2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) | 978-1-6654-3736-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/AIKE52691.2021.00018

A Machine Learning Predictive Model to Classify Severity of Breast Cancer Based on Mammographic Mass Dataset

Mahmud Parvez,[1] Robert B. Eshun,[2] Marwan Bikdash,[2] A.K.M. Kamrul Islam[2†]
[1] Department of Industrial Engineering, University of Louisville, USA
[2] Department of Computational Data Science and Engineering, North Carolina A&T State University, USA
m0parv02@louisville.edu, rbeshun@aggies.ncat.edu, bikdash@ncat.edu, akislam@ncat.edu

Abstract— Mammography which is recognized as a both a diagnostic and screening tool, is an x-ray imaging method used to examine the breast for the early detection of cancer and other breast diseases. Though it's a useful technique for early cancer detection but sometimes its low predictive value resulted unnecessary biopsy of near about 70% benign case. There are also several Computer Aided Diagnostic (CAD) systems already have been developed to reduce this unnecessary breast biopsies. Hence, with this background, this project is directed towards developing a machine learning predictive model based on the "mammographic mass" dataset which contain BI-RADS assessment, the patient's age and three BI-RADS attributes and the ground truth for severity together with the ground truth (the severity field) for 516 benign and 445 malignant masses. Machine learning approach like data cleaning, data transformation, feature engineering and finally several prediction models based on their accuracy are applied. Among all of the prediction models, Linear Discriminant Analysis (LDA) represent highest accuracy of near about 84.50% with reduced attributes after removing missing values and outliers. Finally, Receiver Operating Characteristic (ROC) curve is developed for the finding the classification skill of the developed LDA model and the outcomes shows that the almost 91% area are remained under ROC curve which is really good enough to judge the accuracy of developed model.

Index Terms— Mammogram, Machine Learning (ML), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Receiver Operating Characteristic (ROC) curve**.**

I. INTRODUCTION

A. Background

Breast Cancer is one of the most common cancer diagnosed for women which unfortunately doesn't expose to many symptoms at the early stages. According to the American Cancer Society, more than 40 thousand women are died in breast cancer and among one of the three women having cancer are affected by this breast cancer [1]. Different tests can be used to look for and diagnose breast cancer and among them breast cancer screening with mammography is considered effective in reducing breast cancer–related mortality [2]. A mammogram is a low-dose x-ray imaging technique that allows doctors called radiologists to look for changes in breast tissue to detect breast cancer at early stage even it has the capability to identify cancer several years before physical symptoms are produced. Though mammography is the most effective and safest technique as it outweighs the harms of radiation exposure but sometimes it produces abnormal result. One of the most recent statistics has shown that more or less 5-10% the mammography results are not conspicuous or abnormal that further examination like ultrasound imaging or even lead breast biopsy as ultimate interpretation of the benign or normal breast tissue [1]. The number of breast unnecessary biopsies are as high as only 10%–30% of all breast biopsies actually show a malignant pathology [1][2] which not only causes the physical and mental stress for the patients but also, they have to face a huge amount of monetary loss. So, it is imperative to minimize misses and interpretation errors of visible lesions at digital mammography. To assist the physician for making better decision, several Computer-aided detection (CAD) systemswere introduced for radiologists. But, due to the low specificityof the conventional CAD systems, there is no noticeable improvement appeared in terms of cost-effectiveness of screening. So, the benefit of using CAD in screening is not stillobvious [3][4].

Much research has been done so far to accurately identifythe suspicious lesion seen in a mammogram based on the "Breast Imaging Reporting and Data System" (BIRADS). BIRADS provide the lesion description which helps the physician to make decision about whether to go for breast biopsy or short follow up diagnosis will be well enough. So, the accurate the classification of benign and malignant mass lesion has achieved the utmost importance for exact diagnostics of breast cancer and prevent unnecessary biopsy. Due to having unique critical feature detection advantages, Machine Learning(ML) would be the best choice for the classification and towards development of corresponding prediction model [5] from the complex breast cancer-based dataset. Hence, with thisbackground, in this project, ML approach is applied for developing a predictive model to classify severity (benign or malignant) of Breast Cancer (BC) based on Mammographic Mass Data Set.

