Article - Engineering, Technology and Techniques

Multi-Feature Classification of Breast Cancer Histopathology Images: An Experimental Investigation in Machine Learning and Deep Learning Paradigm

Kushangi Atrey[1]
https://orcid.org/0000-0001-6767-8447

Bikesh Kumar Singh[1*]
https://orcid.org/0000-0002-5052-9768

Narendra Kuber Bodhey[2]
https://orcid.org/0000-0003-1736-3329

[1]National Institute of Technology Raipur, Department of Biomedical Engineering, Raipur, Chhattisgarh, India; [2]All India Institute of Medical Sciences, Department of Radiodiagnosis, Raipur, Chhattisgarh, India.

*Correspondence: bsingh.bme@nitrr.ac.in; Tel.: +91-9826469522 (B.K.S.)

HIGHLIGHTS

- The proposed approach performs the classification of breast tumors in histopathology images.

- The proposed approach evaluates a multi-feature classifier for risk stratification.

- The performance of different classifiers is compared under different data division protocols.

- The highest classification accuracy of 99.85% after feature selection is reported.

Abstract: The existing practice for Breast Cancer (BC) characterization includes histopathological analysis, which is tedious and time-consuming due to massive data analysis. Further, such techniques are subjected to inter-and intra-observer variability due to the non-availability of skilled pathologists, particularly in low resource settings. Thus, we propose a multi-feature classification technique for risk stratification of BC in Histopathology Images (HI) using machine learning strategies and a Long Short-Term Memory (LSTM) based deep learning approach. Experiments are performed on a publicly available HI database from which a total of 658 image features are extracted, while 192 relevant features are obtained after feature selection using genetic algorithm. The highest accuracy of 99.85% using 192 features under the 5-fold data division protocol is obtained with the LSTM approach. The proposed framework for analyzing HI using multiple grayscale and color features showed promising results and can be an effective tool in the histopathology laboratory.

Keywords: Intelligent laboratory analysis; Breast cancer; Feature fusion; Machine learning; Deep learning.

INTRODUCTION

Background

Breast Cancer (BC) is considered one of the most common cancers in women worldwide [1]. The worldwide estimated new cases of female BC are 2,261,419 for the year 2020 [2]. In India, 179,790 BC cases are projected for 2020 [3]. The existing techniques of breast lesion detection include a clinical examination of the breast, mammograms [4], [5], Magnetic Resonance Imaging (MRI), and ultrasound [6]. A biopsy is usually carried out for a final confirmation to determine if a tumor is cancerous (malignant) or not. A biopsy involves the microscopic analysis of Histopathology Images (HI) corresponding to the biopsy sample carried out by a pathologist [7]. In a histopathology laboratory, a particular trail needs a trained specialist with several years of experience to recognize abnormal tissues under microscopes accurately. Such examination needs both tissue diagnosis and analytical estimation based on tissue structure and cell morphology. However, there is a high level of inconsistency in the appearance of the tissue because of irregularity in the staining procedure. Thus, the manual approach of tissue characterization is challenging as it requires experienced and skilled pathologists. Further, such a tedious task consumes more time and is subjected to inter-and intra-observer variability. This led to the emergence of so-called Computer-Aided Diagnosis (CAD), which can decrease misdiagnosis by reducing the workload on pathologists and providing objective evidence for tissue characterization [8]. Further, it can help reduce the errors arising out of inter-and intra-observer variability in diagnosis.

This paper presents and evaluates a machine learning and deep learning strategy for automatic risk stratification of BC in HI. A database consisting of 2015 images including both benign (645 images) and malignant (1370 images) cases are used. A total of 658 image descriptors are extracted and evaluated using machine learning classifiers like Back-Propagation Artificial Neural Network (BPANN), Support Vector Machine (SVM), Discriminant Analysis, Logistic Regression (LR), k- Nearest Neighbor (k-NN), and Naïve Bayes (NB). Also, results are shown after implementing feature selection using a Genetic Algorithm (GA) with 192 selected features. This work also employs a Long Term Short Memory (LSTM) based on Deep Learning (DL) combined with genetic optimization for BC HI analysis. Further, the results are compared with some of the recently reported techniques. To the best of the knowledge of the authors, this is the first study evaluating such a large comprehensive set of descriptors for breast tissue classification in HI.

Related work

In this section, few recent works in the related area are discussed. Dora and coauthors [9] proposed a work on Fine Needle Aspirates (FNA) images of breast tissue. They implemented a Gauss-Newton Representation Based Algorithm (GNRBA) to classify breast tumors.

A deep convolutional neural network (CNN) for the classification of BC using HI is proposed by Wei and coauthors [1]. The proposed method obtained classification accuracy up to 97% for 40X magnification images. Classification using CNNs on a dataset of breast HI is reported by Araújo and coauthors [10]. For the training of the SVM classifier, they used features extracted by CNN. Further, they obtained the optimal parameters using a Radial Basis Function (RBF) kernel with a 3-fold data division on the training data. The best results were obtained using majority voting with accuracies of 77.8% for four classes and 83.3% for carcinoma or non-carcinoma for image-wise classification. Two different sources were used to collect the data by Motlagh and coauthors [11]. Red Green Blue (RGB) color-map was used for preserving the tissue structures of HI. With the DL-based approach and dividing data as 90% training and 10% testing, the ResNet V1 152 achieved a classification accuracy of 98.7%. An approach for the detection and classification of BC in HI using deep CNN is proposed by Rahhal and Mahmoud [12]. They divided the dataset as 70% for training and 30% for testing. The proposed method with the Visual Geometry Group (VGGm) model achieved a high value of classification accuracy 86.80% at the patient level. BC classification on 7909 microscopic images consisting of benign and malignant breast tumors using CNNs is performed by Bardou and coauthors [13]. They used Dense Scale Invariant Feature Transform features and Speeded-Up Robust Features (SURF) as local descriptors. For binary classification, the classification accuracy was achieved between 96.15% and 98.33%, and for multi-class classification, the accuracy obtained was between 83.31% and 88.23.

A CNN, an LSTM, and CNN+LSTM for the classification of BC are reported by Nahid and coauthors [14]. The online dataset consisting of 7909 images was used. They extracted the structural and statistical information from the images. Softmax and SVM layers were used for decision-making after feature extraction. With the 200X dataset, the best accuracy of 91.00% was achieved, the best precision value of 96.00% was achieved on the 40X dataset and for both 40X and 100X datasets, and the best F-Measure value was obtained. A DL-based method to classify the breast tissue images is reported by Golatkar and coauthors [15].

A dataset comprising 400 histology microscopic images were used. The average accuracy of 85% (four classes) and 93% for benign vs. malignant were achieved. Nahid and Kong [16] proposed the classification task of BC HI using CNN. They implemented five different CNN models and achieved the best performance on a 200X dataset with accuracy and F-measure of 97.19% and 98%, respectively.

Alirezazadeh and coauthors [17] performed the classification using a histopathological image dataset. They extracted features using LBP, PFTAS, and local phase quantization (LPQ). These features are used for representation learning and to obtain a projection matrix, after which the classification is carried out. They also evaluated the statistical significance using the paired t-test. The highest accuracy is obtained for 200X magnification as 91%, while an overall highest average accuracy using the proposed method is achieved as 88.5%. The Squeeze-and-Excitation block of the SE-ResNet module, in addition to a new learning rate scheduler to automatically classify BC histology images, is proposed by Jiang and coauthors [18]. For binary classification, highest accuracy values 99.34% (for 200X magnification) and 93.81% (for 100X magnification) for the multi-class classification were achieved.

