International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Hybrid Deep Learning and Optimization Algorithm for Breast Cancer Prediction Using Data Mining

G. Rajasekaran[1], and P. Shanmugapriya[2]

Submitted: 17/10/2022 **Revised**: 22/12/2022 **Accepted**: 02/01/2023

Abstract: Breast Cancer is the uncontrollable growth of cells by abnormal activities of genes. 2 in 10 women in world will be identified with breast cancer in her lifetime. On average, every 5 minutes a woman is identified with breast cancer in the world. So, there is a huge need for intelligent early prediction methods to support a health care peoples for increasing the survival rate of the patients. Recently, data mining approach of Deep Learning (DL) and machine learning (ML) contributes beneficial role in medical field for detection and classifications of diseases. The accuracy of prediction is reduced due to the imbalanced nature of data with unequal distribution of the positive and negative classes. To overcome this issue, the breast cancer prediction is presented by using a Hybrid algorithm such as Linear Discriminant Analysis (LDA), Wild Horse Optimization (WHO) and Advanced Elman Recurrent Neural Network (AERNN) methods in this work. A LDA model is used to remove a features, WHO model is used for feature reduction and tuning a AERNN's hyper parameters and Optimized AERNN model for classifications. The proposed method has outperformed by achieving a result of Precision (98.51%), Recall (98.65%), Accuracy (97.88%) and F1 score (98.32%) and also in the performances of error evaluation of RMSE (1.006) and MAE (1.986) than the prior methods respectively.

Keywords: *Breast cancer prediction, Mining, LDA feature extraction, WHO model, hyper parameters fine-tuning, AERNN model*

1. Introduction

The second most frequent life threatening condition in women is breast cancer. Breast cancer is a un controllable growth of the cells in breast which can spread to other parts of women through blood vessels [1]. The major sign of breast cancer is changes in breast structure and fluid from nipples etc. According to a report, 2.3 million women will be diagnosed with breast cancer worldwide in 2021 and 685,000 people will die from the disease. Early detection of breast cancer increases survival probabilities of 80% or higher. The treatment process includes surgery or radiation therapy on affected areas [2].

For future plan and treatment, the effective prognostic prediction technique is needed by the health care professionals. The well-known aiding approaches are Sprouting angiogenesis, BRCA1, HER-2 and gene expression [3]. Due the development of cloud computing and Internet of Things (IoT) in medical field, the possibility of storing medical records of the patients increases heavily [4]. These data collections, which might be in raw data or image format, provide support for data mining algorithms that are used to make decisions about health.

Data mining is a collection of heuristics and calculations that create models out of unstructured data. The constructed model is used to analyze a medical by extracting important hidden patterns [5]. Also, it increases diagnostic accuracy, save treatment expenses and increase human's lifetime. For example, the C4.5 algorithm, the k-means nearest neighbour (KNN) algorithm, the Support Vector Machine (SVM), the Expectation-Maximization algorithm, the AdaBoost algorithm, algorithm, the Convolutional neural network (CNN), the Back Propagation neural network (BPNN) and the Long Short-Term Memory (LSTM) technique are all used to help breast cancer patients [6].

In this work, the breast cancer prediction is presented by using a Hybrid algorithm such as LDA method, WHO method and AERNN method. Feature extraction is performed by LDA and a significant feature reduction is processed by WHO model. For classification, the hyper parameter of AERNN is fine-tuned by WHO and processed to obtain a breast cancer prediction. The result and discussion are presented for proposed and prior method based on metrics Accuracy, Recall, F1 score, Precision, root mean squared error (RMSE) and mean absolute error (MAE), correspondingly.

The other portions of this work are structured as follows: Section 2 discusses similar works, while Section 3 describes the procedures and materials. The section 4 presented an experimental result of proposed techniques and the section 5 explored a conclusion that followed by references.

2. Related Works

Zhang et al [7] suggested a breast cancer risk prediction model using C5.0Algorithm. Factors of age, habits, first childbearing and Hereditary disease are considered to predict the probability

[1] *Research Scholar, Department of Computer Science Engineering, SCSVMV University, Kancheepuram, Tamilnadu, India, gtrajasekaran@gmail.com*
[2]*Associate Professor, Department of Computer Science Engineering, SCSVMV University, Kancheepuram, Tamilnadu, India, pshanmugapriya@kanchiuniv.ac.in*
**Corresponding Author*

of getting breast cancers. Compared to other ML models, the c5.0 reaches higher accuracy of 97.772%.

To reduce a number of features to train the ML models, a two feature selection techniques of statistical and ensemble feature selection algorithms proposed by Fu, Bo et al [8].Number of features reduced from 52 to 23.Then,XBOOST ML model is used for the prediction of survival score for early breast cancer. In another study, Egwom Onyinyechi Jessica et al [9] applied a Linear Discriminant Analysis for feature reduction. After feature reduction, support Vector Machine (SVM) classifier is used for prediction. Results on Wisconsin Diagnostic Breast Cancer (WDBC) dataset shows that the combined LDA and SVM reach the maximum accuracy of 94.56%.

Polat et al [10] advanced a three-stage prediction model for breast cancer prediction. Initially, median absolute deviation is applied for data preprocessing. Then, k-means clustering is performed for feature selection. Finally, the prediction is achieved using daBoostM1 classifier. A result on data set collected from the city of Coimbra of Portugal shows the three-stage model achieves 92.48% of accuracy which is higher than other previously proposed models.

