International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 6, 1170-1190, 2020
https://doi.org/10.33889/IJMEMS.2020.5.6.089

A Comparison of Artificial Neural Network and Decision Trees with Logistic Regression as Classification Models for Breast Cancer Survival

Venkateswara Rao Mudunuru
Department of Mathematics and Statistics,
University of South Florida, Tampa, FL, USA.
Corresponding author: vmudunur@mail.usf.edu

Leslaw A. Skrzypek
Department of Mathematics and Statistics,
University of South Florida, Tampa, FL, USA.
E-mail: skrzypek@usf.edu

(Received April 10, 2020; Accepted July 3, 2020)

Abstract
In the field of medicine, several recent studies have shown the value of Artificial Neural Networks, decision trees, logistic regression are playing a major role as the predictor, and classification methods. The research has been expanded to estimate the incidence of breast, lung, liver, ovarian, cervical, bladder and skin cancer. The main aim of this paper is to develop models of logistic regression, Artificial Neural Networks, and Decision trees using the same input and output variables and to compare their success in predicting breast cancer survival in woman. To find the best model for breast cancer survival, the sensitivity and specificity of all these models are measured and evaluated with their respective confidence intervals and the ROC values.

Keywords- Artificial neural networks, Logistic Regression, Breast cancer, Decision Trees, Cancer survival.

1. Introduction

In the field of clinical diagnostics, computer models are playing a prominent role in differentiating between a healthy and an ill patient. Such computer model accuracy is held accountable for encouraging correct decisions about the risk of disease based on the patient's characteristics. Numerous data models designed, assessed and improved include both statistical methods as well as non-statistical ones. Every methodology uses different assumptions and, depending on the data context, may or may not produce similar results. Three frequently used methods are regression, decision trees, and artificial neural networks. Regression methodology estimates the relation between the independent and dependent variables. Regression methods are also referred to as dependence analysis techniques (Agresti, 2010). Study of regression models is a crucial part of several research projects. Such models are widely applied in order to assess the survival of severely ill patients admitted to the intensive care unit (Gellar et al., 2014). Linear models and logistic regression models are two main categories of regression methods. The technique of logistic regression is very widely used in data analysis. It is considered a well-known model of classification that enables probabilistic decisions to be made and shows promising results on several issues. A logistic model is often considered a clinically interpretable model for providing best-fit similar results.

Four models are developed in this paper considering the same output and the set of input variables using decision trees, logistic regression and artificial neural networks. Performance of these models is assessed in breast cancer survival prediction. The four models are developed by considering cancer data available in SEER. Surveillance, Epidemiology and End Results (SEER) is the most widely used Medicare database for cancer datasets. This data has been pre-processed (cleansed) to eliminate or account for missing data and repetitions. The final data set considered in this research contained 47,167 records of malignant tumors (Mudunuru, 2016).

In this paper, the survival of a breast cancer woman is estimated using important available attributable variables in the database. Initially, we selected all the independent variables including size of the tumor, age of the patient, stage of the breast cancer, treatment administered, duration, tumor grade, marital status of the patient, and the count of primary tumors recorded. Logistic regression eliminated variables like size of the tumor, grade of the cancer, marital status of the patient to be statistically insignificant in prediction survival of women with breast cancer. We developed our four models by inputting only one of the remaining significant variables at a time. All these models provided us with an output vector with two variables for each case: either 1 (alive/ survived) or 0 (dead/ not survived). A number ranging between 0 and 1 is used to estimate the accuracy of predicted value.

2. Logistic Regression

One of the primary assumptions of linear logistic regression is that the independent covariates are in linear association with their corresponding natural logarithm of odds. The three main components that define logistic function are a systematic part, the related link function and the random experiment. The binary nature of the dependent variable that is eminently suited for modelling data is also another important characteristic of logistic function. As a result of this, the target of logistic regression therefore is slightly different since we estimate the probability that the response variable is equal to provided values of independent variables (McFadden, 1973).

Survival prediction of breast cancer is the output variable in our current research from the given patient’s age, size of tumor, stage of cancer, treatment administered, and the duration.

2.1 Survival Prediction using Logistic Modelling

The two most general and widely used survival analyses techniques are Event history models and logistic regression. Survival time is considered as a continuous variable in the event history model whereas it is considered as discrete in logistic regression models. A dichotomous measure (survived or not) is thus the target. Logistic regression model as a classifier is applied in this section to predict breast cancer survival in women.