Beevi and coauthors [19] performed mitosis detection on a dataset collected from two sources. Using a pre-trained VGGNet CNN model, they extracted deep features and performed the classification task. The results show that they achieved the highest F-score of 89.66% and accuracy of 90%. Singh [20] in the year 2019 implemented various classifiers like SVM, NB, Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), LR, K-NN, and Random Forest (RF) on anthropometric and clinical features and performed the risk prediction for BC. Feature selection and statistical significant analysis are also carried out. With the hold-out data division strategy, an accuracy of 92.105% is achieved with a medium K-NN classifier. Tong and coauthors [21] considered integrating the multi-omics data to improve the survival prediction of BC. Their output shows that combining the DNA methylation and miRNA expression gives the best performance of 0.641 ± 0.031 with concatenation autoencoder (ConcatAE). Another approach is proposed by Wang and coauthors [22] using a Double Deep Transfer Learning (D2TL) and Interactive Cross-task Extreme Learning Machine (ICELM) for feature extraction and classification of breast HI, respectively. Dalwinder and coauthors [23] carried out their work on three different datasets: two of which consist of FNA images, while the third one includes images after positive mammography followed by histological examination. They developed a CAD based system using feature weighting and utilized the Ant Lion Optimization algorithm. With the neural network, they attained a higher value of accuracy.

Hameed and coauthors [24] proposed an ensemble DL based method to classify HI of BC. They adopted a 5-fold data division protocol and implemented VGG16 and VGG19 architectures for this task. With the proposed approach, overall accuracy is obtained as 95.29%. Boumaraf and coauthors [25] implemented several machine learning classifiers and compared performance with DL methods to classify benign and malignant tumors. They carried out their work on two different datasets. The DL based methods achieved the highest accuracy varying from 94.05% to 98.13% (binary) and in the range of 76.77% to 88.95% (eight-class).

Bhowal and coauthors [26] used fuzzy based ensemble of DL models to perform classification of breast HI images using Choquet integral and Information theory. In case of the two-class problem, a test accuracy of 95% was obtained using the Xception model while with the fusion method an accuracy of 96% was achieved.

From the literature review, it is found that shape and texture are the features commonly used for the classification of grayscale images. And other features are calculated using a co-occurrence matrix. Also, various color features like mean, standard deviation, etc., are commonly used. Furthermore, having reviewed the above-mentioned studies, it is found that the classifiers most commonly used for classifying breast HI are CNN, SVM, ANN, k-NN, LDA, etc. However, the reported studies employ different texture and color features for classification. Some studies have reported that the performance of classifiers may vary with image magnification, while some magnification invariant models for histopathology image classification are also reported. The accuracy achieved using CNN with LSTM is 91%, while the accuracy of 95.29% is reported with an ensemble approach [14, 24]. Further using Deep CNN and CNN accuracies of 97 and 97.19% are obtained, respectively [1, 16]. Moreover, the existing studies suffer from the following limitations:

    i.   The features extracted by various authors include GLCM, SURF, texture, and morphological features. However, none of the existing works reported the fusion of multiple features. In this study, we propose a multi-feature approach including a total of 658 color and grayscale features. Exploring such a comprehensive set of features for breast HI classification is not reported earlier in the literature.

    ii.   None of the studies compared the performance of machine learning with DL methods for multiple features. In this work, the proposed multi-feature classification was implemented, and the performance of several machine learning methods and DL techniques are compared.

    iii.   The accuracy reported by different studies still needs to be improved for its clinical acceptance. Further, evaluating various models under different data division protocols is required for a fair comparison.

Despite this significant progress, the comprehensive evaluation of combined gray features and color features is still not explored.

Determining relevant grayscale and color attributes for BC histology classification is still challenging due to a variety of texture, shape, and color measures available in the literature. Due to advantages like relative insensitivity to the gap length of LSTM over other sequence learning methods, LSTM can outperform traditional methods in problems of classification and prediction. Thus, this article comprehensively evaluates several grayscale and color features extracted from BC HI using LSTM based deep learning model with genetic optimization. Further, a comparative investigation of traditional machine learning techniques and the proposed approach is conducted. We hypothesize that the performance of BC CAD systems based on HI can be improved by combining spectral and spatial features. Utilizing a large number of features and selecting the most relevant descriptors for building a machine learning model can result in a high-performance magnification invariant model.

Contributions

The contributions of the present study are summarized as follows:

i. We implement and evaluate a multi-feature classifier for risk stratification in HI using combined temporal, spectral, and color features.

ii. We implement several machine learning techniques and propose LSTM based approach for risk stratification of BC HI. All the techniques are evaluated under different data division protocols.

iii. Another significant contribution of the present study is that we employ and evaluate several temporal, spectral, and color features to identify the most reliable image markers to classify BC HI. We have extracted and evaluated 472 grayscale features based on shape, texture, etc., 186 color features (12 color features based on color moments and 174 color features using wavelet transform). Further, GA is used to determine the subset of the most relevant features.

Organization of the paper

The remaining sections of the paper are organized as follows. Materials and methods used in this study are explained in next section, followed by corresponding results and discussions. In the last section, conclusions are presented.

MATERIAL AND METHODS

Data

A publicly available BreaKHis dataset [27] is utilized in this study consisting of 7909 images (both malignant and benign images) with different resolutions. It includes the images with different magnifications i. e. 40X, 100X, 200X and 400X of each type. However, we have employed images with 40X resolution consisting of 645 benign and 1370 malignant tumors [28]. The resolution of each image is 700×460. Permission from the ethical committee is not required in this study since the open-source database is used. All the experiments in this study are implemented using MATLAB® software.

Feature extraction

Feature extraction aims at extracting the relevant information to characterize each class. It is the process of retrieving the most important raw data attributes by reducing the dimensionality of input data to represent the image efficiently. Features give information associated with color, shape, or texture in an image [29]. Measures such as moment, perimeter, area, and orientation are usually employed to quantify shape [30]. The morphological features are used to determine the shape and margin (smooth or irregular) of lesions. Irregular shapes generally characterize the malignant nuclei, and the edges are spiky. On the other hand, the texture of the image represents the distribution of gray levels. Table 1 shows various texture and shape features extracted from the HI. Features based on First Order Statistics (FOS), Haralick Spatial Gray Level Dependence Matrices (SGLDM), Gray Level Difference Statistics (GLDS), Neighborhood Gray Tone Difference Matrix (NGTDM), Spectral texture of Images(STI), Statistical Feature Matrix (SFM), Laws Texture Energy Measures (TEM), Fractal Dimension Texture Analysis (FDTA), shape, Invariant Moments of Image (IMI), Statistical Measures of Texture in an Image (SMTI), Gray Level Run Length Matrix (GLRLM), Segmentation-based Fractal Texture Analysis (SFTA) are utilized.

The third category of features used is related to color distribution in an image. Compared to shape and texture, features based on color demonstrate improved stability because the color adds more information to an image [31, 32]. Table 2 shows the Color features based extracted from HI.

Wavelets are used in multi-resolution analysis and are utilized extensively for texture measurements. Discrete Wavelet Transform (DWT) is used to generate wavelet coefficients for each image in the database, followed by calculating the mean and standard deviation of these coefficients to construct the feature vector [34]. The coefficients are computed in RGB color space, so for a particular image, 87 coefficients each for mean and standard deviation were extracted, including 29 coefficients for each color channel. Hence, a total of 174 WT based features are used in the classification process. The details of these features are presented in Table 2. Finally, all the extracted features are pooled to generate a feature vector comprising 658 features for each image in the database and are utilized as input for training and testing of the BC risk stratification system using HI.

