PLATFORM - A Journal of Science & Technology

Received: 9 March 2021, Accepted: 27 April 2021, Published: 31 May 2021, Publisher: UTP Press, Creative Commons: CC BY-NC-ND 4.0

CLASSIFICATION OF BREAST CANCER FROM ELECTRICAL IMPEDANCE MEASUREMENTS DATASET IN SAMPLES OF FRESHLY EXCISED BREAST TISSUES

Pranav Verma[1], Sasikala Ramasamy[1*], Dhanapal Durai Dominic Panneer Selvam[2]

[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
[2]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia

*Email: sasikala.ra@vit.ac.in

ABSTRACT

The breast cancerous issues are increasing day by day for various reasons and even it can be found in young women. Early detection decreases the death rate and reduces painful medical treatments such as surgery and chemotherapy. Electrical Impedance Spectroscopy is a powerful and painless, low-cost detection technique, and it can be used with Mammography and MRI scans. The paper analyses the EIS dataset for classifying freshly exercised breast tissues using four different machine learning algorithms: Support Vector Machine, Decision Tree, Random Forest and Modified Random Forest. The results are verified with ANOVA statistical models on four different accuracy results of six classes. The results proved that Modified Random Forest (MRF) works best as providing an accuracy mean value as 99% on 106 datasets with 15% as testing size during the training phase. The one way ANOVA results are also proved that standard error of 0.005, the significance of 0.003 and covariance of 1.14 for MRF.

Keywords: *Machine learning, decision tree, random forest, support vector method, modified random forest*

INTRODUCTION

Breast cancer has a strong propagating nature as it spreads from a set of cells, and as a result of growth, it affects the adjacent tissues. In general, cancer disease spreads as a set of cells exhibit unrestrained growth and drawing adjacent tissues. Since detecting the disease and tissue causing the ailment is very deadly in nature, more concentration towards diagnosis is to be done early [1], as it is a more sensitive disease. Breast cancer is of a type that stems from breast tissue. Breast tissue is an assembly of tissues which is tied to nerves and blood vessels. These tissues are called adipose tissues. There may exist a probability of the existence of other tissue features like fibro-adenoma, mastopathy etc., which can be a result of developing/spread of disease to a more complex extent. Breast cancer can be diagnosed by various ways, such as Mammograms, ultrasound, biopsy, lab tests and MRI. The biopsy is a promising diagnostic procedure that can determine the suspicious area is cancerous [2]. Breast cancer disease is diagnosed by biopsy of tissue samples taken from the breast. For early detection of breast cancer, X-ray-based mammography is used, which is most painful and uncomfortable, damaging the breast tissues.

To avoid inaccuracy and human error, interdisciplinary computer-aided diagnosis makes diagnosis easy and fast [2]. Also, computer-based data analysis produce more effective results in early diagnosis such as stopping disease spreading in early-stage and saves human life. Electrical impedance tomography is an alternative methodology that makes women more comfortable for testing. Also, the system is affordable, which allow widespread use in developing countries. The tissue sample taken from suspected tissues is made to undergo electrical impedance spectroscopy, to calculate complex impedance. By repeating EIS for various factors like water content, cell packing density, and impedance, breast cancer disease can be diagnosed. Through machine learning algorithms [3], EIS is more effective in diagnosis procedure.

LITERATURE SURVEY

A survey on various papers of EIS methodology on breast cancer indications shows those statistical techniques still used for classifying breast tissues of EIS data. Also, ANN-based fully automated computer-based techniques [4] reduce false positive and false negatives rates and enhance accuracy, sensitivity, and specificity values. Reviews on EIS also confirms that it can be used additionally with Mammogram and Ultra scan-based diagnosis techniques [5]. Various papers results are analysed in the following Table 1.

Problem Definition

Many reviews disclosed the significance of EIS on breast cancer diagnosis. Statistical models, Neural Networks and machine learning algorithms on classifying EIS dataset are enhancing clinical diagnosis. However Statistical techniques provides maximum success rate of 80%. The developed classification system based neural networks (BPNN and RBFN) [6] showed overall classification accuracies as obtained from BPNNs and RBFNs are 93.39% for BPNN3 and 94.33% for RBFN2. Improving the accuracy results further and speeding up the process faster, proved machine learning algorithms such as Decision Tree [7], Random Forest techniques are a better option. The accuracy results can also be proved with statistical models. Hence the proposed models apply and compare machine learning algorithms on EIS dataset. The results are proved with one way ANOVA test results.