Similarly, Pawlovsky et al [11] suggested a breast cancer severity prediction using k nearest neighbor algorithm. An efficiency of algorithm is analyzed for the different value of k. the correct value of k achieves better accuracy in prediction.

Liu et al [12] proposed a hybrid survival prediction model using Cox Proportional Hazard Model and XBOOST classifier. The prediction results include low risk, medium risk and high risk. Concordance index and time-dependent Area under Curve were used to compare the suggested approach to other survival prediction models.

Genomic data-based survival prediction algorithm is proposed by Karim, Md et al [13]. Which is the DNA data of organisms like gene methylation and miRNA etc. The problem of over fitting due to limited genomics data is handled by the use of neural network-based models. Additionally, hyper parameter adjustment is made to boost the effectiveness of the prediction model.

A hybrid DL network with SVM classifier is proposed by Sun, Dongdong et al [14]. The hybrid model's final layer employs SVM as a classifiers and the multi-layer feed forwards networks as a feature extractor. The proposed model excellently handles the non-linearity of the data set and extracts features effectively.

Liang et al [15] suggested breast cancer diagnosis system using clustering and fuzzy intelligence. Initially, the data is processed by subtractive clustering. Then, fuzzy membership rules are applied to diagnose the breast cancer.

Sun, Det al [16] explored a multi model DL model for breast cancer prediction. The features extracted from principal component analysis model is used train the DL model for the accurate prediction.

Further, AlGhunaim, S et al [17] analyzed performance three different machine learning algorithm of random forest, SVM and decision tree for breast cancer prediction. Among these, the SVM based prediction shows higher accuracy.

Another method presented by Nurhayati et al [18] proposed a feature selection approach based on Particle Swarm Optimization (PSO) for the diagnosis of breast cancer. Naive Bayes classifiers is employed for the prediction of breast cancer following feature selection.

Next, Osman et al [19] developed a Radial Based Function based neural model for an analysis of breast cancer. To improve its performance further, the Ensemble Boosting Learning is combined.

To solve the low intensity ratio issues on genetic signature-based classification, the hybrid LDA and Auto Encoder Neural Network based hybrid model is suggested by Zhang et al [20]. The purposes of using auto encoder is to extract more relevant features from the data set.

Next, the graph-based prediction model is suggested by Thin Nguyen et al [21] for-breast cancer prediction. Properties of graph are considered as feature for the classification.

Recently, Generative Adversarial Network (GAN) based networks used for data augmentation process in medical images and data's to generate multiple images and data's. GAN based data augmentation is proposed by Hsu, Te-Cheng et al [22] for breast cancer prediction. The integration GAN increase the accuracy of prediction model.

Liu, et al [23] proposed a new feature selection method for survival prediction. The modified infinite feature selection is suggested for feature selection. Also, the random under sampling is applied for data balancing.

Arya et al [24] constructed a stacked ensemble model for breast cancer prognosis prediction. It includes two stage convolutional network while stage one feature extraction and another one for feature selection. Increasing number of hidden layers in the models increases hidden feature extraction capability.

A hybrid feature extraction and selection method are proposed by Raweh et al [25]. The important features are extracted using F-score Feature Selection Method. Using the Kernel Density Estimation Method, the features are chosen. The chosen feature was then categorized by a random forest classifier.

Waseem et al [26] suggested a Reject Option based ML prediction algorithm for cancer prediction. It rejects particular area based on certain thresholds to reduce an error rate. The threshold levels are varied to vary the accuracy level of the prediction.

Likewise, in another work of Zhang et al [27] an auto encoder combined principal component analysis is proposed for feature selection in breast cancer prediction. For classification, the adaboost algorithms is applied. Compared to other algorithms, the proposed algorithm reaches AUC of over 0.824.

3. Materials and Methods

It discusses the sources and methodology used in this paper. The Figure 1 showed the proposed block diagram that has a dataset collection, LDA feature extraction, WHO feature reduction, optimized AERNN classification and performance evaluation that is explained in the following.

3.1. Dataset Description

Breast cancer prediction based on benign or malignant is only identified by normal and affected dataset. The dataset is gathered in a breast cancer UCI repository which is in several categories namely Breast Cancer Diagnosis (BCD), Wisconsin Data set for Diagnostic Breast Cancer (WDBC), Breast Cancer

Original (WBC) and Breast Cancer Prognostic (BCP) respectively. In this work, the WDBC dataset is used that is gathered from the Wisconsin Hospitals university in 1995 [28]. In this dataset, there are 569 samples from 32 patients that has a benign data as 62.74% and malignant data as 37.26%. It consists of 32 attributes such as ID number, Benign and Malignant result and remaining 30 as cancer diagnosis data. All these attributes are collected in the basis of 10 features such as perimeter, Radius, Texture, smoothness, Area, compactness, concavity, concave points, facial dimension and symmetry respectively. Each features have three evaluation values namely mean, standard error and worst value. Meanwhile, there are 30 feature attributes are considered for both training and testing phases.

