EAI Endorsed Transactions
on Scalable Information Systems Research Article **EAI.EU**

A Hybrid Approach for Breast Cancer Classification and Diagnosis

Bibhuprasad Sahu[1,*], Sachi Nandan Mohanty[2] and Saroj Kumar Rout[2]

[1]Research Scholar, North Orissa University, Baripada, Odisha.
[2]Gandhi Institute for Technology, Bhubaneswar, Odisha.

Abstract

Feature selection in breast cancer disease important and risky task for further analysis. Breast cancer is the second leading reason for death among the women. Cancer starts from breast and spread to other part of the body. People are unable to identify their disease before it become dangerous. It can be cured if the disease identified at early stage. Accurate classification of benign tumours can avoid patients undergoing unnecessary treatments. Data Analytics and machine learning methods provides framework for prognostic studies by errorless classification of data instances into relevant based on the cancer severity. In this study we have purposed a prediction model by combining artificial intelligent based learning technique with multivariate statistical method. For automation of the diagnosis process data mining plays an significant role. The data sets available in different repositories are noisy in nature. This study suggests a hybrid feature selection method to be used with PCA (Principal Component Analysis) and Artificial Neural Network (ANN). Pre-processing of data and extracting the most relevant features done by PCA. The proposed algorithm is tested by applying it on Wisconsin Breast Cancer Dataset from UCI Repository of Machine Learning Databases. In classification phase 10 fold cross validation was used. The suggested algorithm was measured against different classifier algorithms on the same database. The evaluation results of the algorithm proposed have achieved better accuracy with sensitivity and F measure comparison with others and by enhancing this concept we can provide a future scope to produce sophisticated learning models for diagnosis.

Keywords: Breast cancer diagnosis, feature selection, PCA, ANN.

Received on 09 October 2018, accepted on 29 December 2018, published on 03 January 2019

doi: 10.4108/eai.19-12-2018.156086

*Corresponding author. Email:prasadnikhil176@gmail.com

Nomenclature
TP=- True Positive
TN -True Negative
FP -False Positive
FN -False Negative
CAD – Computer Aided Tool
PCA –Principle Component Analysis
SVM- Support Vector Machine
ANN – Artificial Neural Network

1. Introduction

In the present generation, breast cancer has become most common disease occurring in females and becomes the reason for death. Early detection and diagnosis of cancer is needed to save the human life. As per the data in 2012 1.7 million women have suffered by cancer [1]. As any healthcare people are unable to identify the disease at early stage which impacts the death rate day by day [2]. Report available at www.breastcanerindia.net, 144937 numbers of women were detected and 70218 numbers have died with breast cancer [3]. Even if different test are done, still it is a difficult to detect the exact diagnosis. So many researchers have identified different automate diagnosis methods to detect the breast cancer. Some cases CAD (Computer Aided Tool) is also used by the physicians. Breast cancer is a common disease which most of the women suffers at urban and rural areas in India. Various machine learning methods are used to detect whether the cancer gene is benign (or) malignant. Benign are non-cancerous in nature (or) non life threatening for human life, but malignant are cancerous and leads the patient towards death. Although technology improved still 20% of women in world die every year [4]. Machine learning is an emerging technique which provides an efficient way to enhance the knowledge in data in order to enhance the performance of disease predictive models [5]. Previous researches reveal the increasing threats of breast cancer which drive to delve into unsolved and problematic issue, which is the main reason for the work reported in this paper. In the early stage of the breast cancer it appear as a malevolent lump and later stage it growths uncontrolled and variable manner .Even if the technology developed still the reason of causing breast cancer is a mystery. The pathologists detect it though some common risk factors which occurs in women. Genetic information of the patient sometimes considered as a attribute for this disease. Breast cancer treatment is normally divided into two types such as local and symmetric [53]. As per the report by World Health Organization, if early diagnosis provide to the cancer patient then 96% of women may live for an average of 5 years.

