EARLY DETECTION OF BREAST CANCER TUMORS USING LINEAR DISCRIMINANT ANALYSIS FEATURE SELECTION WITH DIFFERENT MACHINE LEARNING CLASSIFICATION METHODS

Mazhar Abbas, Hamid Ghous

Department of Computer Science, ISP Multan Pakistan

ABSTRACT

Globally, the frequency of breast cancer and its morality speak to a critical and developing risk for the developing countries. In Asia, Pakistan has the biggest rate of breast cancer. It is evaluated that every year 83,000 cases were reported in Pakistan and over 40,000 deaths are caused by breast cancer. Patients suffering from this malignancy have a better chance of surviving if they are diagnosed early. Many Early identification of breast cancer can be achieved using data mining techniques, allowing preventative treatments to be done. In this research Wisconsin Breast Cancer Dataset (WBCD) and Duke Breast cancer dataset (DBDS) are used with Linear Discriminant Analysis (LDA) feature selection with Support Vector Machine (SVM), Decision Tree (DT), Neural Network and Random Forest (RF) machine learning classifiers to predict breast cancer tumors. The finding of the proposed model is that feature selections through LDA improve the accuracy of detecting tumors and also reduce time duration of executing model. The best machine learning model with LDA feature selection is Neural Network Model with highest accuracy 1.00 among all classification models and also consume less time.

KEYWORD

Breast cancer, LDA, Feature selection, SVM, Decision Tree, WBCD, Duke Breast cancer, Neural Network

1. INTRODUCTION

Breast cancer is the prime reason of death in ladies. It is a multistage process that includes many cells and preventing it remains a challenge around the world. Early diagnosis is one of the most effective ways to prevent it. Because of early detection, the breast cancer survivors' 5-year relative survival score in some developing countries is over 80%.[1] Cancer begins in the cells, which are the primary cause of all structures and vital organs, including the breast, to develop. When the need for a body grows, cells in the body are further separated to create new cells. When regular cells are injured or get mature, they die and are replaced by new cells. This method may be unreliable at times. When a body does not need new cells, old or damage cells do not die. A lump, growth, or tumors is a mass of tissue formed by the development of extra cells.[2]

1.1. Breast Cancer Symptoms

- ♣ New chunk in the your breast
- ♣ Breast skin irritation
- ♣ Flaky skin in nipple section
- ♣ Pain in the vicinity of the breast
- ♣ Bloody nipple secretion in addition to breast milk

- ♣ variations in the breast's size, such as pulling upward or dimpled

1.2. Risk Factor

Anything that poses a "risk factor" is referred to as a "risk factor." that boosts your probability of growing breast cancer. Some factors can be controlled but some factors cannot be controlled. Controlled Factors are Weight, Diet, Exercise, Smoking, and Alcohol Consumption. Gender, age, family background, personal history of breast cancer, race, chest radiation therapy, pregnancy and breastfeeding, reproductive history, and genetic mutations are all variables that cannot be control. Tumors formed as Benign or Malignant. [3]

1.3. Clinical Introduction

Cancer begins with cells, which are the fundamental components of all structures and organs in the body including the breast. When the need for a body grows, cells in the body are further separated to create new cells. When regular cells are injured or get mature, they die and are replaced by new cells. This method may be unreliable at times. When a body does not need new cells, old or removed cells do not die as they should. A lump, growth, or tumors is a mass of tissue formed by the development of extra cells. Tumors formed as Benign or Malignant.[4]

1.3.1. Benign Tumor

The rule of a benign tumor is not destructive/ Harmful. Seldom, invade the tissues around them. A benign Tumor is not spread in the different body part. It can be eliminated and will not regrow.

1.3.2. Malignant Tumor

Malignant can create a danger to life. It has the ability to target nearby organs and tissues. It has the energy to grow to other areas of the body. Frequently can be expelled but in some cases develop back.

2. RELATED WORK

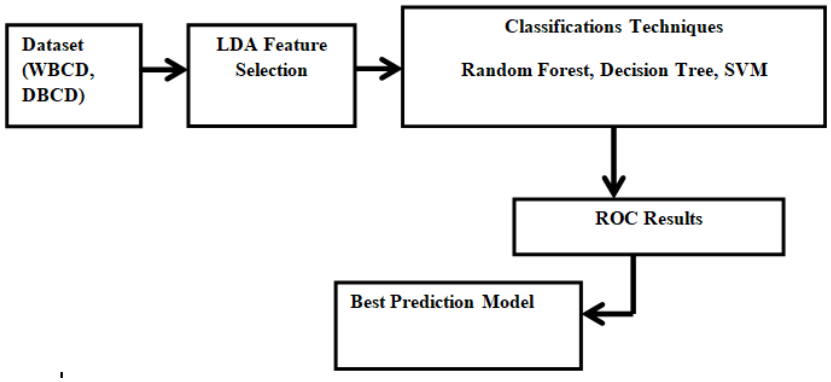
In past, many researchers worked on breast cancer tumor diagnosis and many techniques are used but still, a more accurate classification model is needed for detecting breast tumors that work with the combination of machine learning techniques. Previous work will be discussed in this section. In 2019 Gharibdousti with his team proposed a Hybrid Discriminant Logistic for reducing features model for predict breast cancer tumors. SVM, Nave Bayes, DT, LR, and ANN are applied to assess model results. The UCI Irvine repository's Wisconsin diagnostic repository is included. As a consequence, the Nave Bayes and SVM classification models outperform the competition. Of all versions, a model produced with a Hybrid DA-LR function range performs the best.[5] Aalaei et.al (2016) developed a wrapper approach model using a PS classifier and a GA for selecting features. The proposed feature selection method was applied to various classifiers, including ANN, PS, and GA-based classifiers. WBC and WPBC are the three data sets they use. The analysis shows that feature selection improve all classifiers' quality, consistency, and sensitivity.[6] Chaurasia (2018) developed a breast cancer survivability prediction model. Three of the most common data mining methods are used to create the prediction model Nave Bayes, RBF network, and J48. The 10 fold cross validation model was applied to diagnose results. With 97.36 percent accuracy, the result reveals that Nave Bayes is the best indicator. With 96.77 percentage points, RBF Network took second place. With a result of 93.41 percent, J48 came in eighth.[7] Akay (2009) to increase prediction accuracy uses a vector

