Received: September 11, 2022. Revised: November 7, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Hybrid Convolutional Neural Network with Residual Neural Network for Breast Cancer Prediction Using Mammography Images

Satyabrata Patro[1]* **Jyotirmaya Mishra[1]** **Bhavani Sankar Panda[2]**

[1]*Department of Computer Science and Engineering, GIET University, Gunupur, Odisha, India*
[2]*Department of Computer Science and Engineering,*
Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India
* Corresponding author's Email: satyabrata.patro@giet.edu

Abstract: BC (Breast Cancer) is predominantly prevalent in women and is the leading cause of mortality due to cancers which can be reduced using mammography screenings. The use of CNNs (Convolution Neural Networks), a type of deep learning method has proven to be highly successful in image identifications. However, the quality of acquired mammographic images is found to be low while being used by detection models and hence this work proposed hybrid MLTs (Machine learning Techniques) for overcoming low-quality issues in the prediction of BCs. Initially, Statistical correlation analysis-based pre-processing is introduced for improving classifier performances followed by a hybrid model which predicts BCs effectively. This work also introduces a novel building block called Fuzzy Scoring based Resnets (Residual Networks) and CNNs called FS-Resnet CNNs for optimizing networks. The proposed FS-Resnet CNN model is computationally efficient, less sensitive to noises, and efficient in memory usage. Experimental results show that the proposed model achieves 95% accuracy ,95.45% precision, 93% recall rate, 94.21% f-measure and 18.113(S)-time complexity.

Keywords: Mammography, Statistical correlation analysis, Hybrid deep learning model, Deep convolutional neural networks (CNN), Fuzzy based deep learning model.

1. Introduction

BCs are a type of cancer that threatens women's health and quality of life. Diagnosis and examination of BCs have always been an essential part of healthcare. Technological advancements in medical imaging have been widely used in the screening of BCs and effectively increasing detection of minute breast lesions [1]. Cancer malignancies that are not evident in tiny lesions begin to grow quickly and above anticipated limits in specific body locations. These cancerous masses (Tumours) can be classified as malignant or benign [2]. BCs create lumps in breast cells which grow abnormally in the body causing redness in the breasts. Any type of cancer is dangerous. Most BCs go undetected in women increasing mortality rates though early detections could help in the arrest of these cancerous cells and thus lives [3]. Cancers can be overcome due to their early detections, but unfortunately, most women get affected disproportionately when detected late. BCs are caused by genetic factors, consumption of alcohol, the denseness of breast tissues, exposures to radiations, and many other elements. Since the 1990s, the survival rate of BC patients has improved substantially mainly due to contemporary technologies of treatments and screenings [4]. Studies indicate women with BCs accounted for 2,52,710 cases (2017) with additional 40,610 expected deaths. Attempts to understand the causes for BCs are a major step in decreasing its illness risks [5]. Thus, early screenings can assist in reducing cancer-infringing factors. Further, early identifications are critical to controlling and survival of BC patients where pathology diagnostics is the bottom line. Traditional diagnostics are based on clinicians' experiences where diagnostic outcomes can be subjective with high probabilities [6]. In recent

years, CAD (computer-aided diagnostic) tools, AI (artificial intelligence) based machine learning techniques utilizing quantitative measurements have automated objective judgements of several diseases in healthcare diagnostics. Mammography is frequently used in examining breast disorders, in spite of the equipment's difficulties in detecting tiny lesions, specifically thick breasts. Studies imply CESMs (Contrast-Enhanced Spectral Mammography) have greater specificity in diagnostics of BCs when compared to mammography. CESMs are new approaches that combine intravenous iodine contrast enhancement and digital mammography. High/ low-energy mammography is performed on patients following intravenous contrast injection. Subsequently, recombined images are generated by removing unenhanced tissues in post-processing [7]. Low-energy mammography images show indicative lesions like calcifications or deformed structures. These abnormal regions are maintained in generated recombined images as the degree of lesion enhancements indirectly reflects the part's blood flow.

CNNs are DLTs (Deep Learning Techniques), a part of MLTs (Machine Learning Techniques) and DMTs (Data Mining Techniques) has attracted the attention of researchers and has been used beneficially in video/image recognition. AlexNets were used to classify benign and malignant BCs from images [8] where their classification results showed 6% more recognitions when compared to traditional machine learning approaches. Pre-trained CNN's feature vectors were used for extracting DeCAF features and used as inputs to the classifier [9]. The proposal was an example of multiple instances learning frameworks for CNNs. The study's pooling layer aggregated most informative features found a slide's patches and without necessitating global slide coverage or inter-patch overlaps [10]. Automatic categorization of abnormal BC images using CNNs is a complex challenge as (1) in its deep learning CNNs the parameters grow at a rapid rate resulting in overfitting of the model and a large number of sample BC images which is prohibited [11] are required to be trained to overcome overfitting. As a result, reducing the data samples becomes mandatory by lowering CNNs parameters and employing data augmentation techniques. (2) It is widely known that certain hyperparameters, particularly the learning rate, have a significant impact on the performance of CNNs. Training a Model often necessitates manual adjustments in learning rates for better performance [12] which makes it challenging for non-expert users to implement CNNs in real-world applications. To minimise CNN's training parameters, lightweight CNNs based on features of BCs histological images was developed and subsequently classified.

The primary contribution of this research work is developing a hybrid learning model, a combination of ResNets and CNNs that modifies learning rates during training automatically. This work initially uses statistical correlations for pre-processing data with the aim of enhancing the classifier's performance. CNNs encompass several convolutions and sub-sampling layers with one or multiple FCLs (fully connected Layers). FCLs are multi-layered neural networks where outputs are stored in the final layer. Convolution layers convolve input images with multiple filters (learning) while the pooling layer reduces data's dimensionality. Pooling layers typically execute two functions namely maximum and average pooling. Thus, raw input image pixels get transformed in CNNs multiple stacked layers. Resnets with 22 layers have a high level of complexity and encompasses network, pooling layer, and convolution layers that execute in parallel instead of the traditional sequential executions. Hence, instead of using traditional stochastic gradient descent for optimizations, this work proposes FS-Resnets CNNs (Fuzzy Scoring based Resnets – CNNs) for faster convergences. Moreover, the model aims at computational efficiency and reduced sensitiveness to noises while saving memory usage.

The remainder of the research is divided as follows: The next section examines current approaches for predicting BCs followed by a description of the methodology followed in section three. Section four displays simulated results and discusses simulation findings. This paper concludes with future scope in section five.