In this study, we proposed a hybrid approach PCA+ANN for expecting high accuracy. For minimization of noisy data we applied PCA (Principal Component Analysis) [6] [7] [8][51]. Then reduced data are used for ANN and classification has been performed [52]. In this study we have adopted R and Mat Lab for simulation purpose.

The rest of the paper is organized as follows. Section 2 gives a related work to the predictive techniques that have been applied in breast cancer detection and prognosis. Section 3 described the proposed model of this research and different machine learning methods such as PCA, SVM. Section 4 represents experimental results analysis and Finally Section 5 elaborated conclusion and future scope.

2. Related Work

Tuba Kiyan [9] et al. attempt to implement statistical neural network for breast cancer data set. A comparison between statistical neural network and multilayer perception network is done on WBCN database. For classification radial basic function (RBF), General Regression Neural Network (GRNN) and Probabilistic Neural Network (PNN) were used. In RBF, PNN GRNN & MLP the over performance achieved were 96.18%, 97%, 98.8% and 95.74% respectively .Passlin and Sunthakumaran designed a model implementing Back Propagation Neural Network (BPNN) and achieved 99.28% accruing with Leveaerg-Marquardt algorithm. Median filter is used for Pre-processing and normalization of data using min-max technique [10]. Karabatak and Inus et al. suggested an expert model for breast cancer detection with high accuracy. For diversion reduction they have used association rules. In their analysis, AR1 and AR@ are developed for feature reduction from the original datasets. Out of 9 features AR1 and AR2 reduced to 1 and 5 features respectively. By applying 3 fold cross validation they have achieved 95.6% and 95.2% with AR1 and AR2 respectively [11]. Ahemad and Ahemad et al. used three classification methods such as RBF, PNN with WBC data set. They have proved that among there with accuracy of 96.34% PNN shows better result and MLP is 96.32% [12]. Jhajharix et al. used diversion reduction technique PCA for feature extraction from the data sets. Then they applied it to feed-forward network for classification. A good performance achieved by dividing into training test data [13].Nayak et al. implemented adaptive resource theory (ART) network for data classification. A comparative study is done with PSO-MLP and PSO-BBO algorithm. Result analysis proved that ART performed better with respect to other two classifiers [14]. Nisihi et al. introduced a new methodology combining fuzzy logic with knowledge base system with breast cancer data set. They have used PCA for data reduction. For generation of fuzzy rules CART is used for knowledge based system. For classification fuzzy rule based reasoning method is introduced [15]. Xie-feng Yang et al. presented a comparative analysis on SV classifier. They designed PCA with SVM for breast cancer data classification. Result analysis claims PCA with SVM performed well then SVM [16]. Jihong Liu et al. introduced a methodology by combining two classifiers such as SVM and BPNN with PCA a features extraction algorithm for breast cancer images. SVM proved an the best classifier as per the result achieved [17]. Gouda I. Salama at al. three different classifiers such as C 4.5, SVM and MSP along PCA with 3 breast cancer datasets [18]. Kavitha R et al. proposed a hybrid method for feature selection using PCA and C 4.5 with heart disease datasets. The classification result was quick satisfactory [19]. Yen et al. introduced a new hybrid approach combining discrete PSO with statistical method for mining of breast cancer data at and the classification accuracy achieved is 78.71% [20].

3. Machine Learning Techniques

In machine learning the importance of features selector is inevitable. By the help of the feature selector the datasets is free from ambiguity and dimension of data is also reduced. It solves the over fitting problem of dataset. Identifying the best feature subset from the all features defines the accuracy. Three different features selection method are wrapper method, filter method, embedded method.

3.1.1 Datasets Description

In this study original Wisconsin breast cancer dataset were utilized. Database was downloaded from the University of California, Irvine (UCI) Machine learning repository, which is available through open access. The data set consists of recordings collected from biopsies of real patients in different hospitals of Wisconsin. The data points are grouped chronologically in the order in which the original clinical cases were reported. The dataset has 699 instances or samples characterized by 9 attributes or features. The 1 column presents instance with an ID number while the 11stcolumn of the data table represents class label that describes the severity of the tumour, the labels being benign or malignant. Class labels are originally encoded as 2 for benign and 4 for malignant however for binomial classification purpose, they were suitably considered to 0 and 1 respectively. The 9 attributes of the data (excluding the class label) represent biophysical characteristics of the tumour biopsy and are expressed as integral values.

