

# Comparative Analysis of R Package Classifiers Using Breast Cancer Dataset

Sudhamathy G. <sup>\*1</sup>, Thilagu M. <sup>#2</sup>, Padmavathi G. <sup>#3</sup>

<sup>\*1#2#3</sup> Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore – 641 043, India.

<sup>1</sup> sudhamathy25@gmail.com

<sup>2</sup> thilagu10@gmail.com

<sup>3</sup> ganapathi.padmavathi@gmail.com

**Abstract**—Breast cancer is a very common cancer among women. It remains as the number one form of cancer among women around the globe. Lack of awareness and detection of cancer at an advanced level put patient's life at a very high risk among the cancer affected women. Of the two types, non-invasive and invasive, invasive cancer has the potential to spread to other parts apart from the affected part. This paper attempts to perform breast cancer data analysis using R package. Decision tree is one of the data mining algorithms for classification due to the reason that it is fast, scalable and distributable. Among many tools available for data analysis, R is observed to be better in analyzing the data as it has become popular among the data analysts recently for the study of their large, unstructured and dynamic datasets. The three classifiers taken for study are 'rpath', 'ctree' and 'randomforest'. The algorithms are studied based on their performance measures such as accuracy, precision, recall, sensitivity and specificity. Based on the results, the best classification approach that suits better for cancer data analytics is recommended.

**Keyword**- Breast Cancer, Data Analytics, Classification, R Package

## I. INTRODUCTION

Cancer is a very important disease challenging the human lives all over the world. Out of the cancer affected community, more than 90% of the people have very remote chance of survival. Breast Cancer has been identified as the second primary cause of death among women worldwide. It is also observed to be a common type of cancer among women in both developed and the developing countries. A recent study says that one woman dies in every 13 minutes due to this disease. Breast cancer detection at an early stage has bright chances of saving women; because the survival rate of cancer affected patients is only 5%.

Cancer data present in different sources are heterogeneous in nature and many systems are developed to collect, analyze and learn from data for cancer care. Many open-source tools are available for data analysis. Due to the volume of data involved, data mining techniques are applied for early detection of cancer. Cancer detected at the early stage can promote better diagnosis, treatment and can improve the survival data of patients.

The fundamental data mining methods are classification, clustering, regression, artificial intelligence, association rules and decision tree. The classification methods are the most intensively used methods of data mining in health care. The classification model devised for specific applications is trained with the existing set of data. This learning can help the model in assisting the prediction and classification of new data. This concept can be successfully applied for cancer research as the diagnosis and treatment process involved past data set.

With this idea and motivation, decision tree classification methods available in R package are studied. Three significant classification methods namely, 'rpath', 'ctree' and 'randomforest' algorithms are implemented using R. An analysis of the above three methods is done based on the specific performance metrics such as accuracy of detection, its precision, recall, specificity and sensitivity. The breast cancer data set available in 'mlbench' package of CRAN is taken for testing.

### Significant contributions of this paper:

- i) Study of the three classification methods namely, 'rpath', 'ctree' and 'randomforest'.
- ii) Data analysis using performance metrics for the breast cancer data set taken.

The entire work is organized as follows: "Related Work" Section discusses about the significant literatures available in breast cancer data analysis. "Methodology" Section presents the proposed methodology. "Results and Discussions" Section explains the experimental set up and the results due to experimentation.

## II. RELATED WORK

There has been many works done in the area of Breast Cancer data analysis of which few significant ones have been reviewed here.

The predictive models are discussed in [1] and based on the analysis of their results, it is evident that the integration of multidimensional heterogeneous data, combined with the application of different techniques for feature selection and classification can provide promising tools for inference in the cancer domain. The study [2] clearly shows that data mining techniques is a good method to predict breast cancer recurrence and they present an efficient pre-classification method and discover the information of breast cancer recurrence of SEER dataset. The authors in [3] compare three classification techniques in Weka software and comparison results show that Sequential Minimal Optimization (SMO) has higher prediction accuracy i.e. 96.2% than IBK and BF Tree methods. The work in [4] presents a diagnosis system for detecting breast cancer based on RepTree, RBF Network and Simple Logistic. The outcome of the research in [5] is justified that k-means clustering algorithm and FF algorithm are helpful to early diagnosis of the breast cancer patients. The purpose of the research in [6] is to develop a novel prototype of clinical problem regarding to diagnose and manage patients with breast cancer. Different methods for breast cancer detection are explored in [7] and their accuracies are compared and with their results, they infer that the SVM are more suitable in handling the classification problem of breast cancer prediction.

Given all these literature review it is understood that not much exploration has been so far made using the classification techniques in the R Package which may reveal more interesting performance improvements. Hence, our study focuses on these approaches.