Table 1. Spatial texture and shape attributes extracted from HI

Type of Feature	Number of features	Name of various features
FOS	5	V1-V5: Mean, median, skewness, kurtosis, and standard deviation [29]
SGLDM	26	V6-V31: ASM, contrast, correlation, SOS, IDM, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, and information measures of correlation-2
GLDS	4	V32-V35: Contrast, angular second moment, entropy, and mean
NGTDM	5	V36-V40: Coarseness, contrast, busyness, complexity, and strength
SFM	4	V41-V44: Coarseness, contrast, periodicity, and roughness
TEM	6	V45-V50: LL, EE, SS, LE, ES, and LS
FDTA	4	V51-V54: Hurst coefficient H(k); k=1 to 4
Shape features	3	V55-V57: area, perimeter, and perim^2/area [33]
STI	379	V58-V436: Computes Radial features and angular features [33]
IMI	7	V437-V443: Computes the moment invariants of the image
SMTI	6	V444-V449: Average gray level, average contrast, measure of smoothness, third moment, measure of uniformity, and entropy
GLRLM	11	V450-V460: SRE, LRE, GLN, RLN, RP, LGRE, HGRE, SRLGE, SRHGE, LRLGE and LRHGE
SFTA	12	V461-V472: SFTA based texture features

ASM=Angular Second Moment, SOS=Sum of Squares, IDM= Inverse Difference Moment, SRE= Short Run Emphasis, LRE= Long Run Emphasis, GLN= Gray level non-uniformity, RLN= Run-length non-uniformity, RP= Run percentage, LGRE= Low Gray Level Run Emphasis, HGRE= High Gray Level Run Emphasis, SRLGE= Short Run Low Gray Level Emphasis, SRHGE= Short Run High Gray Level Emphasis, LRLGE= Long Run Low Gray Level Emphasis, LRHGE= Long Run High Gray Level Emphasis

Table 2. Spatial Color features extracted from HI

S. No.	Type of features	Name of Feature	No. of features (Details)	S. No.	Type of features	Name of Feature	No. of features (Details)
1.	Color moments based features	Mean	3 (meanR, meanG, meanB)	4.	Color moments based features	Skewness	3 (yR, yG, yB)
2.		Standard deviation	3 (stdR, stdG, stdB)	5.	Wavelet transform based features	Mean	87 (mean1 to mean87)
3.		Variance	3 (VR, VG, VB)	6.		Standard deviation	87 (std1 to std 87)

Genetic optimization for feature selection

Feature selection aims at choosing a useful subset of attributes from data that is multidimensional and larger in size. It also reduces the computation time needed for classification. There are several techniques available to obtain subsets of features like Genetic Algorithm (GA), Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), etc. We have considered GA as it is known to be an efficient and adaptive method for feature selection [35, 36]. GAs are basically search algorithms based on natural selection and genetics. The GA functions on search space (binary) as bit strings are used as chromosomes. The initial

population is generated depending upon the population size and genome length. A fitness function is to be defined for GA to pick an informative set of features. In this work, k-NN based fitness function is used to evaluate the fitness of each chromosome in the population with k=3. It determines the shortest distance by calculating the Euclidean distance between the data used for training and testing. The new population is created using crossover and mutation. GA stops when it reaches the optimum solution. The highest number of generations and stall generation limit are the stopping conditions used in GA. The other parameters used for the implementation of GA are shown in Table 3. Table 4 gives the details for 192 features selected by GA. The parameter values were selected empirically after several experiments. After applying GA, the numbers of extracted features were reduced to 192 at a population size of 50.

Table 3. Details of parameters used for GA

S.No.	Parameters	Parameter value used	S.No.	Parameters	Parameter value used
1.	Size of population	50	6.	Crossover	Arithmetic
2.	Value of genome length	658	7.	Crossover Probability	0.8
3.	Type of population	Bitstrings	8.	Mutation	Uniform
4.	Fitness function used	k-NN	9.	Mutation Probability	0.1
5.	Number of generations	100	10.	Selection scheme	Tournament of size 2

Table 4. List of selected features obtained using genetic algorithm

Selected features
Spatial texture and shape attributes: V1, V4-V6, V10, V13, V15-V18, V20-V23, V32, V36, V38, V40, V47, V58, V64 , V67, V68, V74, V77, V78, V82, V85, V87, V88, V91, V96, V97, V98, V101, V102, V116, V117, V118, V119, V124, V125, V129, V132, V134, V138, V139, V143, V144, V145, V146, V148, V152, V153, V154, V155, V157, V163, V172, V174, V175, V178, V180,V182, V188, V190, V201, V202 ,V204, V206, V208, V214, V219, V220, V222, V224, V229, V235, V236, V240, V246, V247, V249, V253, V255, V257 ,V258, V260, V263, V268, V273, V283, V284, V286, V287 ,V289, V296, V299, V302, V303, V307, V308, V310, V312, V313, V314, V315, V317, V320, V323, V328, V342, V343, V348, V350, V353, V358, V361, V362, V363, V367, V388, V392, V395, V403, V405, V408, V409, V410, V415, V421, V427, V429, V430, V432, V433, V435, V437 ,V443, V444, V445, V461, V463, V464, V467.
Color features based on color moments: Standard deviation- stdR, stdB; Variance- VR
Spectral color features based on wavelet transform-Mean: mean1, mean2, mean4, mean5, mean9, mean11, mean15, mean17, mean19, mean23, mean27, mean30, mean36, mean40, mean41, mean47, mean49, mean53, mean58, mean60, mean64, mean70, mean71, mean73, mean75, mean77, mean86
Spectral color features based on wavelet transform- Standard deviation: std6, std22, std26, std32, std38, std41, std44, std47, std59, std60, std67, std68, std69, std70, std71, std77, std87

The steps for the same are shown in Algorithm 1.

Algorithm 1: A proposed method for feature selection using genetic optimization

Input: F_E// Input features after feature extraction; Output: F_S // Selected features

Steps: **procedure** Popfcn()   //*Create Initial Population*
pop ← PopulationSize, GenomeLength
Return pop
end procedure
procedure Fitfcn()   //*Fitness Function Evaluation*
Feature index   ← Indices from Chromosome
N_{DS} ← Data set indexed using Feature index; N_F ← No. of elements in Feature index
3 ← No. of Neighbors using kNN   // *NumNeighborskNN*
kNN_{ERR} ←   ClassifierKNN(Data set, ClassInfo, NumNeighborskNN)   // *kNN Error*
Return ← kNN_{ERR}
end procedure
F_S = Selected_features