MACHINE LEARNING ALGORITHMS

For automatic classification of breast tissues, mostly machine learning algorithms are used. EIS is used for data acquisition from breast tissues and on EIS data machines learning algorithms are used for classification, avoiding human error and enhancing diagnosis procedure with higher accuracy. Halter et al. [6] used Back propagation neural network and radial basis function neural networks (RBFN) ANN techniques used for classification [6]. The experimental results shows that RBFN outperforms BPNN techniques in term of accuracy and time complexity. The proposed work demonstrates various machine learning algorithms such as Decision tree, Support vector machine, Random forest and modified random forest for classification of EIS data on breast tissues. For the classification of the exact tissue in the breast tissue, the proposed machine learning method performed well and classified different breast tissue accurately. We have employed Random Forest, Support Vector Method, Decision Tree and Modified Random Forest (varying hyper parameters) and Comparison was conducted after making confusion matrix and performing Oneway ANOVA test.

Decision Tree Algorithm

Decision Tree algorithm is a rule based tree structured algorithm with nodes, branches and leaves. Nodes of a decision Tree denote the attributes in a dataset, branches represent the decision rules and leaves denote the class values. When the rules are formed with respect to Gini Index, the Tree is known as CART decision Tree [7]. ID3 decision is built on Information gain and entropy metrics. A classification method which is non parametric.

Understanding the Algorithm

To conclude a best solution, outcomes of various occurrences are considered. Based on those various occurrences the best or optimal solution is intended. This is the working rule of Decision Tree Algorithm. In this instance the various occurrences of an event happening is considered as an input and those inputs are used to create a set of rules and finally those created rules are used in the calculation of a perfect solution.

Consider a real-time scenario involving persons:

A and B are friends and 'A' asks 'B' regarding joining a higher study course in a reputed institution. For mapping the above idea here, the approach becomes the following:

1. To help A, B asked some set of questions to A.
2. A replied with the answers for those question sets.
3. B recorded those answers as 'Rules'
4. Based on those rules (ie) answers obtained ,'B' helped 'A' to chose a best institution.

Thus in the above instance, B is the Decision Tree, the place which B Recommeded to A is the leaf of the decision tree (Solution).Thus here the solution is obtained from a single decision tree. There are many different type of Decision tree algorithms ranging

Table 1 Analysis on EIS based classification of Breast Tissues

Title	Author	Technology or Method used	Objective	Techniques and Algorithms used	No. of cases	Results and conclusion	Year of publication
Electrical Impedance Of Breast's Tissue Classification By Using Bootstrap Aggregating	Narumol Chumuang; Patiyuth Pramkeaw; Adil Farooq	electrical impedance of tissue from the breast	Bootstrap Aggregating	10-folds cross validation	106 objects	the result is that the accuracy of 74.47%.	2019 [1]
Electrical Impedance Mammography: the key to low-cost, portable and non-invasive breast cancer screening?	Sebu, Cristiana	EIT	Study on key success of EIT in breast cancer detection	Study work	-	Portable, low cost, little patient discomfort, non-invasive	2017 [2]
Phantom experiments using soft-prior regularization EIT for breast cancer imaging	Murphy EK, Mahara A, Wu X, Halter RJ	soft-prior regularization - EIT	Assist to reconstruct and tumour	Soft-prior	unknows	reduce false positives	2017 [3]
Electrical impedance spectroscopy for breast cancer diagnosis: Clinical study	Z. Haeri, M. Shokoufi, M. Jenab, R. Janzen, F.Golnaraghi	EIS-Probe and the EIS-Hand-Breast (EIS-HB)	Minimal invasive techniques	least square method (LSM), and least absolute deviation (LAD) method	10	sensitivity of 77.8% in EIS Probe and EIS-Hand-Breast	2016 [4]
A Review on Breast Electrical Impedance Tomography Clinical Accuracy	Norhayati Mohd Zain, Kanaga Kumari	Review on EIT	Clinical Accuracy Sensitivity and Specifity	study	10 Full length papers	Can be used as adjunct screening with MG and USG	2015 [5]
Real-Time Electrical Impedance Variations in Women With and Without Breast Cancer	Ryan J Halter, Alex Hartov, Steven P Poplack, Roberta diFlorio-Alexander, Wendy A Wells, Kari M Rosenkranz, Richard J Barth, Peter A Kaufman, Keith D Paulsen	high-speed electrical impedance tomography (EIT)	to differentiate malignant from benign regions within the breast	Statistics	19	Sensitivity 77% and specifity 81%	2015 [6]