Age and size of the tumor are used in the first model (model-1). Age, size of the tumor and stage of the cancer are used in the second model (model-2). Along with the three variables selected in model-2, treatment is added to develop the third model (model-3). Age, size of the tumor, cancer stage, treatment and duration are all used in the fourth model (model-4). Our primary aim is to measure or estimate the women's breast cancer survival.

The overall accuracy of the models is attributed as a reliability measure of the given estimate. Specificity, sensitivity along with 95% confidence intervals and overall accuracy for the developed logistic models are given in the Table 1.

Table 1. Sensitivity, specificity and overall results of Logistic regression models

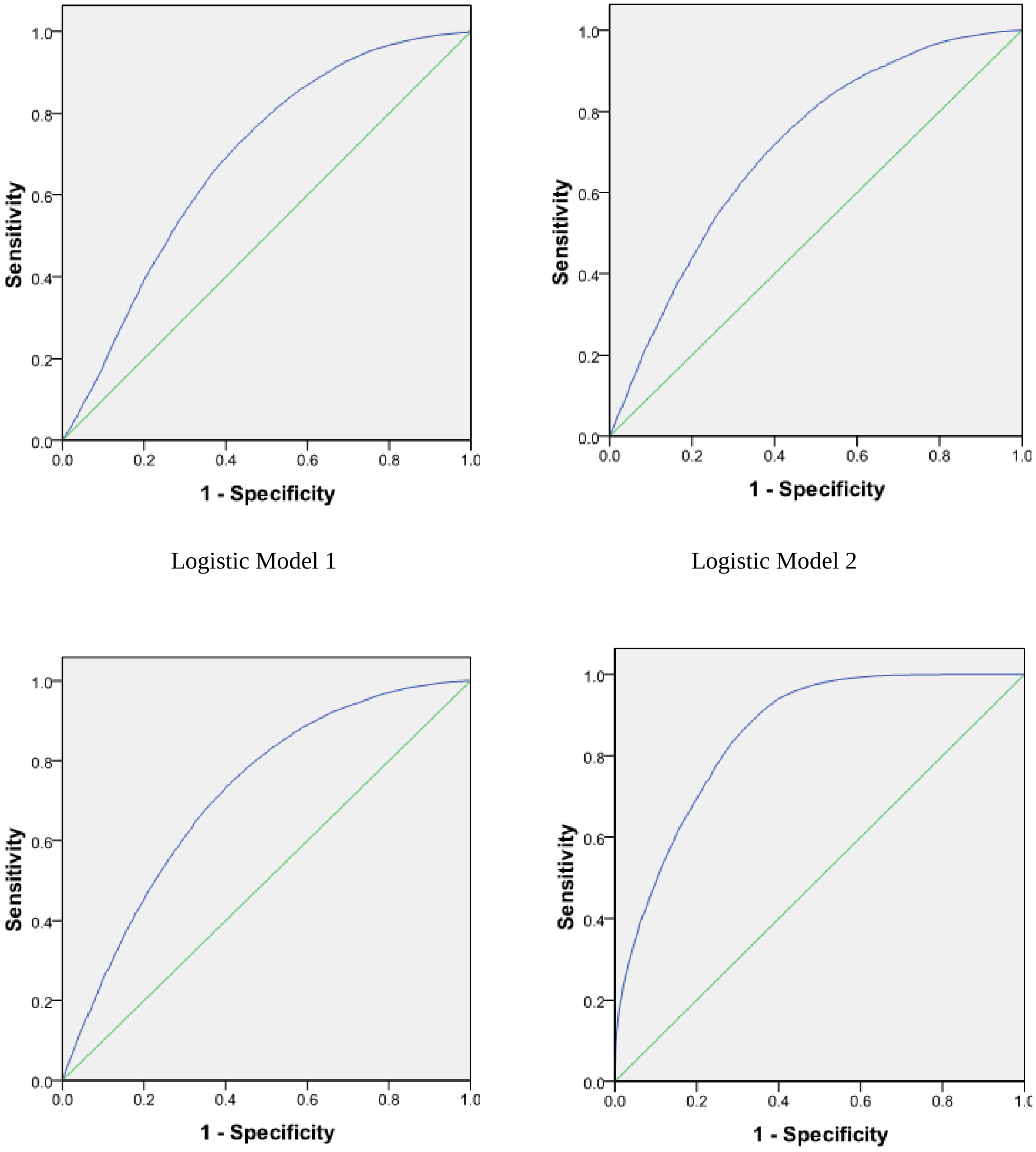
Logistic Regression Models	Sensitivity		Specificity		Accuracy
	Value	95% Confidence Interval	Value	95% Confidence Interval	
Logistic model 1	68.01%	67.04 - 68.96	69.45%	68.99 - 69.91	69.2%
Logistic model 2	67.31%	66.39 - 68.21	70.47%	70.00 - 70.93	69.78%
Logistic model 3	67.69%	66.80 - 68.55	71.26%	70.78 - 71.72	70.42%
Logistic model 4	76.82%	76.14 - 77.47	81.54%	81.11 - 81.97	79.98%

The ROC area values along with specificity at 95% sensitivity for the developed logistic models are given in Table 2.