Proposed framework for multi-feature classification using machine learning

Figure 1 shows the schematic block diagram of the proposed approach. The images are initially divided into training and test dataset using different data division strategies. A dotted line separates the training and testing phase of the proposed machine learning framework. The left side of the dotted line incorporates a supervised learning strategy for training and building the machine learning and deep learning model. This part is referred to as an offline system. On the contrary, the evaluation of the proposed model using test images is referred to as an online system and represented on the right side of dotted lines. Various grayscale and color features are combined to perform feature fusion. To eliminate the redundant and misguiding HI features, GA is used for feature selection. Various traditional techniques such as Support Vector Machine (SVM), Back-Propagation Artificial Neural Networks (BPANN), k-Nearest Neighbors (k-NN), Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Naive Bayes (NB) are implemented and evaluated for the classification of HI into benign or malignant using image markers selected by GA. Only relevant features are extracted from test HI in the online phase to save computational time.
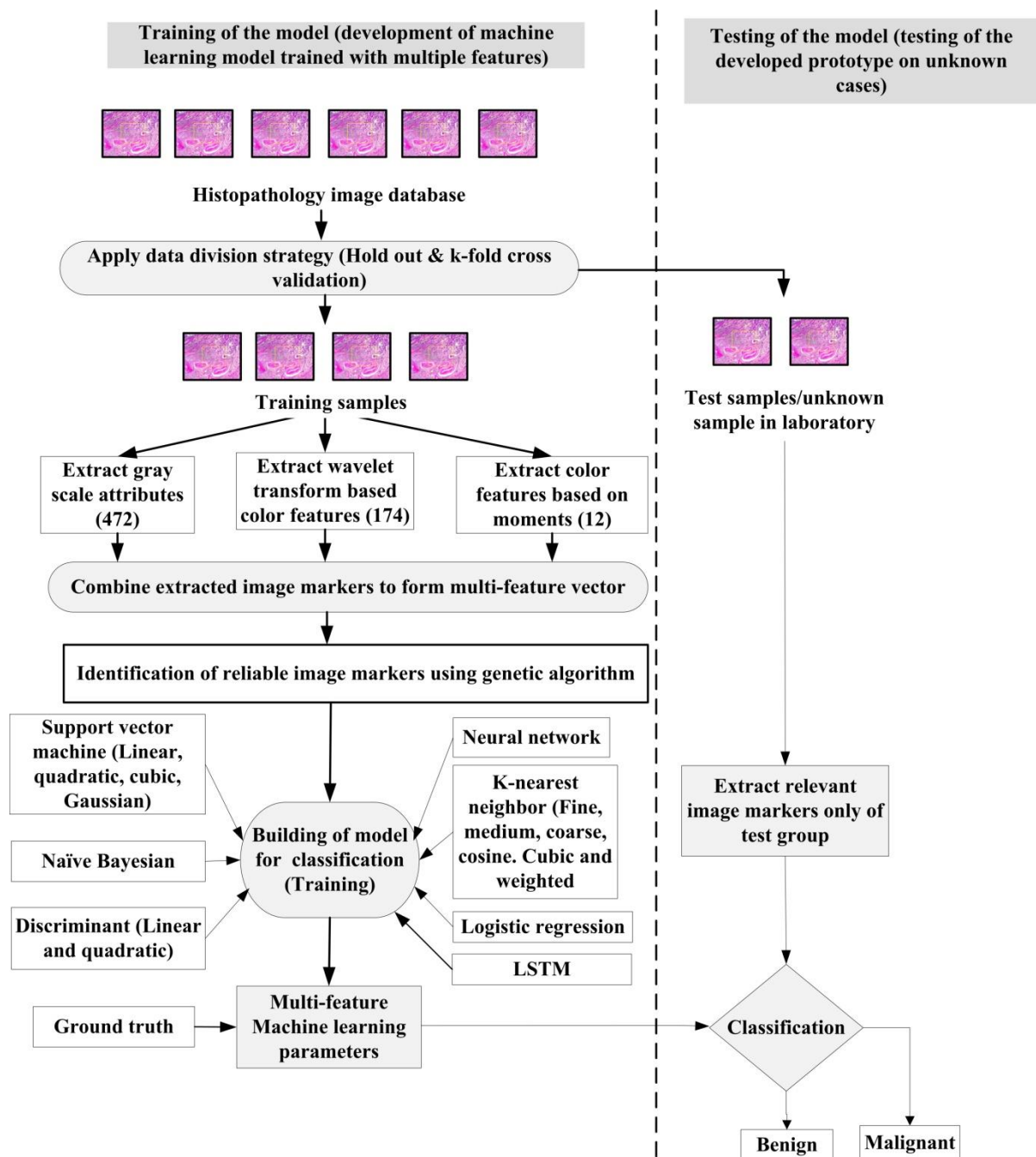

Figure 1. Block diagram of the proposed approach

Proposed framework for multi-feature classification using LSTM based deep learning

The DL is considered a powerful technology in the machine learning area [37]. It is a type of neural network that allows multiple hidden layers and has achieved good results in various research problems related to image classification, speech recognition, etc. In this study, work on HI of BC is carried out using deep recurrent neural networks implementing the LSTM model for classification. Recurrent neural networks (RNN) use weights to the present and the previous input. LSTM networks enable RNN to remember their inputs for a long duration of time. LSTM has three gates: the input, forget, and output gate [38]. The sigmoid function is used as a gate activation function, while a hyperbolic tangent is applied for block input and output activation function. The output is connected again to the block input and all the gates in a recurrent manner that replaces the usual hidden units of ordinary recurrent networks [39]. The value of the input feature evaluated using an artificial neuron unit can be accumulated if the sigmoidal input gate allows it. The forget gate controls the weight of the self-loop. Multiplier input and output gate units are connected to act as a buffer for storing the memory contents to avoid irrelevant inputs [38].

We have designed the deep neural network by specifying the input size as 658 (all features) and 192 (selected features), respectively. The bidirectional LSTM layer maps the input features and then connects the output to a fully connected (FC) layer. FC layer works like the feed-forward layer in a neural network. This layer multiplies the input by a weight W_M and thereby adds it to a bias value b_i given by:

$$W_M X_t + b_i \tag{1}$$

Where X_t denotes time step t of X. It combines all the information learned by the previous layers. As we are performing the classification task in this work, the number of classes specified in the FC layer is 2. FC layer is followed by the Softmax layer defined as:

$$\pi_k(x) = \frac{\exp(a_k^\pi)}{\sum_{l=1}^{K} \exp(a_l^\pi)} \tag{2}$$

where $0 \leq \pi_k(x) \leq 1$ and $\sum_{k=1}^{K} \pi_k(x) = 1$. The softmax activation function determines the probabilities of each class being classified as benign or malignant [39]. The next layer is the classification layer that computes the categorical label for each sample. It uses cross-entropy as a loss function. To perform the classification task, we have used Adaptive Moment Estimation (ADAM) [40] to optimize network parameters. It is computationally efficient, requires less memory, and performs better on large datasets. Another advantage of using ADAM as an optimizer is that updating the parameters is invariant to gradient rescaling. Further, the value of maximum epochs is empirically chosen to be 20. An initial learning rate of 0.01 is used. The mini batch size is set to 150, i.e., it instructs the network to consider 150 training samples at a time. Training and testing of developed models were done using the following data division protocols: (i) Hold-out: the dataset was split into 67% training, and the remaining 33% was used for testing. (ii) K-fold: 5-folds and 10-folds cross-validation were used for the proposed network. The developed machine learning and deep learning framework are evaluated in terms of overall accuracy (A), i.e., the total percentage of rightly categorized samples. Several other performance measures, namely, sensitivity (S_e), specificity (S_p), the area under the receiver operating characteristic curve (AUC), and Matthew's correlation coefficient (MCC) are also used.

RESULTS AND DISCUSSIONS

This section presents the results of various steps included in the proposed machine learning and deep learning framework.

Results of proposed machine learning approach

The results of the machine learning approach are calculated for the following combination of features at classifier input: (i) four hundred seventy two grayscale features; (ii) twelve color features based on moments and (iii) one hundred seventy-four color features based on WT and (iv) total of 658 features by combining grayscale and color features (v) selected 192 features obtained by feature selection using GA. Table 5 shows the performance measures (in %) of the BPANN model using holdout, 10-fold, and 5-fold data division protocols, respectively, for grayscale and color features separately. Also, in Table 5, the results for the two best performing SVM classifier types (medium gaussian and coarse gaussian SVM) are presented.