Table 1 Analysis on EIS based classification of Breast Tissues (*Cont.*)

Title	Author	Technology or Method used	Objective	Techniques and Algorithms used	No. of cases	Results and conclusion	Year of publication
A 4.9mΩ- Sensitivity Mobile Electrical Impedance Tomography IC for Early Breast-Cancer Detection System	Sunjoo Hong; Kwonjoon Lee; Unsoo Ha; Hyunki Kim; Yongsu Lee; Youchang Kim; Hoi-Jun Yoo	Mobile based EIT IC	Convenient and Compact breast cancer detection system fabricated by P-FCB	weighted back-projection algorithm	-	breast cancer can be detected at home in early stage	2014 [7]
Breast Imaging Using Electrical Impedance Tomography: Correlation of Quantitative Assessment with Visual Interpretation	Norhayati Mohd Zain, Kanaga Kumari Chelliah	EIT	to correlate the quantitative assessment and visual interpretation	ANOVA t-test	150	Quantitative assessment of electrical impedance tomography was significantly related with visual interpretation of images of the breast ($p<0.05$)	2014 [8]
Breast Imaging using 3D Electrical Impedence Tomography	Sachin N Prasad, Dana Houserkova, Jan Campbell	3D EIT	efficiency of diagnosis using 3D Electrical Impedance Tomography is compared with Mammography Ultrasonography (USG)	Statistics	88	Cases involving cysts, EIT has 100% sensitivity where MG and USG have 81%	2008 [9]
Classification of breast tissue by electrical impedance spectroscopy	J. Estrela da Silva, J. P. Marques de Sá & J. Jossinet	Electrical impedance spectroscopy	Minimal invasive technique	Statistical techniques	106 cases	classification efficiency of ~92% with carcinoma discrimination > 86%	2000 [10]

from ID3, C4.5, C5.0 and CART i.e. Classification and Regression tree, here we are using an optimized version of CART present in scikit-learn library of python. CART produces trees depending on the dependent variable which can be numerical or categorical. Based on the data variables in the dataset the decision tress formation takes place, these rules are defines based on the variables values to find the best split when it is notices that no further gain can be made then the algorithm splitting stops.

Basically in a decision tree algorithm, a tree like structure is formed where leaves are the guess that a particular variable represent *'c'* or *'y'*. The path constructed from the root to the leaves form the classification rules. It can also have multiple conditions to put on the variables. It is also a non parametric method. The Decision Tree Classifier in Scikit-learn can perform Multi-Class Classification on a dataset. After the fitting of the data the model can predict class of samples. Decision tree is a type of supervised machine learning algorithm where a parameter is used to constantly split the data. The leaves gives the decision of the final outcome.

Decision Tree algorithm is applied to classify freshly exercised breast tissues EIS dataset in the range of 488 Hz to 1 Mhz into 6 classes of breast tissues: carcinoma, fibro-adenoma, mastopathy, glandular, connective, or adipose. The dataset is taken from publically available UCI machine Learning Repository. The below python snippets from Scikit-Learn, a python machine learning library split the proportion of test data set size as 0.15 and invokes DecisionTreeClassifier. The accuracy score of decision tree algorithm is 0.933 and confusion matrix for 6 classes are printed.

x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.15)

dt = DecisionTreeClassifier()

dt.fit(x_train, y_train)

print dt.score(x, y)

y_pred = dt.predict(x)

print confusion_matrix(y,y_pred)

target_names = ['class 1', 'class 2', 'class 3','class 4', 'class 5', 'class 6']

print(classification_report(y, y_pred, target_names= target_names))