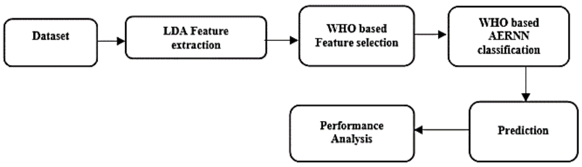

Fig. 1. Proposed Block Diagram

3.2 LDA Based Feature Extraction

After the pre-processing, the feature extractions are carried out to access the relevant feature data to gather an information from it. In this work, the feature extraction is performed by using a LDA methodology [29]. The LDA model is used for identifying a linear feature combination to classify two or more classes. The LDA has been identified the maximum linear classes separating directions that is given in Figure 2. The finding a direction is provided an optimal result in data discrimination.

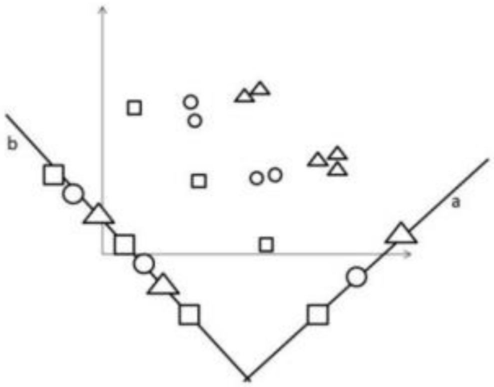

Fig. 2. Direction of Linear class separation

Consider high separation coefficient (F) to evaluate the direction is expressed in the following equation (1):

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

Where S_m denotes between class scatter, S_w indicates a within class scatter. Maximum value of F would be higher probability of classes' separation.

The S_c and$S_w$is evaluated for the class c is expressed in equation (2-3):

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

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

Where μ_c denotes the mean value, n_i indicates number of x_i in every class and N denotes a total number of classes.

Class scatter (S_B^C)for class c is expressed as:

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

Where μ_i denotes the mean of x_i for class i, μ indicates mean of every x_i observations for all classes.

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

By providing a maximum Eigen value for matrix (S), the eigenvectors provided a direction of best class separation that is expressed as:

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

The Eigen vector evaluation is difficult due to the non-symmetric S-matrix. To solve this issue, the generalized Eigen value is used. The transformed data set is given by

$$y = x^TW \tag{7}$$

Where W is matrix of w1,w2,..,wM with the M eigenvectors and highest Eigen values.

LDA is used to minimize the real feature space dimension to M. The dataset y is generated a linear combination of every feature input (x) with weight (W). It has 184 features and only two LDA are fetched for investigation for data transmission given in Figure 3.

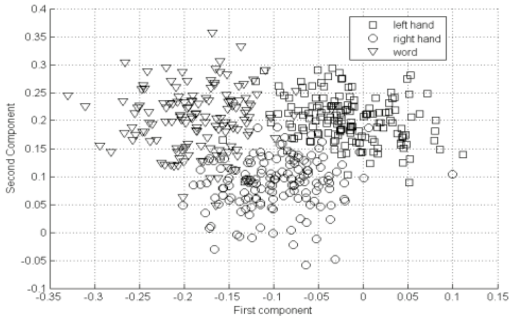

Fig. 3. Data after LDA transform

Though the LDA has some reduction in feature extraction, there are huge number of features are used. Only a few features have information that can be used to differentiate classes. A new, improved feature set with lower dimensions was discovered to be possible by the linear combination of characteristics. The authors employed the WHO to feature reductions in the following stage.

3.3. WHO Based Feature Selection

Optimization is process of identifying the best solution among search space. There are two types of optimization algorithm available in literature: conventional and meat heuristic. Conventional algorithm achieves optimal solutions. But it takes hundreds of years to solve the certain problems and failed to achieve a best solution for large scale problems. Meta heuristic algorithm guarantees best solution for large scale problems with minimum iterations. It is based on enhancing prior algorithms or was motivated by animal behavior and physics rules. The WHO method is one of the meta heuristic optimization algorithms inspired by social behavior of horse [30]. Social behavior of horses like gazing, mating and leadership quality mathematically modelled to solve the optimization problems. This work uses WHO to select best features and tune the parameters of classifier for improve performance.

Then-size population initialization is evaluated using this Equation (8):

$$\vec{X} = \{\vec{x}_1, \vec{x}_2, \dots., \vec{x}_n\} \tag{8}$$

The wild horse grazing character by fixing stallion as grazing location centre and the other are members in a ground location. The location of members is expressed in Equation (9):

$$\vec{X}_{i,g}^{j} = 2Zcos(2\pi RZ)\left(S^{j} - X_{i,g}^{j}\right) + S^{j} \quad (9)$$

Where R denotes a random number in range of [−2, 2],$X_{i,g}^{j}$denotes mares present location in j-group, S^{j}indicates a stallion in j-group.

The adaptive parameter (Z) can be expressed in equation (10):

$$Z = R_2 \ominus IDX + \vec{R}_3 \ominus (\sim IDX), IDX = (P == 0); \quad (10)$$

$$P = \vec{R}_1 < TDR$$

Where P denotes a vector 0 and 1,R_2rrepresents the random vector in [0, 1],$\vec{R}_1$and$\vec{R}_3$ denotes a random vector in [0, 1],IDX indicates $\vec{R}_1$indices for (P==0) and TDR represents an adaptive parameter that is expressed as:

$$TDR = 1 - t/t_{maximum} \quad (11)$$

Where ‘t’ denotes a present iteration and$t_{maximum}$ indicates a maximum iteration number.