2. Literature review

Researchers have proposed new techniques/models and tools to prevent errors in reasoning while identifying BCs [13]. distinguished patients with or without BCs using hybrid DMTs. In the first phase, a statistical approach pre-processed input data where unimportant characteristics were removed. This reduced computing complexity while speeding DMTs executions. In the second phase, their novel DMTs based on conventional PSOs (particle swarm optimizations) called DPSOs (discrete PSOs) was used. Their scheme created new viable PSOs by transforming particle values into positive integers. Their proposed DPSOs outperformed prior approaches in terms of accuracy and resilience by scoring increased accuracy (98.71%), sensitivity (100%) and specificity (98.21%). The study's suggested DPSOs were

referred by hospitals making decisions and also served as a reference to researches. FOAs (fly optimization algorithms) were used by Huang, H., Zhou, S., Jiang, J., Chen, H., Li, Y., & Li, C.Huang et al. [14] who proposed advanced MLTs for diagnosing BCs. The study used FOAs with LFs (Levy flights) to optimise two important parameters of SVMs (support vector machines) and thus develop LFOA-based SVMs (LFOA-SVMs) for diagnosing BCs. For the first time, high-level characteristics were abstracted from volunteers diagnosed BCs. The study compared other methods including FOA-SVM, PSO-SVM, GAs (Genetic Algorithms) based GA-SVM, RFs (random forests), and BPNNs (backpropagation neural networks) using the 10-fold cross-validation method. The proposed LFOA-SVMs improved the quality of FOAs optimizations and convergence rates in addition to avoiding local optimums. PSOs were used for feature selections by Sakri et al. [15] and tested on classifiers: NBs (Naive Bayes), KNNs (K-nearest neighbours), and fast decision tree learners.

GONNs (Genetically Optimized Neural Networks) were proposed [16] for classification, where NNs (Neural Networks) were genetically evolved to improve the network's architecture in terms of shape and weights. The destructive character of operators was also reduced. The study's crossover and mutation operators differed from normal GAs crossover and mutation operators. The proposed GONNs determined if BCs were benign or malignant. GONNs were tested on the UCI Machine Learning repository's WBCD database which was splitting the dataset into training and test datasets and compared with BPNNs for its performance evaluations. GONNs achieved a classification accuracy of 98.24 percent, 99.63 percent, and 100 percent, respectively in the splits and for 10 fold cross-validation, it achieved a classification accuracy of 100 percent. Thus, the GONNs were viable alternatives to detecting BCs. Alzubaidi et al. [17] proposed TLA (Transfer learning approach) where DLTs were trained and fine-tuned. Their study used TLAs in two different ways: Initial training on the same domain and target dataset, followed by different domain and target datasets. The study found TLAs could improve performances. Their scheme, running in parallel convolution and residual connection executions, differentiated BC biopsy images into normal, invasive cancers, carcinomas, and benign cancers. Their scheme's performances on validation sets showed superior performances when compared to other approaches, scoring 97.4% and 96.1% accuracy in training and testing sets respectively while classifying BCs from images of the ICIAR-2018 dataset.

DLTs extracted important visual characteristics in the study [18], proposed a BC classification model. Their DLTs extract better features when compared to handmade feature extraction methods. The study used a unique boosting method where efficiency was enhanced by gradually merging weak classifiers with stronger classifiers. The study categorised eosin and hematoxylin stained biopsy images of the breast into carcinomas/non-carcinomas while classifying them into normal, benign and carcinomas. The scheme's application to histopathological BC images that were complex for diagnosis using traditional methods, showed that the classifier with boosting DLTs performed better than other techniques by a substantial margin. Three distinct machine learning methods, KNNs, NBs, and SVMs, were examined using Weka software [19] was the study found k-NNs and SVMs were more accurate as they had identical confusion matrices and accuracy values. DTs (Decision trees), RFs, NNs, extreme boost, LRs (logistic regressions), and SVMs were used [20] for predictions. The study grouped information was based on the receptor status of BC patients determined by immune histories and advanced modelling using RFs. Key elements were prioritised on Wisconsin Breast Cancer datasets by RFs variable selections and finally, DTs were used and validated for patient's survival analysis. In [21] evaluated a multitude of MLTs including SVMs, DTs (C4.5), NBs, and k-NNs to evaluate the algorithmic efficiency and effectiveness of classifications, where SVMs scored 97.13% inaccuracy and minimal error rates when simulated on WEKA. Scores for the likelihood of BCs were proposed [22] in their Breast Imaging Reporting and Data System. Clinicians had to just focus on suspect breast areas to obtain regional scores and thus deduce lesions/tumors/abnormalities to confirm the existence of BCs. The study's analysis of variances was compared in terms of AUCs (Area under characteristic) curves, specificity, sensitivity, and reading time and generalised linear models for repeated measurements. In [23], introduced DCNNs (Deep CNNs) for screening of BCs using 1000000 images and with over 200000 tests. Their evaluations of screening resulted in an AUC of 0.895 in predicting the existence of BCs. The study's ResNets tuned high-resolution medical images in terms of depth and breadth. Classifications from screening are similar to processing high noisy labels. For selecting the best approach to combine various input perspectives from multiple options, 14 readers reading 720 screened mammography images were used to verify the proposed model where it was equal

to expert radiologist references in terms of accuracy. The hybrid model combined radiologists' prognosis of malignancy with predictions from the proposed NNs for matching diagnostics. The proposed network's performance was assessed on several subpopulations of screening as well as the model's architecture, training method, mistakes, and internal representation characteristics. In [24], a new classification system based on enhanced neural networks and correlated-contours fuzzy rules (EC-FR) is used. By using an optimization approach known as a bat, more significant aspects from the input samples are extracted. An effective framework is built-in according to fuzzy rules extraction, with similarity-based directional components of data partitioning and cloud data creation used in the second layer. Wavelet functions are used to compute the weight and bias values of neurons are used in the system. To evaluate fuzzy sets, a novel shape able membership feature with adaptive shape is used to outline contours with different shapes. Following that, the derived fuzzy rules parameters are fine-tuned using the hybrid optimization approach [25].

BI-RADS density classification using MIAS, based on a lightweight Convolutional Neural Networks (CNNs) architecture is presented [26] . This is a small data problem as MIAS has only 322 images with ground truth, so we use image pre-processing and augmentation to solve the problem. Five-fold cross validation is used to evaluate the proposed approach, and has achieved a test accuracy of 83.6% on average. This suggests that deep learning has the potential to address the small data problem in mammography, which is prevalent in many medical image analysis tasks. The experience we have, especially in how to optimize the deep learning architecture, will benefit other researchers and medical practitioners.

Thus, the primary shortcoming of present designsis that information beyond ROIs (regions of interests) are ignored. Certain studies used patch-level classifiers as the first level of feature extractions over which further layers executed in models. Most model's trainings do not fit into GPU memory, a drawback as the size of minibatch is restricted utilized (typically to one) while in terms of depth at the patch level they are densely applied.Hence, this work proposes hybrid model using CNNs, and Resnetsfor improving detection accuracyof BCs.