Table 2. ROC Areas for Logistic models

Logistic Regression Models	ROC	At 95% Sensitivity
		Specificity
Logistic model 1	68.8%	25%
Logistic model 2	71.0%	30%
Logistic model 3	71.8%	29%
Logistic model 4	85.5%	61%

The results showed that for model-3, the overall accuracy is 70.42% that increases to 80% for model-4. As expected, the tumor duration has large significance for predicting accurate survival during the study period. The ROC area of logistic regression model-1 is 68.8% and a survival sensitivity of 95% yielded a specificity of just 25%, model-2 on the other hand with a ROC area of 71% and survival sensitivity of 95% provided 30% specificity rate. The ROC area values for the remaining two models are 71.8% and 85.5% respectively, and the 95% survival sensitivity rate has a 29% and 61% overall specificity, respectively. Highly attractive is the performance of fourth logistic model offering 80% overall accuracy. The specificity and sensitivity for this model are 81.54%, 76.82% respectively. At a 95 % sensitivity this model has 61% specificity. Figure 1 displays the ROC graphs for the logistic models.



Logistic Model 1

Logistic Model 2

Logistic Model 3

Logistic Model 4

Figure 1. ROC graphs of Logistic Regression models

3. ANN Perceptron Classification and Back Propagation

A perceptron is an arrangement of one McCulloch-Pitts neuron input layer feeding forward into one McCulloch-Pitts neuron output layer. A feedforward neural network of at least two layers is characterized as a multilayer perceptron. Growing of the perceptron layers is utilized by splitting them into small linearly separable parts of data given to solve nonlinear separable problems (Widrow and Lehr, 1990). Hard-limiting function (step function) is the most preferred activation function for producing the outputs. The second most important function is a sigmoid function. Sigmoid function prevents the information of the inputs to overflow into the inner neurons. In modelling a complex data, step function is replaced with a sigmoid function. The combination of each individual perceptron with a sigmoid function combined with another series of perceptrons serves as the output of a multilayer perceptron. The architecture and design of the network is intended to be more flexible, for example, no immediate connection among input and target layers as well as associations between layers is assumed, the number of outputs need not be equivalent to the number of inputs, and there is no limitation on the number of hidden layers or units. Hidden units can even be more than or less than input and output units (Jabri and Flower, 1992).

The most widely used learning algorithm for training feedforward artificial neural networks is the backpropagation algorithm. In this technique, corresponding weights of previous layers neurons are multiplied separately along with receiving signals in the current layer. Inputs of one or more previous neurons are weighted independently and added. The weights between neurons are optimized using the backpropagation technique to generate the best network. Thus, multilayer perceptrons have two essential characteristics, generalization and fault tolerance. Neural networks are extremely tolerant to faults. This feature is also called graceful degradation (Mohr et al., 2000). Even if certain interconnections between certain neurons within the layers fail, the neural networks keep working. We developed ANN models in this research to fulfil the above characteristics.

3.1 Survival Prediction using ANN Modelling

When modelling non-linear data, artificial neural networks (ANN) are efficient estimators. Constructing an ANN requires minimum domain awareness in the fields of mathematics and statistics. The type of ANN used in this study is called the multilayer perceptron (MLP) or multilayer feed-forward network that propagates input signals forward and returns error signals. During the process the weights are changed to make prediction more accurate. This method is vulnerable to problematic overfitting. Training to the network is provided with some of the data values to prevent overfitting and then test its performance by testing the trained network with the values of the remaining data. We split the data into 70% - 30%. 70% for training our ANN models and the 30% data for testing and validation.

The ANN models consist of an input layer, a hidden layer and an output layer. Table 3 summarizes the training results of ANN models including specificity, sensitivity along with 95% confidence interval and overall accuracy of the model performance.