It is found that BPANN with 174 wavelet-based color features under 10-fold data division protocol achieved the highest classification accuracy of 98.56%. On the other hand, BPANN with 472 grayscale features as input under 10-fold data division protocol achieved a classification accuracy of 98.46%, which are very close to that using 174 color WT features. For BPANN with 12 features based on color moments as

input under the 10-fold data division protocol, the classification accuracy achieved is 98.17%. It is thus concluded that wavelet-based color features outperform others in classifying benign and malignant cases when BPANN is used as a classifier.

The highest classification accuracy achieved using medium and coarse Gaussian SVM classifiers is 93.85%, when color features based on color moments under 10-fold data division protocol are used. From the results of Table 5, it is concluded that color features based on color moments and wavelet transform are more significant for classifying BC in HI.

Table 5. Results of BPANN, medium and coarse gaussian SVM classifier with different data division protocols and feature combinations

Data division	Performances (%)	BPANN classifier			Medium Gaussian SVM classifier			Coarse Gaussian SVM classifier		
		Features			Features			Features		
		472 grayscale features	12 Color Moments	174 WT features	472 grayscale features	12 Color Moments	174 WT features	472 grayscale features	12 Color Moments	174 WT features
	A	86.02	91.28	88.12	77.56	92.32	86.75	70.48	92.17	86.75
	S_e	92.29	91.93	92.57	97.57	96.24	95.35	100	96.24	95.58
Holdout	S_p	73.66	89.69	79.19	34.91	83.96	68.40	7.55	83.49	67.93
	AUC	82.98	90.81	85.88	66.24	90.10	81.88	53.77	89.87	81.75
	MCC	68.07	79.62	72.90	45.31	82.10	68.64	22.94	81.74	68.64
	A	94.94	97.57	96.03	78.86	93.10	88.19	71.32	93.15	88.14
	S_e	96.57	98.25	97.08	97.74	96.13	95.11	100	95.99	95.18
5-fold	S_p	91.47	96.12	93.79	38.76	86.67	73.49	10.39	87.13	73.18
	AUC	94.02	97.19	95.44	68.25	91.40	84.30	55.19	91.56	84.18
	MCC	88.28	94.39	90.94	49.15	84.00	72.21	27.03	84.13	72.09
	A	**98.46**	**98.17**	**98.56**	79.21	**93.85**	89.28	71.22	**93.85**	88.93
	S_e	98.98	98.76	99.27	97.45	96.28	95.33	100	96.06	95.18
10-fold	S_p	97.34	96.91	97.04	40.47	88.68	76.43	10.08	89.15	75.66
	AUC	98.16	97.84	98.15	68.96	92.48	85.88	55.04	92.60	85.42
	MCC	96.41	95.79	96.61	49.96	85.76	74.86	26.61	85.78	74.02

Finally, the classifier's performance is evaluated by generating a feature vector using a combination of all grayscale and color features resulting in a total of 658 features. Table 6 shows the classification results using BPANN and different types of SVM classifiers when all 658 features are used as input. It is observed that the overall performance of classifiers improves when all 658 features are supplied to their input. The highest classification accuracy achieved is 99.26% using the 10-fold data division protocol for BPANN.

Table 6. Results of classification after combining all grayscale and color features (total 658 features)

Data Division	Performance measures	Classifiers						
			SVM					
		BPANN	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
	A	93.53	87.80	94.88	94.58	82.08	91.57	77.11
	S_e	95.58	94.91	96.68	94.91	99.34	96.90	98.67
Hold-out	S_p	89.15	72.64	91.04	93.87	45.28	80.19	31.13
	AUC	92.37	83.78	93.86	94.39	72.31	88.55	64.90
	MCC	85.06	71.22	88.17	87.73	58.40	80.30	44.69
	A	98.81	87.84	94.14	94.54	82.53	91.71	79.40
	S_e	98.98	94.89	96.35	95.99	99.56	96.50	97.74
5-fold	S_p	98.45	72.87	89.46	91.47	46.36	81.55	40.47
	AUC	98.71	83.88	92.90	93.73	72.96	89.02	69.10
	MCC	97.29	71.37	86.46	87.46	59.77	80.67	50.63
	A	**99.26**	87.89	94.99	95.09	83.08	92.16	80.89
	S_e	99.27	94.74	97.15	96.64	99.64	96.72	97.74
10-fold	S_p	99.23	73.33	90.39	91.78	47.91	82.48	45.12
	AUC	99.25	84.04	93.77	94.21	73.77	89.60	71.43
	MCC	98.35	71.50	88.41	88.68	61.15	81.73	54.56

Table 7 (a & b) shows the results for classification using BPANN, several types of SVM classifiers, NB classifier, LDA, QDA, LR, and different types of k-NN when 192 selected features were applied at the input of the classifiers. After applying feature selection using GA, the highest accuracy is obtained as 99.75% (for 192 features subset) using the 10-fold data division protocol for BPANN.

Table 7a. Results of hold out classification after applying feature selection (192 features)

Classification technique	Performance measures (%)				
	A	S_e	S_p	AUC	MCC
BPANN	92.18	94.87	86.64	90.75	82.1
Linear SVM	88.10	92.92	77.83	85.38	72.21
Quadratic SVM	94.13	95.80	90.57	93.18	86.47
Cubic SVM	94.58	96.02	91.51	93.76	87.53
Fine Gaussian SVM	93.22	97.57	83.96	90.76	84.23
Medium Gaussian SVM	93.37	97.57	84.43	91.00	84.58
Coarse Gaussian SVM	93.37	97.57	84.43	91.00	84.58
Naïve Bayesian	67.62	88.72	22.64	55.68	14.87
Linear discriminant	65.06	77.43	38.68	58.06	16.78
Quadratic discriminant	64.46	84.51	21.70	53.11	7.63
Logistic regression	85.99	90.49	76.42	83.45	67.51
Fine k-NN	92.92	93.36	91.98	92.67	84.06
Medium k-NN	90.36	94.25	82.08	88.16	77.56
Coarse k-NN	80.12	93.81	50.94	72.37	51.69
Cosine k-NN	87.50	92.70	76.42	84.56	70.75
Cubic k-NN	89.76	94.69	79.25	86.97	76.03
Weighted k-NN	91.87	93.58	88.21	90.90	81.39

Table 7b. Results of 5-fold and 10-fold classification after applying feature selection (192 features)