B. Literature Review

Due to the advancement of ML, many ML techniques have been applied by the researcher to classify the benign from malignant tumors and to predict prognosis. Among all of the machine learning problems, Classification is a kind of complex optimization problem [3] which is widely applied in the medical field. The artificial neural network (ANNs), support vector machine (SVMs), decision tree (DTs) and k-nearest neighbor (k-NNs) techniques as they are the main methods used in Breast Cancer (BC) diagnosis and prognosis. Over the time there are several datasets has been introduced in the literature like Wisconsin breast cancer diagnosis (WBCD) Prognostic Breast Cancer Chemotherapy (WPBCC), Wisconsin Diagnostic Breast Cancer (WDBC) and so on. The ML techniques that have applied on that dataset for provide the accuracy ranged between 94.36% to 99.90% [3]. At the beginning of 2004, the two most popular machine learning

978-1-6654-3736-3/21/$31.00 ©2021 IEEE
DOI 10.1109/AIKE52691.2021.00018

classification methods like Decision Tree (DT) and Artificial Neural Network (ANN) were applied for predicting the survival rate of a dataset having more than 200000 cases and the result of ML methods were compared with the simple statistical linear regression method. The results revealed that both ML methods has the accuracy of 93.6% (DT) and 91.2% (ANN) whereas the linear regression end up with the accuracy of 89.2% [6][7]. In the early 90's, a three-layer feedforward ANN network back propagation algorithm was applied for predicting about lesion whether it's malignant or benign by experimenting mammography atlas of breast cancer from a database of 133 cases [8]-[10]. The result exposed significant improvement in accuracy by almost 3-5% than traditional statistical methods. Another ML methods Support Vector Machine (SVM) was applied for WBCD and almost 97.2% classification accuracy was reported [11]-[14]. A J48 decision tree method was applied to classify the WBCD, and this method shows the classification accuracy of 94.56% [15][16]. To have the advantage of smaller construction and higher predictive accuracy, a modified C4.5 decision tree was applied by Quinlan to almost 20 UCI database which also provides higher accuracy than the conventional methods [17]. One of the most widely used algorithm, k-NN have also number of applications in BC diagnosis and prognosis where the values of the "k" ranges between 1 to 15 and the achieved accuracy was almost 98% fortest and testing dataset [9][18].

Some of the past studies has also carried out their analysis on the same mammographic mass dataset, which is an open-source dataset plugged into the UCI machine learning repository. From the result, it was reported as ROC performance values achieved for the different classifiers such as for Decision Tree (DT) it was 0.838±0.017, Case-based reasoning it was 0.857±0.016 and for ANN approach it was about 0.847±0.017 [1].

II. METHODOLOGY

A. Dataset and Proposed Method:

The dataset used for this project is titled as mammographic mass dataset collected by the Institute of Radiology of the University Erlangen-Nuremberg between 2003 and 2006 and it is an open-source dataset which available for public at the well-known UCI machine learning repository. This data set is used in this project develop the prediction model to classify and predict the severity (benign or malignant) of a mammographic mass lesion from BI-RADS attributes and the age of the patients. The dataset contains 961 instances and 6 attributes: 6 (1 goal field, 1 non-predictive, 4 predictive attributes). The dataset also contains missing values in the form of "?" and "0" for some attributes. Among 6 attributes, 5 acts as input and the one which represent the class is the output or response variable. Five input variables are BI-RADS assessment: ranging from 1 (definitely benign) to 5 (highly suggestive of malignancy) (ordinal) , Age: patient's age in years (integer), Shape: mass shape: round=1 oval=2 lobular=3 irregular=4 (nominal), Margin: mass margin: circumscribed=1 micro lobulated=2 obscured=3, ill-defined=4 spiculated=5(nominal), Density: mass density high=1 iso=2 low=3 fat-containing=4 (ordinal) Severity: benign=0 or malignant=1 (binominal). Among these 961 observations in the dataset, 516 masses are identified as benign and 445 are identified as malignant on full field digital mammogram [10].