machine with a feature selection process. The WBCD is used to measure performance. The classification efficiency, positive negative predict value, specificity, ROC curves, and misunderstanding matrix are used to assess the model's results. The best accuracy is seen in the output is 99.51% obtained by the SVM model with it is compared with other classification models.[8] Bibhuprasad Sahu (2019) work on integrating artificial intelligence-based learning techniques with a multivariate statistical approach, a model was formulated. A hybrid feature selection method that combines PCA and ANN. Upon this WBCD is used, 10 fold cross-validations were used during the classification process. The proposed model classifies diagnosis cases into either cancerous or non-cancerous with a high amount of accuracy as compared with other existing models.[9] Joana Diz (2016) comparing the two datasets to detect the most accurate tools for predicting whether a tumors is benign or malignant. Texture matrixes were used to remove features, which were then categorized using DM algorithms. For the most part, Random Forest is the best while NB was the best at identifying mass texture.[10] Chinnaiyan Ponnuraja (2017) Using data mining methods, we built a model to predict patient survival. The DT algorithm employs defined features. The decision tree algorithm yields a path with the best rate of survival (96.4%) and the highest rate of dying (96.4%).[11] Uma Ojha (2017) Compare Clustering and Classification Algorithm on Dataset (Decision tree, SVM, Fuzzy c-mean). The experiment of classification algorithms are best prediction model than the clustering algorithm. According to the findings, the DT and SVM have the best prediction model with 81% accuracy and fuzzy c-mean with the low accuracy with 37%.[12] Emina (2017) developed a model for breast cancer detection. The extraction of insightful and important features is done using genetic algorithms. In the second level, different human and multiple classification models were examine to build an effective source for breast cancer classification. The RF model with GA most accurate with 99.48%.[13] Seral Sahana (2007) suggested a novel mixed algorithm for breast cancer detection based on a fuzzy artificial immune system and the KNN algorithm. The WBCD was used to develop a tool for solving this diagnosis issue. As contrasted to other research, In terms of classification precision, our suggested method outperforms them. The best classification accuracy ever achieved was 99.14 percent. Using 10-fold cross-validation, the classification accuracy was determined.[14] Yılmaz kaya (2013) Centered on a rough system and an intense learning engine, a breast cancer detection with a new intelligent classification has been suggested. Breast cancer was diagnosed using a combination of rough collection (RS) and extreme learning machine (ELM) approaches. For research, the WBCD was used. The RS + ELM model was found to have a 100 percent success rate.[15] F. M. Okikiola (2019) proposed a fuzzy specialist framework for breast cancer diagnosis and treatment recommendations. The JAVA, MATLAB, and the SQLite database were used to create the framework. The fuzzy inference was used to increase the system's accuracy and precision improve 0.1 % according to the assessment. The method is simple to use and has a high degree of validation.[16] Perviz Abdulmaleki (2001) ANN was used to propose a classification model with extraction of features for breast cancer with resonance imaging. The Jackknife approach was used to plan, network train, and test a neural network system and compared with Radiologists. The diagnostic accuracy of this network is 89% compared with radiologists' 79%.[17] Carlos Andres (1999) proposed a fuzzy genetic approach to diagnose breast cancer by Fuzzy processes and evolutionary algorithms are combined in this study. The fuzzy-genetic approach produces systems that attain high classification performance and involve a few simple rules.[18] Sabri Boughorbel (2016) AU ROC accuracy for various diagnose times was used to assess the performance of many prediction strategies for breast cancer prognosis. SVM, RF, ANN, KNN, and Boosted Trees are contrasted to the GLM, GLM-Net, PLS, SVM, RF, Neural Networks, KNN, and Boosted Trees.[19] Pakizah Saqib (2019) proposed a range of hybrid features approach with the aim of generating the most appropriate feature and improving result prediction. A wrapper is used, which consists of a GA, a heuristic search function, and a RF forecasting model. The proposed method was tested on the WBCD, and it achieved a prediction accuracy of 99.04 percent.[20] Davi Carvalho (2016) proposed a hybrid approach to aid in breast cancer early detection. A Bayesian Network is built hybrid method to

measure the conditional the likelihood of a certain person developing cancer, and MCDA is used. Automated develop a tool to provide a precise detect in events in general and to assist in the decision-making process in more complex cases.[21] Zohaib Mushtaq (2019) explores the efficiency of KNNs by experimenting with K values and distance function order to find an effective KNN. Chi-square-based feature selection, feature selection without a SVM, feature selection without a SVM The findings showed that, as opposed to current versions, the Chi square based feature selection methodology has the highest accuracy levels.[22] Reem Alyami (2017) ANN SVM and along with feature extraction was used to propose a model for diagnosing breast cancer. SVM and ANN are used to pick features depending on the correlation coefficient. In observational experiments comparing SVM and ANN, the results showed that SVM performed well compared ANN with classification precision of 97.14 and 96.71, respectively.[23] T.Sridevi (2014) proposed a hybrid model K-mean clustering and SVM on WBCD. SVM is examine to partition the given information system using the K means clustering algorithm. A hybrid model's performance rate is calculated by 99 percent.[24] Cuong Nguyen (2013) proposed a machine learning system built on a RF classifier with feature selection for diagnosing and prognosticating breast cancer. A model proposed by merging RF and method of selecting features. In the best-case scenario, accuracy of classifier is 100 percent, with an average of 99.8%.[25] Abeer Alzubaidi (2016) proposed a model to diagnose breast cancer through GA and mutual information. GA based on the MI technique with two classification model, namely the KNN, and SVM. The proposed model perform well when it both these model are combined.[26] Dr. J. Arunadevi (2019) GLM and RF used Feature selection three classifiers KNN, SVM, and ANN are also examine to detect breast cancer. As a result feature selection method improves the accuracy of the classification model. GLM performed very well as compared to other model.[27] Abderrahmane Eddy (2020) for the classification of breast cancer, affiliation rules and help vector machines are used on a reduced feature set. To exclude trivial features, Association Rules (AR) is used. To separate the incoming tumors, multiple classifiers are used in the second step. For eight attributes, the SVM model with AR 98.4 % which is best among all classification while for four properties it is 96.14 percent.[28] Adel S Assiri and Saima Nazir (2020) focused on a plurality voting system, suggested an ensemble classification mechanism. LR learning, SVM learning with gradient descent optimization, and multilayer networks are used in ensemble classification with a voting system. The WBCD (Wisconsin Breast Cancer Dataset) is used. The majority-based voting mechanism outperforms the state-of-the-art WBCD algorithm with a score of 99.42 percent.[29] Ahmed Abdullah (2020) developed a hybrid feature selection technique that blends the benefits of feature selection methods with an optimized OGA to choose the right features for results improvement. As a fitness function, the C4.5 DT classifier is used. The output is that for optimal feature collection, Single filter methods and PCA are outperformed by the proposed hybrid feature selection.[30]

3. PROPOSED FRAMEWORK

The publically available WBCD from UCI Repository and Duke Breast cancer dataset are used. LDA feature selection with different Classifiers to predict Breast cancer. SVM, Decision Tree, and Random forest classifiers are used in this prediction model.



