

Original Article

# Effective Breast Cancer Prediction Based on Feature Extraction, Fusion and Selection using Hybrid Methodologies

G. Rajasekaran<sup>1</sup>, C. Sunitha Ram<sup>2</sup>

<sup>1,2</sup>Department of Computer Science and Engineering, SCSVMV University, Kancheepuram, Tamilnadu, India.

<sup>1</sup>Corresponding Author : [gtrajasekaran2@gmail.com](mailto:gtrajasekaran2@gmail.com)

Received: 12 March 2023

Revised: 19 April 2023

Accepted: 14 May 2023

Published: 29 May 2023

**Abstract** - Women have been affected by many diseases for decades due to their low immune systems. Especially breast cancer is the second largest and most harmful disease for a woman that, leads to death. The earlier prediction of breast cancer can help for easy treatment and Save lives. Deep Learning methods are implemented in medical fields to attain an effective prediction and higher prediction accuracy. Therefore, in this work, the novel hybrid prediction topology is implemented two feature extractions, feature fusion, feature selection and classification with several algorithms such as Linear Discriminant Analysis (LDA), Canonical Correlation Analysis (CCA), Convolutional Neural Network (CNN) Long Short-Term Memory (LSTM), Transit Search Optimization (TSO), Support Vector Machine (SVM) respectively. The LDA and CNN-LSTM method is used to execute a two-feature extraction, and the CCA model is used to merge the two extracted features, a feature fusion task. Next, the TSO-based feature selection is implemented to minimize the redundant features. At last, the classification is carried out using an SVM method to acquire an efficient breast cancer prediction. The proposed hybrid method has achieved more accurate decision-making and efficient training than the conventional methods. The experimental result of the novel hybrid method acquired a superior performance in terms of precision, Recall, Accuracy, F1 Score, RMSE and MAE as 98.81%, 99.02%, 98.18%, 98.85%, 1.001 and 1.016, respectively.

**Keywords** - Breast cancer prediction, Feature extraction, ML/DL methods, CCA based feature fusion, TSO-based selection, Classification, Performance accuracy.

## 1. Introduction

Breast cancer is the most drastically developing disease in these decades., becomes a significant public health problem for women in the world. A survey-based on the Cancer Society 2009 stated that nearly 269,800 women are affected and die by cancer in every country, and 15% of deaths were caused by breast cancer [1]. There 27% of breast cancer were diagnosed among 713,220 cases in every country. Breast cancer has been categorized into two main types: benign and malignant [2].

In recent times, there has been a massive growth of technological inventions implemented. Computer-Aided Detection (CAD) systems are used in various fields like agriculture, medical, military, communication and education [3]. The medical field has benefitted from this CAD system, which can effectively detect an abnormality in human health data. There are a lot of medical applications such as Brain Tumour, Breast Cancer, Kidney Stones, Lung cancer, Alzheimer, Heart disease, skin cancer, and so on that are implemented for prediction using CAD systems [4]. In this

work, Breast cancer predictions are necessary to detect it early and accurately using a CAD model.

The conventional CAD system has been processed using pre-processing, feature extraction and classification [5]. However, in some cases, the Feature selection is also considered to select the required feature and avoid an extra feature [6]. These steps are easy to identify the cancer types, benign or malignant, in terms of statistical properties, textures, fractal domain, spatial domain, and wavelet bases [7]. Some of the CAD systems of Deep Learning (DL) and Machine Learning (ML) models used for medical technologies are k-means Nearest Neighbor (KNN) Algorithm, Support Vector Machine (SVM), AdaBoost Algorithm, CNN-LSTM, MobileNet, AlexNet, ShuffleNet, DenseNet and Back Propagation Neural Network (BPNN) respectively [8].

Due to the classification complexity, metaheuristics methods are used to finetune the hyperparameter of DL models to provide an optimal result in classification [9]. The



metaheuristics algorithm has overcome the optimization issues and reduced the complex computations. A few frequently used metaheuristics methods are simulated annealing, iterated local search, genetic algorithm, tabu search, variable neighbourhood search, evolutionary computation, ant colony optimization, Ant Lion optimizer (ALO), Particle Swarm Optimization (PSO), Memetic algorithms, Firefly Optimization Algorithm (FOA) and rider optimization algorithm etc are implemented for optimal prediction [10].

In this work, the novel prediction topology is presented with various combinations of algorithms. This work has the novelty of performing a two-feature extraction task using various algorithms, such as the LDA and CNN-LSTM methods. Then these two features are fused using a Canonical Correlation Analysis (CCA) model to combine all attained features. Next, feature selection is carried out by a Metaheuristics TSO method to reduce the redundant features. Then the SVM-based classification is used to predict breast cancer with accurate decision-making and efficient training. Thus, the proposed method has achieved superior performance in Precision, Recall, F1 score, Accuracy, MAE and RMSE, respectively.

## 2. Related Works

In recent decades, the medical fields have been highly advanced due to an ML/DL methods invention. The related works are discussed for breast cancer prediction-based ML/DL methods. Huang et al. [11] described an SVM method for small- and large-scale datasets for a breast cancer diagnosis. The SVM has attained good training and delay overhead of 99.52% accuracy, 0.876 ROC, and 99.5% F1 score, respectively. Also, a few cases discussed the probability of breast cancer death using several methods like DT, RF, SVM, NN, and LR, respectively. It was developed by Li et al. [12] and has processed multiple attributes, and the result showed that the RF method has a superior AUC of 0.989 to other methods. Besides, Sivakami et al. [13] implement the hybrid DT-SVM methods for breast cancer feature Extraction to attain an Information Treatment.

The classification accuracy of DT – SVM is 91% which outperformed the conventional methods. Moreover, extreme learning machines (ELM) has used for breast cancer feature selection based on their types of normal, benign and malignant by Melekoodappattu et al. [14]. It has effectively attained the performances in terms of sensitivity (97.5%), precision (100%) and accuracy (99.04%), respectively.

Some metaheuristics models are also used for a prediction that Sangaiah et al. [15] explored an entropy-based genetic method based on Relief attribute reduction. This method has achieved 85.89% of accuracy in breast cancer diagnosis. Another method has been presented SVM-

ANN method for breast cancer prediction by Lim et al. [16], using the MIAS dataset comprising 40 benign and 40 malignant data. This hybrid model's performance accuracy of 98% is attained with effectiveness. Also, the Tri-Branch DL method is presented by Chen et al. [17] using ultrasound data for immunohistochemical HER2 prediction. It attained an accuracy of 86.23% for HER2 prediction. Alongside this, the hybrid method of Ada Boost-SVM-Kmeans is used for breast cancer classification [18, 19]. This hybrid method achieved a maximum accuracy of 98.85%, providing an earlier diagnosis.