## III. METHODOLOGY

The major objective of the work is to perform the breast cancer data analysis using specific classifiers. The secondary objective of the work is to compare the performance and identify the best suited decision tree classification algorithm out of the three algorithms taken for study. Fig. 1 shows the methodology used for study. The work has been implemented using the decision tree algorithms '*rpart*', '*ctree*' and '*randomforest*' available in R Tool using the *BreastCancer* Dataset of the '*mlbench*' Package. The reason for choosing decision trees among the available classification techniques in R is that the results are fast, scalable and distributable.

### A. About R

R is a software environment used by data analysts for data mining and statistical analysis. R has recently become more popular among data analysts for the study of their large and dynamic databases. R was originally developed by GNU and its source code was written in C, Fortran and R itself. R has both command line interface and graphical interface. R Tool has by default many packages installed in it and each of these packages has a set of functions that performs certain analysis or graphical representations.

### B. Package "mlbench"

Some packages in R Tool are available by default and some needs to be installed by using the Install Package option in R. The package '*mlbench*' needs to be installed in R and this package has only datasets and no functions in it. This package consists of a set of artificial and real-world machine learning datasets with many of them taken from the UCI repository.

### C. About the Dataset

The dataset used for this work is the *BreastCancer* dataset that has 683 observations and it has the below list of attributes as shown in Table I. The dataset is divided into the training and the test dataset in the ratio of 80% and 20% respectively.

### D. About 'rpart'

The decision tree technique '*rpart*' means recursive partitioning for classification and regression trees. These algorithms build classification or regression models of a very general structure using a two stage procedure.

### E. About 'ctree'

The decision tree technique '*ctree*' means conditional inference trees that are available in the package '*party*'. It is a non parametric class of regression trees embedding tree-structured regression models into a well defined theory of conditional inference procedures.

### F. About 'randomforest'

The decision tree technique '*randomforest*' means classification and regression based on a forest of trees using random inputs. This method implements *Breiman's* random forest algorithm for classification and regression. It can also be used in unsupervised mode for assessing proximities among data points.

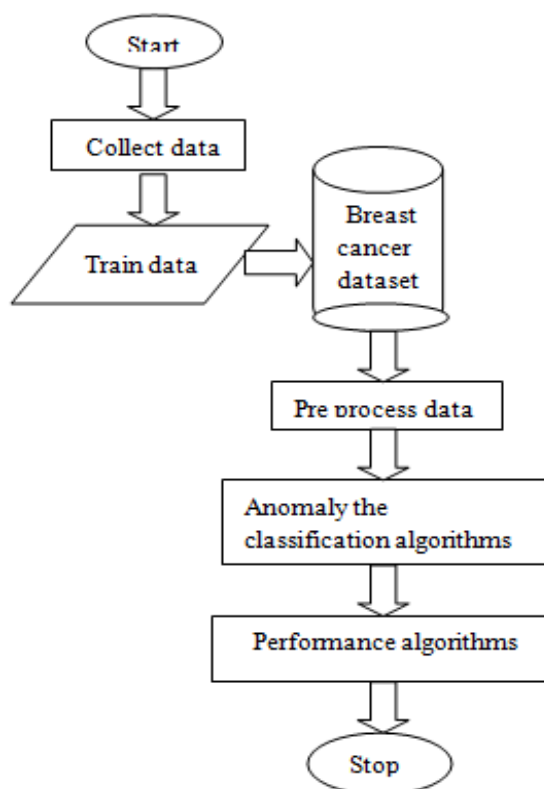


Fig. 1. Steps in Proposed Methodology

TABLE I. Dataset Attributes

Attribute Names	Attribute Description	Values
Cl.thickness	<i>Clump Thickness</i> – Benign cells are mono layered Malignant Cells are multi layered	0 – 10 (Numeric)
Cell.size	<i>Cell Size</i> – Cancerous cells vary in size	0 – 10 (Numeric)
Cell.shape	<i>Cell Shape</i> – Cancerous cells vary in shape	0 – 10 (Numeric)
Marg.adhesion	<i>Marginal Adhesion</i> Adhesion means cells sticking together In cancerous cells there will be loss of adhesion	0 – 10 (Numeric)
Epith.c.size	<i>Single Epithelial Cell Size</i> Significantly enlarged Epithelial cells are Malignant cells	0 – 10 (Numeric)
Bare.nuclei	<i>Bare nuclei</i> Nuclei not surrounded by cytoplasm are Benign cells	0 – 10 (Numeric)
Bl.cromatin	<i>Bland Chromatin</i> Chromatin tends to be coarser in cancerous cells and with uniform texture in Benign cells	0 – 10 (Numeric)
Normal.nucleoli	<i>Normal nucleoli</i> In cancer cells the nucleoli become more prominent In normal cells nucleoli is very small	0 – 10 (Numeric)
Mitoses	<i>Mitoses</i> – Rate of cell division	0 – 10 (Numeric)
Class	<i>Types of Tumors</i> Benign Cells – Not Cancerous Malignant – Cancerous	Benign, Malignant