The confusion matrix is an error matrix in supervised learning algorithms. The rows are represented as instances of predicted classes (6), and columns represented as actual class instances. Here the decision tree algorithm predicts class 1 and 2 correctly. However, out of 14 instances, 13 are predicted correctly as class 3, 1 as wrongly predicted. On the whole, the Decision tree algorithm, class 4 – 2, class 5 – 3 and class – 1 are wrongly predicted. Hence the accuracy score would be 0.94. The other derivations (*precision, recall, F1 score, support*) from the confusion matrix are mentioned below:

$$Precision\ ratio = t_p / (t_p + f_p)$$

Where t_p is number of true positives and f_p is number of false positives,

$$recall = t_p / (t_p + f_n)$$

where f_n is false negatives.

F1 score is harmonic means of recall and precision. The *F1* score is between 1 and 0 where 1 is best score and 0 is worst score. The support is the number of occurrences of each class in *y_true*.

The dataset was taken from UCI Machine Learning Repository for training and testing. The dataset has 106 instances of electrical impedance measurements of freshly excised breast tissues. The dataset has 10 attributes and the breast tissues are classified as Car-Carcinoma(class 1), Fad-Fibro-adenoma(class 2), Mas-Mastopathy(class 3), Gla-Glandular(class 4), Con-Connective(class 5) and Adi-Adipose(class 6). The data set can be used for predicting the above classes with the testing data size as 15% out of 106 instances. The accuracy score of the decision tree algorithm in Figure 1 is 0.933, and the above results show that the decision tree classifies on breast tissues provides precision as 0.94, recall 0.93, f1 score 0.93 and support value as 106.

Support Vector Machine

Support vector machine is a function based classifier that aims to separate classes with a hyper plane divider [8]. If a line function is used as hyperplane,

```
0.9339622641509434
[[22  0  0  0  0  0]
 [ 0 21  0  0  0  0]
 [ 1  0 13  0  0  0]
 [ 0  1  0 13  0  1]
 [ 0  0  0  0 13  3]
 [ 0  1  0  0  0 17]]
             precision    recall  f1-score   support

    class 1       0.96      1.00      0.98        22
    class 2       0.91      1.00      0.95        21
    class 3       1.00      0.93      0.96        14
    class 4       1.00      0.87      0.93        15
    class 5       1.00      0.81      0.90        16
    class 6       0.81      0.94      0.87        18

avg / total       0.94      0.93      0.93       106

Pranavs-MacBook-Pro:desktop pranav$
```

Figure 1 Decision Tree performance score

then the SVM is known as linear SVM. There are various other kernel functions available in SVM to perform multi-class classification task. In SVM hyper planes are constructed and the classification takes place. First the data is arranged in two dimension then particularly a smart hyper plane is constructed which can differentiate the both the classes very well by the algorithm separating the data for Svc classification. A good classification is produced by a hyper plane that has the distance as largest. It is a very good machine learning algorithm used in classification as well as regression. Different axis can be taken in account according to the particular data in the data set. SVM is a supervised learning method which we are using here for classification. Using svc class we perform multi class classification of the dataset. This classification is handled using one - vs - one scheme.

A hyper plane is constructed by svm or a set of hyper-planes in an infinite or high dimensional space, which we are mainly using for classification here. A hyper plane that has distance as largest produce a *good separation.*

Understanding SVM

Like Decision Trees and Random Vector, SVM too can be induced for both Regression and classificational Purposes. SVM uses the ideology of finding a subspace that divides a data into two class modules. Consider the diagram below:

The major line in Figure 2, which divides datasets, is called Hyper planes or subspace and the points

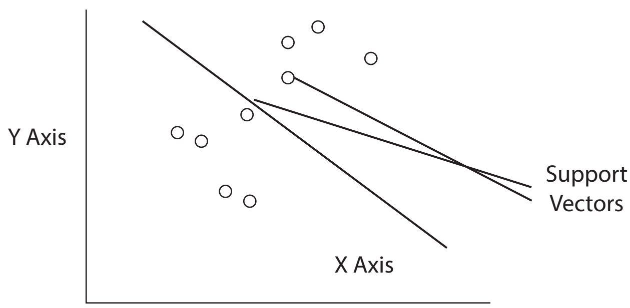

Figure 2 Data set classification

(or) data that lie closer to subspace is called Support Vectors. The selection of best support vectors is based on the identification of hyperplane or subspaces. It's tedious to select a perfect hyper plane for more complex situations where data is nonseparable.