3.1. Data Preprocessing

Data collection is an important component of Research. In this research, WBCD is used for the diagnosis of breast cancer tumors. The dataset is gathered from UCI Repository which is publically available on the Internet. Second Dataset Duke Breast cancer dataset is available on kaggle website. WBCD dataset consist of 699 records and each record has 9 attributes. Each feature value range is 1-10. The target variable is 0 and 1, 0 for benign tumors and 1 for malignant tumors. Second dataset is used in this research is duke breast cancer. DBCD is developed by Shirish Krishnaj Shevade and S. Sathiya Keerthi in 2003. Dataset is publically available at LIBSVM (Library for Support Vector Machine) Data repository. Dataset consist of 86 instances and 7130 attributes. The dataset contain zero missing value. Class distribution value is 0 and 1.

3.2. Feature Selection

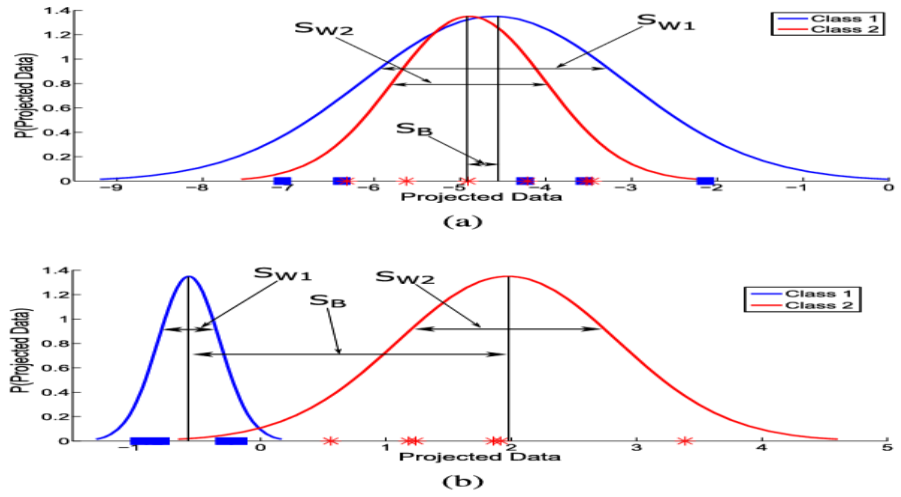
In this research R, 4.0.3 is used. LDA feature selection is used in R to select a feature from WBCD. Feature selection a machine learning technique which is used to select the relevant feature and ignore irrelevant feature for better accuracy of a model. Every feature selection technique defined attributes of a dataset into two groups firstly is the relevant feature and secondly is the irrelevant feature. In this research LDA feature selection method is used for the detection of breast tumors.

$P(X|\ \pi_i)= 1(2\ \pi)p/2|1/2exp[-12(X-Ui)\Sigma-1(X-\mu i)]$

3.2.1. Linear Discriminant Analysis (LDA)

Linear Discriminate analysis is used different in fields to locate a combination of the features and also separate different classes of objects. This is a machine learning technique applied for dividing data into two groups. The LDA approach works by projecting the original data matrix into a lower-dimensional environment. Three stages were required to accomplish this goal.

1. The first step is to determine the between-class variance, which is the separability between various classes.
2. The within-class variance is calculated by calculating the distance between the mean and the samples of each class.
3. The final stage is to create a lower-dimensional space that optimizes between-class variance while minimizing within-class variance.



[31]

3.3. Machine Learning Techniques

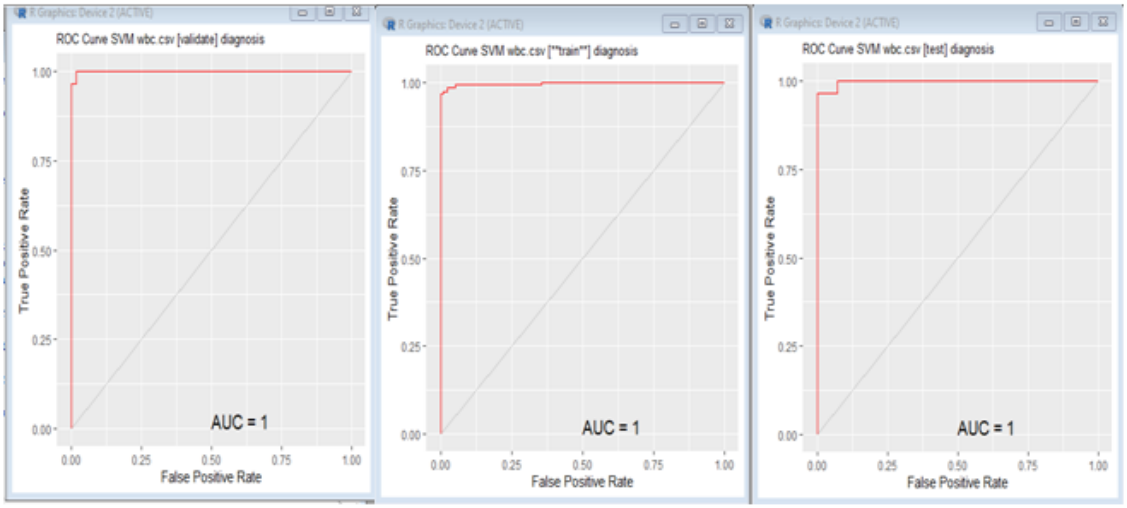
After preprocessing and feature selection from WBCD and DBCD, now Machine learning classifiers are applied to the data to predict tumors. SVM, DT, and RF Classifiers are applied.

4. RESULTS

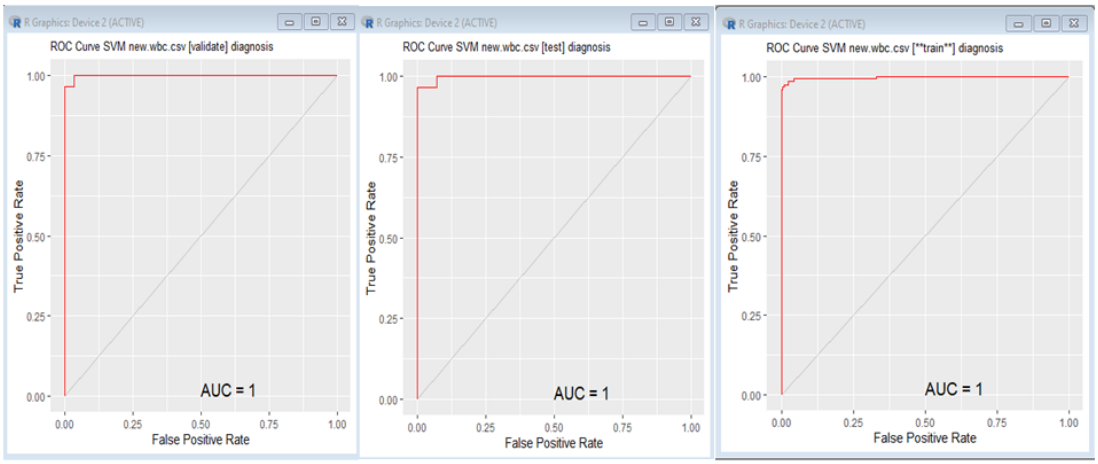
Results are calculated on two datasets. WBCD and DBCD are used.