Table 1. Cell Nuclei Characteristics

Cell Nuclei Characteristics
1. Radius [mean of distances from centre to points on the perimeter]
2. Texture [standard deviation of grey-scale values]
3. Perimeter
4. Area
5. Smoothness [local variation in radius lengths]
6. Compactness [((perimeter)2 / area) -1]
7. Concavity [severity of concave portions of the contour]
8. Concave points [number of concave portions of the contour]
9. Symmetry
10. Fractal dimension ["coastline approximation" - 1]

3.1.2 Principal Component Analysis

In machine learning PCA (a statically method) used for data analysis. Basically it is a dimension reduction technique which includes the relate features. So that the huge dataset can be reduced and can be expressed within few no of variables. When PCA is applied to correlated attributes it gives better result. In this study, PCA is applied to both test and training attributes of breast cancer dataset. The main function of PCA is to detect the patterns in dataset and find similarity and differences between each individual attributes. Variance of breast cancer dataset can be determined as [25, 26]

After PCA was applied, the fundamental variable with high variance is considered as first variable and other converted variables are called as principal component to first variable. Except first variable other variables are represented is descending order according to other variance values [21, 24]. Explanation of PCA is stated as follows.

Consider Z is a P-dimensional dataset. There are n principal axes G_1, G_2, G_3 …….. G_n& 1<=n<=t are presented as orthogonal axes which retains the highest variance in the projected space [22]. GenerallyG_1, G_2, G_3 …….. G_n can be represented by the n highest Eigen vectors of the sample covariance matrix.

Here $x_k \epsilon$ Z and x defines the mean of samples and L represents the no of samples.

So we can represent

$$UG_k = V_k G_k,\ K \in 1 \ldots.. n \tag{4}$$

Where

V_k= K^{th} highest Eigen value of U

N = principal component of given observation vector x_k€ z can be represented as

$$A = [q_1\ q_2\ q_3 \ldots.\ q_n] = [G_1^T{}_x, G_2^T{}_x, G_3^T{}_x \ldots\ldots.. G_n^T{}_x] = G^T{}_x \tag{5}$$

Here ‘A’ represents x principal component of a [23].

3.1.3 Artificial Neural Networks (ANN)

An artificial neural network functions like human brain. Last few years, application of ANN has increased rapidly and has become most active research area [26-29]. ANNs have achieved a great success for classification and diagnosis of breast cancer at early stage [30-37]. A basis ANN model consist of 3 hierarchy layers such as input, hidden and output layer [38]. Each layer is interconnected with neurons and each contains an activation function to improve nonlinear expression ability. The data flow starts from the input layer to output layer through the hidden layer. Finally the output layer produces the classification result. The no of stages varies within the problem description of BPNN (Back Propagation Neural Network) is a gradient descent algorithm used in various domains [39]. In real world problem BPNN is never used as it is extensive which results in slow training [49][50]. Now a day’s researchers have developed different hybrid methods with combination of ANN, which produces better outcomes.

3.1.4 Support Vector Machine (SVM)

A statically learning method proposed by Vapnik, now SVM plays important role in machine learning technique. At early stage SVM was designed for binary problem but now it is comfortable with multi class problem [40-41].