In such cases, the representation is switched into a 3D view as below:

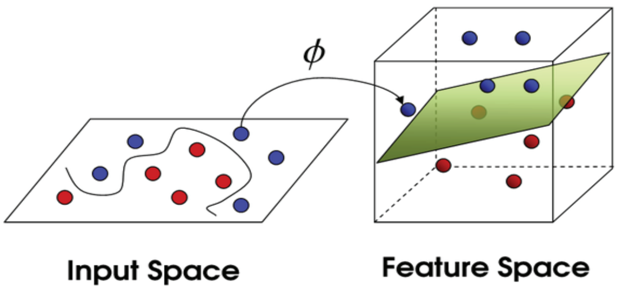

Figure 3 3D view

Since the representation in Figure 3 is mapped into 3D, here the line which separated the data (subspace) here becomes the plane. Its easier to plot the points or data sets that are closer as the data is in plane. In large data sets, what far the points lie from the space, the classes assigned to them decide and precedes the place where hyper plane is to be chosen and how two data sets or points are related to eachother. Thus SVM algorithm here is used to explain the instance of classification of EIS Breast tissue data. The python snippet calculates SVM score, confusion matrix, prediction, recall, F1 score and support.

```
ds = pd.read_csv('./BreastTissue.csv')
ds.head()
x = ds.drop(columns=['Class','Case #'],axis=1)
x.head()
y = ds['Class']
y.head()
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.15)
svc = svm.SVC()
svc.fit(x_train,y_train)
print svc.score(x,y)
y_pred = svc.predict(x)
print confusion_matrix(y,y_pred)
target_names = ['class 1', 'class 2', 'class 3','class 4', 'class 5', 'class 6']
print(classification_report(y, y_pred, target_names=target_names))
```

The results show that precision as 0.92, recall 0.87, f1 score 0.88 and support as 106. Hence the SVM classifier accuracy rates are low compare to the Decision tree algorithm. The confusion matrix identifies that SVM prediction is a little lower than the Decision tree algorithm. For example, class 1 is truly classified as class 1 for 18 cases and falsely classified as 4 cases. But in the decision tree, it is class 1 is classified as class 1 for all 22 cases.

Random Forest Algorithm

Random Forest is an ensemble machine learning algorithm as shown in Figure 4 [9]. As the name suggests, it is a forest of random decision tree models built on a dataset. If a Random Forest classifier contains n decision trees, then n different predictions are obtained. Based on the voting mechanism, the prediction class that obtained maximum frequency is returned as output. The random forest algorithm used is a further extension of decision tree and work for some complex dataset, here the algorithm first makes some random sets from the dataset. Then it forms its own decision tree for each of the data set.

```
Pranavs-MacBook-Pro:desktop pranav$ python /Users/pranav/Desktop/projectx.py
0.8679245283018868
[[18  4  0  0  0  0]
 [ 0 21  0  0  0  0]
 [ 0  4 10  0  0  0]
 [ 0  2  0 13  0  0]
 [ 0  1  0  0 15  0]
 [ 0  3  0  0  0 15]]
             precision    recall  f1-score   support

    class 1       1.00      0.82      0.90        22
    class 2       0.60      1.00      0.75        21
    class 3       1.00      0.71      0.83        14
    class 4       1.00      0.87      0.93        15
    class 5       1.00      0.94      0.97        16
    class 6       1.00      0.83      0.91        18

avg / total       0.92      0.87      0.88       106

Pranavs-MacBook-Pro:desktop pranav$
```

Figure 4 SVM performance scores

Finally, voting is conducted and the best possible outcome is selected, and a prediction is made.

Understanding Random Forest Algorithm

A meta estimator that fits a various number of decision trees classifier on sub-samples of the dataset and improve the predictive accuracy by averaging and controls over-fitting. The original input sample size is always the same as sub-sample size. N-estimators are the number of trees in the forest. Each tree is formed by a sample which is a bootstrap sample, drawn from the training set. The working of the random forest algorithm is as follows:

1. It selects random samples from the dataset. Then it forms a decision tree from each sample and get a prediction from each decision tree.
2. Then performing a vote for each predicted result. The most voted prediction result is selected.