4.1. Wisconsin Breast Cancer Dataset (WBCD)

We are using R 4.0.3 software for classification. Features are selected through LDA which is also implemented in R. SVM, DT, and RF classification is applied on the dataset. These models are used for training, testing, and validation. Firstly SVM model is implemented.



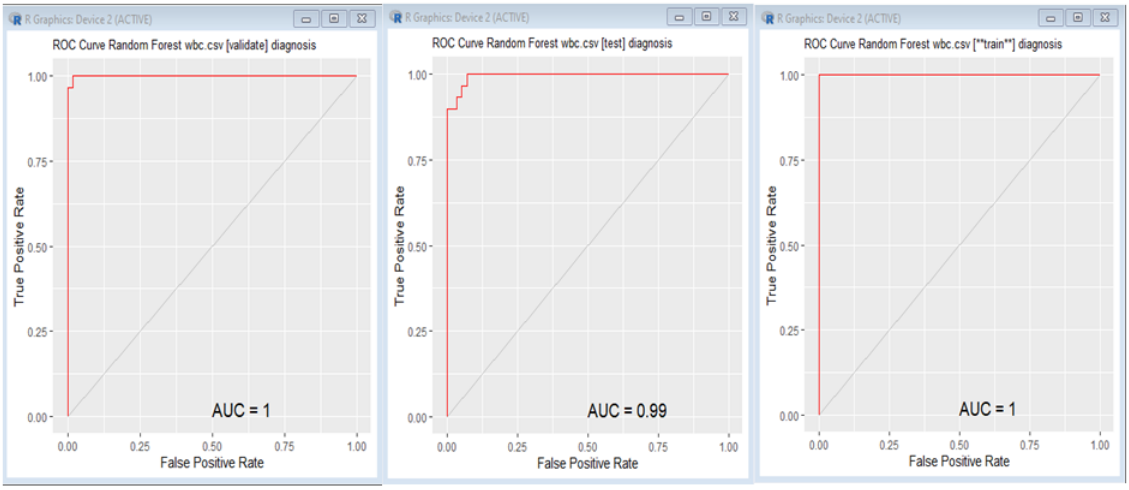
The above ROC Curve on SVM model without LDA on WBCD and bellow are the roc curve on SVM model with LDA. (Training, Testing, Validation)



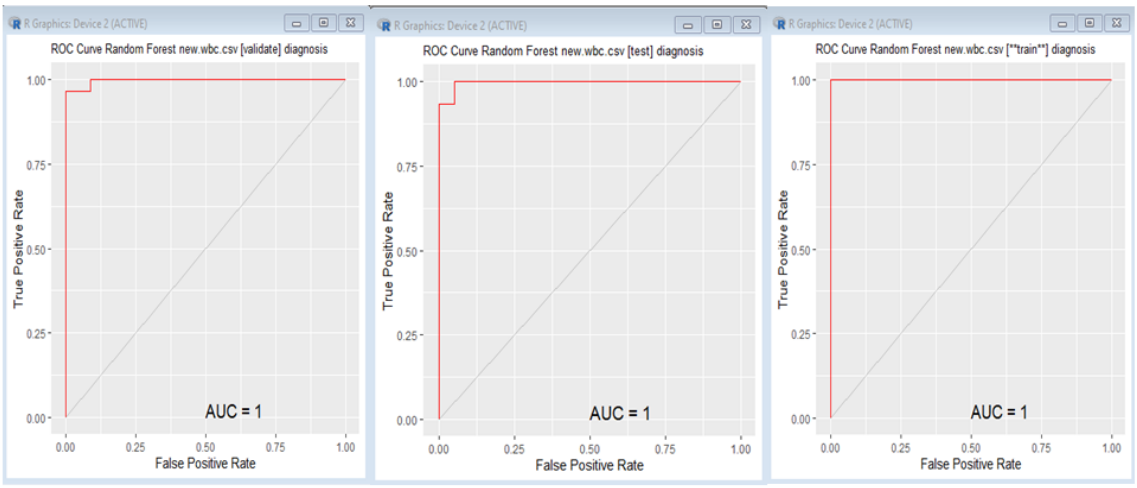
Results are shown below on SVM model with LDA and without LDA

Dataset	Classification Model	Time Taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
WBCD without LDA	SVM	1.08	0.9970	0.9976	0.9994
WBCD with LDA	SVM	0.11	0.9971	0.9976	0.9988

Secondly Random Forest model with LDA and without LDA on Training, Testing, Validation



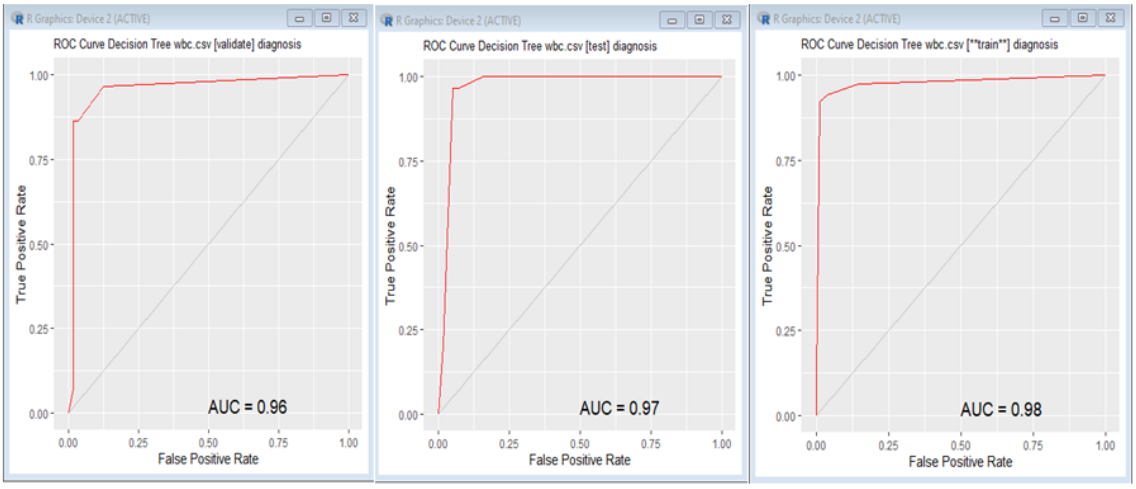
Above ROC curves of Random Forset are without LDA and below with LDA