3. Proposed methodology

Models increase network depths to enhance performances while assigning the complete image to a class, but increasing network's depth decreases performances when allocating each pixel to a certain class. Hence, this paper, proposes a hybrid MLT for prediction of BCs. This study uses CNNs to recognize images by working on entire images by integrating basic concepts of ResNet and CNNs for categorization of BCs and representing pixels as classes. This paper's proposed novel methodology is also assessed in terms of performance metrics in its evaluations. The suggested prediction model's procedure is depicted as Fig. 1.

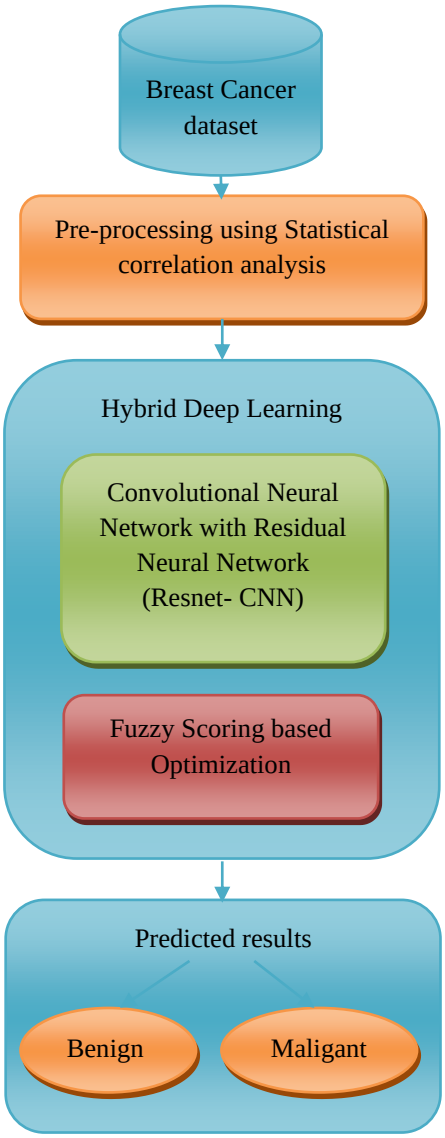


Figure. 1 The process flow of this work's proposed BC prediction model

The flow of this work can be summarized as follows:

- SCCs (Statistical correlation Coefficients) are used for pre-processing image data.
- The hybrid model proposed for predicting BCs combines core concepts of ResNets and CNNs into a single model for classify BC related data.

Table 1. Notation list

Symbol/Notation	Description
$\hat{y}$	*y's* predicted value
$\bar{y}$	*y's*- mean value
x	Independent variable (Expressed as a percentage).
H(x)	Mapping for stacked layers and x is the input of first layers
F	Residual function
F(x)+x	Original function
x	Input layer vector,
y	output layer vector
$F(\mathbf{x}, \{Wi\})$	Learned residual mappings
σ	ReLU with the omission of biases
Ws	Linear projection
T	Training set
x_i	Data variable
y_i	Class label
m	category count
$E(y_i x_i)$	Likelihood of x_i
W_i	Weight of ith layer
$b_i.$	bias of i^{th} layer
a	Negative growths
b	Positive growths
i	Highest category

- A novel Fuzzy Scoring based building block is introduced (FS-Resnet CNN) for optimizing the networks.

Table 1. Lists Notations and their description for the proposed model.

3.1 Pre-processing using SCCs

The degree of connection between two variables may be defined as the mean where variables representing outcomes, dependency, responses and explained variables are considered in computations. If a variable x is covariate to another variable y, then x, y are continuous when their values fluctuate together. Correlation coefficients assess linear connection strengths instead of predictions or r_{xy}(r).

Though many numerical methods are available in addition to graphing for assessment of data's regression fits, determining coefficientsare useful statistics to examine, R^2, when data has to be fitted into linear regressions irrespective of the count of predictors and where R^2 squares total computed for regressions divided by the number of squares counted for by (SS_{Reg}) equals the sum of the squares of departure from the mean (S_{YY}) for a model with constant term (homoscedastic case, w_i=1)

$$R^2 = \frac{SS_{Reg}}{S_{YY}} = \frac{S_{YY}-SSE}{S_{YY}} = 1 - \frac{SSE}{S_{YY}} = 1 - \frac{(y_i-\hat{y}_i)^2}{(y_i-\bar{y}_i)^2} \quad (1)$$

Where, $\hat{y}$–*y's* predicted value, $\bar{y}$ - *y's*- mean value where both totals are over *i*=1,2, . . . n. *SSE* - residual sum of squares.

Models without constants, R^2=1−*SSE/SST*, where stands for cumulative sum of squares y^2. R^2 in Eq. (1) computes total variation of $\bar{y}$ obtained fromregression. When R^2 is large it implies variations of $\bar{y}$ *get* reduced due to *x the* independent variable (Expressed as a percentage). Since $0 \leq SSE \leq SS_{YY}$, this implies $0 \leq R^2 \leq 1$ and *y* and $\hat{y}$are correlated by *R*

$$R = r_{y\hat{y}} = \frac{\sum(y_i-\hat{y}_i)(y_i-\bar{y}_i)}{[\sum(y_i-\hat{y}_i)]\,[(y_i-\bar{y}_i)]^{1/2}} \quad (2)$$

The multiple correlation coefficients, $R^2$are generated with varied coefficients by various equations should not be compared if they are generated from the same data set. Irrespective of R^2 in a regression output, determining correlations between coefficient squares of x, y with a constant can be explained as a notation for in simple regressions:

$$r_{xy} = \pm\sqrt{R^2} = \sqrt{1 - \frac{S_{YY}-a_1^2 S_{XX}}{S_{YY}}} = a_1\sqrt{\frac{S_{XX}}{S_{YY}}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} \quad (3)$$

The outputs can be positive or negative based on the regression line (slope a1) fitted in the data. Variance is explained when $R^2$is 1 as the slope fits perfectly with all points resting on the slope. Alternative when R^2 is 0 the line is horizontal implying y cannot be a function of x. Coefficients formed in regressions can also be connected as r_{xy} for generic regressions. Covariance in a normal distribution is the deviations total from anticipated values of random variables x and y and calculated as follows:

$$cov(x,y) = \frac{1}{n-1}\sum(x_i-\bar{x})(y_i-\bar{y}) \quad (4)$$

$$cov(x,y) \leq s_x s_y \quad (5)$$

Where $r \leq 1$ is implied and linearly linked. Covariance values are +ve or –ve based on the connection's slope and equals zero when x, y are independent. Also there is a possibility of random variables that are highly dependent have negligible covariance (correlation). The variance is a particular

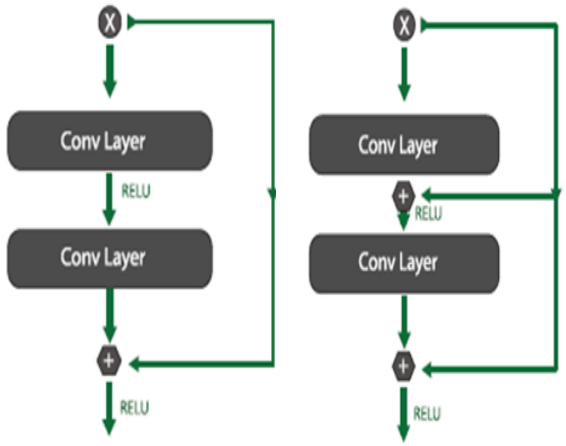
instance of the covariance of a random variable with itself, which is commonly overlooked in beginning textbooks. The standard deviation (σ)for samples s is the square root of the variance and always positive.