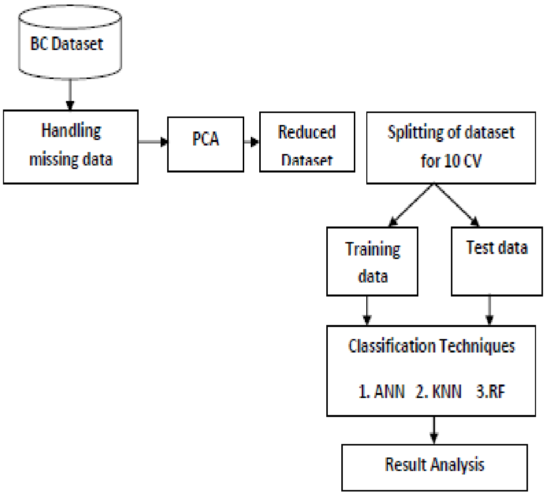

Figure 1. Proposed Model

4. Proposed System

The proposed method uses PCA for classification purpose and ANN is used for identifying a subset of features from the provided dataset. The stepwise involvement in the proposed model is discussed in figure-1.

1. To remove noise from the instances from WBC breast cancer data set pre-processing should done.
2. PCA is applied for reducing the dimension of data from the huge dataset.
3. Reduced dataset is divided into 2 parts as training set and testing set, ANN is applied to gain high accuracy of classification.

5. Experimental Results

As per the original dataset the value provided in the database is 2 and 4 for benign and malignant respectively. As we are considering our problem as a binary class problem so we have to convert all values of class to 0 and no instance have missing values, if exist then missing values was removed which reduces the error rate of the designed system. Remaining 568 instances are used for feature selection.

In PCA dimension reduction process, initially 32 features are provided as input .The objective of using PCA is to reduce no of attributes without losing the main objective information from the original dataset. After PCA is successfully applied to the dataset on the basis of the mean and the covariance factor we have remove the PC component. Then we will apply different classification technique like NN to the PCA output. We have compared with the classifier performance of PCA+RF, PCA+KNN, PCA+NN.

The performance of the model is calculated through the confusion matrix. The classified and non classified rate of the system can be easily found out using the confusion matrix. Calculating the accuracy of the system we can easily indentify effectiveness and performance. The accuracy can be defined as

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

Sensitivity is a measure to find out ratio between correctly indentified instance and actual positive.

Sensitivity= $\frac{TP}{TP+FN}$

Specificity can be measured to find out the ratio between correctly indentified instance and actual negative.

Specificity=$\frac{TN}{TN+FP}$

To find the accuracy of the designed model F-score is used. To find out F-score we find that precision and recall are needed to calculate.

Precision= $\frac{TP}{TP+FP}$

Recall=$\frac{TP}{TP+FN}$

F-score=$\frac{(\beta^2+1)*precision*recall}{\beta^2*precison+recall}$

We have trained the network model with the training dataset 80% and test set with 20% of instances is applied to find out the accuracy rate of the classifier. We have done comparison between different hybrid methods like PCA+RF, PCA+KNN, PCA+NN.As expected this model performs better and provides a promising result of ANN for 70-30% partition spilt. Table-2 shares the confusion matrix of different classifiers used in this research work. By the help of the confusion matrix we can summarize the performance of the classification algorithm. If you have unequal no of observations or if it is a multi class problem then classification accuracy may mislead alone. Confusion matrix provides a better idea of whether the classification models provide right or not and what type error it is making. And figure 3 show the comparative study of different with other hybrid methods. From this comparison table we can easily analyse KNN, ANN, PCA+ANN provide the accuracy of 97%, where as PCA, PCA+RF and NB provide accuracy of 95% and 91% respectively. if we comparison between KNN,ANN and PCA+KNN in term of all evaluation criteria such as Kappa statistics, Sensitivity, Specificity, Prevalence PCA+KNN performs better. So PCA+ANN considered as a better classifier as compared to others.