Thawkar et al. [20] investigated a Butterfly optimization algorithm (BOA)-ALO method for a feature selection to provide an optimal result in prediction. This BOA-ALO method has presented an accuracy of 96.82% than the previous methods. In some cases, to attain an efficient objective function, Melekoodappattu et al. [21] used a Glowworm Swarm Optimization (GSO).

This method can be used by several methods like GSO, ELM, and FOA for breast cancer prediction and performances achieved of precision at 100% and sensitivity of 97.91% and an accuracy of 99.15%, respectively. Moreover, the ANN and Bayesian methods were implemented for breast cancer detection by Choi et al. [22] and acquired a better prognosis evaluation of 87.2% accuracy, 93.3% sensitivity and 83.1% specificity.

Another method, Mohebian et al. [23], discussed the Hybrid Predictor of Breast Cancer Recurrence (HPBCR) using an SVM, Multilayer Perceptron (MLP) and DT methods. The performance result has attained 77% sensitivity, 93% specificity, 95% precision and 85% accuracy, respectively.

## 3. Proposed Methodologies

In this section, the proposed methodology is discussed with detailed explanations. The proposed block diagram is presented in Figure 1, which comprises the dataset, feature extraction, i.e., LDA-based Feature A and CNN-LSTM-based Feature B, CCA-based Fusion, TSO-based Feature selection and SVM classification.

### 3.1. Dataset Description

The UCI repository is the popular dataset for breast cancer diagnosis. The datasets are categorized as standard and cancer-affected data based on benignancy and malignancy, only identified by the affected and standard datasets. In this article, the Wisconsin Data set for Diagnostic Breast Cancer (WDBC) has 32 patients' data of 569 patterns. All these patterns have provided several attributes that are shown in Table 1. From the table, the attribute (4-13) is classified into mean, standard error and worst result, containing 30 attributes, respectively.

**Table 1. Dataset attributes**

S.NO	Attributes
1	ID number
2	Benign result
3	Malignant result
4	Radius
5	Perimeter
6	Texture
7	Smoothness
8	Area
9	Compactness
10	Concave points
11	Concavity
12	Facial dimension
13	Symmetry

**3.2. Feature Extraction**

To obtain appropriate feature data and extract information from it, feature extractions are performed.

**3.3. Feature Extraction-A Based on LDA**

The LDA model is used as the first part of feature extraction, denoted as feature A. This model classifies linear feature combinations into more than one class. The maximum number of linearity classes are used to partition the directions

shown in Figure2. Selecting the best outcome in a specific direction assists with data discrimination [24].

To calculate a direction, the high separation coefficient (F) is expressed as the ratio between class scatter ( $S_m$ ) and within-class scatter ( $S_w$ ), which is given in Equation (1). The higher separation can be carried out by attaining a maximum value of F.

$$F = \frac{\text{tr}(S_m)}{\text{tr}(S_w)} \tag{1}$$

Based on class c, the  $S_c$  and  $S_w$  is estimated in terms of overall classes (N), mean value ( $\mu_c$ ) and the number of  $x_i$  ( $n_i$ ) for every class that is given in the below equation (2-3):

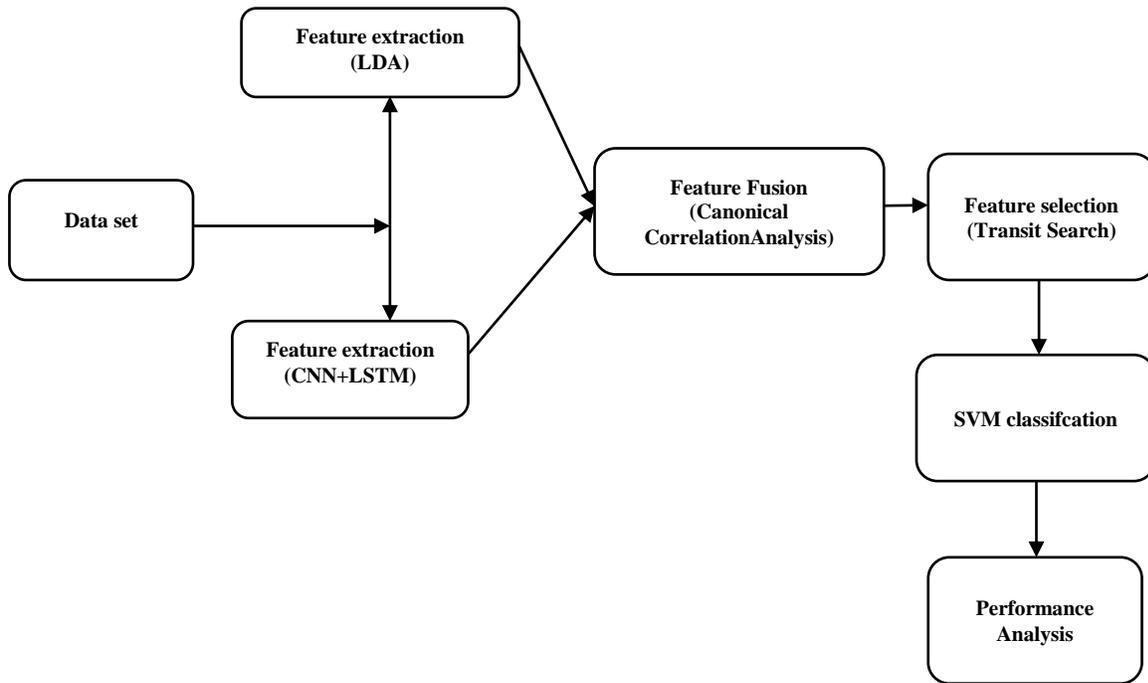
$$S_c = \sum_{i=1}^N (x_{ic} - \mu_c)(x_{ic} - \mu_c)^T \tag{2}$$

$$S_w = \sum_{i=1}^C \frac{n_i}{N} S^i \tag{3}$$

Then the value of the class scatter ( $S_B^C$ ) for class c is defined as the difference between mean of  $x_i$  for class i ( $\mu_i$ ) and the mean of all  $x_i$  ( $\mu$ ) that is in equation (4).