3.2 Hybrid DLTs

Formulation of BC categorizations as a learning problem using multiple learning frameworks is necessary as breasts might have both malignant and benign lesions. When deeper networks begin to converge, degradation problems emerge and as network depth increases, accuracy becomes saturated (predictable) and then rapidly declines which may not be caused by over fitting and addition of more layers to a sufficiently deep model leads to increased training errors as seen from experiments. The difference between shallower and its deeper version is the base for addition of more layers [27]. Deeper models map identities while copying learning data from shallower models and thus reduce their training in comparison to less deeper models. This paper introduces DLT to overcome the degradation problem. Three convolution layers followed by one average pooling layer are used in networks. Each layer's outputs become the inputs for the next layer as illustrated in Fig. 2(b), the design consists of a single completely linked cascaded residual block. Each convolution layer in this residual model takes input from all preceding convolution layers. Three convolution layers were found to be optimum after empirical evaluation of the model. Before feeding the data to the classifier, the convolution layers execute convolutions on input data while average pooling is applied in the pooling layer. Resnets, more sophisticated with many layers uses network, pooling, and big and tiny convolution layers that compute in parallel. Hence, to optimise networks this study uses a unique building block FS-Resnet CNN instead of stochastic gradient descents [28].

3.2.1. Resnets learning

DRNs (Deep residual networks) and CNNs have demonstrated their capacities to scale up to volumes of data by adding layers correspondingly while resulting in improved efficiency of results. CNNs [29] are made up of several convolution networks. Several convolution and sub-sampling layers are followed by one or more FCs in a standard CNN model where FCs are multi-layered NNs. The outputs (Class scores) are stored in the final FC layer. Convolution layers down sample data while pooling layers convolve input images using multiple filters (learnable weights). Pooling layers typically execute



(a) (b)

Figure. 2: (a) Core building block of DRNs and (b) Modified version of the residual block: with fully connected cascaded layers

two functions: maximum and average pooling. The original raw pixels are converted to final class scores by CNNs using multiple stacked layers. Several semantic segmentation models have employed CNNs as their building pieces. ResNets and CNNs are sophisticated DLTs with a lot of capability as detailed before. The fundamental building element of the DRNs is shown in Fig. 2.

The network has 152 layers and a novel design as illustrated in Fig. 2(a), that adds residual blocks to solve the challenge of training deep architectures using identity skip connections. The layer inputs are copied and forwarded to the next layer by leftover blocks. The identity skip connection is used to overcome the problem of disappearing gradients and ensures subsequent layers learn something new from inputs.

3.2.1.1. Residual representation

If $H(x)$ is mapping for stacked layers and x is the input of first layer, it can be hypothesized that non-linear layers approximate residual functions or $H(x) - x$ (when dimensionally inputs equal outputs). $H(x)$gets approximated to residual function $F(x) := H(x) - x$ while the original function is $F(x)+x$. In spite of both function's approximations are asymptotical and approach the required functions (as anticipated), the ease with which they may be learned may differ. The paradoxical events surrounding the deterioration problem have prompted this reformulation (Fig. 2(a), left). The addition of layers in deep models results in very minimal errors when compared to shallower versions. Issues of degradations arise when non-linear layers cannot be approximated. When identities mapped are optimal, models add weights to make non-linear layers zero and thus use residual

learning. When optimal values near identity instead of 0 in mappings, discovering identity perturbations is easier than learning. Experiments (Fig. 3) show that learnt residual functions show very limited responses or mapping of identities result in high predictability and apply residual learning to every few stacked layers. Fig. 3 depicts a construction block of this work defined as:

$$y = F(x, \{W_i\}) + x \tag{6}$$

Where, **x** – input layer vector, **y** - output layer vector and $F(\mathbf{x}, \{Wi\})$ – learned residual mappings like in Fig. 3there are two layers. In $F = W2\sigma(W1\mathbf{x})$, σ stands for ReLU with the omission of biases to obtain simpler notations.$F + \mathbf{x}$ is executed using shortcut connections and element additions. A second non-linearity condition is used after addition ($\sigma(\mathbf{y})$ - Fig. 3).The shortcut connections in Eq. (6) avoid addition of parameters as it is critical to reduce model's complexity and for comparing simple with residual networks in operations which are equal in terms of parameter counts, depths, widths, and computational costs at the same time (except for negligible additions of each element). The dimensions of x, F must be the same as per Eq. (6), but when they are not the same due to input/output channels, shortcut connections are used to execute a linear projection Ws to match the dimensions:

$$y = F(x, \{W_i\}) + W_s x \tag{7}$$

Ws is only needed when matching dimensions. The residual function F can take any shape. This research uses a function F with two or three levels (Fig. 3), however additional layers are feasible. However, if F only has one layer, Eq. (6) behaves similarly to a linear layer: $\mathbf{y} = W1\mathbf{x}+\mathbf{x}$, For which no benefits have been shown. It can be noted that the afore described examples can also be used for multiple convolutions represented by $F(\mathbf{x}, \{Wi\})$ where channel by channel additions on feature maps are executed each and every element[30].