#### IV. RESULTS AND DISCUSSIONS

##### A. Code and Output

The code for the 'rpart' classification method goes as below. Before executing the code in the R Tool it is required to install the package *mlbench* and call the dataset *BreastCancer*. The dataset is then divided into the training and the test dataset in the ratio of 80% and 20% respectively.

```
> require(mlbench)
> data(BreastCancer)
> set.seed(2)
> ind <- sample(2, nrow(BreastCancer), replace = TRUE, prob=c(0.8, 0.2))
> require(rpart)
> x.rp <- rpart(Class ~ ., data=BreastCancer[ind == 1,])
> x.rp.pred <- predict(x.rp, type="class", newdata=BreastCancer[ind == 2,])
> x.rp.prob <- predict(x.rp, type="prob", newdata=BreastCancer[ind == 2,])
> x11()
> plot(x.rp, main="Decision tree created using rpart")
> text(x.rp, use.n=TRUE, all=TRUE, cex=.6)
```

The result for this code is shown in the Fig. 2. The interpretation of the resultant tree is as below where T stands for total number of records, M stands for records with class Malignant and B stands for records with class Benign. The level numbers shows the levels in the tree and the depth of conditions used.

- Level – 0: All (Root) [T = 535; M = 183; B = 352]
- Level – 1: Cell.Size = 1, 2, 3 [T = 375; M = 31; B = 344]
- Level – 2: Bare.Nuclei = 1, 2, 3, 4 & Cell.Size = 1, 2, 3 [T = 344; M = 6; B = 338]
- Level – 3: Bare.Nuclei = 1, 2 & Cell.Size = 1, 2, 3 [T = 323; M = 0; B = 323]
- Level – 3: Bare.Nuclei = 3, 4 & Cell.Size = 1, 2, 3 [T = 21; M = 6; B = 15]
- Level – 4: Cl.Thickness = 1, 2, 3, 4 & Bare.Nuclei = 3, 4 & Cell.Size = 1, 2, 3 [T = 14; M = 0; B = 14]
- Level – 4: Cl.Thickness = 5, 6, 7, 8, 9, 10 & Bare.Nuclei = 3, 4 & Cell.Size = 1, 2, 3 [T=7; M=6; B=1]
- Level – 2: Bare.Nuclei = 5, 6, 7, 8, 9, 10 & Cell.Size = 1, 2, 3 [T = 31; M = 25; B = 6]
- Level – 3: Cell.Shape = 1 & Bare.Nuclei = 5, 6, 7, 8, 9, 10 & Cell.Size = 1, 2, 3 [T = 8; M = 2; B = 6]
- Level – 3: Cell.Shape = 2,3,4,5,6,7,8,9,10 & Bare.Nuclei = 5,6,7,8,9,10 & Cell.Size = 1,2,3 [T=23; M=23; B=0]
- Level – 1: Cell.Size = 4, 5, 6, 7, 8, 9, 10 [T = 160; M = 152; B = 8]

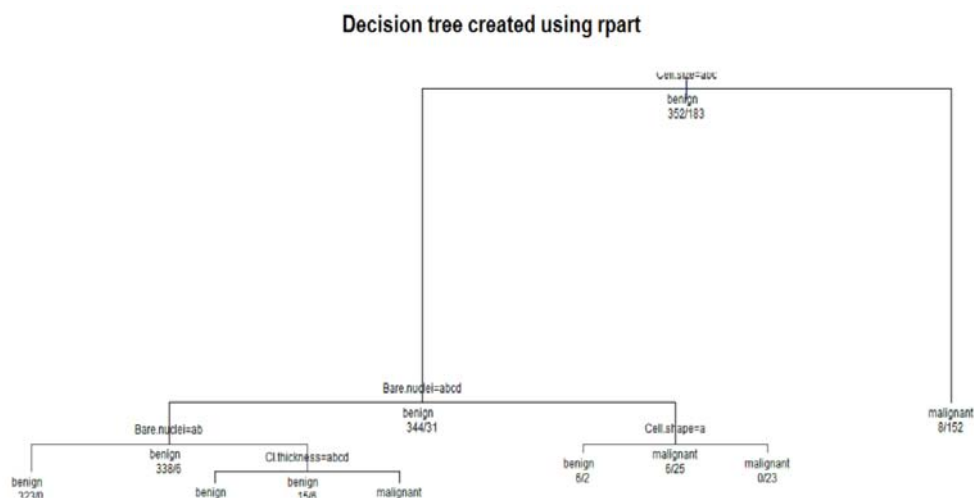


Fig. 2. Output of 'rpart' Method

The code for the 'ctree' classification method goes as below. Before executing the code in the R Tool it is required to install the package 'party'.

```
> require(party)
> x.ct <- ctree(Class ~ ., data=BreastCancer[ind == 1,])
> x.ct.pred <- predict(x.ct, newdata=BreastCancer[ind == 2,])
> x.ct.prob <- 1 - unlist(treeresponse(x.ct, BreastCancer[ind == 2,]),
use.names=F)[seq(1,nrow(BreastCancer[ind == 2,])*2,2)]
> plot(x.ct, main="Decision tree created using ctree", type="simple", cex=.6)
```

The result for this code is shown in the Fig. 3. The interpretation of the resultant tree is as below where T stands for total number of records, M stands for records with class Malignant and B stands for records with class Benign. The level numbers shows the levels in the tree and the depth of conditions used. The values given with the brackets '(')' represents the probability of the Benign and Malignant class records.