The features of intermarriage behavior of young horse by assuming the i-male and j-female but it can mate after adultery. The np-horse of location in kth group formatting and leaving behavior is expressed in equation:

$$X_{g,k}^{p} = Mean\left(X_{g,i}^{q}, X_{g,i}^{Z}\right), i \neq j \neq k, p = q = end \quad (12)$$

Where $X_{g,i}^{q}$ denotes the q-th location for mating in ith group and $X_{g,i}^{Z}$indicates the z-th location for mating in ith group.

The location updating in i-th group leader (S_{gi}) is expressed as equation (13):

$$S_{gi} = \begin{cases} 2Zcos(2\pi RZ)\left(X_{wh} - S_{gi}\right) + X_{wh}R_3 > 0.5 \\ 2Zcos(2\pi RZ)\left(X_{wh} - S_{gi}\right) + X_{wh}R_3 \leq 0.5 \end{cases} \quad (13)$$

Where X_{wh} represents location of water hole and S_{gi}denotes the leader’s location.

When the fitness $f\left(X_{g,i}\right)$ is greater than $f(S_{g,i})$, then the $S_{gi}$is updated using the Equation (14):

$$S_{gi} = \begin{cases} X_{g,i}\; f\left(X_{g,i}\right) < f(S_{g,i}) \\ S_{g,i}\; f\left(X_{g,i}\right) > f(S_{g,i}) \end{cases} \quad (14)$$

Consequently, the algorithm 1 provides a pseudo code for the WHO approach.

Algorithm1: Pseudo code of WHO

```
Input: n, t_maximum, PC and PS and number of runs (N).
Set upper bound and lower bound limits
Fix t = 1 and run = 1.
while run is greater than N
do
while t greater than t_maximum
do
Initialize using equation (8)
Evaluate fitness function
For i = 1: PS
For j = 1: N for young horse
When randomness lesser than PS
Estimate $X_{g,k}^{p}$using Equation (9).
else
Evaluate $X_{g,k}^{p}$using Equation (12)
end if
Evaluate the corresponding fitness values
if j greater than number of young horse
j = j+1
else
Estimate $S_{gi}$using Equation (13)
end if
end for
choose leader using Equation (14).
When i greater than N_stallion, then i = i+1
else
t = t+1
end if
end for
end while
run = run+1
end while
Return the parameters
```

The feature minimization is to be done by selecting significant features are done by WHO model. This WHO method is presented to provide an optimal result in feature selection.

3.4. Optimized AERNN Model for Classification

The selected features are performed a classification method to provide an accurate prediction and performance in breast cancer. In this work, the WHO based AERNN is used to perform a classification. Initially the hyper parameter of AERNN such as weight and biases are fine-tuned using the WHO method and then the AERNN is implemented for breast cancer prediction.

3.4.1. AERNN Methodology

The AERNN is a unique learning methodology that is based on the BPNN model for a maximum distance data. The Figure 4 shows the structure of AERNN that has three main layers namely input, output, hidden and recurrent layer. There are activation function, output and inputs are presented in every neuron in AERNN. The dataset is collected in input layer and moved to hidden layer. The data of previous terms are stored in hidden layer and saved in recurrent layer. The output layer is used to display the corresponding result for an input data [31].

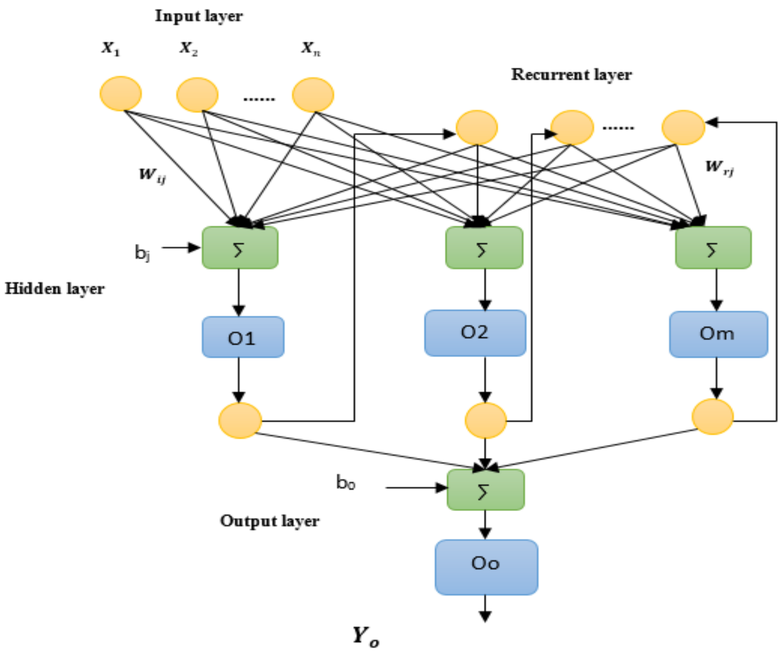

Fig. 4. Structure of AERNN method

Assume Number of recurrent neurons as r=1,2…..m number of inputs as i=1,2….n, network's weight as Wij, W_{rj} and W_{jo} , and Number of hidden neurons as j=1,2…..m respectively.

The hidden layer output ($H_o(\boldsymbol{t})$)at t is given in Equation (15-16).

$$H_{oj}(t) = \sum_{i=1}^{n}\sum_{j=1}^{m}(W_{ij} * X_i(t)) + \sum_{r=1}^{m}\sum_{j=1}^{m}(W_{rj} * O_j(t-1)) + b_j \quad (15)$$

$$Y_{oj}\ (t) = T(H_{oj}(t)) \quad (16)$$

Where i denotes a number of inputs i.e., i=1,2….n, j denotes a Number of hidden neurons i.e., j=1,2…..m, r represents the recurrent neurons i.e., r=1,2…..m, W indicates weight of networks i.e, Wij, W_{rj} and W_{jo}, b indicates the bias and represents hyperbolic tangent function.