$$S_B^C = \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T \tag{4}$$

$$S_m = \sum_{i=1}^C \frac{n_i}{N} S_B^i \tag{5}$$



**Fig. 1 Proposed architecture for breast cancer prediction**

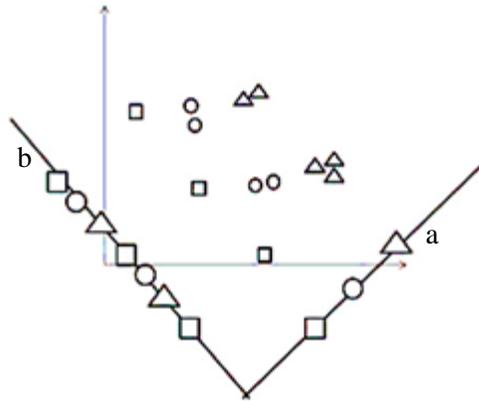


Fig. 2 Direction partitioning of linear class

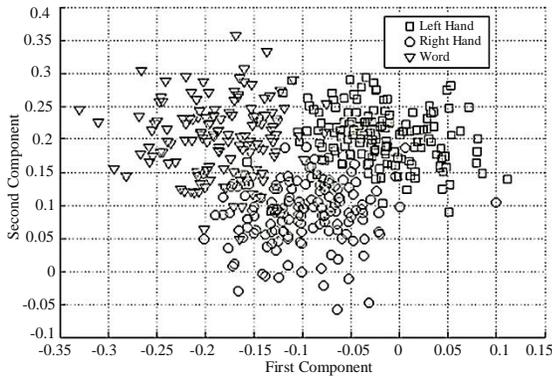


Fig. 3 Data following LDA transformation

Assume Matrix (S) as the maximum eigenvalue, the best class direction partitioning is evaluated as in Equation (6).

$$S = S_w^{-1} S_m \tag{6}$$

To generalize an eigenvalue based on an eigenvector, the transformed data set (y) is expressed in terms of a weighted matrix (W) with a Eigenvector (M), i.e.,  $w_1, w_2, \dots, w_M$  and transformed linear feature input ( $x^T$ ) that is given in Equation (7).

$$y = x^T W \tag{7}$$

The LDA method has provided 184 data transmission features, showing two LDA data shown in Figure 3. The selected linear feature information from various classes is obtained with a better feature set reduction in dimensions.

### 3.4. CNN-LSTM-Based Feature Extraction-B

Feature B is carried out using a hybrid CNN and LSTM methods which belong to a DL model. A block diagram of the CNN-LSTM technique is presented in Figure 4, which combined both the feature layers in CNN and LSTM to attain an adequate accuracy in prediction [25-27]. From the figure, the first part of the block is implemented by CNN layers, and an LSTM layer presents the second part of the block. The CNN method extracts the features, and the LSTM method is used for feature interpretation across time steps.

The CNN-LSTM method comprises various layers: input, flatten, pooling, convolutional, LSTM and fully connected. The data is fetched by an input layer and sent for convolution. The convolution extracts the spatial features at adjacent time points with a matrix connection of similar weights. The feature maps are performed as a training set using a CNN parameter.

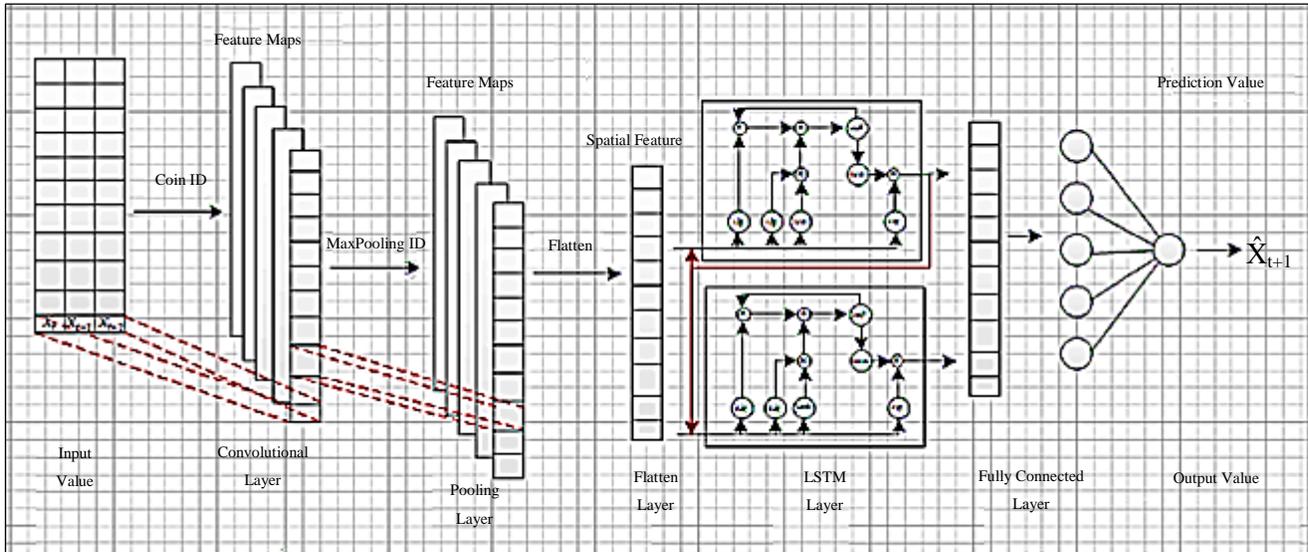


Fig. 4 CNN-LSTM architecture

Then the signal features are observed and sent to the next layer. The pooling is used for downsampling to provide an output of the feature map. Then the 1-Dimensional (1D) vectors are provided as an output from a flattened layer. The 1D vectors are fetched by an LSTM layer based on long-term time series in sequential modelling.

The LSTM benefits from vanishing gradient issues with high decision-making in a sequential series. It has three main gates such as forgets, input, and output gates. The hidden features are stored in the memory block where the temporal data is stored. The multiplicative units are processed by these gates that can be controlled the input, save and output in the memory. The memory blocks are provided with a feature and transferred to a fully connected layer that executes a feature map's prediction. Thus, the dense layer showed the feature maps across time with a superior prediction result.

### 3.5. CCA-Based Feature Fusion

Feature fusion merges the extracted output features of CNN-LSTM and LDA methods. A CCA method, an efficient fusion tool, implements this feature fusion. The CCA method evaluates the two random vector's correlation issues: A and B. The A denotes the extracted features of A, and B denotes the extracted features of B. where  $A = \{A_1, A_2, \dots, A_N\}$  and  $B = \{B_1, B_2, \dots, B_N\}$ .

The CCA is used to identify the directions of a and b to enhance the correlation of  $\alpha^T A_i$  and  $\beta^T B_i$  and simultaneously reduce the correlation among  $\alpha^T A_i$  and  $\beta^T B_i$ . The optimal x and y results are evaluated with the maximum correlation  $\text{Corr}(\alpha^T A_i, \beta^T B_i)$ . Thus, the fitness function is expressed to reduce these issues in the following Equation (8).

$$f(\alpha, \beta) = \text{Corr}(\alpha^T A_i, \beta^T B_i) = \frac{\text{COV}(\alpha^T A, \beta^T B)}{[\text{Var}(\alpha^T A) * \text{Var}(\beta^T B)]^{1/2}} \quad (8)$$

On simplifying, we get Equation (9)

$$f(\alpha, \beta) = \frac{\alpha^T S_{xy} \beta}{(\alpha^T S_{AA} \alpha \beta^T S_{BB} \beta)^{1/2}} \quad (9)$$

Where  $A_i$  and  $B_i$  indicates the  $i^{\text{th}}$  feature samples,  $S_{AA}$  indicates the covariance matrix of A and  $S_{BB}$  denotes the covariance matrix of B and also the  $S_{AB}$  symbolizes the covariance matrix of A and B.