- Level - 0: All (Root)[T = 535]
- Level - 1: Bare.nuclei == { 1, 2} [T = 335]
- Level - 2: Bare.nuclei == { 1, 2} & Cell.size <= 3 [T = 323; B = 323 (1); M = 0 (0)]
- Level - 2: Bare.nuclei == { 1, 2} & Cell.size > 3 [T=12; B = 0 (0); M = 12 (1)]
- Level - 1: Bare.nuclei == { 3, 4, 5, 6, 7, 8, 9, 10} [T = 200]
- Level - 2: Bare.nuclei == { 3, 4, 5, 6, 7, 8, 9, 10} & Cl.thickness <= 4 [T = 42]
- Level - 3: Bare.nuclei == { 3, 4, 5, 6, 7, 8, 9, 10} & Cl.thickness <= 4 & Cell.shape <= 2 [T = 21; B = 20 (0.952); M = 1 (0.048)]
- Level - 3: Bare.nuclei == { 3, 4, 5, 6, 7, 8, 9, 10} & Cl.thickness <= 4 & Cell.shape > 2 [T = 21; B = 3 (0.143); M = 18 (0.857)]
- Level - 2: Bare.nuclei == { 3, 4, 5, 6, 7, 8, 9, 10} & Cl.thickness > 4 [T = 158; B = 6 (0.038); M = 152 (0.962)]

The code for the 'randomforest' classification method goes as below. Before executing the code in the R Tool it is required to install the package 'randomForest'.

```
> require(randomForest)
> x.cf <- randomForest(Class ~ ., data=BreastCancer[ind == 1,], ntree=100, proximity=TRUE)
> x.cf.pred <- predict(x.cf, newdata=BreastCancer[ind == 2,])
> x.cf.prob <- 1 - unlist(treeresponse(x.cf, BreastCancer[ind == 2,]),
use.names=F)[seq(1,nrow(BreastCancer[ind == 2,])*2,2)]
> plot(x.cf1, main="Plot created using randomForest", cex=.6)
```

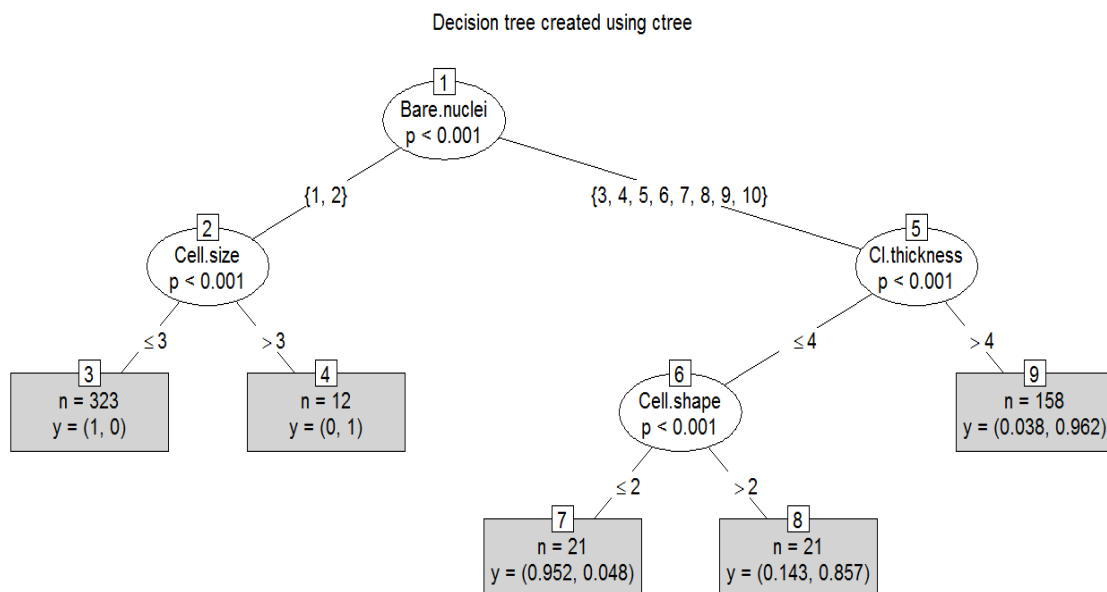


Fig. 3. Output of 'ctree' Method

The result for this code is shown in the Fig. 4. The interpretation of the resultant graph is given using the confusion matrix below that shows the class error and the number of records that were classified correctly and those that were classified wrongly (error rate).