Classification technique	Performance measures (%) for 5-fold data division					Performance measures (%) for 10-fold data division				
	A	S_e	S_p	AUC	MCC	A	S_e	S_p	AUC	MCC
BPANN	98.41	99.05	97.05	98.05	96.32	**99.75**	99.85	99.53	99.69	99.43
Linear SVM	87.79	94.09	74.42	84.25	71.30	88.24	94.38	75.19	84.79	72.37
Quadratic SVM	93.65	95.77	89.15	92.46	85.34	94.04	95.84	90.23	93.04	86.29
Cubic SVM	94.59	95.84	91.94	93.89	87.60	94.89	96.35	91.78	94.07	88.24
Fine Gaussian SVM	94.64	97.81	87.91	92.86	87.58	95.14	98.03	88.99	93.51	88.74
Medium Gaussian SVM	94.89	97.88	88.53	93.21	88.16	95.14	97.96	89.15	93.55	88.74
Coarse Gaussian SVM	94.99	97.88	88.84	93.36	88.39	95.19	98.03	89.15	93.59	88.85
Naïve Bayesian	67.64	89.20	21.86	55.53	14.82	67.39	89.05	21.40	55.23	13.71
Linear discriminant	68.68	76.64	51.78	64.21	28.32	68.59	76.72	51.32	64.02	27.98
Quadratic discriminant	66.70	84.96	27.91	56.44	15.26	66.25	82.77	31.16	56.97	15.78
Logistic regression	85.71	90.07	76.43	83.25	66.96	86.50	90.73	77.52	84.12	68.77
Fine k-NN	94.00	95.69	90.39	93.04	86.19	94.54	95.77	91.94	93.85	87.49
Medium k-NN	91.61	93.72	87.13	90.43	80.76	91.91	93.87	87.75	90.81	81.45
Coarse k-NN	81.49	95.40	51.94	73.67	55.47	81.99	95.40	53.49	74.44	56.76
Cosine k-NN	90.22	92.34	85.74	89.04	77.67	90.67	92.99	85.74	89.36	78.60
Cubic k-NN	91.22	93.87	85.58	89.73	79.75	91.91	94.23	86.98	90.61	81.38
Weighted k-NN	93.05	96.06	86.67	91.36	83.88	93.65	96.35	87.91	92.13	85.28

Results of proposed LSTM-based deep learning approach

Table 8 shows the results for the deep learning approach that uses the LSTM network. The performance can be seen in terms of accuracy, sensitivity, specificity, AUC, and MCC when (i) all combined 658 grayscale and color features; and (ii) one ninety-two features selected using GA are applied at the input of the deep neural network. In Table 8, it can be observed that the highest classification accuracy of 99.85% is obtained

with selected 192 features under a 5-fold data division protocol. On the other hand, an accuracy of 99.83% is obtained with all 658 features under holdout data division protocol when the average of 10 readings is considered.

As seen from the results, the accuracy is reduced using the 10-fold cross-validation for the selected 192 features. This has happened possibly due to the class imbalance present in the dataset. While using 10-fold cross validation, the data is partitioned into subsets 10 times (for 10-fold) for evaluating the performance of the model. Hence, one class of data might have been overrepresented due to which the accuracy has reduced with 10-fold cross validation.

Table 8. Performance of deep learning using LSTM with different data division protocols and feature combinations

Deep learning	Performance measures (%)	All 658 features	Selected 192 features
Holdout	A	99.83	93.25
	S_e	99.85	95.12
	S_p	99.81	95.75
	AUC	99.83	95.43
	MCC	99.62	84.43
5-fold	A	99.70	**99.85**
	S_e	99.57	99.93
	S_p	100.00	99.69
	AUC	99.78	99.81
	MCC	99.32	99.66
10-fold	A	98.16	84.89
	S_e	97.43	92.72
	S_p	100.00	75.39
	AUC	98.71	84.05
	MCC	95.82	68.21

The training progress for best performing protocol using LSTM is shown in Figure 2 (a, b, c, & d) for all combined 658 features using holdout data division protocol (10 readings); and selected 192 features using GA under 5-fold data division protocol (5 readings) respectively. The plots in Figure 2 (a & b) represent the variation of training accuracy with the number of iterations, i.e., the classification accuracy on each mini batch for all combined 658 features using holdout data division protocol and selected 192 features under 5-fold data division protocol respectively. Each iteration represents an update of the network parameters. It is observed that training accuracy reaches 100% as the training progresses. Training loss for all combined 658 features using holdout data division protocol and selected 192 features under 5-fold data division protocol is shown in Figure 2 (c & d), respectively. As desired, the training loss decreases to zero as training progresses. The faint line indicates the training process, while the bold line represents the smoothened version of the same in all the plots. It takes a few seconds on a single GPU to obtain results with a learning rate of 0.01 and a mini-batch size of 150. To help a network learn better, the mini-batch size or initial learning rate values can be decreased, but it may result in a longer training time. From the results, we can interpret that the proposed approach (multi-feature + GA + LSTM) can effectively classify BC as benign or malignant in HI.

We now compare our proposed work vis-à-vis other relevant research contributions reported in the literature. Table 9 compares the performance of the proposed method with some of the recently reported studies. Some of the reported work utilized CNN [12-16] and deep CNN [1, 22] as classifiers. Few works are reported on transfer learning based classification of breast HI in [19, 24-25]. The size of the dataset reported in [24] is small, and they performed classification on two-class only. Most of the recent findings on the same dataset analyze only the accuracy [17]. We calculated other parameters like specificity, sensitivity, MCC, and accuracy to assess our model. In [17], traditional methods were used to extract features, and an accuracy of 88.5% was obtained, which is low compared to our proposed model. However, the performance of the proposed model is higher than in existing studies due to the large number of texture and color features used in this study. The highest accuracy obtained by the proposed model is 99.85% which is higher than reported by other authors in Table 9.

The use of multiple features adds more information to the descriptor set, resulting in improved efficiency of automated disease diagnosis systems. Further, such approaches can reduce the workload of medical professionals and reduce variability among their observations. The removal of redundant attributes is turned out to be beneficial for accurate classification. The results show that the classification accuracy is improved after eliminating irrelevant and redundant datasets. This also signifies that irrelevant features can misguide

the classifiers and thus deteriorate their overall performance. A lesser number of attributes also reduces the computation time involved in the feature extraction process. Though the results of the proposed study are promising, it has certain limitations. Firstly, HI of only one resolution is considered in this study. In the future, work on the higher resolution can be explored. The class imbalance present in the dataset is also a limitation of the present study. Another limitation is the use of a single modality to classify breast tumors as benign or malignant. The use of the multi-modality approach can be considered further, and advanced machine learning techniques based on a hybrid/ensemble approach can be explored to classify BC using relevant markers identified from multimodal data. The proposed approach can also be implemented to detect cancers of skin [41], lung [42], and brain tumor [43].
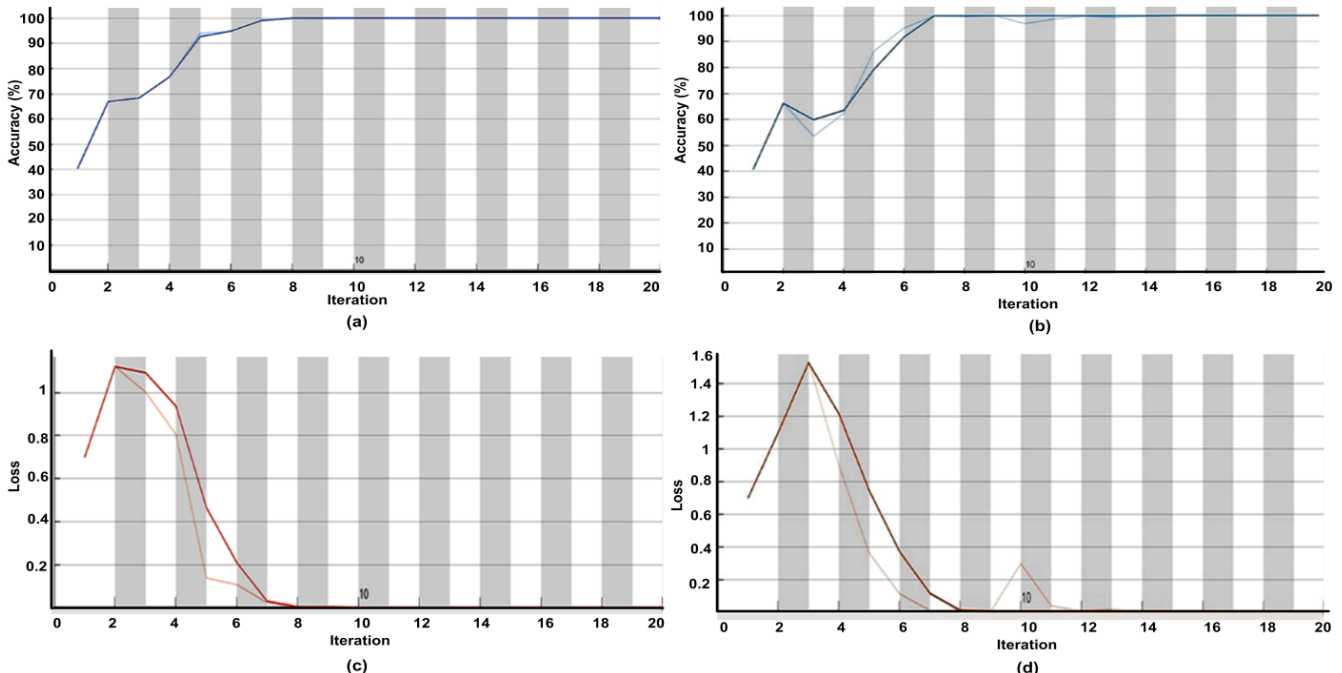