Therefore, an Eigen Function is attained on simplification based on  $\alpha$  and  $\beta$  relation, and then we get Equation (10)

$$\begin{cases} S_{AA}^{-1} S_{AB} S_{BB}^{-1} S_{BA}, \alpha = \lambda^2 \alpha \\ S_{BB}^{-1} S_{BA} S_{AA}^{-1} S_{AB}, \beta = \lambda^2 \beta \end{cases} \quad (10)$$

Consider x and y are the eigenvector, then the matrix  $X_d = \{\alpha_1, \alpha_2, \dots, \alpha_d\}$  and  $Y_d = \{\beta_1, \beta_2, \dots, \beta_d\}$  corresponding to the highest function of eigenvalues d. Based on the eigenfunctions, the overall feature fusion is expressed in Equation (11).

$$z = \begin{pmatrix} X_d & 0 \\ 0 & Y_d \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad (11)$$

Where the  $S_{AA}$  and  $S_{BB}$  are required to be non-singular, which is a full rank matrix. Therefore, the CCA method-based feature fusions are achieved with practical analysis and evaluation. Next, the feature selection is to be carried out.

### 3.6. TSO-Based Feature Selection

The feature selection process is necessary to construct an efficient classification model. This process reduces the number of extracted features based on their redundancy and requirement. This section uses the TSO method to provide a more accurate selection for classification. The TSO method is discussed below.

#### 3.6.1. Transit Search (TS)

The TSO method is a popular physics-based meta-heuristic algorithm [28]. The transit is discovered by more than 3800 planets using a space telescope database. The Transit model has identified the radial velocity of more than 915 planets until March 2022. By detecting a small cosmic scale planet, the efficiency of this Transit method in astrophysics is superior in the result. Based on this idea, the stars emit light at certain intervals, and luminosity changes are observed. When the luminosity of light is reduced, it is declared that the planets are crossing over the stars. This process can be performed in several phases: Galaxy, Transit, Planet, Neighbour, and Exploitation, respectively.

Galaxy phase: This phase chooses a galaxy to randomize the search space location for a galaxy centre selection. The galaxy's habitable zones are estimated after identifying the location.

Transit phase: In this phase, the transit is to be measured. The amount of received light is measured again from the star to notice its light signal reduction.

Planet phase: This phase carries the process of the star's light being received by the observer (telescope) and then observing a reduction value of light from transit. The planet is determined by light reduction due to its crossing between the star and the telescope, the light gets reduced.

Neighbour phase: In this phase, if the present planet does not occur any transit, then the priorly determined neighbourhood planet is replaced for a better condition of host life

Exploitation phase: In this phase, the detected planet evaluates the planet's characteristics and host life conditions.

Based on these phases, the TSO method is implemented for this feature selection work which is discussed in Algorithm 1 and for more clarification, refer to [28].

Algorithm 1 : Pseudocode of TSO Algorithm

Input : Population of host stars  $n$ , SNR, iteration quantity  $n_i$   
Output : overall Planet Location ( $L_B$ ), the Fitness value

```

1: Initialize galaxy location
2: do
3: Evaluate Galaxy Phase
4: Return Best Stars
5: While (stop condition satisfied) do
6: Evaluate the Transit Phase
7: for  $i=1:n$ 
8: If transit determined
9: do
10: Determine Planet phase
11: else
12: do
13: Evaluate Neighbour Phase
14: end
15: end
16: Return present Planet Location and present Fitness
17: end
18: Return  $L_B$  and overall Fitness value

```

### 3.7. Classification

After executing the feature selection process, limited and required features are shortlisted for the following classification task. The classification task is performed by an SVM method to provide an accurate prediction in it.

#### 3.7.1. SVM classifier

The SVM is based on the ML method, the most popular classifier model with a binary class [29-31]. These classes are insisted on to identify the hyperplane boundaries among the classes. This hyperplane is used to enhance the margin in acquired training data. The hyperplanes with a training sample near the boundary are support vectors. The margin is the estimated distance between hyperplane boundaries and support vectors. Based on these features, the SVM can be defined as a decision boundary by using a decision plane. Thus, the data classification based on the SVM method can be executed for training and testing data with few data instances, such as target class labels and feature attributes.

The SVM model maps the output data from feature selection into kernel space. In the kernel spaces, various types are presented, such as linear (uses dot product), polynomial, quadratic, Multilayer Perceptron kernel, Radial Basis Function kernel etc. several methods like minimal sequential optimization, quadratic programming and least

squares are used to execute SVM model. The kernel data is modified and selected based on the attained support vectors. The final result is acquired by a modified kernel result for a prediction.

## 4. Experimental Results

The experimental result of the proposed method is discussed in this section. The proposed method is executed and evaluates the feature extraction, fusion, selection and classification with various hybridization methods. The 32-attribute WDBC dataset is used in this work for both the training and testing phases. There are 65% of the data is used for training, and 35% of the data is used for testing. Classification performance is determined using precision, Accuracy, Recall, F1 score, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The proposed method's performance result is compared with the conventional method to prove its effectiveness. Some conventional methods used for comparisons are CNN-RNN, BPNN-LSTM, CNN-LSTM, PSO-CNN, RF-SVM and ACO-SVM, respectively. The following equations are used to evaluate the performance metrics [32].

$$\text{Precision} = \frac{T^+}{T^+ + F^+} \quad (12)$$

$$\text{Recall} = \frac{T^+}{T^+ + F^-} \quad (13)$$

$$\text{Accuracy} = \frac{T^+ + T^-}{T^+ + F^+ + T^- + F^-} \quad (14)$$

$$\text{F1 score} = \frac{2\text{Precision} \times \text{Recall}}{\text{Recall} + \text{precision}} \quad (15)$$