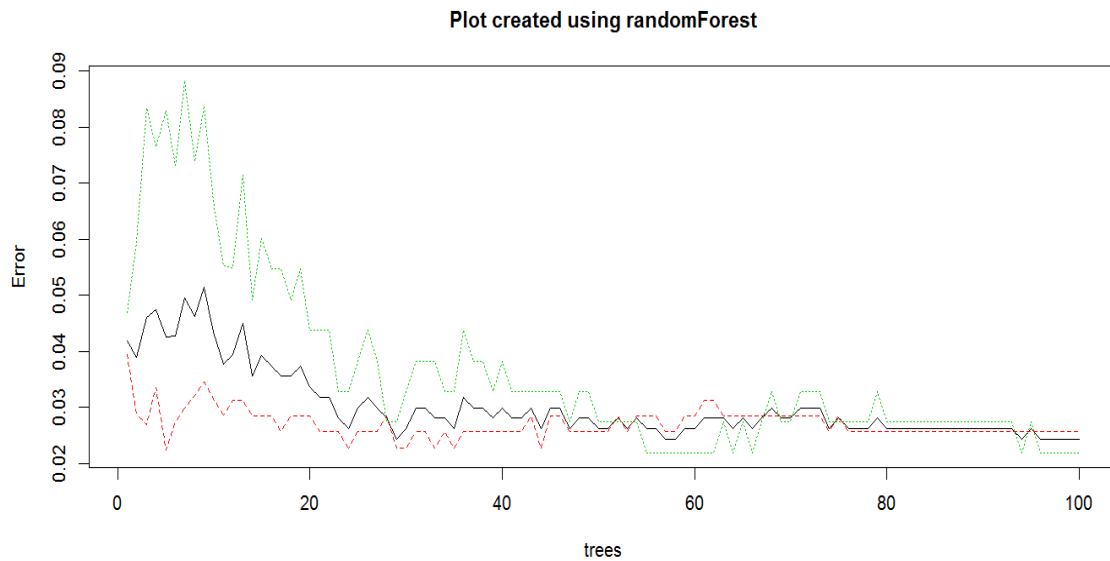


Fig. 4. Output of 'randomforest' Method

No. of variables tried at each split: 3  
 OOB estimate of error rate: 2.62%

Confusion Matrix:

**Benign Malignant class.error535**  
**Benign** 343 9 0.02556818 352  
**Malignant**.5.178.0.02732240.183

### B. Performance Measuring

The package ROCR has to be installed to get the performance measures of the three decision tree algorithms. The performance measures used in this study are *Accuracy*, *Precision*, *Recall*, *Sensitivity* and *Specificity*. The code and the graphs (Fig. 5, Fig. 6, Fig. 7, Fig. 8 and Fig. 9) that take the readings for these performance measures are given as below.

```
> require(ROCR)
> x.rp.prob.rocr <- prediction(x.rp.prob[,2], BreastCancer[ind == 2,'Class'])
> x.rp.perf1 <- performance(x.rp.prob.rocr, "acc","cutoff")
> plot(x.rp.perf1, col=2, lwd=3, main="Comparing Accuracy of the Classification Methods")
> x.ct.prob.rocr <- prediction(x.ct.prob, BreastCancer[ind == 2,'Class'])
> x.ct.perf1 <- performance(x.ct.prob.rocr, "acc","cutoff")
> plot(x.ct.perf1, col=3, lwd=3, add=TRUE)
> x.cf.prob.rocr <- prediction(x.cf.prob, BreastCancer[ind == 2,'Class'])
> x.cf.perf1 <- performance(x.cf.prob.rocr, "acc","cutoff")
> plot(x.cf.perf1, col=4, lwd=3, add=TRUE)
> legend(0.6, 0.6, c('rpart', 'ctree', 'cforest'), 2:4)
```