To reach the goal of this project the applied ML methods were categorized as data preprocessing, feature engineering, model building and evaluate the prediction. After preprocessing operation on the dataset, data visualization techniques are applied to compare the distribution of the dataset and also some of the popular machine learning algorithms like, Linear Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Decision Tree Classifier (DT), Gaussian Naïve Bayes (NB), Support Vector Machine (SVM) are used check the accuracy at every processing steps. A schematic overview of details methodology presented in fig. 1.

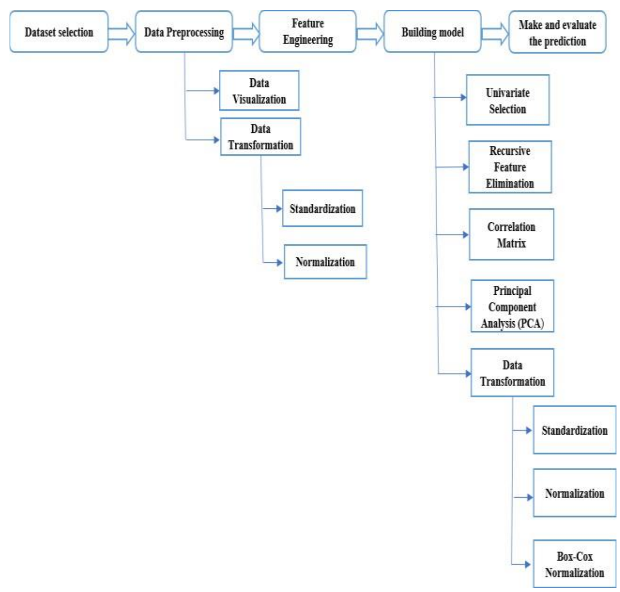

Fig. 1: Overview of Methodology

B. Data Preprocessing and Transformation:

Preprocessing is an essential step but as the assumptions about the data varies algorithm wise, so, sometimes different transformation is required for dataset. As the original dataset contains missing values (0) and outliers (data points beyond the ±3σ limits), data preprocessing is done for best representation of dataset towards building machine learning classification model. To do this for the very first step, missing values from each column were figured out which demonstrate in fig. 2. After finding the missing values two-

```
BIRADS        7
Age           5
Shape        31
Margin       48
Density      76
Severity      0
dtype: int64
```

Fig. 2. Missing values for each attribute

approaches were used. First one, was to remove all missing values rows and second one was to replace the missing values by corresponding mean of that particular columns. After cleaning the missing values, we were proceeded towards finding outliers an removing the rows contained outliers for each of the previous mentioned approach. From the output comparison fig. 3, it is observed that after removing missing values rows the dataset contains more outliers than after replacing the missing values with means.

Outliers after removing missing values	Columns with outliers: [0, 4] Rows with outliers: [340, 150, 151, 181, 288, 289, 364, 595, 814]
Outliers after replacing missing values with means	Columns with outliers: [0] Rows with outliers: [340]

Fig. 3. Index of outlier

As data visualization is an important skill in applied statistics and machine learning to have some qualitative gain of that dataset like identification of pattern, missing values and outliers and so and so forth. So, for better understating and comparison about the distribution of the dataset along with the detecting missing values and outliers several data visualization plots like Box and Whisker, Histogram, Scatter Plot matrix and Density were also introduced. The plots are compared pattern of that particular dataset for three different case like, original dataset, dataset after removing missing value rows and outliers and the dataset after replacing missing values with mean of each individual attributes with removing outliers.

Histogram is one of the data visualizations plots which grouped the data into bins to provides quick understanding about the distribution of the dataset, i.e, whether the corresponding attributes follows like, Gaussian, skewed or even an exponential distribution [11]. The histogram plot for each of the attributes for three above mentioned cases are demonstrated as in the following fig. 4. From the histogram, except **Age** and **BIRADS** all of the attributes show multiple peaks. **BIRADS** show a larger right tail and narrower peak also some accumulation of values near zero (0), which represents as missing values. It will be visualized better from corresponding density plot.