An output layer is given in the Equation (17-18).

$$H_o\ (t) = \sum_{i=1}^{n} \mathrm{Wjo} * Y_{oj}\ (t) + b_o \quad (17)$$

$$Y_o\ (t) = P(O_o\ (t)) \quad (18)$$

Where Pindicates the purelin function

The AERNN has disadvantages of minimum speed of convergence and low generalization performance. The AERNN hyper parameters of weight and bias are finetunedusing WHO method to avoid these issues. Therefore, the WHO based AERNN has executed a greater accuracy in prediction and performances than prior models.

4. Experimental Results

In this section, the result and discussion are presented for proposed and prior method based on metrics Accuracy, Precision, root mean squared error (RMSE),Recall, F1 score and mean absolute error (MAE)correspondingly. In this work, the WDBC dataset is processed to train the data that contains 60% of total data and for test data contains 40% of total data for prediction. In this section, comparison between the suggested method and prior methods such as Extreme Learning Machine (ELM), AdaBoost, BPNN, LSTM, AutoEncoder, CNN, Random Forest-Decision Tree (RFDT) are presented. The performance metrics are evaluated using the following equations.

$$Precision = \frac{True^{+}}{True^{+}+False^{+}} \quad (19)$$

$$Recall = \frac{True^{+}}{True^{+}+False^{-}} \quad (20)$$

$$Accuracy = \frac{True^{+}+True^{-}}{True^{+}+False^{+}+True^{-}+False^{-}} \quad (21)$$

$$\text{F1 value} = \frac{Recall \times 2Precision}{precision+Recall} \quad (22)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{a=1}^{n}(\text{determined cases} - \text{total cases})^2} \quad (23)$$

$$MAE = \sqrt{\frac{1}{n}\sum_{a=1}^{n}|\text{determined cases} - \text{total cases}|^2} \quad (24)$$

Where $True^{-}$indicates the Truly negative,$True^{+}$ denotes the truly positive,$False^{+}$ represents the Falsely positive, $False^{-}$represents the Falsely negative, n represents the entire patients respectively.

Table 1. Performance metrics result of proposed and prior methods

Techniques	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**F1score (%)**	**RMSE**	**MAE**
Proposed	98.51	98.65	97.88	98.32	1.006	1.986
ELM	97.72	96.21	95.50	94.87	3.129	3.006
AdaBoost	96.62	95.26	95.50	93.45	3.304	3.270
BPNN	93.40	94.58	95.03	92.15	3.982	4.761

LSTM	92.48	94.12	94.32	91.21	4.043	4.860
Auto Encoder	91.90	93.66	94.12	91.05	4.675	5.352
CNN	92.21	93.23	93.79	90.54	5.006	6.005
RFDT	90.77	92.20	93.21	89.21	5.569	6.842