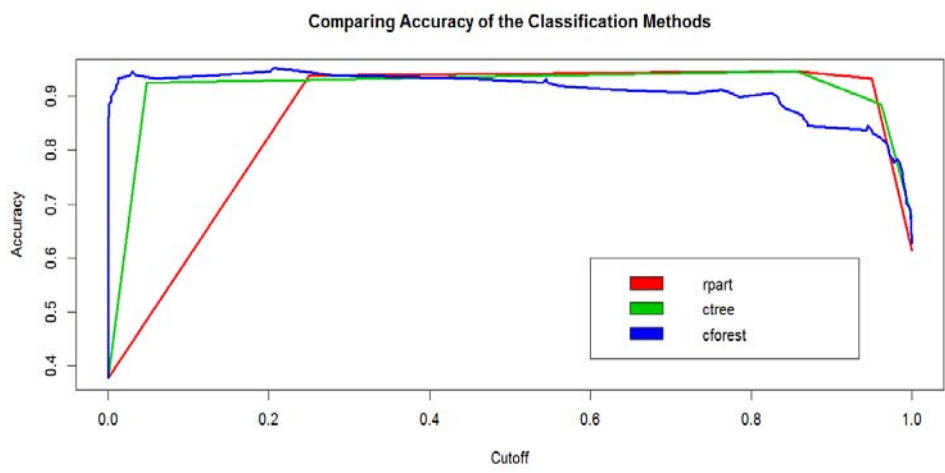


Fig. 5. Comparison of Accuracy

```
> x.rp.perf2 <- performance(x.rp.prob.rocr, "prec", "cutoff")
> plot(x.rp.perf2, col=2, lwd=3, main=" Comparing Precision of the Classification Methods ")
> x.ct.perf2 <- performance(x.ct.prob.rocr, "prec", "cutoff")
> plot(x.ct.perf2, col=3, lwd=3, add=TRUE)
> x.cf.perf2 <- performance(x.cf.prob.rocr, "prec", "cutoff")
> plot(x.cf.perf2, col=4, lwd=3, add=TRUE)
> legend(0.6, 0.6, c('rpart', 'ctree', 'cforest'), 2:4)
> x.rp.perf3 <- performance(x.rp.prob.rocr, "rec", "cutoff")
> plot(x.rp.perf3, col=2, lwd=3, main=" Comparing Recall of the Classification Methods ")
> x.ct.perf3 <- performance(x.ct.prob.rocr, "rec", "cutoff")
> plot(x.ct.perf3, col=3, lwd=3, add=TRUE)
> x.cf.perf3 <- performance(x.cf.prob.rocr, "rec", "cutoff")
> plot(x.cf.perf3, col=4, lwd=3, add=TRUE)
> legend(0.6, 0.6, c('rpart', 'ctree', 'cforest'), 2:4)
```