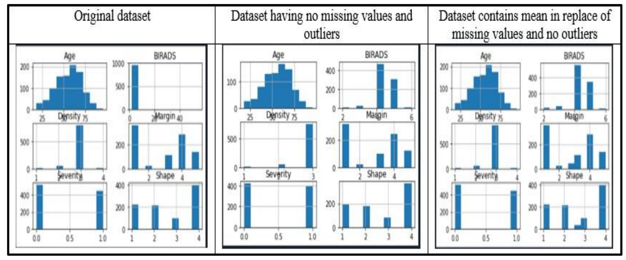

Fig. 4. Histogram Plot

After removing missing values and outliers, for the **BIRADS** attributes missing values are reduced but still there are multiple peaks for all of the attributes except **Age**. It can also be visualized better from the corresponding density plot. Almost same observation obtained for the outlier free dataset having mean instead of missing values.

Density plots are abstraction of the histogram with a smooth curve drawn through the top of each bin fig. 5 [11].

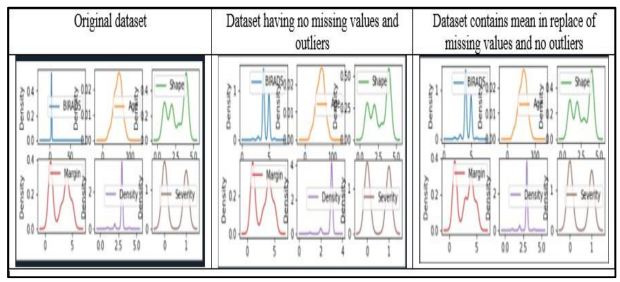

Fig. 5. Density Plot

From density plot, which attributes has narrower peak, having long tails also having small peak along with large peak that contains outliers and missing values. The **BIRADS** have narrower peak and have long right tails towards zero (0) represents the present of missing values. Other attributes like, **Shape, Margin and Density** also have multiple peaks and tails towards zero (0) indicates the present of missing values and outliers. After removing the missing values and outliers, not so much improvement can be figured out in terms missing values and outliers for all attributes except **BIRADS**, where the missing values (0) are reduced significantly. After replacing the missing values with means and removing outliers the visualization shows almost similar output like removing missing value rows.

Box and Whisker Plots is another data visualization plot, which summarizes the distribution of each attribute by drawing a line for the median (middle value) and a box around the 25th and 75th percentiles (the middle 50% of the data) [11]. The whiskers provide a qualitative estimate about the spread of the data and dots outside of the whiskers show candidate outlier values. The Box and Whisker plot for this dataset demonstrated in the fig. 6. From the Box and Whisker plot, for the original dataset, there represents some missing values in the form of zero (0) **BIRADS, Age and Density** columns. After removing missing values and outliers, number of missing values are reduced for **BIRADS, Age and Density.**

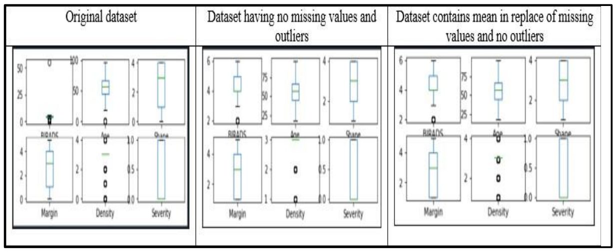

Fig. 6. Box and Whisker Plot

After replacing the missing values with means and removing outliers, number of missing values reductions for **BIRADS andAge** are almost same as the before steps and **Density** has moremissing values than before steps.

Scatter plot matrix is another way of data visualization which shows the relationship between two variables as dots in two dimensions, one axis for each attribute [11]. The Scatter plot matrix for this dataset demonstrated in the fig. 7.