3.2.1.1.1. Shortcut connections.

Models using shortcut connections have been practised earlier in MLPs (multi-layer Perceptrons) trainings where inputs and outputs of layers are linked. Shortcut connections are used to centre layer replies, gradients, and propagated errors. Resents have shortcuts and deeper branches in its layers. They are like Highway networks where shortcut connections have gating features working in tandem.Their one important difference is that these

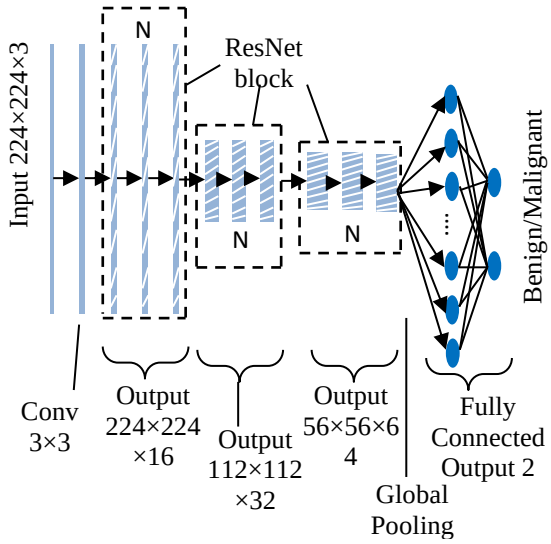


Figure. 3 Architecture of proposed Resnet -CNN model

gates are parameterized and dependent on data while identity short cur connections are free of parameters. Resnet layers are non-residual functions and the gates close while approaching 0 while in this study's formulation, identity shortcuts do not close thus transmitting information for learning further residues. Furthermore, with greatly increased depth, high-way networks have not exhibited accuracy increases (e.g., over 100 layers). Resnet has a high level of complexity, with 22 layers, and its design has network, pooling, and big/tiny convolution layers which execute in parallel. The proposed FS-Resnet CNNs are less sensitive to noises, optimise networks and are computationally with efficient usage of memory.

3.2.1.2. Fuzzy scoring and structure of fuzzy fully connected layer

Assuming training set $T = [(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)]$, and x_i stands for data variable with a corresponding label y_i where i=1,2,...,n in n training samples. This work partitioned samples into m categories for scoring and defined as $S = [S_1, S_2, \ldots, S_m]$real scores. To obtain accurate evaluations, an estimated score $\tilde{S}$=[$\tilde{S}$1,$\tilde{S}$2,...,$\tilde{S}m$] is followed with decimal parts.

Fuzzy Function: Influences between related categories are minimized in this work using a fuzzy function which uses extensional scores. CNN's generated probabilities using a sigmoid function are:

$$P(x_i|y_i = i) = \frac{e^{-E(y_i x_i)}}{\sum_{y_1}^{y_m} e^{-E(y_i x_i)}} \tag{8}$$

and its left and right neighbours:

$$P(x_{i\pm1}|y_{i\pm1}=i\pm1)=\frac{e^{-E(y_{i\pm1}x_{i\pm1})}}{\sum_{y_1}^{ym}e^{-E(y_{i\pm1}x_{i\pm1})}} \quad (9)$$

Where,$E(y_ix_i)$– likelihood of x_i predicted as y_i, m– category count. Based on studies for CNNs($x_i \mid y_i$)∈(0,1) where its probabilistic distributions can be equated as,

$$P(x_i|y_i=i)=\sigma(W_ix_i+b_i) \quad (10)$$

$$P(x_{i\pm1}|y_{i\pm1}=i\pm1)=\sigma(W_{i\pm1}x_{i\pm1}+b_{i\pm1}) \quad (11)$$

Where,W_i–weight of ith layer, b_i – bias of th layer. For minimizing neighbour conglutinations/ remote classes, iteration scores were computed using:

$$\widetilde{V_0}=\frac{i\sum_{i-a}^{i+b}P(x_i|y_i)}{\sum_{i-a}^{i+b}P(x_i|y_i)} \quad (12)$$

Where, a- negative growths and b- positive growths i.e. normal trends indicate healthiness while negative trends indicate cancer. Hence, outputs changed by these operators tend towards global averages and redistributed probabilities from Eq. (8) get optimized as

$$\tilde{P}(x_i|y_i=i)=\frac{|i-V_0|}{b-a}\times\sum_{j=i-a}^{i+b}\frac{e^{-E(y_j,x_j)}}{\sum_{y_1}^{ym}e^{-E(y_j,x_j)}} \quad (13)$$

Errors between estimated and real back propagations was modified from

$$\varepsilon_i=y_i-P(x_i|y_i=i) \quad (14)$$

to

$$\tilde{\varepsilon}_i=y_i-\tilde{P}(x_i|y_i=i) \quad (15)$$

Where modified $\tilde{\varepsilon}_i$ ‘s influence can computed using:

$$\phi=\tilde{\varepsilon}_i-\varepsilon_i \quad (16)$$

then,

$$\phi=\frac{|i-V_0|}{b-a}\times\sum_{j=i-a}^{i+b}\frac{e^{-E(y_j,x_j)}}{\sum_{y_1}^{ym}e^{-E(y_j,x_j)}}-\frac{e^{-E(y_i,x_i)}}{\sum_{y_1}^{ym}e^{-E(y_i,x_i)}} \quad (17)$$

As, are variables the between them is a constant and the left side of the formula turns into

$$\frac{|i-V_0|}{b-a}<1 \quad (18)$$

When category$ii-aa$- category $ii+bb$ probabilities are similar then it can be discovered that:

$$\phi=\left(\frac{|i-V_0|}{b-a}-1\right)\times\frac{e^{-E(y_i,x_i)}}{\sum_{y_1}^{ym}e^{-E(y_i,x_i)}} \quad (19)$$

Thus, ≤0. Though category $i-a$ to category $i+b$probabilities are not the same, category i is defined as the highest category and:

$$\sum_{j=i-a}^{i+b}\frac{e^{-E(y_j,x_j)}}{\sum_{y_1}^{ym}e^{-E(y_j,x_j)}}\le\frac{e^{-E(y_i,x_i)}}{\sum_{y_1}^{ym}e^{-E(y_i,x_i)}} \quad (20)$$

It can be concluded that $\phi\le0$ in the two above described conditions implying that$\tilde{\varepsilon}$ will have reduced influence in NNs while taking into account FCs global optimal strategy. Prior to assembling FS into Resnet-CNNs, specific category parameters are updated based on traditionally possible errors where the errors could induce a chain reaction in all layers, thus leading them away from global optimum. For example, the error $\varepsilon1$appearing in Grade1 point which presents in Fig.4. go through every layer from Grade1 point to input layer. The thick green arrow shows the back propagation of $\varepsilon1$ before assembling FS into Resnet-CNNs, and referring to the typical ReLU functions which can open or close the