Random forest algorithm also can be explained by the above strategy. Consider the same scenario. 'A' asks several of his friends to help him out to join in a reputed institution. Various friends asked 'A' various questions (about interests, domain field etc) and they suggested various recommendations to 'A'. Then, A calculates the recommendations he received from various sources. Based on the maximum number of count he will select the institution. For example if maximum number of people suggests 'x' institution, then he will select 'x'. Hence, the solution obtained is being decided by various members. Each friend is a tree and various of those trees will form a forest.

A model pesudocode:
1. Stocastially select 'r' values from 's' values (where r < s).
2. Among "r" values calculate the node "d" using the best point of split.
3. Split the main node into sub nodes.
4. Repeat the above steps and build a forest for "n" numberof trees by repeating the process for n times.

Modified Random Forest

By changing hyper parameters specifically N estimator gives the best result for the classification. By default scikit-learn use N-estimator 10 they are the collection of fitted sub-estimators. N estimator = 15, the number of trees fixed to 15 gives the best results of all the random forest. Modified random forest in this particular approach listed here the n_estimators in the algorithm of random forest listed above some changes were made and the best change which yielded the best value consistently was fixing the n_estimators (i.e. the number of the trees which are also the hyper parameters to 15) by default the skikit-learn uses 10. After fixing this value and applying the random forest algorithm maximum accuracy was obtained.

rf = RandomForestClassifier()

dt = RandomForestClassifier(n_estimators=15)

ONEWAY ANOVA RESULTS AND DISCUSSIONS

One-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups [10]. ANOVA is also useful for comparing more than two groups for statistical significance. Comparing more than one group on one factor using Anova is called one way _ANOVA. Four machine learning algorithms accuracy mean values are taken for comparison: SVM, Decision Tree, Random Forest and Modified Random Forest.

The below ANOVA results generates descriptive statistics such as Mean, STD deviation and STD error etc. Standard error calculates the measure of the precision in which sample mean approximates the true mean value. From the Table 2, the standard error indicates that MRF algorithm produces better prediction than other algorithms. Confidence interval gives the practical significance of the results. Here the more the interval, the higher the confidence. Mean square is a sum of squares divided by df and F value is ratio of mean squares.

Table 3 shows that the significance is 0.003, far less than 0.05, which is the standard significance. Hence we can reject the null hypothesis, and we can conclude that not all the population means are equal. When variance is not equal, Welch statistics and Brown Forsythe statistics are more reliable than F Tests in ANOVA. Hence Welch and Brown Forsythe tests are conducted and results are taken.

```
Pranavs-MacBook-Pro:desktop pranav$ python /Users/pranav/Desktop/projectx.py
0.9716981132075472
[[22  0  0  0  0  0]
 [ 0 20  0  0  0  1]
 [ 1  0 13  0  0  0]
 [ 0  0  0 15  0  0]
 [ 0  0  0  0 16  0]
 [ 0  1  0  0  0 17]]
             precision    recall  f1-score   support

    class 1       0.96      1.00      0.98        22
    class 2       0.95      0.95      0.95        21
    class 3       1.00      0.93      0.96        14
    class 4       1.00      1.00      1.00        15
    class 5       1.00      1.00      1.00        16
    class 6       0.94      0.94      0.94        18

avg / total       0.97      0.97      0.97       106
```

```
Pranavs-MacBook-Pro:desktop pranav$ python /Users/pranav/Desktop/projectx.py
0.9905660377358491
[[21  0  1  0  0  0]
 [ 0 21  0  0  0  0]
 [ 0  0 14  0  0  0]
 [ 0  0  0 15  0  0]
 [ 0  0  0  0 16  0]
 [ 0  0  0  0  0 18]]
             precision    recall  f1-score   support

    class 1       1.00      0.95      0.98        22
    class 2       1.00      1.00      1.00        21
    class 3       0.93      1.00      0.97        14
    class 4       1.00      1.00      1.00        15
    class 5       1.00      1.00      1.00        16
    class 6       1.00      1.00      1.00        18

avg / total       0.99      0.99      0.99       106
```