Table 3. ANN training results: sensitivity, specificity and accuracy

Artificial Neural Network Models	Sensitivity		Specificity		Accuracy
	Value	95% Confidence Interval	Value	95% Confidence Interval	
Model 1	66.78%	65.70 - 67.83	70.76%	70.19 - 71.31	69.85%
Model 2	67.25%	66.19 - 68.27	71.48%	70.91 - 72.03	70.46%
Model 3	68.23%	67.20 - 69.23	72.08%	71.51 - 72.63	71.12%
Model 4	88.95%	88.27 - 89.59	80.60%	80.09 - 81.09	82.80%

The specificity, sensitivity along with 95% confidence intervals and overall accuracy results of the developed ANN models when testing is summarized in Table 4.

Table 4. ANN testing results: sensitivity, specificity and accuracy

Artificial Neural Network Models	Sensitivity		Specificity		Accuracy
	Value	95% Confidence Interval	Value	95% Confidence Interval	
Model 1	66.18%	64.52 - 67.80	70.66%	69.79 - 71.51	69.63%
Model 2	68.66%	67.08 - 70.20	72.00%	71.14 - 72.84	71.20%
Model 3	66.87%	65.26 - 68.43	71.89%	58.27 - 59.97	70.67%
Model 4	89.36%	88.32 - 90.32	80.97%	80.20 - 81.72	83.20%

Architecture of developed artificial neural network models with their ROC values along with specificity at 95% sensitivity are given in Table 5.

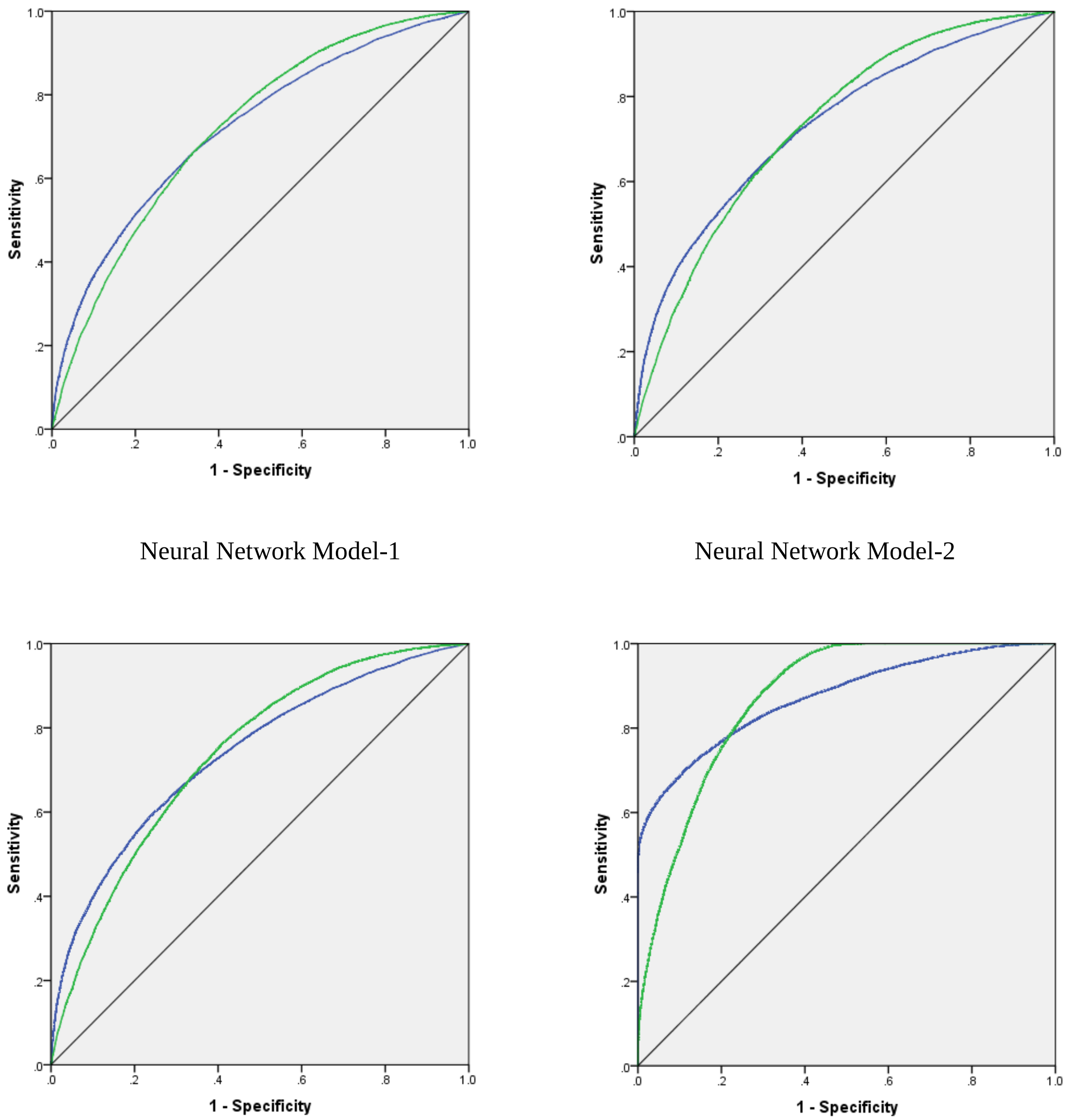
Table 5. Architecture of fitted ANN models and ROC AUC values