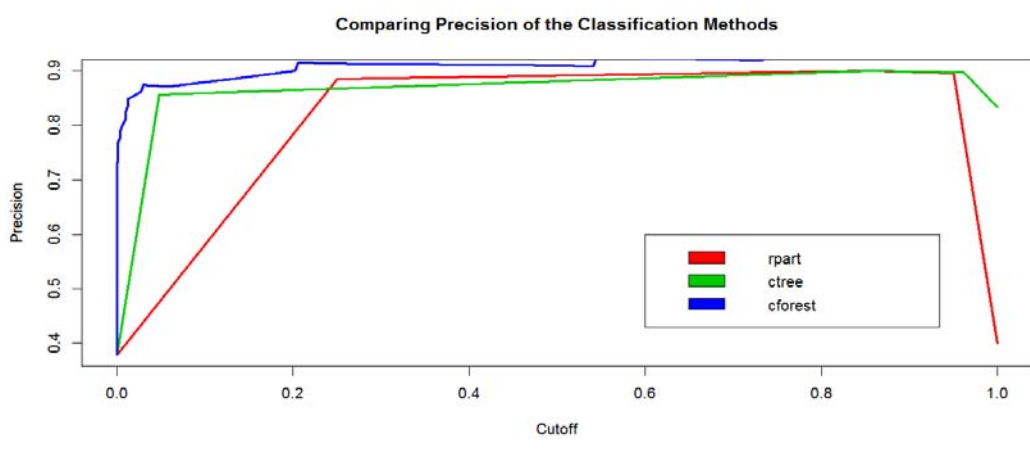


Fig. 6. Comparison of Precision

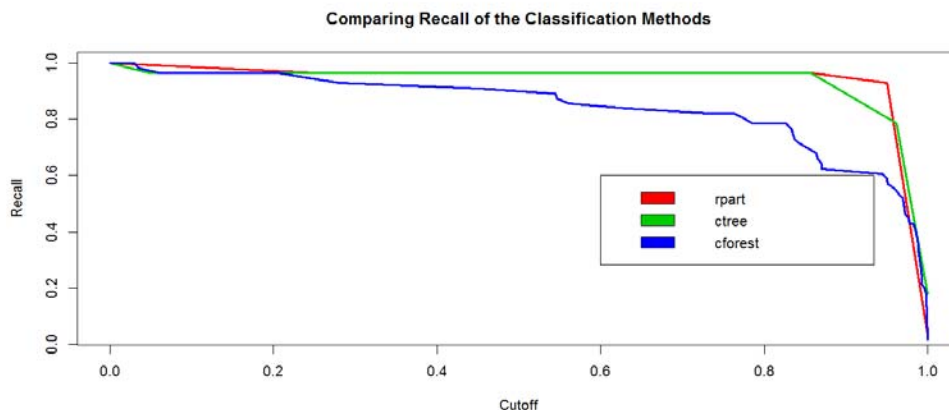


Fig. 7. Comparison of Recall

```
> x.rp.perf4 <- performance(x.rp.prob.rocr, "sens", "cutoff")
> plot(x.rp.perf4, col=2, lwd=3, main=" Comparing Sensitivity of the Classification Methods ")
> x.ct.perf4 <- performance(x.ct.prob.rocr, "sens", "cutoff")
> plot(x.ct.perf4, col=3, lwd=3, add=TRUE)
> x.cf.perf4 <- performance(x.cf.prob.rocr, "sens", "cutoff")
> plot(x.cf.perf4, col=4, lwd=3, add=TRUE)
> legend(0.6, 0.6, c('rpart', 'ctree', 'cforest'), 2:4)
> x.rp.perf5 <- performance(x.rp.prob.rocr, "spec", "cutoff")
> plot(x.rp.perf5, col=2, lwd=3, main=" Comparing Specificity of the Classification Methods ")
> x.ct.perf5 <- performance(x.ct.prob.rocr, "spec", "cutoff")
> plot(x.ct.perf5, col=3, lwd=3, add=TRUE)
> x.cf.perf5 <- performance(x.cf.prob.rocr, "spec", "cutoff")
> plot(x.cf.perf5, col=4, lwd=3, add=TRUE)
> legend(0.6, 0.6, c('rpart', 'ctree', 'cforest'), 2:4)
```

All the above said performance measures are based on the Positive Class (P), Negative Class (N), True Positive (TP) samples, True Negative (TN) samples, False Positive (FP) samples and False Negative (FN) samples whose definitions are as given below.

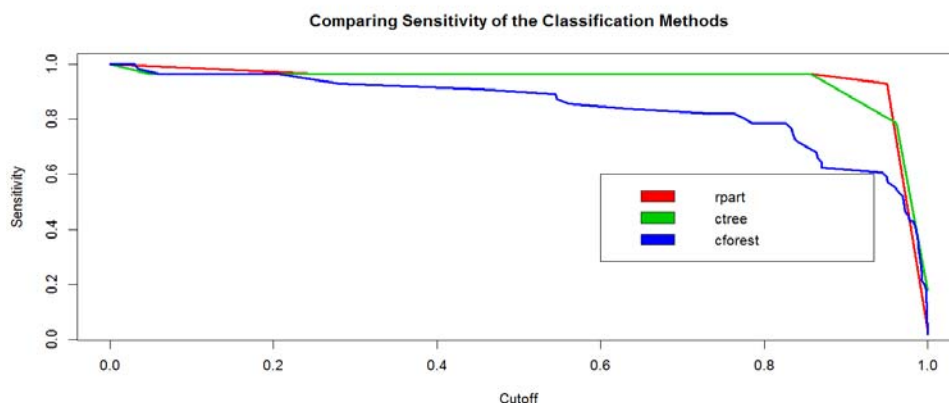


Fig. 8. Comparison of Sensitivity



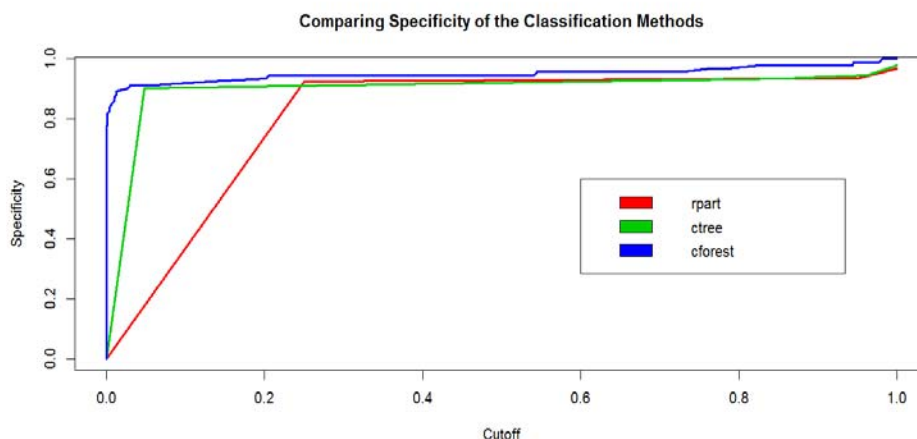


Fig. 9. Comparison of Specificity

P – Positive Class = TP + FN = Samples with Class Benign  
 N – Negative Class = FP + TN = Sample with Class Malignant

TABLE II. Definition of Samples

Actual / Predicted		Predicted	
		Benign	Malignant
Actual	Benign	TP	FN
	Malignant	FP	TN

Based on the above factors the definitions of the performance measures under discussion are given as below.

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{P + N} & \text{Precision} &= \frac{TP}{TP + FP} \\
 \text{Recall} &= \frac{TP}{TP + FN} & \text{Sensitivity} &= \frac{TP}{P} & \text{Specificity} &= \frac{TN}{N}
 \end{aligned}$$

C. Observations

Taking the readings from the graphs for the performance measures the Table III and Table IV are arrived at that gives the minimum value, maximum value and the standard deviation (SD) of the readings.