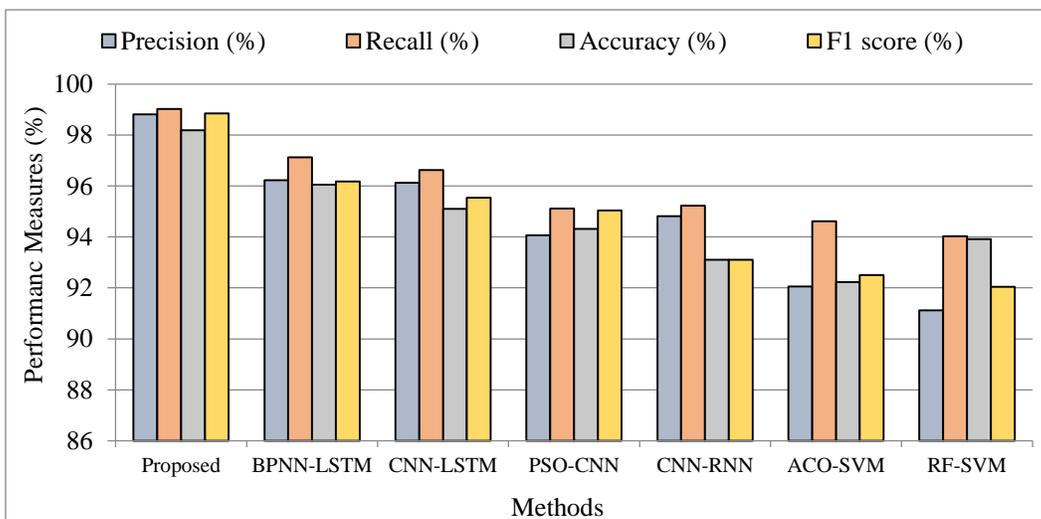
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{a=1}^n (\text{predicted cases} - \text{overall cases})^2} \quad (16)$$

$$\text{MAE} = \sqrt{\frac{1}{n} \sum_{a=1}^n |\text{predicted cases} - \text{overall cases}|} \quad (17)$$

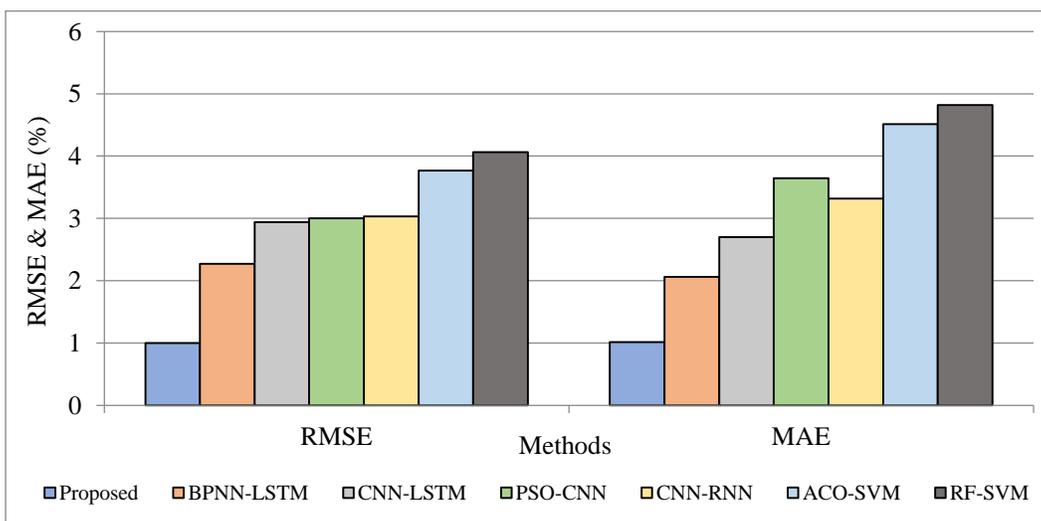
Where,  $T^+$  represents the accurately positive,  $T^-$  denotes the accurate negative,  $F^+$  indicates the inaccurate positive result,  $F^-$  denotes the inaccurate negative result, and  $n$  indicates overall patients. Table 2 and Figure 5 show a Classification result of the proposed and conventional methods. The proposed hybrid method has achieved a classification metrics value of precision, Recall, Accuracy, and F1 Score as 98.81%, 99.02%, 98.18% and 98.85%, respectively. The result showed that the proposed model performs more effectively in classification than the conventional methods. Table 2 and Figure 6 show the error validation result of the proposed and conventional methods. The Error validation result of the proposed hybrid method attained by RMSE and MAE is 1.001 and 1.016, respectively. Thus the result of the proposed model has achieved a minimum error than the conventional methods.

**Table 2. Overall metrics results**

Methods	Precision (%)	Recall (%)	Accuracy (%)	F <sub>1</sub> score (%)	RMSE	MAE
Proposed	98.81	99.02	98.18	98.85	1.001	1.016
BPNN-LSTM	96.23	97.12	96.05	96.17	2.270	2.062
CNN-LSTM	96.12	96.62	95.10	95.54	2.94	2.701
PSO-CNN	94.07	95.12	94.32	95.04	3.002	3.643
CNN-RNN	94.81	95.23	93.11	93.11	3.031	3.319
ACO-SVM	92.05	94.62	92.23	92.50	3.765	4.512
RF-SVM	91.12	94.03	93.91	92.04	4.061	4.821