Figure 2. Performance plots of LSTM-based deep learning approach: (a) training for all 658 features under holdout data division, (b) training for selected 192 features under 5-fold data division, (c) loss for all 658 features, and (d) loss for selected 192 features

CONCLUSION

This article proposed a deep learning based approach that utilizes multi-feature space, LSTM network, genetic optimization, and softmax function for the classification of BC HI. Further, various classical machine learning models were also implemented and evaluated using combined grayscale and color descriptors extracted from the spatial and spectral domain of BC HI of 40X resolution. The comparative evaluation was carried out under different data division protocols such as holdout, 5-fold, and 10-fold using performance measures like accuracy, sensitivity, specificity, AUC, and MCC. The main findings are summarized as follows- (i) when all the 658 features are applied at the input, the proposed LSTM based approach achieved a classification accuracy of 99.83% under repeated holdout data division protocol, while the accuracy achieved by the classical BPANN model is 99.26% under 10-fold data division protocol; (ii) when 192 most reliable features elected using GA are applied at the input, the proposed LSTM based approach achieved a classification accuracy of 99.85% under 5-fold data division protocol while the accuracy achieved by classical BPANN model is 99.75% under 10-fold data division protocol. The performance reported by both the deep learning based LSTM approach and BPANN is higher than other reported studies on the same database. We conclude that the integration of multiple features from different domains improves the overall classification accuracy of BC HI. Thus, the proposed method will help to reduce the load on the medical practitioners and also reduce the errors due to inter- and intra-observer variability. The proposed classification and feature selection scheme can also prove beneficial for other areas such as, the detection of skin cancer, lung cancer, and brain tumor. In future, this work can be extended to multi-modal analysis and hybrid/ensemble approach can be used for the classification of BC using multimodal data. Also, the size of dataset can be increased in future to eliminate the class imbalance problem.

Table 9. Comparison of performance of the proposed method with recently reported studies

Author and year	Size of the dataset	Classifier/ Architecture used	Performance Accuracy
Bhowal et al. (2022) [26]	400	Fuzzy based ensemble of DL models	96% (two class)
Boumaraf et al. (2021) [25]	7909	VGG-19	98.13% (Binary) 88.95% (Multi-class)
Hameed et al. (2020) [24]	544	VGG16 and VGG19	95.29%
Wang et al. (2020) [22]	1125	D^{2}TL and ICELM	98.18% (Normal vs. Malignant)
Beevi et al. (2019) [19]	53+45	VGGNet	90%
Jiang et al. (2019) [18]	7909	CNN	99.34% (Binary) 93.81% (Multi-class)
Motlagh et al. (2018) [11]	6,402 + 7909	Inception and ResNet	98.7%
Rahhal and Mahmoud (2018) [12]	9109	CNN: VGGm	86.80%
Bardou et al. (2018) [13]	7909	CNN	98.33%
Nahid et al. (2018) [14]	7909	CNN, LSTM, and CNN+LSTM	91%
Golatkar et al. (2018) [15]	400	Fine-tuning Inception-v3 CNN	93%
Nahid and Kong (2018) [16]	7909	CNN	97.19%
Alirezazadeh et al. (2018) [17]	7909	Unsupervised domain adaptation, representation learning	88.5% (Average)
Wei et al. (2017) [1]	7909	Deep CNN: Softmax	97%
		472 grayscale features +BPANN	98.46%
		12 color moments + BPANN	98.17%
		174 wavelet-based features + BPANN	98.56%
		All features combined (658 features) + BPANN	99.26%
Proposed method	2015	192 features selected using genetic algorithm+ BPANN	99.75%
		All features combined using deep learning (LSTM) (658 features)	99.83%
		With deep learning (LSTM) using 192 selected features	99.85%

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Wei B, Han Z, He X, Yin Y. Deep learning model based breast cancer histopathological image classification. In 2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA), 2017 Apr 28 (pp. 348-353). IEEE. doi: 10.1109/ICCCBDA.2017.7951937

2. GLOBOCAN 2020: New Global Cancer Data [Internet]. [place unknown: publisher unknown]; [updated 2021 Feb 15; cited 2021 March 17]. Available from: GLOBOCAN 2020: New Global Cancer Data | UICC15

3. Maurya AP, Brahmachari S. Current status of breast cancer management in India. Indian J Surg. 2021 Jun;83(2):316-21. doi: https://doi.org/10.1007/s12262-020-02388-4

4. Ting FF, Tan YJ, Sim KS. Convolutional neural network improvement for breast cancer classification. Expert Syst. Appl. 2019 Apr 15;120:103-15. doi: https://doi.org/10.1016/j.eswa.2018.11.008

5. Celaya-Padilla JM, Guzmán-Valdivia CH, Galván-Tejada CE, Galván-Tejada JI, Gamboa-Rosales H, Garza-Veloz I, et al. Contralateral asymmetry for breast cancer detection: a CADx approach. Biocybern Biomed Eng. 2018 Jan 1;38(1):115-25. doi: https://doi.org/10.1016/j.bbe.2017.10.005

6. Byra M. Discriminant analysis of neural style representations for breast lesion classification in ultrasound. Biocybern Biomed Eng. 2018 Jan 1;38(3):684-90. doi: https://doi.org/10.1016/j.bbe.2018.05.003

7. Aswathy MA, Jagannath M. Detection of breast cancer on digital histopathology images: Present status and future possibilities. Inform. Med. Unlocked. 2017 Jan 1;8:74-9. doi: https://doi.org/10.1016/j.imu.2016.11.001

8. Jimenez-del-Toro O, Otálora S, Andersson M, Eurén K, Hedlund M, Rousson M, et al. Analysis of histopathology images: From traditional machine learning to deep learning. In Biomedical Texture Analysis 2017 Jan 1 (pp. 281-314). Academic Press.