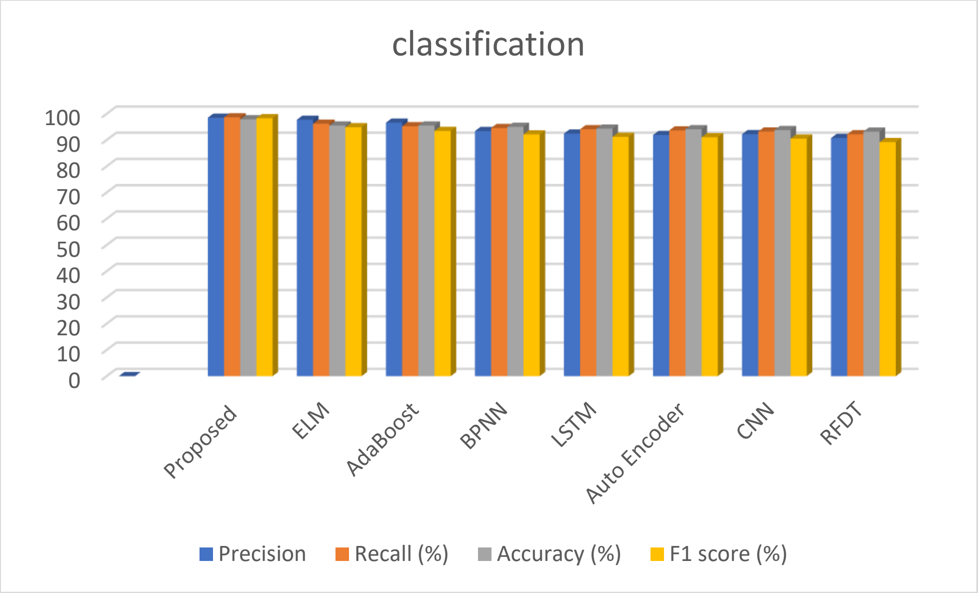

Fig. 5. Comparison table of proposed and prior model classification result

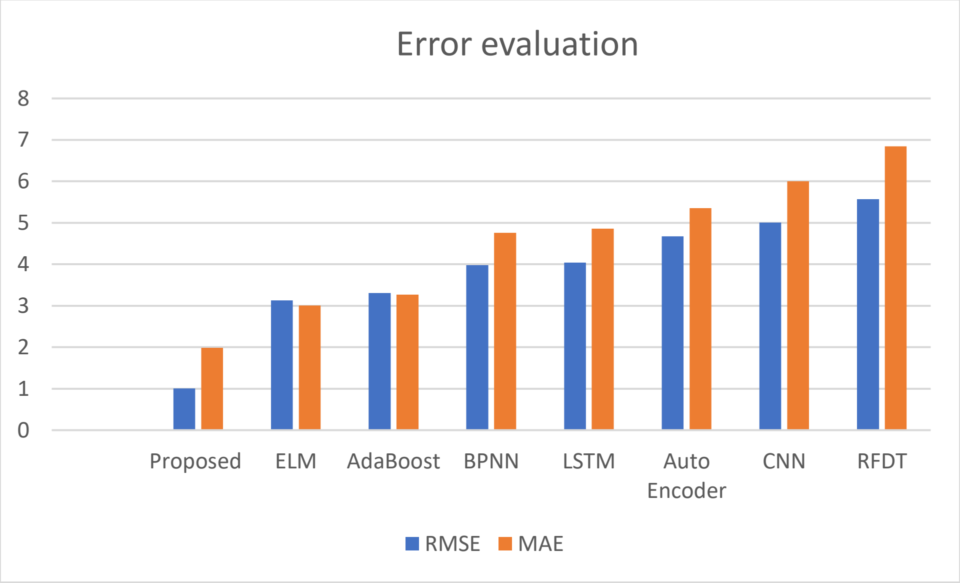

Fig. 6. Comparison of Error evaluation result for proposed and prior techniques

The overall performance metrics for a proposed and prior methods are tabulated in Table 1 and Figured in Figure 5 and 6. The Figure 5 represents the classification results and Figure 6 presented the error evaluation results. The proposed method has outperformed than the prior methods by achieving a result of Precision (98.51%), Recall (98.65%), Accuracy (97.88%) and F1 score (98.32%) and also in the performances of error evaluation of RMSE (1.006) and MAE (1.986) respectively.

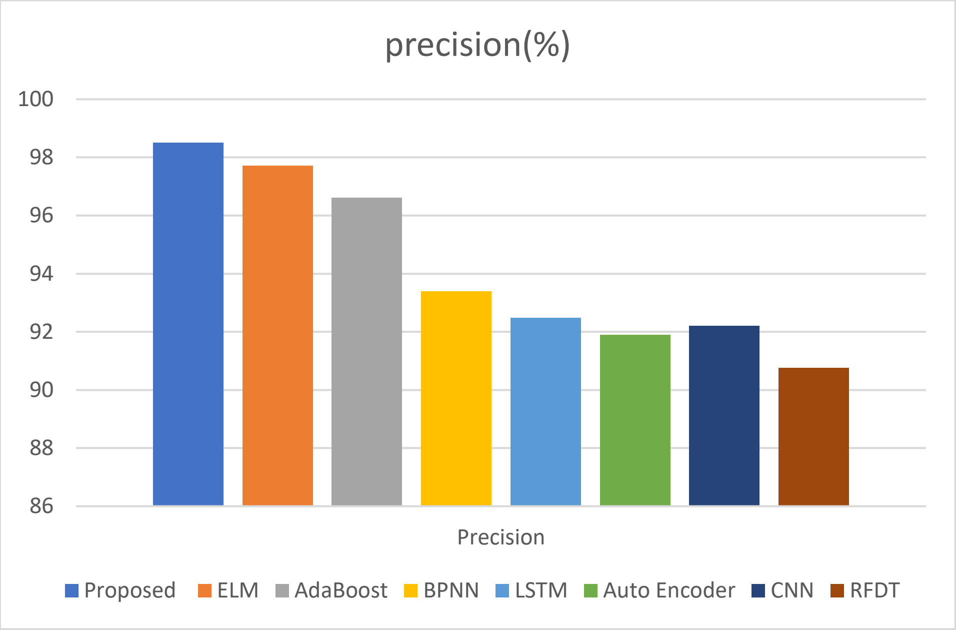

Fig. 7. Comparison chart of precision result

The comparison chart of precision result for proposed and prior is shown in Figure 7. A precision value of proposed (98.51%), ELM (97.72%), AdaBoost (96.62%), BPNN (93.40%), LSTM (92.48%), Auto Encoder (91.90%), CNN (92.21%) and RFDT (90.77%) are achieved whereas the proposed method has superior performance than all priors respectively.

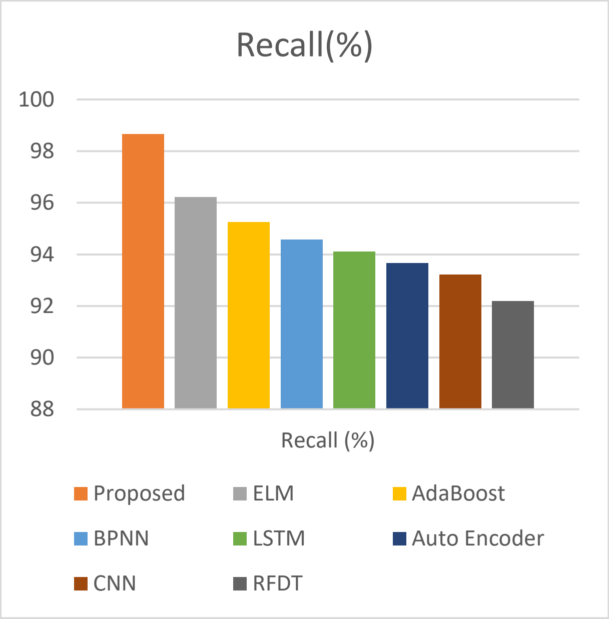

Fig. 8. Comparison chart of Recall result

The comparison chart of Recall result for proposed and prior is shown in Figure 8. The Recall value of proposed (98.65%), ELM (96.21%), AdaBoost (95.26%), BPNN (94.58%), LSTM (94.12%), Auto Encoder (93.66%), CNN (93.23%) and RFDT (92.2%) are achieved whereas the proposed method has greater recall performance than all priors correspondingly.