TABLE III Readings of Accuracy, Precision and Recall

Method / Parameter	Accuracy			Precision			Recall		
	Min	Max	SD	Min	Max	SD	Min	Max	SD
<b>RPart</b>	0.38	0.95	0.22	0.38	0.90	0.28	0.00	1.0	0.49
<b>Party</b>	0.38	0.95	0.22	0.38	0.90	0.22	0.00	1.0	0.44
<b>Random Forest</b>	0.38	0.95	0.12	0.38	1.00	0.15	0.00	1.0	0.29

TABLE IV Readings of Sensitivity and Specificity

Method / Parameter	Sensitivity			Specificity		
	Min	Max	SD	Min	Max	SD
<b>RPart</b>	0.00	1.00	0.49	0.00	1.00	0.39
<b>Party</b>	0.00	1.00	0.44	0.00	1.00	0.39
<b>Random Forest</b>	0.00	1.00	0.30	0.00	1.00	0.22

#### D. Discussion

It is evident from the readings that minimum and maximum values of all the five measures are the same except for that of Precision in which case the maximum value of Precision for the 'randomforest' method is 1.00. But, when the standard deviations of the measures are considered, they vary between the three methods compared. The standard deviation of the values produced by the 'randomforest' method is minimum, when compared to the other methods. This reveals the fact that the deviation between the values produced by the 'randomforest' method is lesser. With this evidence, we can arrive at a decision that the 'randomforest' decision tree classification method performs better as it provides the maximum precision and the standard deviation is lesser. Hence for classifying Breast Cancer data, we can use the 'randomforest' technique available in R to get accurate predictions.

#### V. CONCLUSION

Early detection of breast cancer can be predicted accurately by using classification methods in R Package. This may result in the decrease of health cost and may enhance time required for a patient to receive treatment. In this paper three decision tree classification methods were compared to suggest one best that can have better performance than others when considering breast cancer data analysis. The software tool used for this purpose is R, which is one of the most popular among data analysts of the current days and that has tremendous performance so that the run times are significantly very less. It is observed that among the three techniques compared namely 'rpart', 'ctree' and 'randomforest', 'randomforest' is the best based on the performance measure precision which distinctly differs from the rest of measures used. The method of 'randomforest' shows the highest precision, which is 1 and hence we conclude saying that usage of 'randomforest' method best suits cancer analytics. As a future scope we shall be exploring the other classification techniques such as SVM, KNN and Naïve Bayes and comparing their performances with that of 'randomforest' decision tree approach.

#### REFERENCES

- [1] K. Konstantina, P.E. Themis, P.E. Konstantinos, V.K. Michalis, and I.F. Dimitrios, "Machine Learning Applications in Cancer Prognosis and Prediction", Computational and Structural Biotechnology Journal, vol. 13, pp. 8-17, 2015.
- [2] D.R. Umesh, and B. Ramachandra, "Predicting Breast Cancer Recurrence Using Machine Learning Techniques", International Journal of Latest Trends in Engineering and Technology, vol. 5(1), pp. 255-259, 2015.
- [3] V. Chaurasia, and S. Pal, "A Novel Approach for Breast Cancer Detection using Data Mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering, vol. 2(1), pp. 2456-2465, 2014.
- [4] V. Chaurasia, and S. Pal, "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability", International Journal of Computer Science and Mobile Computing, vol. 3(1), pp. 0-22, 2014.
- [5] J. Jahanvi, D. Rinal, and P. Jigar, "Diagnosis of Breast Cancer using Clustering Data Mining Approach", International Journal of Computer Applications, vol. 101(10), pp. 13-17, 2014.
- [6] J. Jahanvi, D. Rinal, and P. Jigar, "Diagnosis and Prognosis Breast Cancer Using Classification Rules", International Journal of Engineering Research and General Science, vol. 2(6), pp. 315-323, 2014.
- [7] G. Ravi Kumar, G.A. Ramachandra, and K. Nagamani, "An Efficient Prediction of Breast Cancer Data using Data Mining Techniques", International Journal of Innovations in Engineering and Technology, vol. 2(4), pp. 139-144, 2013.

#### AUTHOR PROFILE

Dr. G. Padmavathi is the Professor and Head of the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. She has 28 years of Teaching experience and one year of Industrial experience. Her areas of interest include Real Time Communication, Wireless Communication, Network Security and Cryptography. She has significant number of publications in peer reviewed International and National Journals. She is the life member of CSI, ISTE, WSEAS, AACE and ACRS.

Dr. G. Sudhamathy, She is currently working as Assistant professor in the Department of Computer Science in Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamilnadu, India. She has a rich industrial experience of around 11 years working in various multinational Information Technology companies like Cognizant Technologies Solutions, L&T Infotech, etc. She has worked with international clients and has worked in London for a year. She also has around 7 years of academic and research experience. Her research interests are in Web Usage Mining and Applications of Data Mining using R. She is the life member of CSI and ISTE.

Dr. M. Thilagu is the Assistant Professor in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. She has 18 years of teaching experience and her areas of interest include Web Log Mining, DBMS and Object Oriented Programming.