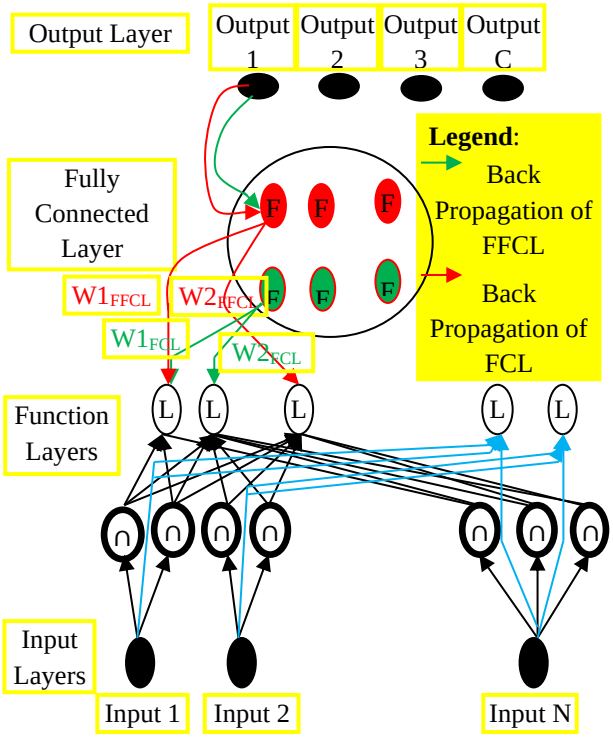


Figure. 4 Resnet-CNNs provide a conceptual understanding of the FS's structure. There are four layers namely input, function, fuzzy transformation and output layers

connection between previous and current layers, $Fa1$ connects with $L1$ and $L2$. However, FSis embedded into Resnet-CNNs, $\varepsilon 1$'s influence switches to global optimized errors and connections are changed to $F1$ with $L1$ and $L3$, as hyperparameter $W1_{FCL}$is nearer to adispersive solution in CNN's training and optimized by the proposed fuzzy strategy. Redundant functions/blocks ($Lz-1$ and Lz) appear in NNs due to NN's attributes. Figure four's function layers include traditional structures like convolution, ReLU and pooling layers. The proposed NNs structure was modified, or refined for effectively predicting BCs.

4. Results and discussion

Two assessment measures were used in this work to assess the efficiency of the suggested model in distinguishing distinct classes of BCs from images. In all the trials, training set optimized model's parameters, while the validation set tuned hyperparameters and training methods of the model. Breast Cancer dataset images were used for the data sets. Several criteria evaluated the efficacy of techniques in the prediction of BCs from data sets. TPs (True positives), FPs (False Positives), TNs (True negatives) and FNs (false negatives) were the base for performance metric evaluations. Performance metrics of precision, Recall were used as the base for evaluating the suggested work.Proposed model is compared with NN[20], SVM[20] ,GONN[16] and BI-RADS[26] methods.

Despite the contradictory nature of precision and recall, a combination of these two measures with equal weights called F-measure was obtained as a single metric. The last performance indicator was accuracy defined as the fraction of properly predicted occurrences in comparison to all anticipated instances.

Precision is defined as the ratio of correctly found positive observations to all of the expected positive observations.

$$\text{Precision} = TP/TP+FP \quad (21)$$

Sensitivity or Recall is defined the ratio of correctly identified positive observations to the over-all observations.

$$\text{Recall} = TP/TP+FN \quad (22)$$

F - measure is defined as the weighted average of Precision as well as Recall. As a result, it takes false positives and false negatives.

F1 Score =

2 X (Recall X Precision) / (Recall + Precision) (23)

Table 2. Comparison table for proposed and existing methods

Metrics	NN	SVM	GONN	BI-RADS	FS-ResnetCNN
Accuracy	74	80	84	90	95
Precision	78	79	80	91.667	95.45
Recall	75	79	82	90	93
F-measure	78.5	79.5	81	90.82	94.21
Time complexity	42.12	38.53	34.11	26.17	18.113

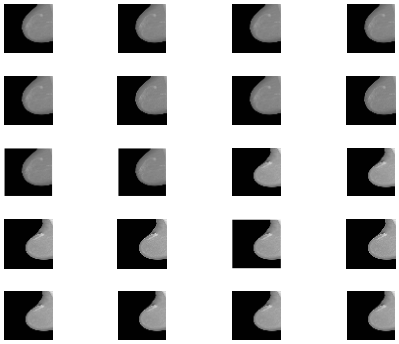


Figure. 5 Input images

Accuracy is calculated in terms of positives and negatives as follows:

$$\text{Accuracy} = (TP+FP)/(TP+TN+FP+FN) \quad (24)$$

Table 2 lists comparative results for the proposed and existing methods while the Fig. 5 illustrates input image used for analysis.

Fig. 6 illustrates pre-processed BC image while Fig. 7 depicts classified image of this work.

The precision comparisons between proposed FS-ResnetCNN approach and existing NN, SVM, GONN and BI-RADS methods are shown in fig.8. Proposed model uses Statistical correlation analysis for pre-processing by which precision results improved. The findings show that the proposed FS-ResnetCNN approach outperforms existing classification algorithms in terms of precision. Proposed FS-ResnetCNN approach achieves 95.45% and the existing NN, SVM, GONN and BI-RADS methods produces 78%,79%, 80% and 91.667% accordingly.

Fig. 9 Shows the recall comparisons between proposed FS-ResnetCNN approach and existing NN, SVM, GONN and BI-RADS methods. From the above results it is observed that the proposed FS-

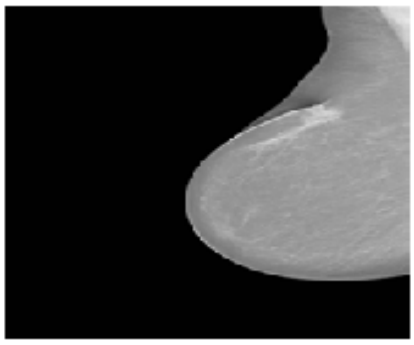


Figure. 6 Pre-processed BC image

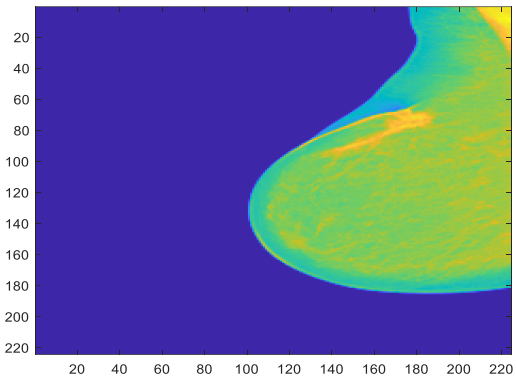


Figure. 7 Classified BC image

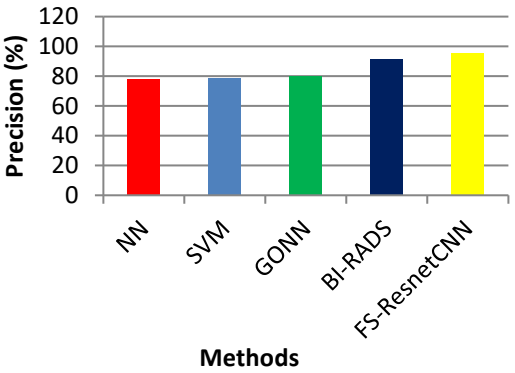


Figure. 8 Comparisons between suggested and current methods in categorising BCs based on precision values