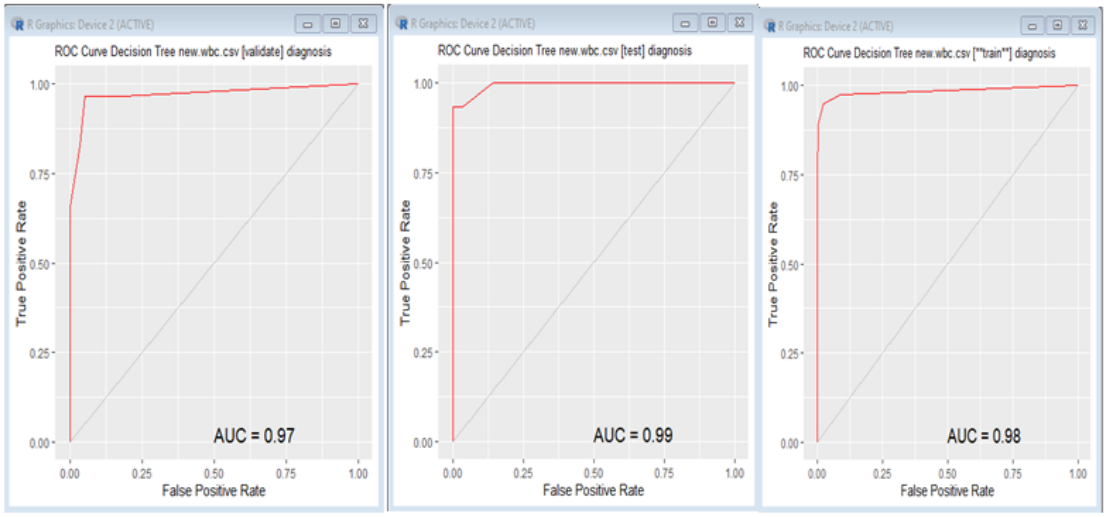
Results of Random Forest with LDA and without LDA

Dataset	Classification Model	Time Taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
WBCD without LDA	Random Forest	0.84	1.0000	0.9946	0.9994
WBCD with LDA	Random Forest	0.33	1.0000	0.9964	0.9969

Now Third Classification model Decision Tree with LDA and without LDA



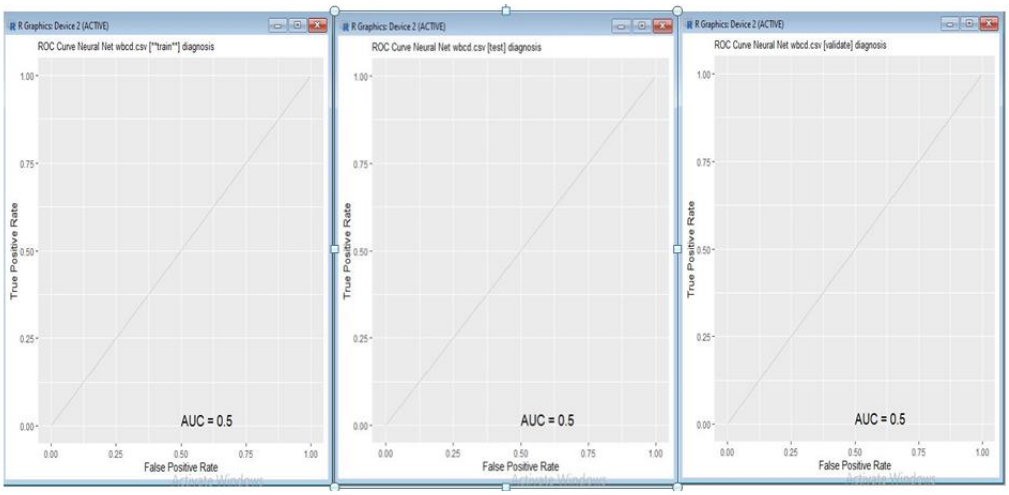
The above ROC curves of Decision Tree without LDA and bellow with LDA



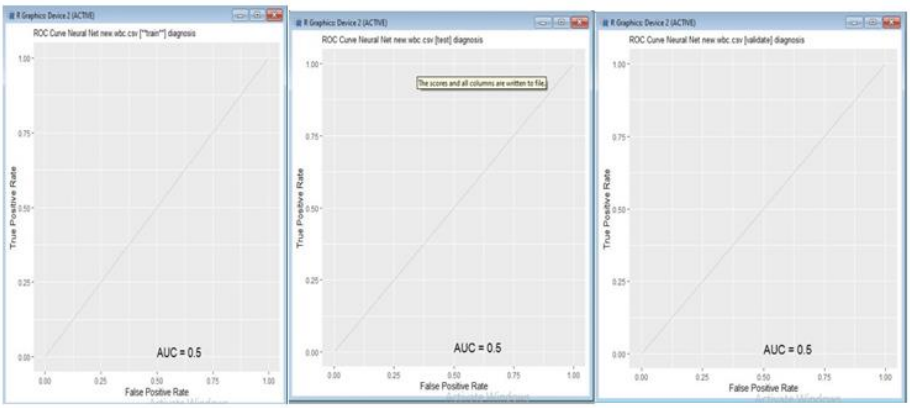
The results of Decision Tree with LDA and without LDA

Dataset	Classification Model	Time Taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
WBCD without LDA	Decision Tree	0.08	0.9765	0.9670	0.9575
WBCD with LDA	Decision Tree	0.04	0.9831	0.9940	0.9704

The Result of Neural Network Model with LDA and without LDA



The above ROC of Neural network model without LDA and below ROC with LDA



The results of Neural Netwok Model with LDA and without LDA

Dataset	Classification Model	Time Taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
WBCD without LDA	Neural Netwok	0.6	0.5	0.5	0.5
WBCD with LDA	Neural Netwok	0.4	0.5	0.5	0.5

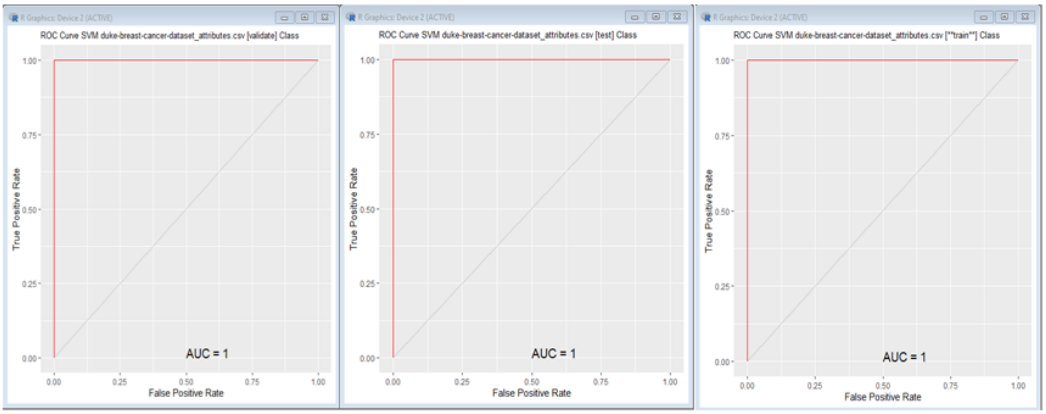
These are the results of different machine learning classification techniques with LDA with without LDA feature selection.