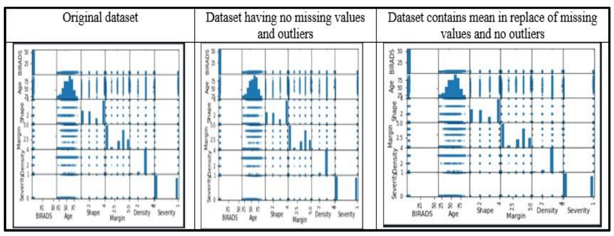

Fig. 7. Scatter plot matrix Plot

From scatter- plot, the data points that are far away from the accumulated data points are recognized as outliers and accumulation of data points near the missing values, i.e. zero (0) in this case, represent as the presence of missing values. From the above plots, it's hard to explicitly find out the presence of missing values and outliers. After removing missing values and outliers, not so much improvement can be observed and explained from here. Other plot also represents the same visualization as removing missing values with rows.

After data visualization the data transformation was done for each of all attributes. During this phase data transformationwas done by using standardization, normalization and Box-cox normalization approach. Standardizations, a well-recognized technique for transforming the attributes with a Gaussian distribution and differing means and standard deviations to a standard Gaussian distribution with a mean of 0 and a standard deviation of 1 [11]. Normalization is another rescaling technique, refers to rescaling an input variable to the range between 0 and 1 and this is pre-processing techniques can be useful for the dataset having lots of zeros (0) called as sparse dataset and also the dataset with attributes of varying scales. Box-Cox normalization which is a power transformation technique also applied for data transformation. It assumes the values of the input variable to which it is applied are **strictly positive** that implies, 0 and negative values are not supported. This transformation technique using the Power Transformer class and setting the "method" argument to "box-cox". After box cox normalization for each attributes the distribution looks bit more Gaussian than the original data.

At the end of the preprocessing stage, accuracies were compared between the original data and preprocessed data of above describing methods for each of the previously mentioned six machine learning spot checking algorithm and accuracy comparison is demonstrated in the table. 1.

Machine Learning Algorithm	Original dataset	Removing rows with missing values and outliers	Replacing missing values with mean and removing outliers	Accuracy after normalization of all attributes	Accuracy after standardization of all attributes	Accuracy of models after Box-cox Transformations of data
Logistic Regression	80.46%	82.34%	82.03%	82.96%	82.81%	81.51%
Linear Discriminant Analysis	78.90%	84.35%	83.71%	84.35%	84.35%	82.27%
K-Neighbors Classifier	79.03%	80.36%	79.03%	81.14%	81.29%	80.45%
Decision Tree Classifier	74.48%	75.76%	75.13%	76.07%	75.76%	76.96%
Gaussian Naïve Bayes	78.37%	83.73%	82.15%	83.73%	83.73%	80.15%
Support Vector Classifier	80.34%	80.67%	80.34%	82.50%	82.81%	82.87%

Table 1: Accuracy of ML algorithm after preprocessing

From the above result, it can be visualized that after preprocessing the dataset the accuracy significantly improved for all over the algorithms. Among all of the four preprocessing techniques, the dataset having no outliers and missing values, normalized and standardized dataset provides same highest accuracy of 84.35% for Linear Discriminant Analysis (LDA) model. So, we can choose either one among these three for the next feature engineering stages. For simplicity, this project proceeds with the dataset having no missing values and outliers.

C. Feature Engineering:

Feature engineering or feature selection is the process of selecting the most contributed features to the prediction or output response variable. It helps to improve the accuracy of the dataset by not only exposing and eliminating the least contributed features but also reduce the overfitting of the dataset. For this particular project different feature engineering techniques like, univariate selection, Recursive feature elimination, Principal Component Analysis (PCA) and Correlation matrix plots are used to rank the features. Feature engineering is done on missing values and outliers' free dataset. From the univariate feature selection technique, according to the fit scores of all attributes**, 'Age', 'Shape', 'Margin'** is figured out as most contributed features to the output variables. RFE identified three features are **'BIRADS', 'Shape', 'Margin'** and the least contributed features are **"Age"** and **"Density**. From the correlation matrix plot, fig.8, demonstrated the similar output like RFE where the **"Age" and "Density"** has lowest correlation with other attributes. For PCA, output of the reduced dimension of the dataset and explained variance is observed for both two and three principal components and it observed that three components are almost explain (0.983+0.013+0.002) or 99.8% and two components two components are almost explained (0.983+0.013) or 99.6% variance of this dataset.