Figure 5 Modified Random Forest performance scores

Table 2 ANOVA results: descriptive statistics

Alg.	Classes	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
DT	6	0.9717	0.02563	0.01046	0.9448	0.9986	0.94	1.00
SVM	6	0.8817	0.07910	0.03229	0.7987	0.9647	0.75	0.97
RF	6	0.9317	0.04070	0.01662	0.8890	0.9744	0.87	0.98
MRF	6	0.9917	0.01329	0.00543	0.9777	1.0056	0.97	1.00

Table 3 ANOVA test results

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.042	3	0.014	6.471	0.003
Within Groups	0.044	20	0.002		
Total	0.086	23			

Table 4 Welch and Brown statistics

	Statistic	df1	df2	Sig.
Welch	6.730	3	9.803	0.010
Brown-Forsythe	6.471	3	9.032	0.013

Table 5 Covariance calculations

Algorithms	Covariance
SVM	8.97
DT	4.37
RF	2.64
MRF	1.14

Tables 4 and 5 are the Welch and Brown Forsythe tests conducted, and results were taken to derive the conclusion.

CONCLUSION

While in place of a neural network, if we use machine learning algorithms, such as random forest, support vector machine, decision trees and modified random forest with a particular n-estimator, the results obtained had significant accuracy. On a certain hypothesis generated by training 85% and testing on the entire dataset, the Random forest gave an accuracy of 97%, S.V.M. gave an accuracy of 86%, the decision tree of 93% and modified random forest of 99%. The one way ANOVA results are also proved that standard error of 0.005, the significance of 0.003 and covariance of 1.14 for MRF.

REFERENCES

[1] N. Chumuang, P. Pramkeaw, & A. Farooq, "Electrical Impedance Of Breast's Tissue Classification By Using Bootstrap Aggregating", *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Sorrento, Italy 2019*, pp. 551-556, 2019.

[2] C. Sebu, "Electrical Impedance Mammography: the key to low-cost, portable and non-invasive breast cancer screening?", *Xjenza*, 5, 2, pp. 154-157, 2017.

[3] E.K. Murphy, A. Mahara, X. Wu, & R.J. Halter, "Phantom experiments using soft-prior regularization EIT for breast cancer imaging", *Physiological Measurement*, 38, 6, p. 1262, 2017.

[4] Z. Haeri, M. Shokoufi, M. Jenab, R. Janzen, & F. Golnaraghi, "Electrical impedance spectroscopy for breast cancer diagnosis: Clinical study", *Integrative Cancer Science and Therapeutics*, 3, 6, p. 1-6, 2016.

[5] N. Zain, & K. Kanaga, "A review on breast electrical impedance tomography clinical accuracy", *ARPN Journal of Engineering and Applied Sciences*, 10, 1, pp. 6230-6234, 2015.

[6] R.J. Halter, A. Hartov, S.P. Poplack, W.A. Wells, K.M. Rosenkranz, R.J. Barth, P.A. Kaufman, & K.D. Paulsen, "Real-time electrical impedance variations in women with and without breast cancer", *IEEE Transactions on Medical Imaging*, 34, 1, pp. 38-48, 2014.

[7] S. Hong, K. Lee, U. Ha, H. Kim, Y. Lee, Y. Kim, & H.-J. Yoo, "A 4.9 mΩ-sensitivity mobile electrical impedance tomography IC for early breast-cancer detection system", *IEEE Journal of Solid-State Circuits*, 50, 1, pp. 245-257, 2014.

[8] N.M. Zain, & K.K. Chelliah, "Breast imaging using electrical impedance tomography: correlation of quantitative assessment with visual interpretation", *Asian Pacific Journal of Cancer Prevention*, 15, 3, pp. 1327-1331, 2014.

[9] S.N. Prasad, D. Houserkov, & J. Campbell, "Breast imaging using 3D electrical impedence tomography", *Biomedical Papers of the Medical Faculty of Palacky University in Olomouc*, 152, 1, pp. 151-154, 2008.

[10] J.E. Da Silva, J.M. De Sá, & J. Jossinet, "Classification of breast tissue by electrical impedance spectroscopy", *Medical and Biological Engineering and Computing*, 38, 1, pp. 26-30, 2000.