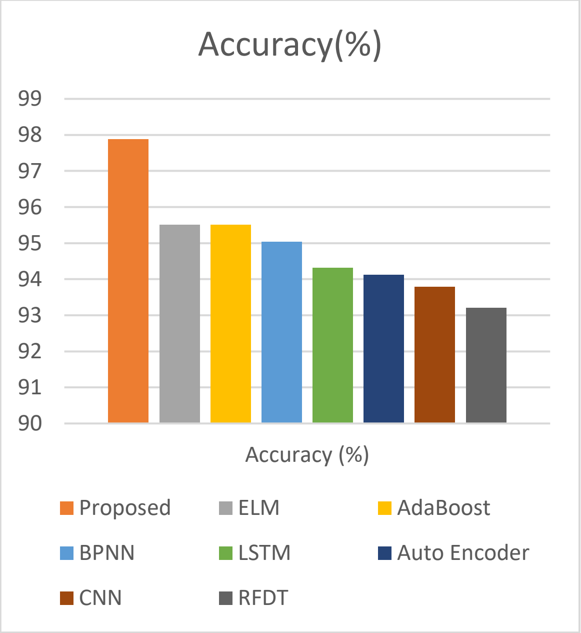

Fig. 9. Comparison chart of Accuracy result

The comparison chart of accuracy result for proposed and prior is shown in Figure 9. A precision value of proposed (97.88%), ELM (95.50%), AdaBoost (95.50%), BPNN (95.03%), LSTM (94.32%), Auto Encoder (94.12%), CNN (93.79%) and RFDT (93.21%) are achieved whereas the proposed method has maximum accuracy than all prior methods.

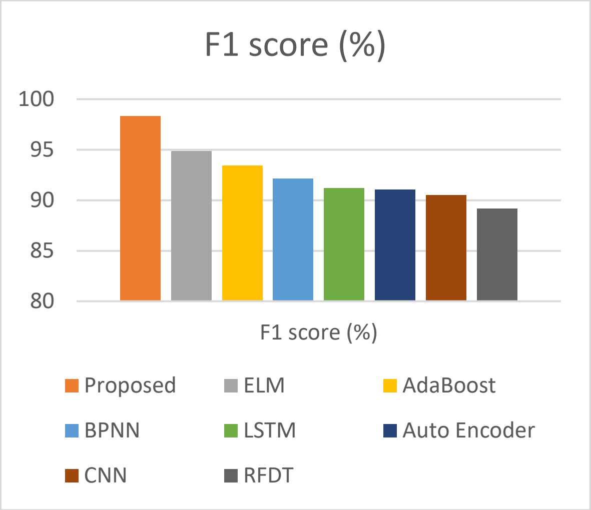

Fig. 10. comparison chart of F1 value

The comparison chart of F1 value result for proposed and prior is shown in Figure 10. F1 score value of proposed (98.32%), ELM (94.87%), AdaBoost (93.45%), BPNN (92.15%), LSTM (91.21%), Auto Encoder (91.05%), CNN (90.54%) and RFDT (89.21%) are achieved whereas the proposed method has maximum performance than all prior methods.

5. Conclusion

Early detection of breast cancer increases survival probabilities of 80% or higher. The data mining of WDBC dataset is processed for training and testing as 60% and 40% respectively. In this work, the LDA based feature extraction is done and the WHO model is used for reducing the features and allowed a significant feature. This process made a classification task easier with a minimum feature and the optimized AERNN model is performed for classification. The AERNN's weight and bias are fine-tuned by WHO model to provide an efficient and optimal result in breast cancer prediction. The performance metrics of proposed method has attained a Precision (98.51%), Recall (98.65%), Accuracy (97.88%) and F1 score (98.32%) and also in the performances of error evaluation of RMSE (1.006) and MAE (1.986) respectively. The outcome shown that the proposed hybrid optimized strategy improved an earlier approaches.

References

[1] A. H. Osman, "An Enhanced Breast Cancer Diagnosis Scheme based on Two-Step-SVM Technique," Int. J. Adv. Comput. Sci. Appl., 2017.

[2] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," Front. Genet., 2018, DOI: 10.3389/fgene.2018.00515.

[3] Nurhayati and A. N. Rahman, "Implementation of Naive Bayes and K-Nearest Neighbor Algorithm for Diagnosis of Diabetes Mellitus," Proc. 13th Int. Conf. Appl. Comput. Appl. Comput. Sci. (ACACOS '14), pp. 117–120, 2014.

[4] B. Alić, L. Gurbeta, and A. Badnjević, "Machine learning techniques for classification of diabetes and cardiovascular diseases," 2017, DOI: 10.1109/MECO.2017.7977152.