Table 2. Confusion matrix for proposed model

Classifiers	Prediction	Reference	
		B	M
RF	B	105	5
	M	2	58
PCA+RF	B	104	5
	M	3	58
KNN	B	107	5
	M	0	58
ANN	B	107	5
	M	0	58
PCA+ANN	B	105	3
	M	2	60
NB	B	99	7
	M	8	56

Table 3. Performance for the proposed system (Accuracy)

Classifier	Kappa statistics	Accuracy	Sensitivity	Specificity	Prevalenc
RF	0.91	0.95	0.92	0.98	0.37
PCA+RF	0.95	0.95	0.92	0.97	0.37
KNN	0.93	0.97	0.92	1	0.37
ANN	0.93	0.97	0.92	1	0.37
PCA+ANN	0.93	0.97	0.95	0.98	0.37
NB	0.81	0.91	0.88	0.92	0.37

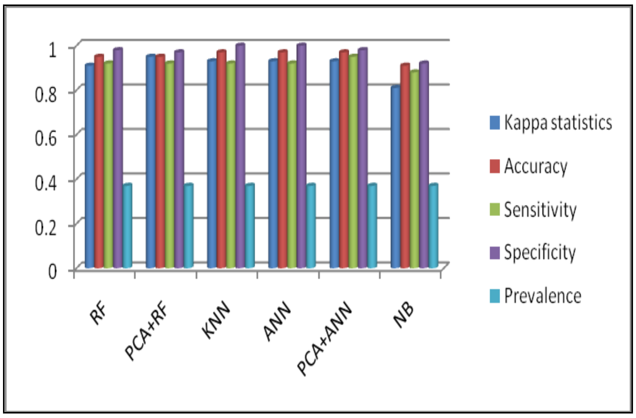

Figure 2. Performance of different classifiers

Table 3 and Figure 2 help to visualize the accuracy of the proposed method. Figure 4 shows graphical representation of the performance of the proposed method mention in Table 3 in term of accuracy, sensitivity and specificity with different ranges of train-test spilt partitioning. Figure-3 shows ROC (Receiver Operating Characteristic) curve to analyze the performance of the model.

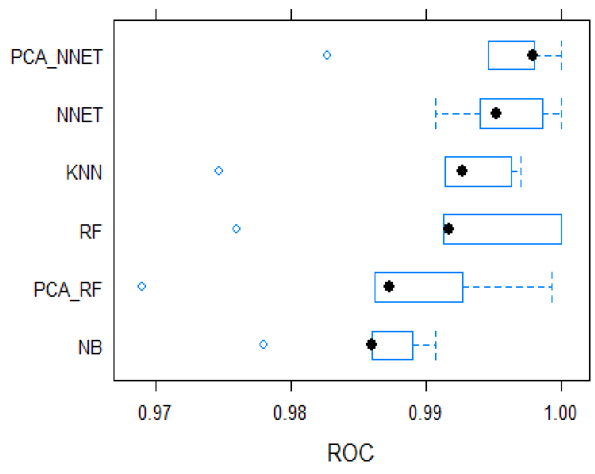

Figure 3. Models evaluation with ROC

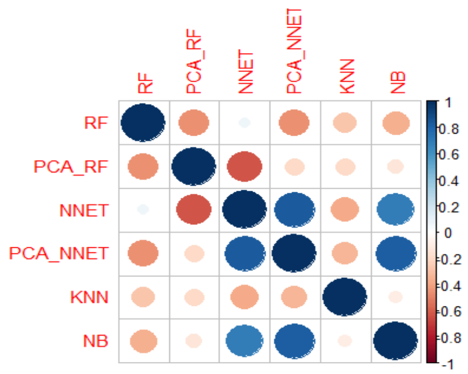

Figure 4. Performance Evaluation (Correlation)

Conclusion

A predictive model of diagnosis of cancer has be proposed in this research study. Multivariate statistical and machine learning techniques are incorporated for better accuracy. Though PCA results a significant contribution towards defining the variance in data, we have kept the original attributes of data for our study. The performance classification accuracy has been encouraging comparison with various existing classifiers implemented in the literature. As expected the proposed model classifies diagnosis cases into either cancerous or non cancerous with high amount of accuracy. Even if technology has developed, still lots of people are facing many issues with modern age diseases. Breast cancer has become one of the most common deadliest diseases rising over days among all countries in world. Ratio of this disease increases due to lack of awareness and late identification. Our result reveals ANN (machine learning) plays measure factor for detection of cancer diagnosis to save the human life from the dangerous disease.