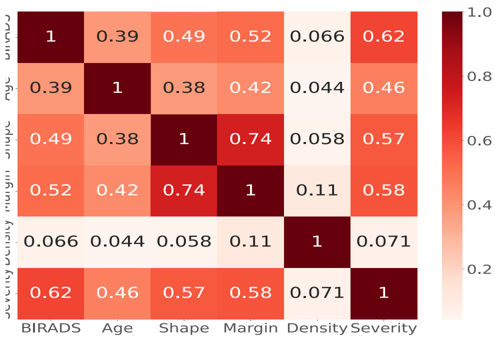

Fig. 8. Correlation Matrix

Finally, accuracies for each of the feature engineering techniques mentioned before are discovered and compared for each of the machine learning spot checking algorithms which listed as table 2.

Machine Learning Model	Original dataset	Removing rows with missing values and outliers	Accuracy after removing two least contributed features, Univariate	Accuracy after removing two least contributed features, RFE	Accuracy after removing two least contributed features, Correlation Matrix	Accuracy for three principal components	Accuracy for two principal components
Logistic Regression	80.46%	82.34%	80.97%	82.50%	82.50%	81.13%	81.28%
Linear Discriminant Analysis	78.90%	84.35%	80.81%	84.50%	84.50%	81.28%	80.51%
K-Neighbors Classifier	79.03%	80.36%	79.13%	82.05%	82.05%	79.13%	79.28%
Decision Tree Classifier	74.48%	75.76%	71.77%	82.97%	82.97%	72.69%	73.16%
Gaussian Naïve Bayes	78.37%	83.73%	80.05%	82.96%	82.96%	80.05%	80.20%
Support Vector Classifier	80.34%	80.67%	80.05%	83.73%	83.73%	80.51%	80.67%

Table 2: Accuracy of ML algorithm after feature engineering

To sum up, from the above table, the accuracy of the model after removing two least attributes **("Age" and "Density")** by using RFE and Correlation matrix plot, shows highest accuracy for almost all of the algorithms and LDA model gives higher accuracy compared to other models.

III. Model Building And Prediction

A. Select The Best Model:

As the dataset after reducing two least contributed attributes provides higher accuracy, so this reduced dataset is used for building ML model. To build the ML model, six machine learning algorithms are tested which are mentioned before such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB) Support Vector Machines (SVM). For selecting the best model, accuracies of each of the models are estimated and compared as listed in the following table 3.

Machine Learning Model	Accuracy of models with reduced dataset	Standard deviation of models with reduced dataset
Logistic Regression	82.50%	0.042
Linear Discriminant Analysis	84.50%	0.037
K-Neighbors Classifier	82.06%	0.036
Decision Tree Classifier	82.82%	0.052
Gaussian Naïve Bayes	82.96%	0.033
Support Vector Classifier	83.74%	0.039

Table 3: Accuracy of ML model

The above accuracy table demonstrates that, among all of the models, Linear Discriminant Analysis (LDA) has the largest estimated accuracy score at about 0.8450 or 84.50% and standard deviation is also reasonable.

B. Make and Evaluate Predictions:

Prediction can be accomplished by choosing an algorithm. As the results in the previous section suggest that the LDA was perhaps the most accurate model, so, this (LDA) model will use as final model for this dataset. To get the idea of the accuracy of the model on our validation set, it is necessary to carried out an independent final check on the accuracy of the best model. So, the predictions can be made on the validation dataset by fitting the LDA model on the entire training dataset.