9. Dora L, Agrawal S, Panda R, Abraham A. Optimal breast cancer classification using Gauss–Newton representation based algorithm. Expert Syst. Appl. 2017 Nov 1;85:134-45. doi: https://doi.org/10.1016/j.eswa.2017.05.035

10. Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of breast cancer histology images using convolutional neural networks. PloS one. 2017 Jun 1;12(6):e0177544. doi: https://doi.org/10.1371/journal.pone.0177544

11. Motlagh NH, Jannesary M, Aboulkheyr H, Khosravi P, Elemento O, Totonchi M, et al. Breast cancer histopathological image classification: A deep learning approach. bioRxiv. 2018;1-8:242818. doi: https://doi.org/10.1101/242818

12. Rahhal A, Mahmoud M. Breast Cancer Classification in Histopathological Images using Convolutional Neural Network. Int J Adv Comput Sci Appl. 2018;9(3):64-8.

13. Bardou D, Zhang K, Ahmad SM. Classification of breast cancer based on histology images using convolutional neural networks. IEEE Access. 2018 May 1;6:24680-93. doi: 10.1109/ACCESS.2018.2831280

14. Nahid AA, Mehrabi MA, Kong Y. Histopathological breast Cancer image classification by deep neural network techniques guided by local clustering. BioMed Res. Int. 2018 Mar 7;1-20. doi: https://doi.org/10.1155/2018/2362108

15. Golatkar A, Anand D, Sethi A. Classification of breast cancer histology using deep learning. InInternational Conference Image Analysis and Recognition 2018 Jun 27 (pp. 837-844). Springer, Cham. doi: https://doi.org/10.1007/978-3-319-93000-8_95

16. Nahid AA, Kong Y. Histopathological breast-image classification using local and frequency domains by convolutional neural network. Information. 2018 Jan 16;9(1):19. doi: https://doi.org/10.3390/info9010019

17. Alirezazadeh P, Hejrati B, Monsef-Esfahani A, Fathi A. Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images. Biocybern Biomed Eng. 2018 Jan 1;38(3):671-83. doi: https://doi.org/10.1016/j.bbe.2018.04.008

18. Jiang Y, Chen L, Zhang H, Xiao X. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module, PLoS One. 2019 Mar 29;14(3): e0214587. doi: https://doi.org/10.1371/journal.pone.0214587

19. Beevi KS, Nair MS, Bindu GR. Automatic mitosis detection in breast histopathology images using convolutional neural network based deep transfer learning. Biocybern Biomed Eng. 2019 Jan 1;39(1):214-23. doi: https://doi.org/10.1016/j.bbe.2018.10.007

20. Singh BK. Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: A comparative investigation in machine learning paradigm. Biocybern Biomed Eng. 2019 Apr 1;39(2):393-409. doi: https://doi.org/10.1016/j.bbe.2019.03.001

21. Tong L, Mitchel J, Chatlin K, Wang MD. Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis. BMC Medical Inform. Decis. Mak. 2020;20(1):1-2. doi: https://doi.org/10.1186/s12911-020-01225-8

22. Wang P, Song Q, Li Y, Lv S, Wang J, Li L, et al. Cross-task extreme learning machine for breast cancer image classification with deep convolutional features. Biomed Signal Process Control. 2020 Mar 1;57:101789. doi: https://doi.org/10.1016/j.bspc.2019.101789

23. Dalwinder S, Birmohan S, Manpreet K. Simultaneous feature weighting and parameter determination of neural networks using ant lion optimization for the classification of breast cancer. Biocybern Biomed Eng. 2020 Jan 1;40(1):337-51. doi: https://doi.org/10.1016/j.bbe.2019.12.004

24. Hameed Z, Zahia S, Garcia-Zapirain B, Javier Aguirre J, María Vanegas A. Breast cancer histopathology image classification using an ensemble of deep learning models. Sensors. 2020 Aug 5;20(16):4373. doi: https://doi.org/10.3390/s20164373

25. Boumaraf S, Liu X, Wan Y, Zheng Z, Ferkous C, Ma X, et al. Conventional Machine Learning versus Deep Learning for Magnification Dependent Histopathological Breast Cancer Image Classification: A Comparative Study with Visual Explanation. Diagnostics. 2021 Mar 16;11(3):528. doi: https://doi.org/10.3390/diagnostics11030528

26. Bhowal P, Sen S, Velasquez JD, Sarkar R. Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification. Expert Syst. Appl. 2022 Mar 15;190:116167. doi: https://doi.org/10.1016/j.eswa.2021.116167

27. [dataset] [27] Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. IEEE. Trans. Biomed. Eng. 2016; 63(7):1455-62. doi: 10.1109/TBME.2015.2496264

28. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. In 2016 international joint conference on neural networks (IJCNN) 2016 Jul 24 (pp. 2560-2567). IEEE. doi: 10.1109/IJCNN.2016.7727519

29. Singh BK, Verma K, Panigrahi L, Thoke AS. Integrating radiologist feedback with computer aided diagnostic systems for breast cancer risk prediction in ultrasonic images: An experimental investigation in machine learning paradigm. Expert Syst. Appl. 2017 Dec 30;90:209-23. doi: https://doi.org/10.1016/j.eswa.2017.08.020

30. Singh BK, Verma K, Thoke AS. Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images. Expert Syst. Appl. 2016 Dec 30;66: 114-23. doi: https://doi.org/10.1016/j.eswa.2016.09.006

31. Kumar S, Chauhan A. Feature extraction techniques based on color images. InSpecial conference issue: National conference on cloud computing & big data 2013 (pp. 208-214).

32. Anusha V, Reddy VU, Ramashri T. Content based image retrieval using color moments and texture. Int. J. Eng. Res. Technol. 2014 Feb;3(2):2812-5.

33. Gonzalez RC, Woods RE. Digital Image Processing Using MATLAB, 2nd ed. Prentice Hall, 2010.

34. Ashraf R, Ahmed M, Jabbar S, Khalid S, Ahmad A, Din S, et al. Content based image retrieval by using color descriptor and discrete wavelet transform. J Med. Syst. 2018 Mar;42(3):1-2. doi: https://doi.org/10.1007/s10916-017-0880-7

35. Babatunde OH, Armstrong L, Leng J, Diepeveen D. A genetic algorithm-based feature selection. International Journal of Electronics Communication and Computer Engineering. 2014;5(4):2278–4209.

36. Bethapudi P, Reddy ES, Sitamahalakshmi T, Varma KV. Feature Analysis and Classification of BI-RADS Breast Cancer Using Genetic Algorithm, Int. J Sci.Eng. Res. Feb. 2015;6(2):750-56.

37. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. IEEE Trans Neural Netw Learn Syst. 2016 Jul 8;28(10):2222-32. doi: 10.1109/TNNLS.2016.2582924

38. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997 Nov 15;9(8):1735-80. doi: 10.1162/neco.1997.9.8.1735

39. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press, 2016 Nov 10.

40. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014 Dec 22;1-5.

41. Javaid A, Sadiq M, Akram F. Skin Cancer Classification Using Image Processing and Machine Learning. In 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST) 2021 Jan 12 (pp. 439-444). IEEE. doi: 10.1109/IBCAST51254.2021.9393198

42. Maurer A. An Early Prediction of Lung Cancer using CT Scan Images. Journal of Computing and Natural Science. 2021 Apr:39-44. doi: https://doi.org/10.53759/181X/JCNS202101008

43. Ramkumar G, Prabu RT, Singh NP, Maheswaran U. Experimental analysis of brain tumor detection system using Machine learning approach. Mater. Today: Proc. 2021 Feb 25. doi: https://doi.org/10.1016/j.matpr.2021.01.246