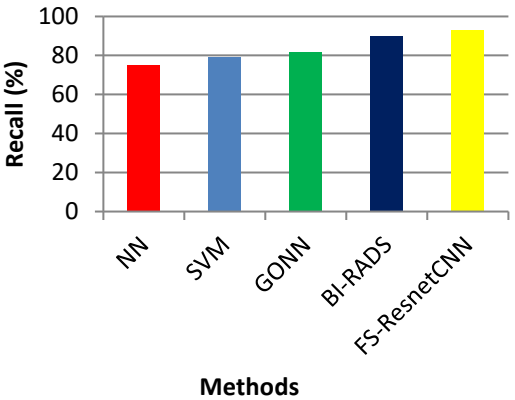


Figure. 9 Recall comparison values of the proposed and current methods for categorising BCs

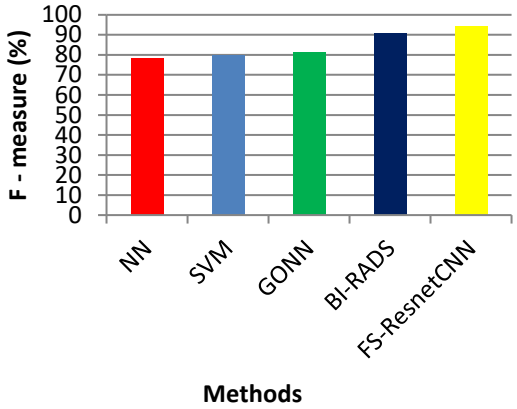


Figure. 10 Comparative F-measure values for categorising BCs

ResnetCNN approach outperforms existing classification algorithms in terms of recall. Proposed FS-ResnetCNN approach achieves 93% recall result and the existing NN, SVM, GONN and BI-RADS methods produces 75%,79%, 82% and 90% accordingly.

Performance comparisons of FS-ResnetCNN approach and existing NN, SVM, GONN and BI-RADS methods are shown in fig.10 in terms of F-measure. The findings show that the proposed FS-ResnetCNN approach outperforms existing classification algorithms in terms of F-measure. Proposed FS-ResnetCNN approach achieves 94.21% and the existing NN, SVM, GONN and BI-RADS methods produces 78.5%,79.5%, 81% and 90.82% accordingly.

The accuracy comparisons between FS-ResnetCNN approach and existing classification NN, SVM, GONN and BI-RADS methods for categorising BCs data are shown in Fig. 11. Proposed work uses deep learning for classification and it increases the accuracy results.The findings show that the proposed FS-ResnetCNN approach outperforms existing classification algorithms in terms of accuracy. Proposed FS-ResnetCNN approach achieves 95% accuracy result and the existing NN, SVM, GONN and BI-RADS methods produces 74%, 80%, 84% and 90% accordingly.

The time complexity of each algorithm is shown in Fig. 12, where the vertical axis represents the time complexity and the horizontal axis represents the methods. Fig. 12. indicate that the proposed ResnetCNN algorithm produces lower time complexity results of 18.113(S) and the existing NN, SVM, GONN and BI-RADS methods produces 42.12(s),38.53(s),34.11(s) and 26.17(s).

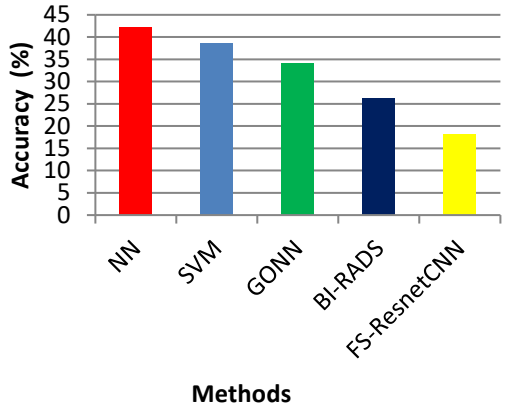


Figure. 11 Suggested and current methods comparative accuracy scores in categorising BCs

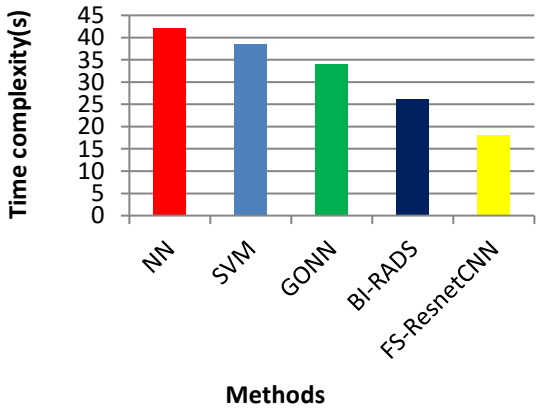


Figure. 12 Time complexity analysis between the proposed and existing algorithms

5. Conclusion

ResNet and CNN, which are powerful, sophisticated deep learning models for categorising BCs data, are combined in this study effort. This study confirmed that the visual enhancement approach does not significantly increase classification performance, and also demonstrated transfer learning can extract characteristics from medical data. Semantic scores of the medical images were generated by an integrated fully connected layer in this work's FS-Resnet CNN model. To begin, this paper introduces a pre-processing method based on statistical correlation analysis for enhancing the performance of the class. Several convolution and sub-sampling layers are followed by one or more fully connected FCs in a standard CNN model. FCs are multi-layered NNs in a traditional sense. The output, which is the class score, is stored in the final FC layer. The convolutional layer's job is to down sample the data while the pooling layer's job is to convolve the input picture using multiple filters (learnable weights). The pooling layer typically has two functions: maximum pooling and average pooling. From the original raw pixels to the final class scores, the CNN transforms the input image through multiple stacked layers. Resnets have a high level of complexity, with 22 layers, and its design uses network, pooling, and big and tiny convolution layers that are computed calculated in parallel. As a result, instead of using the traditional stochastic gradient descent to optimise the network, this study proposes a unique building block known as the Fuzzy Scoring based Resnet - CNN (FS-Resnet CNN) model. In addition, the suggested Fuzzy Scoring based Resnet - CNN (FS-Resnet CNN) model is computationally efficient, noise-resistant, and memory-efficient. The suggested FS-ResnetCNN architecture has the benefit of reducing the impact of ambiguous and imprecise semantic descriptions on medical diagnosis. Although this design can handle classification problems with overlaps between two neighbour classes in a favourable way, it is more likely to decrease the impact of semantic conglutination. Further, this research focuses on applying optimization techniques to ignore overlaps between two neighbouring classes. Experimental results shows that the proposed model produces better performance than other existing models proposed model obtains 95% accuracy ,95.45% precision, 93% recall rate, 94.21% f-measure and 18.113(S)-time complexity for breast cancer detection.

Funding Details

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Ethical Approval

This article does not contain any studies involving human and animals performed by any of the authors.

Authorship Contributions:

Satyabrata Patro and Jyotirmaya Mishra designed and performed the experiments, derived the models and comparative results for the proposed and existing methods. SatyabrataPatro wrote the manuscript in consultation with Jyotirmaya Mishra and Bhavani Sankar Panda.