References

[1] Sahu, B. (2018). A Combo Feature Selection Method (Filter+ Wrapper) for Microarray Gene Classification. International Journal of Pure and Applied Mathematics, 118(16), 389-401.

[2] www.breastcancerindia.net, Statistics of Breast Cancer in India.

[3] Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. International journal on computer science and engineering, 3(5), 1787-1797.

[4] Naseriparsa, M., & Kashani, M. M. R. (2014). Combination of PCA with SMOTE resampling to boost the prediction rate in lung cancer dataset. arXiv preprint arXiv:1403.1949.

[5] Napoleon, D., & Pavalakodi, S. (2011). A new method for dimensionality reduction using k-means clustering algorithm for high dimensional data set. International Journal of Computer Applications, 13(7), 41-46.

[6] Raschka Sebastian and Mirjalili (2017),"Python Machine Learning",2nd edition, Packt Publishing Ltd.,ISBN:978-1-78712-593-3.

[7] Kara, S., Güven, A., & İçer, S. (2007). Classification of macular and optic nerve disease by principal component analysis. Computers in Biology and Medicine, 37(6), 836-841.

[8] Kiyan, T., & Yildirim, T. (2004). Breast cancer diagnosis using statistical neural networks. IU-Journal of Electrical & Electronics Engineering, 4(2), 1149-1153.

[9] Paulin, F., & Santhakumaran, A. (2011). Classification of breast cancer by comparing back propagation training algorithms. International Journal on Computer Science and Engineering, 3(1), 327–332.

[10] Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications, 36(2), 3465–3469.

[11] Azar, A. T., & El-Said, S. A. (2014). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. Neural Computing and Applications,24(5), 1163–1177.

[12] Jhajharia, S., Varshney, H. K., Verma, S., & Kumar, R. (2016). A neural network based breast cancer prognosis model with pca processed features. In: 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1896–1901.

[13] Nayak, T., Dash, T., Rao, D. C., & Sahu, P. K. (2016, August). Evolutionary neural networks versus adaptive resonance theory net for breast cancer diagnosis. In Proceedings of the International Conference on Informatics and Analytics (p. 97). ACM.

[14] Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics and Informatics, 34(4), 133–144.

[15] Virmani, J., Dey, N., & Kumar, V. (2016). PCA-PNN and PCA-SVM based CAD systems for breast density classification. In Applications of intelligent optimization in biology and medicine(pp. 159-180). Springer, Cham.

[16] Liu, J., & Ma, W. (2008, January). An effective recognition method of breast cancer based on PCA and SVM algorithm. In International Conference on Medical Biometrics (pp. 57-64). Springer, Berlin, Heidelberg.

[17] Salama, G. I., Abdelhalim, M., & Zeid, M. A. E. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC), 32(569), 2.

[18] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal, 15, 104-116.

[19] Jolliffe, I. (2011). Principal component analysis. In International encyclopedia of statistical science (pp. 1094-1096). Springer, Berlin, Heidelberg.

[20] Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition

using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720

[21] Calisir D, Dogantekin E (2011) A new intelligent hepatitis diagnosis system: PCA-LSSVM. Expert Syst Appl 38(8):10705– 10708. doi:10.1016/j.eswa.2011.01.014.

[22] Kotu V and Deshpande B (2015) Predictive Analytics and Data Mining (Waltham: Morgan Kaufmann)

[23] Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

[24] Fu, Q., Luo, Y., Liu, J., Bi, J., Qiu, S., Cao, Y., & Ding, X. (2017, October). Improving learning algorithm performance for spiking neural networks. In Communication Technology (ICCT), 2017 IEEE 17th International Conference on (pp. 1916-1919). IEEE.

[25] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.

[26] Wang, L., Wang, Z., Wei, G., & Alsaadi, F. E. (2018). Finite-time state estimation for recurrent delayed neural networks with component-based event-triggering protocol. IEEE transactions on neural networks and learning systems, 29(4), 1046-1057.

[27] Yang, F., Dong, H., Wang, Z., Ren, W., & Alsaadi, F. E. (2016). A new approach to non-fragile state estimation for continuous neural networks with time-delays. Neurocomputing, 197, 205-211.