**Fig. 5 Classification result of proposed and conventional**



**Fig. 6 Error validation result of proposed and conventional**

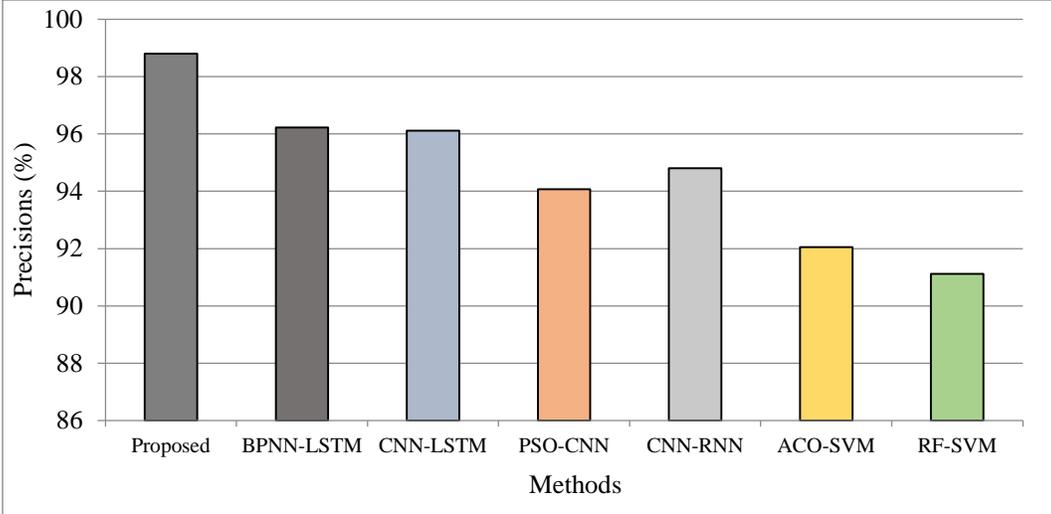


Fig. 7 Precision result of proposed and conventional

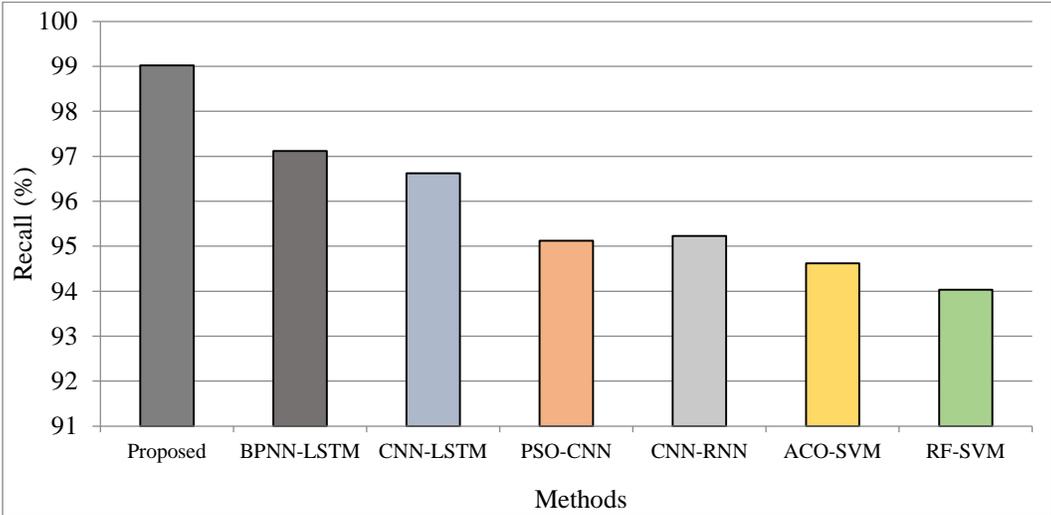


Fig. 8 Recall the result of the proposed and conventional

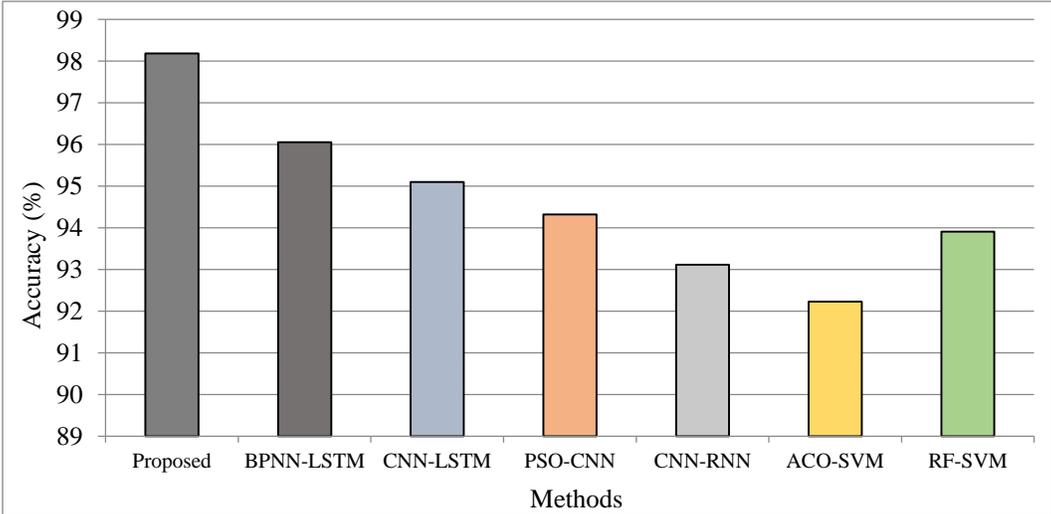


Fig. 9 Accuracy result of proposed and conventional

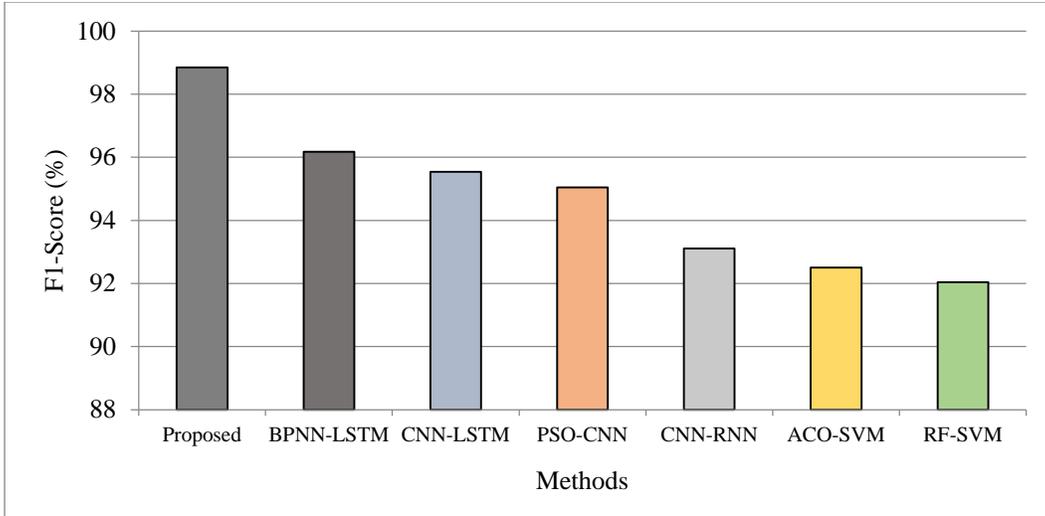


Fig. 10 F1 score result of proposed and conventional

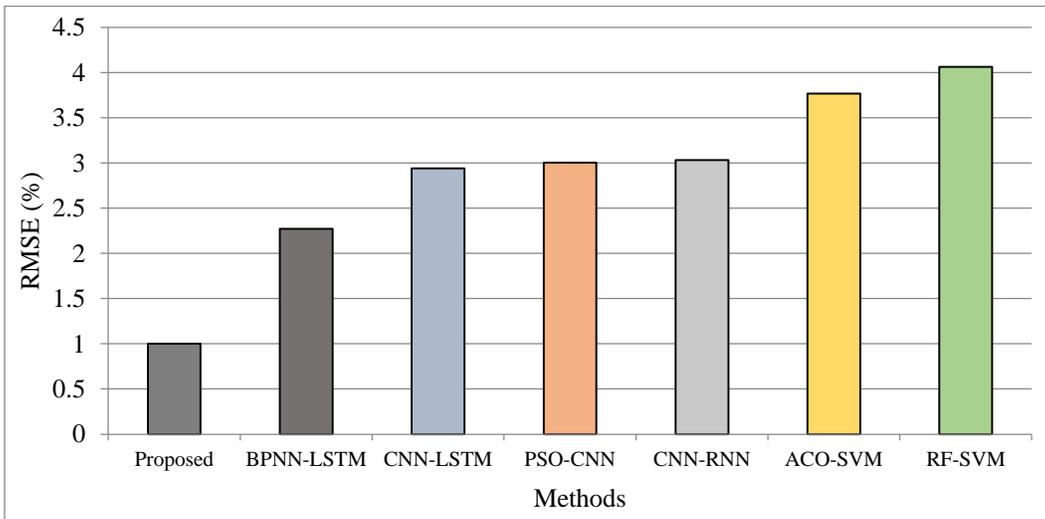


Fig. 11 RMSE result of proposed and conventional

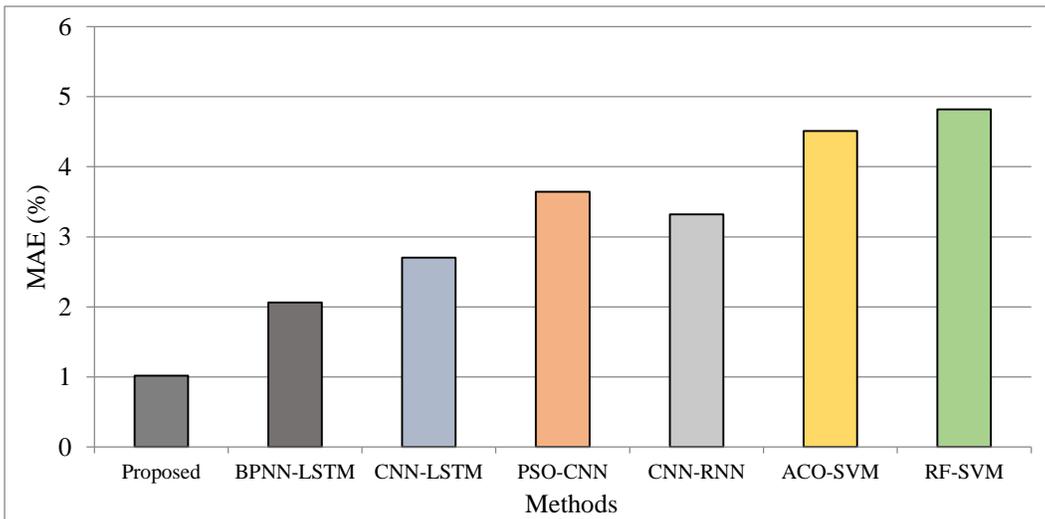


Fig. 12 MAE result of proposed and conventional

Figure 7 shows the Precision results of the proposed and conventional methods. The precision results of these techniques achieved by proposed BPNN-LSTM, CNN-LSTM, PSO-CNN, CNN-RNN, ACO-SVM and RF-SVM are 98.81%, 96.23%, 96.12%, 94.07%, 94.81%, 92.05% and 91.12%. From the result, the proposed result has attained a superior to the conventional methods.

Figure 8 shows the Recall Result of the proposed and conventional methods. Recall the result of these techniques achieved by the proposed BPNN-LSTM, CNN-LSTM, PSO-CNN, CNN-RNN, ACO-SVM and RF-SVM is 99.02%, 97.12%, 96.62%, 95.12%, 95.23%, 94.62% and 94.03%. From the result, the proposed Recall result has acquired a better performance than the conventional methods. Figure 9 shows the Accuracy Result of the proposed and conventional methods. The Accuracy result is attained as proposed (98.18%), BPNN-LSTM (96.05%), CNN-LSTM (95.10%), PSO-CNN (94.32%), CNN-RNN (93.11%), ACO-SVM (92.23%) and RF-SVM (93.91%) respectively. The proposed method has more accuracy than conventional methods.

Figure 10 shows the F1 score Result in the proposed and conventional methods. F1 score result of proposed BPNN-LSTM, CNN-LSTM, PSO-CNN, CNN-RNN, ACO-SVM and RF-SVM has attained 98.85%, 96.17%, 95.54%, 95.04%, 93.11%, 92.50% and 92.04% respectively. The proposed method has attained a higher F1 score than conventional methods.