References

[1] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. H. Ng, "Computer-aided breast cancer detection using mammograms: a review", *IEEE Reviews in Biomedical Engineering*, Vol. 6, pp. 77-98, 2012.

[2] C. Thirumalai and R. Manzoor, "Cost optimization using normal linear regression method for breast cancer Type I skin", In: *Proc. of International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, Vol. 2, pp. 264-268, 2017.

[3] R. Krithiga and P. Geetha, "Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review", *Archives of Computational Methods in Engineering*, Vol. 28, No. 4, pp. 2607-2619, 2021.

[4] Y. Khourdifi and M. Bahaj, "Applying best machine learning algorithms for breast cancer prediction and classification", In: *Proc. of International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pp. 1-5, 2018.

[5] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. M. Hasan, and M. N. Kabir, "Breast cancer prediction: a comparative study using machine learning techniques", *SN Computer Science*, Vol. 1, No. 5, pp. 1-14, 2020.

[6] G. S. Tewolde and D. M. Hanna, "Particle swarm optimization for classification of breast cancer data using single and multisurface methods of data separation", In: *Proc. of IEEE International Conference on Electro/ Information Technology*, pp. 443-446, 2007.

[7] B. M. Gayathri, C. P. Sumathi, and T. Santhanam, "Breast cancer diagnosis using machine learning algorithms-a survey", *International Journal of Distributed and Parallel Systems*, Vol. 4, No. 3, p. 105, 2013.

[8] K. R. Gandhi, M. Karnan, and S. Kannan, "Classification rule construction using particle swarm optimization algorithm for breast cancer data sets", In: *Proc. of International Conference on Signal Acquisition and Processing*, pp. 233-237, 2010.

[9] P. Patro, K. Kumar, and G. S. Kumar, "Neuro Fuzzy System with Hybrid Ant Colony Particle Swarm Optimization (HASO) and Robust Activation", *Jour of Adv Research in Dynamical & Control Systems*, Vol. 12, 03-Special Issue pp. 741-750, 2020.

[10] M. Tahmooresi, A. Afshar, B. B. Rad, K. B. Nowshath, and M. A. Bamiah, "Early detection of breast cancer using machine learning techniques", *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, Vol. 10, No. 3-2, pp. 21-27, 2018.

[11] A. Arafi, R. Fajr, and A. Bouroumi, "Breast cancer data analysis using support vector machines and particle swarm optimization", In: *Proc. of Second World Conference on Complex Systems (WCCS)* pp. 1-6, 2014.

[12] J. M. Jerez, I. Molina, P. J. G. Laencina, E. Alba, N. Ribelles, M. Martín, and L. Franco, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem", *Artificial Intelligence in Medicine*, Vol. 50, No. 2, pp. 105-115, 2010.

[13] W. C. Yeh, W. W. Chang, and Y. Y. Chung, "A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method", *Expert Systems with Applications*, Vol. 36, No. 4, pp. 8204-8211, 2009.

[14] H. Huang, S. Zhou, J. Jiang, H. Chen, Y. Li, and C. Li, "A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features", *BMC Bioinformatics*, Vol. 20, No. 8, pp. 1-14, 2019.

[15] S. B. Sakri, N. B. A. Rashid, and Z. M. Zain, "Particle swarm optimization feature selection for breast cancer recurrence prediction", *IEEE Access*, Vol. 6, pp. 29637-29647, 2018.

[16] A. Bhardwaj and A. Tiwari, "Breast cancer diagnosis using genetically optimized neural network model", *Expert Systems with Applications*, Vol. 42, No. 10, pp. 4611-4620, 2015.

[17] L. Alzubaidi, O. A. Shamma, M. A. Fadhel, L. Farhan, J. Zhang, and Y. Duan, "Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model", *Electronics*, Vol. 9, No. 3, p. 445, 2020.

[18] D. M. Vo, N. Q. Nguyen, and S. W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks", *Information Sciences*, Vol. 482, pp. 123-138, 2019.

[19] B. Akbugday, "Classification of breast cancer data using machine learning algorithms", *Medical Technologies Congress (TIPTEKNO)*, pp. 1-4, 2019.

[20] M. D. Ganggayah, N. A. Taib, Y. C. Har, P. Lio, and S. K. Dhillon, "Predicting factors for survival of breast cancer patients using machine learning techniques", *BMC Medical Informatics*

and Decision Making, Vol. 19, No. 1, pp. 1-17, 2019.

[21] H. H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis", *Procedia Computer Science*, Vol. 83, pp. 1064-1069, 2016.

[22] A. R. Ruiz, E. Krupinski, J. J. Mordang, K. Schilling, S. H. H. Köbrunner, I. Sechopoulos, and R. M. Mann, "Detection of breast cancer with mammography: effect of an artificial intelligence support system", *Radiology*, Vol. 290, No. 2, pp. 305-314, 2019.

[23] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, and K. J. Geras, "Deep neural networks improve radiologists' performance in breast cancer screening", *IEEE Transactions on Medical Imaging*, Vol. 39, No. 4, pp. 1184-1194, 2019.

[24] P. Patro, K. Kumar, G. S. Kumar, and G. Swain, "Similarity and wavelet transform based data partitioning and parameter learning for fuzzy neural network", *Journal of King Saud University –Computer and Information Sciences*, Vol. 34, Issue 6, Part-B, pp. 3424-3432, 2022.

[25] P. Patro, K. Kumar, and G. S. Kumar, "Optimized Hybridization of Ant Colony Optimization and Genetic Algorithm (HACOGA) Based IC-FNN Classifier for Abalone", *Journal of Computational and Theoretical Nanoscience*, Vol. 17, pp. 2756-2764, 2020.

[26] P. Shi, C. Wu, J. Zhong, and H. Wang, "Deep learning from small dataset for BI-RADS density classification of mammography images.", In: *Proc. of 10th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 102-109, 2019.

[27] X. Xiong, Y. Kim, Y. Baek, D. W. Rhee, and S. H. Kim, "Analysis of breast cancer using data mining & statistical techniques", In: Proc. of *Sixth International Conference on Software Engineering" Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Network*, pp. 82-87, 2005.

[28] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Completely automated CNN architecture design based on blocks", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 4, pp. 1242-1254, 2019.

[29] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images", *IEEE Transactions on Information Forensics and Security*, Vol. 14, No. 5, pp. 1181-1193, 2018.

[30] C. Kang, X. Yu, S. H. Wang, D. S. Guttery, H. M. Pandey, Y. Tian, and Y. D. Zhang, "A heuristic neural network structure relying on fuzzy logic for images scoring", *IEEE Transactions on Fuzzy Systems*, Vol. 29, No. 1, pp. 34-45, 2020.