[5] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," World J. Eng. Technol., 2018, DOI: 10.4236/wjet.2018.64057.

[6] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Computational and Structural Biotechnology Journal. 2015, DOI: 10.1016/j.csbj.2014.11.005.

[7] Zhang X, Sun Y. Breast cancer risk prediction model based on C5. 0 algorithm for postmenopausal women. In2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC) 2018 Dec 14, pp. 321-325. IEEE.

[8] Fu, B., Liu, P., Lin, J., Deng, L., Hu, K. and Zheng, H., 2018. Predicting invasive disease-free survival for early stage breast cancer patients using follow-up clinical data. *IEEE Transactions on Biomedical Engineering*, 66(7), pp.2053-2064.

[9] Jessica EO, Hamada M, Yusuf SI, Hassan M. The Role of Linear Discriminant Analysis for Accurate Prediction of Breast Cancer. In2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC) 2021 Dec 20, pp. 340-344. IEEE.

[10] Polat K, Sentürk U. A novel ML approach to prediction of breast cancer: combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier. In2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2018 Oct 19, pp. 1-4. Ieee.

[11] Pawlovsky, A.P. and Nagahashi, M., 2014, June. A method to select a good setting for the kNN algorithm when using it for breast cancer prognosis. In *IEEE-EMBS International conference on biomedical and health informatics (BHI)*, pp. 189-192. IEEE.

[12] Liu, P., Fu, B., Yang, S.X., Deng, L., Zhong, X. and Zheng, H., 2020. Optimizing survival analysis of XGBoost for ties to predict disease progression of breast cancer. *IEEE Transactions on Biomedical Engineering*, 68(1), pp.148-160.

[13] Karim, M.R., Wicaksono, G., Costa, I.G., Decker, S. and Beyan, O., 2019. Prognostically relevant subtypes and survival prediction for breast cancer based on multimodal genomics data. *IEEE Access*, 7, pp.133850-133864.

[14] Sun, D., Wang, M., Feng, H. and Li, A., 2017, October. Prognosis prediction of human breast cancer by integrating deep neural network and support vector machine: supervised feature extraction and classification for breast cancer prognosis prediction. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1-5. IEEE.

[15] Liang, M., Huang, L. and Ahmad, W., 2017, December. Breast cancer intelligent diagnosis based on subtractive clustering adaptive neural fuzzy inference system and information gain. In *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, pp. 152-156. Ieee.

[16] Sun, D., Wang, M. and Li, A., 2018. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional

data. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(3), pp.841-850.

[17] Alghunaim, S. and Al-Baity, H.H., 2019. On the scalability of machine-learning algorithms for breast cancer prediction in big data context. *IEEE Access*, 7, pp. 91535-91546.

[18] Agustian, F. and Lubis, M.D.I., 2020, October. Particle swarm optimization feature selection for breast cancer prediction. In *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1-6. IEEE.

[19] Osman, A.H. and Aljahdali, H.M.A., 2020. An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model. *IEEE Access*, 8, pp.39165-39174.

[20] Zhang, X., He, D., Zheng, Y., Huo, H., Li, S., Chai, R. and Liu, T., 2020. Deep learning based analysis of breast cancer using advanced ensemble classifier and linear discriminant analysis. *IEEE Access*, 8, pp.120208-120217.

[21] Nguyen, T., Lee, S.C., Quinn, T.P., Truong, B., Li, X., Tran, T., Venkatesh, S. and Le, T.D., 2021. PAN: Personalized Annotation-based Networks for the Prediction of Breast Cancer Relapse. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6), pp.2841-2847.

[22] Hsu, T.C. and Lin, C., 2020, July. Generative adversarial networks for robust breast cancer prognosis prediction with limited data size. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5669-5672. IEEE.

[23] Liu, P. and Fei, S., 2020. Two-stage prediction of comorbid cancer patient survivability based on improved infinite feature selection. *IEEE Access*, 8, pp.169559-169567.

[24] Arya, N. and Saha, S., 2020. Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. *IEEE/ACM transactions on computational biology and bioinformatics*.

[25] Raweh, A.A., Nassef, M. and Badr, A., 2018. A hybridized feature selection and extraction approach for enhancing cancer prediction based on DNA methylation. *IEEE Access*, 6, pp.15212-15223.

[26] Waseem, M.H., Nadeem, M.S.A., Abbas, A., Shaheen, A., Aziz, W., Anjum, A., Manzoor, U., Balubaid, M.A. and Shim, S.O., 2019. On the feature selection methods and reject option classifiers for robust cancer prediction. *IEEE Access*, 7, pp.141072-141082.

[27] Zhang, D., Zou, L., Zhou, X. and He, F., 2018. Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. *IEEE Access*, 6, pp.28936-28944.

[28] Agarap, A.F.M., 2018, February. On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing*, pp. 5-9.

[29] Izenman, A. J., 2013. Linear discriminant analysis. In *Modern multivariate statistical techniques*, pp. 237-280. Springer, New York, NY.

[30] Naruei, I., & Keynia, F., 2021. Wild horse optimizer: A new meta-heuristic algorithm for solving engineering optimization problems. *Engineering with Computers*, 1-32.

[31] Übeyli, E. D., 2009. Combining recurrent neural networks with eigenvector methods for classification of ECG beats. *Digital Signal Processing*, 19(2), 320-329.