Figure 11 shows the RMSE validation results for both the proposed and conventional methods. The RMSE result of the proposed BPNN-LSTM, CNN-LSTM, PSO-CNN, CNN-RNN, ACO-SVM and RF-SVM has achieved 1.001, 2.270,

2.94, 3.002, 3.031, 3.765 and 4.061 respectively. The proposed method has attained a minimum RMSE error than conventional methods.

Figure 12 shows the MAE validation results for the proposed and conventional methods. The RMSE result of the proposed BPNN-LSTM, CNN-LSTM, PSO-CNN, CNN-RNN, ACO-SVM and RF-SVM has achieved as 1.016, 2.062, 2.701, 3.643, 3.319, 4.512 and 4.821 correspondingly. The proposed method has attained a minimum MAE error than conventional methods.

## 5. Conclusion

This work presents different hybrid techniques for breast cancer prediction that have improved prediction accuracy. The proposed method has carried various tasks such as two feature extractions (A and B), feature Fusion, selection and classification. Each task is performed by several methods of feature extractions by the LDA model and CNN-LSTM model, the CCA model does fusion, selection is carried by the TSO model and classification is performed by the SVM model. The proposed work has been processed using the WDBC dataset for practical training and testing. The experimental results are validated and achieved based on the metrics like precision, Accuracy, Recall, and F1 Score as 98.81%, 99.02%, 98.18% and 98.85%, respectively. Also, the error validation is evaluated as RMSE and MAE as 1.001 and 1.016, respectively. Thus, the result of the proposed model has achieved a minimum error than the conventional methods. Therefore, the proposed method has attained an efficient prediction with a higher accuracy that can improve the survival rate of patients. The job of doctors is to identify cancer and process the remedy very early to cure the patient.