Prediction can be evaluated the by comparing them to the expected results in the validation set, then calculate classification accuracy, as well as a confusion matrix and a classification report presented in the following figure. 9

```
[816 rows x 4 columns]
0.7865853658536586
[[69 22]
 [13 60]]
              precision    recall  f1-score   support

         0.0       0.84      0.76      0.80        91
         1.0       0.73      0.82      0.77        73

    accuracy                           0.79       164
   macro avg       0.79      0.79      0.79       164
weighted avg       0.79      0.79      0.79       164
```

Fig. 9. Confusion Matrix

From the above output, it can be seen that the accuracy is about

0.7865 or 78.66% dataset on the hold out dataset for selected LDA model. The confusion matrix provides an indication of the errors made. Finally, the classification report provides a breakdown of each class by precision, recall, f1-score, and support showing modest results so far as the validation dataset is small enough.

C. Receiver Operating Characteristics Curve (ROC)

In Machine Learning, to compare models that predict probabilities for two-class problems, i.e., when predicting the probability of a binary outcome, the Receiver Operating Characteristic curve or ROC curve is used in general. It is a plot of the false positive rate (x-axis) versus the true positive rate (y-axis) for a number of different candidate threshold values between 0.0 and 1.0. A skillful model will assign a higher probability to a randomly chosen real positive occurrence than a negative occurrence on average. This is mean is that skillful models are represented by curves that bow up to the top left of the plot. The ROC plot for our model is represented in the fig.10 for original dataset and attributes reduced dataset.

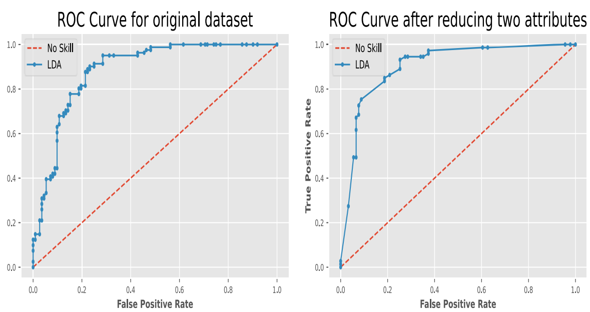

Fig. 10. Area under ROC Curve Plot for a No Skill Classifier and Linear Discriminant Analysis model

For the original dataset, the area under the ROC was 88.5% but after using machine learning process the area under the ROC curve increased significantly to (90.4%) which indicates that the developed LDA model is skill enough to accurately predict and classify the outcome of lesion about whether it will be benign (1) or malignant (0) and this will further help the physician to make decision about the necessity of surgery for that particular case.

IV. Results And Discussions

This project is based on the several steps and accuracies are found for almost all of the steps by using six ML spot checking algorithm. All of the accuracies which listed on the tables are coherence kappa accuracy. Firstly, accuracy of the original dataset was discovered, and Logistic Regression provides highest accuracy of 80.46%. After that accuracy was compared for different preprocessing states removing the missing value rows and removing outliers, replacing the missing values with means and removing the outliers. Missing values and outliers' free data provide the highest accuracy of 84.35% for Linear Discriminant Analysis (LDA). During the data transformation box-cox normalization techniques provides lowest accuracy for LDA model of 82.27% compared to other two data transformation techniques such as Normalization(84.35%) and also standardization (84.35%).

During the feature engineering phase accuracies are compared for different techniques. Univariate feature selection provides the highest accuracy of 80.81% for LDA after reducing two least contributed attributes. The dimensionality reduction by using two and three Principal Component Analysis provides the highest accuracy of 80.51% and 81.28% respectively for LDA model. However, RFE and Correlation matrix plot provides the similar output about least contributed attributes like **"Age" and "Density"** and provide the overall highest accuracy of 84.50% for LDA model and this LDA is selected as the best ML prediction model for classifying the severity of the mammographic mass lesion. Confusion matrix provides the accuracy of 0.7865 or 78.66% on the hold out dataset and the classification report provide a breakdown of each class by precision, recall, f1-score and support showing modest results so far as the validation dataset is small enough. Finally, analysis, the area under the ROC curve shows 90.4% for developed LDA model which represents the competency of developed LDA model.