4.2. Duke Breast Cancer Dataset (DBCD)

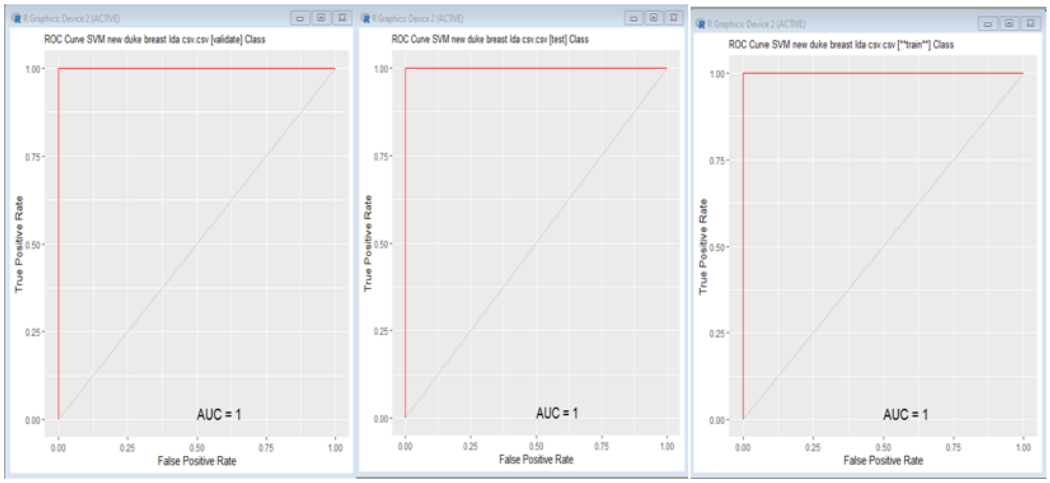
Second dataset is used in this research is duke breast cancer. DBCD is developed by Shirish Krishnaj Shevade and S. Sathiya Keerthi in 2003. Dataset is publically available at LIBSVM (Library for Support Vector Machine) Data repository. Dataset consist of 86 instances and 7130 attributes. The dataset contain zero missing value. Class distribution value is 0 and 1.

4.2.1. SVM Model

Roc Cure without LDA on SVM model.
Below ROC of SVM model on training, Testing and validation of the Support Vector Machine on Wisconsin Breast Cancer Dataset (WBCD).



The above ROC curve without LDA on SVM model with DBCD and bellow with LDA

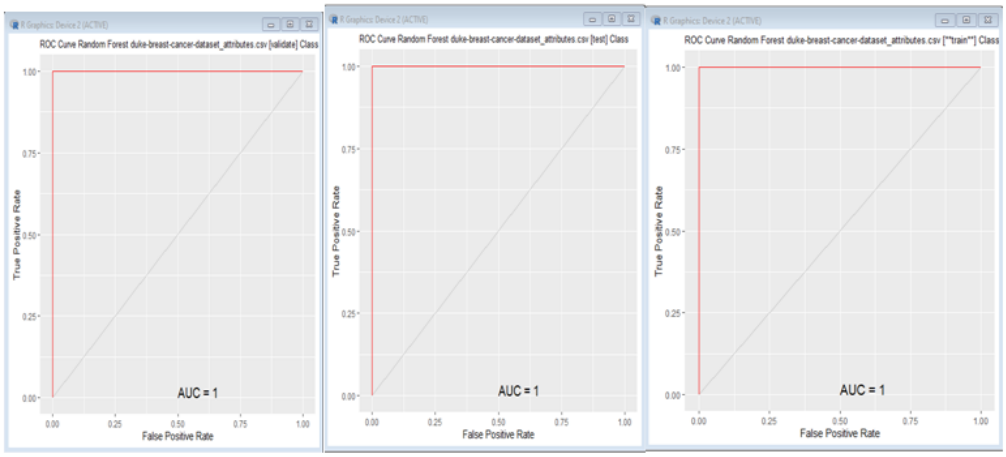


The Results of SVM model on DBCD with LDA and without LDA

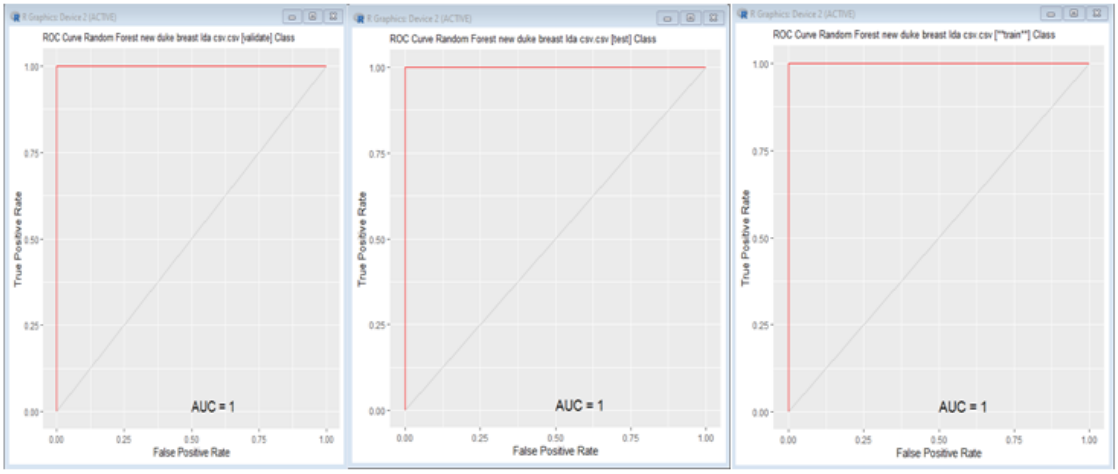
Dataset	Classification Model	Time taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
DBCD without LDA	SVM	7.16	1.0000	1.0000	1.0000
DBCD with LDA	SVM	1.75	1.0000	1.0000	1.0000