## References

- [1] Ahmedin Jema et al., "Cancer Statistics 2005," *CA: A Cancer Journal for Clinicians*, vol. 55, no. 1, pp. 10-30, 2005. [[Google Scholar](#)][[Publisher Link](#)]
- [2] Kemal Polat, and Salih Güneş, "Breast Cancer Diagnosis using Least Square Support Vector Machine," *Digital Signal Processing*, vol. 17, no. 4, pp. 694-701, 2007. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [3] Mehmet Fatih Akay, "Support Vector Machines Combined with Feature Selection for Breast Cancer Diagnosis," *Expert systems with Applications*, vol. 36, no. 2, pp. 3240-3247, 2009. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [4] Wei-Chang Yeh, Wei-Wen Chang, and Yuk Ying 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. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [5] A. Marciano-Cedeño, J. Quintanilla-Domínguez, and D. Andina, "WBCD Breast Cancer Database Classification Applying Artificial Meta Plasticity Neural Network," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9573-9579, 2011. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [6] Yılmaz Kaya, and Murat Uyar, "A Hybrid Decision Support System Based on Rough Set and Extreme Learning Machine for Diagnosis of Hepatitis Disease," *Applied Soft Computing*, vol. 13, no. 8, pp. 3429-3438. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [7] Kindie Biredagn Nahato, Khanna Nehemiah Harichandran, and Kannan Arputharaj, "Knowledge Mining from Clinical Datasets using Rough Sets And Back Propagation Neural Network," *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1-13, 2015. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]

- [8] Lijuan Liu, and Mingrong Deng, "An Evolutionary Artificial Neural Network Approach for Breast Cancer Diagnosis," *In Third International Conference on Knowledge Discovery and Data Mining*, pp. 593-596, 2010. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [9] Hui-Ling Chen et al., "A Support Vector Machine Classifier with Rough Set-Based Feature Selection for Breast Cancer Diagnosis," *Expert Systems with Applications*, vol. 38, no. 7, pp. 9014-9022, 2011. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [10] Noreen Fatima et al., "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," *IEEE Access*, vol. 8, pp. 150360-150376, 2020. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [11] Min-Wei Huang et al., "SVM and SVM Ensembles in Breast Cancer Prediction," *PloS One*, vol. 12, no. 1, pp. e0161501, 2017. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [12] Yixuan Li, and Zixuan Chen, "Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction," *Applied and Computational Mathematics*, vol. 7, no. 4, pp. 212-216, 2018. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [13] K. Sivakami, "Mining Big Data: Breast Cancer Prediction using DT-SVM Hybrid Model," *International Journal of Scientific Engineering and Applied Science*, vol. 1, no. 5, pp. 418-429, 2015. [[Google Scholar](#)][[Publisher Link](#)]
- [14] Jayesh George Melekoodappattu, and Perumal Sankar Subbian, "Automated Breast Cancer Detection using Hybrid Extreme Learning Machine Classifier," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 5489-5498, 2023. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [15] Ilangovan Sangaiah, and A. Vincent Antony Kumar, "Improving Medical Diagnosis Performance using Hybrid Feature Selection via Relief and Entropy-Based Genetic Search (RF-EGA) Approach: Application to Breast Cancer Prediction," *Cluster Computing*, vol. 22, no. 3, pp. 6899-6906, 2019. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [16] Tze Sheng Lim et al., "Breast Cancer Diagnosis System using Hybrid Support Vector Machine-Artificial Neural Network," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 3059-3069, 2021. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [17] Fang Chen et al., "Immunohisto Chemical Index Prediction of Breast Cancer by using Hybrid Ultrasound Data," *In IEEE Transactions on Biomedical Engineering*, vol. 70, no. 4, pp. 1401-1412, 2023. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [18] R. Senkamalavalli, and T. Bhuvaneshwari, "Improved Classification of Breast Cancer Data using Hybrid Techniques," *International Journal of Advanced Engineering Research and Science*, vol. 5, no. 5, pp. 77-81, 2017. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [19] Shobha Gusain, "Modifiable Risk Factors for Breast and Cervical Cancer and their Association with Sample Characteristics," *SSRG International Journal of Nursing and Health Science*, vol. 6, no. 1, pp. 10-16, 2020. [[CrossRef](#)][[Publisher Link](#)]
- [20] Shankar Thawkar et al., "Breast Cancer Prediction using A Hybrid Method Based on Butterfly Optimization Algorithm and Ant Lion Optimizer," *Computers in Biology and Medicine*, vol. 139, p. 104968, 2021. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [21] Jayesh George Melekoodappattu, Perumal Sankar Subbian, and M. P. Flower Queen, "Detection and Classification of Breast Cancer from Digital Mammograms using Hybrid Extreme Learning Machine Classifier," *International Journal of Imaging Systems and Technology*, vol. 31, no. 2, pp. 909-920, 2021. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [22] Jong Pill Choi, Tae Hwa Han, and Rae Woong Park, "A Hybrid Bayesian Network Model for Predicting Breast Cancer Prognosis," *Journal of Korean Society of Medical Informatics*, vol. 15, no. 1, pp. 49-57, 2009. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [23] Mohammad R. Mohebian et al., "A Hybrid Computer-Aided-Diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) using Optimized Ensemble Learning," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 75-85, 2017. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [24] Youness Aliyari Ghassabeh, Frank Rudzicz, and Hamid Abrishami Moghaddam, "Fast Incremental LDA Feature Extraction," *Pattern Recognition*, vol. 48, no. 6, pp. 1999-2012, 2015. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [25] Almas Begum et al., "A Combined Deep CNN: LSTM with a Random Forest Approach for Breast Cancer Diagnosis," *Complexity*, vol. 2022, no. 1-9, 2022. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [26] Mariam Shadan et al., "Histological Categorization of Stromal Desmoplasia in Breast Cancer and Its Diagnostic and Prognostic Utility," *SSRG International Journal of Medical Science*, vol. 4, no. 6, pp. 8-11, 2017. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [27] B. Johnson, P. Keerthi Vasan, and V. Thillaivendan, "Performance of Hyperthermia for Breast Cancer," *SSRG International Journal of Applied Physics*, vol. 3, no. 2, pp. 1-5, 2016. [[CrossRef](#)][[Publisher Link](#)]
- [28] Masoomeh Mirrashid, and Hosein Naderpour, "Transit Search: An Optimization Algorithm Based on Exoplanet Exploration," *Results in Control and Optimization*, vol. 7, 2022. [[CrossRef](#)][[Google Scholar](#)][[Publisher Link](#)]
- [29] G. Ravi Kumar, G. A. Ramachandra, and K. Nagamani, "An Efficient Prediction of Breast Cancer Data using Data Mining Techniques," *International Journal of Innovations in Engineering and Technology*, vol. 2, no. 4, pp. 139-144, 2013. [[Google Scholar](#)][[Publisher Link](#)]
- [30] Brotobor Deliverance, Edeawe Isaac Osahogie, and Brotobor Onoriode, "Awareness and Practice of Student Nurses on Breast Self-Examination: A Risk Assessment Tool for Breast Cancer," *SSRG International Journal of Nursing and Health Science*, vol. 6, no. 1, pp. 66-69, 2020. [[CrossRef](#)][[Publisher Link](#)]

- [31] Nidhi Mongoriya, and Vinod Patel, "Review the Breast Cancer Detection Technique using Hybrid Machine Learning," *SSRG International Journal of Computer Science and Engineering*, vol. 8, no. 6, pp. 5-8, 2021. [[CrossRef](#)][[Publisher Link](#)]
- [32] M. Hossin, and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *International Journal of Data Mining and Knowledge Management Process*, vol. 5, no. 2, pp. 1-11, 2015. [[Google Scholar](#)][[Publisher Link](#)]