Mihir Narayan Mohanty can be contacted at: mihir.n.mohanty@gmail.com