V. Conclusion

This projected presented a Machine Learning classification model based on BI-RADS assessment, the patient's age and three BI-RADS attributes for predicting the severity of mammographic mass lesion to assist the physician regarding making decision about whether to go for biopsy or not. The developed LDA machine learning model provides the highest performance value of 90.4% compared to the other past studies done on the same dataset which were 84.7% for Artificial Neural Network (ANN), 85.7% for Case-Based Reasoning (CBR) method. Though our developed model provides modest accuracy, but it might be possible to improve more by including more attributes as well as data preprocessing techniques.

References

[1] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process" Medical Physics 34, 4164 (2007); doi: 10.1118/1.2786864

[2] R. Alejandro et al., "Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System" Radiology 2019; 00:1–10, https://doi.org/10.1148/radiol.2018181371.

[3] Y. Wenbin et al., “Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis”, Designs 2018, 2, 13; doi:10.3390/designs2020013

[4] Delen, D.; Walker, G.; Kadam, A., “Predicting breast cancer survivability: A comparison of three data mining methods”, Artif. Intell. Med. **2005**, 34, 113–127. [CrossRef] [PubMed]

[5] Wu, Y.; Giger,M.L.; Doi, K.; Vyborny, C.J.; Schmidt, R.A.;Metz, C.E. “Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer”, Radiology **1993**, 187, 81–87. [CrossRef] [PubMed]

[6] Bennett, K.P.; Blue, J.A., “A support vector machine approach to decision trees”, In Proceedings of the IEEE International Joint Conference on Neural Networks, Anchorage, AK, USA, 4–9 May 1998; Volume 3, pp. 2396–2401.

[7] Witten, I.H.; Frank, E.; Trigg, L.E.; Hall, M.A.; Holmes, G.; Cunningham, S.J. “Weka: Practical Machine Learning Tools and Techniques with JavaImplementations”. https://researchcommonswaikato.ac.nz/handle/10289/1040 (accessed on 10 January 2018).

[8] Quinlan, J.R. “Improved use of continuous attributes in C4.5”. J. Artif. Intell. Res. 1996, 4, 77–90

[9] Bagui, S.C.; Bagui, S.; Pal, K.; Pal, N.R. “Breast cancer detection using rank nearest neighbor classification rules. Pattern Recognition”, 2003, 36, 25–34. [CrossRef]

[10] UCI Machine Learning Repository (n.d.). MammographicMass Data Set. Available: https://archive.ics.uci.edu/ml/datasets/Mammographic%20 Mass

[11] J. Brownlee, “Machine learning mastery with Python: understand your data, create accurate models and workprojects end-to-end”, Melbourne, Australia: Jason Brownlee, 2016.

[12] Eshun, Robert; Islam, AKM Kamrul; Bikdash, Marwan, “Identification of Significantly Expressed Gene Mutations for Automated Classification of Benign and Malignant Prostate Cancer”, IEEE EMBC conference, 2021.

[13] Rabby, MKM; Islam, AKM Kamrul; Bikdash; Marwan, Belkasim, Saeid, “Wavelet Transform-based Feature Extraction Approach for Epileptic Seizure Classification”, ACM Southeast conference, 2021

[14] Eshun, Robert; Rabby, MKM; Islam, AKM Kamrul; Bikdash, Marwan, “Histological Classification of Non-small Cell Lung Cancer with RNA-seq Data Using Machine Learning Models”, ACM BCB conference, 2021.

[15] Rabby, MKM; Islam, AKM Kamrul; Bikdash; Marwan, Belkasim, Saeid, “Epileptic Seizures Classification in EEG using PCA based Genetic Algorithm through Machine Learning”, ACM Southeast conference, 2021.

[16] Islam, AKM Kamrul; Belkasim, Saeid, “Ensemble of SVM for Colorectal Cancer Classification from Microarray Gene Expression Data ”, ISBRA conference, 2020.

[17] Akter, Morium; Rahman, Atiqur; Islam, AKM Kamrul, “An Improved Method of Automatic Exudates Detection in Retinal Image”, International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), Volume 2, Issue 5, 2014

[18] Wu, Jiande; Hicks, Chindo, “Breast Cancer Type Classification using Machine Learning”, MDPI Journal of Personalized Medicine, 2021.