[28] Furundzic, D., Djordjevic, M., & Bekic, A. J. (1998). Neural networks approach to early breast cancer detection. Journal of systems architecture, 44(8), 617-633.

[29] Floyd Jr, C. E., Lo, J. Y., Yun, A. J., Sullivan, D. C., & Kornguth, P. J. (1994). Prediction of breast cancer malignancy using an artificial neural network. Cancer: Interdisciplinary International Journal of the American Cancer Society, 74(11), 2944-2948.

[30] Fogel, D. B., Wasson III, E. C., & Boughton, E. M. (1995). Evolving neural networks for detecting breast cancer. Cancer letters, 96(1), 49-53.

[31] Fogel, D. B., Wasson, E. C., Boughton, E. M., Porto, V. W., & Angeline, P. J. (1998). Linear and neural models for classifying breast masses. IEEE transactions on medical imaging, 17(3), 485-488.

[32] Pendharkar, P. C., Rodger, J. A., Yaverbaum, G. J., Herman, N., & Benner, M. (1999). Association, statistical, mathematical and neural approaches for mining breast cancer patterns. Expert Systems with Applications, 17(3), 223-232.

[33] Setiono, R. Extracting rules from pruned neural networks for breast cancer diagnosis. Artif. Intell. Med. 1996, 8, 37–51. [CrossRef]

[34] Wilding, P., Morgan, M. A., Grygotis, A. E., Shoffner, M. A., & Rosato, E. F. (1994). Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. Cancer Letters, 77(2-3), 145-153.

[35] Wu, Y., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., & Metz, C. E. (1993). Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology, 187(1), 81-87.

[36] Minsky, M., & Papert, S. A. (2017). Perceptrons: An introduction to computational geometry. MIT press.

[37] Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: explorations in the microstructure of cognition. volume 1. foundations.

[38] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[39] Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

[40] Scholkopf, B., Sung, K. K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T., & Vapnik, V. (1997). Comparing support vector machines with Gaussian kernels to radial basis function

classifiers. IEEE transactions on Signal Processing, 45(11), 2758-2765.

[41] Wan, V., & Campbell, W. M. (2000). Support vector machines for speaker verification and identification. In Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop (Vol. 2, pp. 775-784). IEEE.

[42] De Mántaras, R. L. (1991). A distance-based attribute selection measure for decision tree induction. Machine learning, 6(1), 81-92.

[43] Mingers, J. (1989). An empirical comparison of selection measures for decision-tree induction. Machine learning, 3(4), 319-342.

[44] Apté, C., & Weiss, S. (1997). Data mining with decision trees and decision rules. Future generation computer systems, 13(2-3), 197-210.

[45] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.

[46] TP, L., & Parthiban, L. (2017). Abnormality detection using weighed particle swarm optimization and smooth support vector machine. Biomedical Research (0970-938X), 28(11).

[47] Prevention Control: Center for Diseases Control Prevention(2014).

[48] Sharma, M., Singh, G., & Singh, R. (2018). Clinical decision support system query optimizer using hybrid Firefly and controlled Genetic Algorithm. Journal of King Saud University-Computer and Information Sciences.

[49] Sharma, M., Sharma, S., & Singh, G. (2018). Performance Analysis of Statistical and Supervised Learning Techniques in Stock Data Mining. Data, 3(4), 54.

[50] Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. Computational and structural biotechnology journal, 15, 75-85.

[51] Kaur, P., & Sharma, M. (2017). A Survey on Using Nature Inspired Computing for Fatal Disease Diagnosis. International Journal of Information System Modeling and Design (IJISMD), 8(2), 70-91.

[52] O'Malley, C. D., Le, G. M., Glaser, S. L., Shema, S. J., & West, D. W. (2003). Socioeconomic status and breast carcinoma survival in four racial/ethnic groups: a population-based study. Cancer: Interdisciplinary International Journal of the American Cancer Society, 97(5), 1303-1311.