4.2.2. Random Forest

The RoC with LDA and without LDA



Above ROC without LDA and bellow ROC with LDA

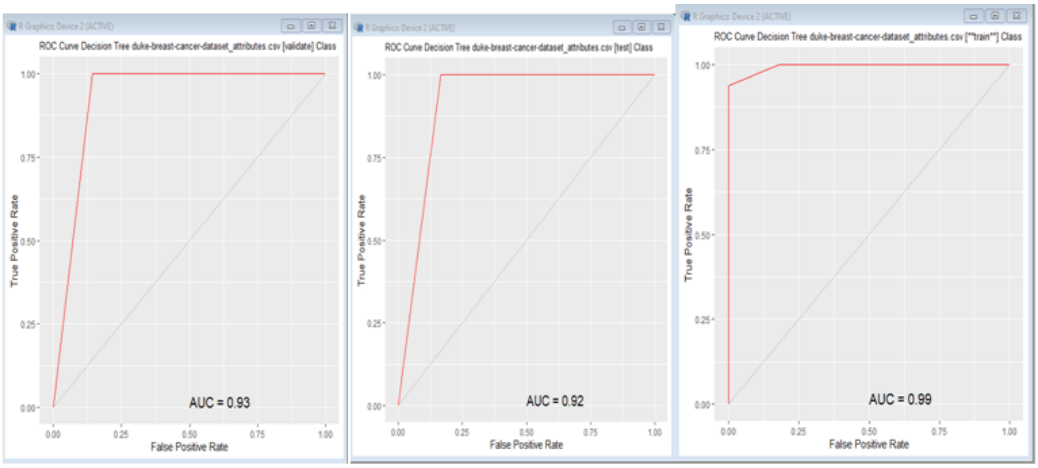


The Results of Random Forset with LDA and without LDA on DBCD

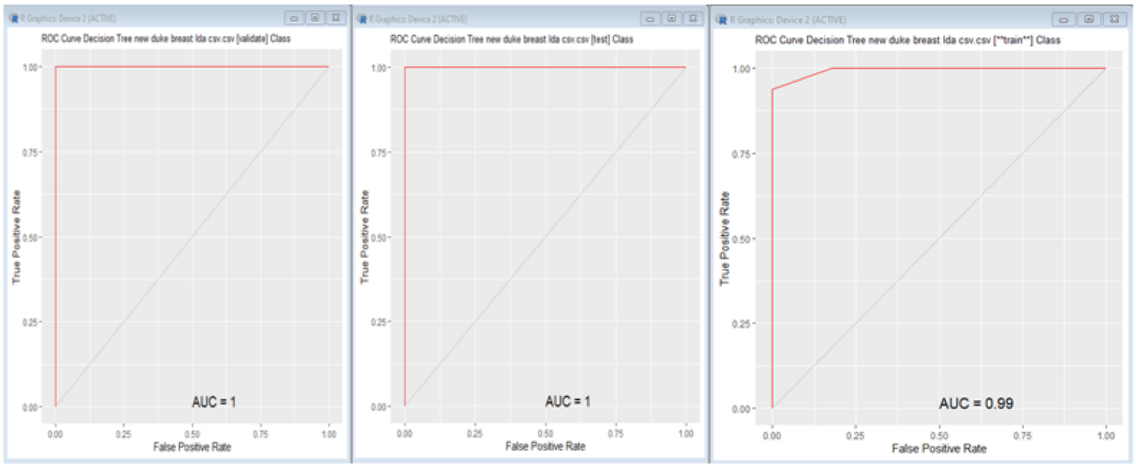
Dataset	Classification Model	Time taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
DBCD without LDA	Random Forest	41.56	1.0000	1.0000	1.0000
DBCD with LDA	Random Forest	1.55	1.0000	1.0000	1.0000

4.2.3. Decision Tree

The RoC of Decision Tree without LDA and without LDA



The above ROC without LDA and bellow curves are with LDA

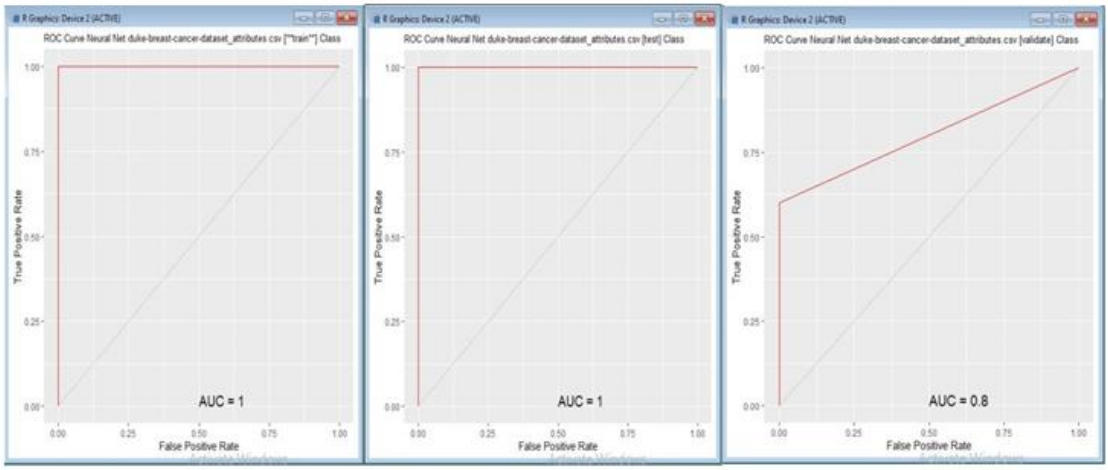


The Results of Decision Tree with LDA and without LDA on DBCD

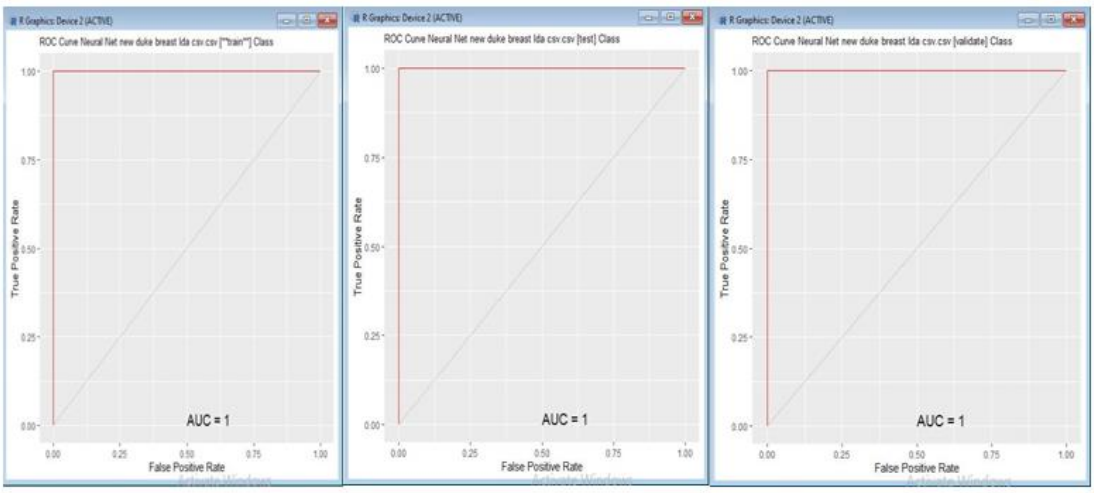
Dataset	Classification Model	Time taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
DBCD without LDA	Decision Tree	4.72	0.9944	0.0167	0.9286
DBCD with LDA	Decision Tree	0.17	0.9944	1.0000	1.0000

4.2.4. Neural Network Model

The RoC of NN without LDA and without LDA on DBCD



The above ROC without LDA and below are with LDA



The Results of Neural Network with LDA and without LDA on DBCD

Dataset	Classification Model	Time taken (Seconds)	Training (ROC)	Testing (ROC)	Validation (ROC)
DBCD without LDA	Neural Network	18.06	1.0000	1.000	0.80
DBCD with LDA	Neural Network	0.57	1.0000	1.0000	1.0000

These are the results of different machine learning techniques on duke breast cancer dataset with LDA feature selection and without feature selection.

5. CONCLUSION

In this thesis, two dataset are used for classification without feature selection and with feature selection. Two dataset are WBCD and Duke Breast Cancer Dataset. In first experiment different classification techniques like SVM, DT and RF are implemented without feature selection method. These models provide higher accuracy rate almost 1.0000 at training, testing and validation. But the Issue is that these entire models spend more time for execution. To reduce the time linear discriminant analysis feature selection is applied to select positive feature. After selecting these feature classification method implemented. This feature selection reduces more than half time. After developing this LDA model one more dataset is used to check the entire working of the model. Duke breast cancer dataset is used for second dataset. Similarly, as in previous experiment this dataset is used with feature selection and without feature selection. This dataset is more complex than previous one because this dataset have 7130 attributes. Its take a lot of times to load in rattle. Its structure is complex and taking a lot of time to run classification model. This dataset on these classification models performed well on training, testing and validation. But the problem is that it's taking a lot of time to execute. The proposed model is implemented on that dataset to select positive features. After selecting these feature 80% time is reduced in all classification model. Different feature selection and extraction can be implemented to improve the accuracy; also different classification models can be implemented. The best machine learning model is Neural Network model with LDA feature selection with accuracy of 1.000; also improve the validation accuracy of the model.

REFERENCES

[1] Maajani, K., et al., Survival rates of patients with breast cancer in the Eastern Mediterranean Region: a systematic review and meta-analysis. East Mediterr Health J, 2019. 25.

[2] Intlekofer, A.M. and L.W. Finley, Metabolic signatures of cancer cells and stem cells. Nature metabolism, 2019. 1(2): p. 177-188.

[3] Bravi, F., A. Decarli, and A.G. Russo, Risk factors for breast cancer in a cohort of mammographic screening program: a nested case–control study within the FR iCaM study. Cancer medicine, 2018. 7(5): p. 2145-2152.

[4] Beger, H.G., et al., New onset of diabetes and pancreatic exocrine insufficiency after pancreaticoduodenectomy for benign and malignant tumors. Annals of surgery, 2018. 267(2): p. 259-270.

[5] Gharibdousti, M.S., et al., Breast cancer diagnosis using feature extraction techniques with supervised and unsupervised classification algorithms. Applied Medical Informatics., 2019. 41(1): p. 40-52.

[6] Rani, R.R. and D. Ramyachitra. Krill herd optimization algorithm for cancer feature selection and random forest technique for classification. in 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). 2017. IEEE.

[7] Chaurasia, V., S. Pal, and B. Tiwari, Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 2018. 12(2): p. 119-126.

[8] Akay, M.F., Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 2009. 36(2): p. 3240-3247.

[9] Sahu, B., S. Mohanty, and S. Rout, A hybrid approach for breast cancer classification and diagnosis. EAI Endorsed Transactions on Scalable Information Systems, 2019. 6(20).

[10] Diz, J., G. Marreiros, and A. Freitas, Applying data mining techniques to improve breast cancer diagnosis. Journal of medical systems, 2016. 40(9): p. 1-7.

[11] Ponnuraja, C., et al., Decision tree classification and model evaluation for breast cancer survivability: A data mining approach. Biomed Pharmacol J, 2017. 10: p. 281-289.

[12] Ojha, U. and S. Goel. A study on prediction of breast cancer recurrence using data mining techniques. in 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence. 2017. IEEE.

[13] Aličković, E. and A. Subasi, Breast cancer diagnosis using GA feature selection and Rotation Forest. Neural Computing and Applications, 2017. 28(4): p. 753-763.

[14] Şahan, S., et al., A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. Computers in Biology and Medicine, 2007. 37(3): p. 415-423.

[15] Kaya, Y., A new intelligent classifier for breast cancer diagnosis based on a rough set and extreme learning machine: RS+ ELM. Turkish Journal of Electrical Engineering & Computer Sciences, 2013. 21(Sup. 1): p. 2079-2091.

[16] Okikiola, F., et al., Design and Implementation of a Fuzzy Expert System for Diagnosing Breast Cancer. Journal of Advances in Mathematics and Computer Science, 2019: p. 1-14.

[17] Abdolmaleki, P., L.D. Buadu, and H. Naderimansh, Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using artificial neural network. Cancer letters, 2001. 171(2): p. 183-191.

[18] Pena-Reyes, C.A. and M. Sipper, A fuzzy-genetic approach to breast cancer diagnosis. Artificial intelligence in medicine, 1999. 17(2): p. 131-155.

[19] Boughorbel, S., R. Al-Ali, and N. Elkum, Model comparison for breast cancer prognosis based on clinical data. PloS one, 2016. 11(1): p. e0146413.

[20] Saqib, P., et al. Hybrid of Filters and Genetic Algorithm-Random Forests Based Wrapper Approach for Feature Selection and Prediction. in Intelligent Computing-Proceedings of the Computing Conference. 2019. Springer.

[21] Carvalho, D., P.R. Pinheiro, and M.C.D. Pinheiro, A hybrid model to support the early diagnosis of breast cancer. Procedia Computer Science, 2016. 91: p. 927-934.

[22] Mushtaq, Z., et al., Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets. Journal of the Chinese Institute of Engineers, 2020. 43(1): p. 80-92.

[23] Alyami, R., et al. Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural network and support vector machines. in 2017 International Conference on Informatics, Health & Technology (ICIHT). 2017. IEEE.

[24] Sridevi, T. and A. Murugan, An intelligent classifier for breast cancer diagnosis based on K-Means clustering and rough set. International Journal of Computer Applications, 2014. 85(11).
[25] Nguyen, C., Y. Wang, and H.N. Nguyen, Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. 2013.
[26] Alzubaidi, A., et al. Breast cancer diagnosis using a hybrid genetic algorithm for feature selection based on mutual information. in 2016 International Conference on Interactive Technologies and Games (ITAG). 2016. IEEE.
[27] Arunadevi, J. and K. Ganeshamoorthi. Feature Selection Facilitated Classification For Breast Cancer Prediction. in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). 2019. IEEE.
[28] Ed-daoudy, A. and K. Maalmi, Breast cancer classification with reduced feature set using association rules and support vector machine. Network Modeling Analysis in Health Informatics and Bioinformatics, 2020. 9: p. 1-10.
[29] Assiri, A.S., S. Nazir, and S.A. Velastin, Breast tumor classification using an ensemble machine learning method. Journal of Imaging, 2020. 6(6): p. 39.
[30] Farid, A.A., G. Selim, and H. Khater, A Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm For Breast Cancer Detection. 2020.
[31] Tharwat, A., et al., Linear discriminant analysis: A detailed tutorial. AI Commun., 2017. 30: p. 169-190.