Artificial Neural Network Models	Architecture			ROC	At 95% sensitivity
	I	H	O		Specificity
Model 1	2	7	2	72.1%	30%
Model 2	6	3	2	73.1%	32%
Model 3	10	6	2	73.8%	39%
Model 4	11	3	2	87.4%	66%

I, H, O are the number of inputs, hidden and output nodes respectively.

From the training results we notice that the overall accuracy for model-3 to model-4 increased from 71.12% to 82.80%. The neural network model-1 yielded a ROC area of 72.1% and a survival sensitivity of 95% yielded a specificity of 30%, model-2 reported ROC area under curve of 73.1% and a survival sensitivity of 95% provided 32% specificity. The ROC region for the remaining two models is 73.8% and 87.4% respectively, and the 95% survival sensitivity yielded 39% and 66% specificity rates, respectively.

Comparing at 95% sensitivity, all four artificial neural network models has a higher specificity than logistic models. Figure 2 has the area under curve ROC graphs for the four neural network models.



Neural Network Model-1

Neural Network Model-2

Neural Network Model-3

Neural Network Model-4

Figure 2. ROC graphs of ANN models

4. Decision Tree Classification

The effective classification approach is combining data mining tools with decision trees which assists physicians and medical specialists with a clear and easy understanding of classification rules. False positive and false negative decisions can be minimized with the help of methods of data mining. (Quinlan, 1987).

A classification method or a classifier to assess acceptable action is referred as a Decision tree. Root, internal, or test nodes and leaf nodes form the core part of a simple decision tree. The final decisions for the target class are obtained on the leaf nodes by performing split tests within the internal nodes. The leaf node contains, in complex cases, target value or a probability vector for the result. (Zantema and Bodlaender, 2000).

The decision tree is usually composed of continuous or nominal attributes. One outcome is given for the target value when working with nominal attributes, while the continuous attributes will have two outcomes, one for each interval. Decision-makers typically prefer a less complex, decision-making tree. Any path from root to leaf of the decision tree generates a rule by calculating tests along the path that grant terminal node class prediction (Friedman et al., 1996).

4.1 Splitting Techniques

Decision trees also employ univariate separating, i.e., separating steps at each internal node is centred on the single attribute. Upon completion of splitting, the inducer is looking for the best attribute at the internal node. The splitting procedures are used in numerous ways, depending on the tree's initial measure and the level of the tree. For a univariate splitting, the choice of splitting criteria does not affect the performance of the tree.

Multivariate splits are usually based on the linear combination of the input variables. Methods used to find optimal splitting include the greedy search method (Breiman et al., 1984) linear programming (Duda et al., 2012), linear discriminant analysis (Friedman, 1977) and many more.

An alternative approach for maximizing the tree is to allow it to grow and replenish using certain methods of pruning. Pruning methods aim to yield simple trees at relatively small cost of reducing accuracy. There are different pruning methods, such as cost-complexion, reduced and minimal error, optimal, etc.

4.2 Survival Prediction using Decision Trees

The training and testing results of specificity, sensitivity along with 95% confidence intervals and overall accuracy results using both CHAID and CRT methods of the four decision tree models are summarized in Table 6 and Table 7.

Table 6. Decision trees training results: sensitivity, specificity and accuracy

Decision Tree Training		Sensitivity		Specificity		Accuracy
		Value	95% CI	Value	95% CI	
Decision Tree Model 1	CHAID	64.77%	63.73 - 65.8	71.12%	70.60 - 71.73	69.6%
	CRT	66.83%	65.74 - 67.91	70.75%	70.19 - 71.30	69.9%
Decision Tree Model 2	CHAID	69.33%	68.27 - 70.37	71.21%	70.65 - 71.76	70.8%
	CRT	69.2%	68.14 - 70.24	71.2%	70.63 - 71.74	70.7%
Decision Tree Model 3	CHAID	67.59%	66.58 - 68.59	72.25%	71.69 - 72.81	71.1%
	CRT	66.22%	65.22 - 67.21	72.64%	72.07 - 73.20	70.9%
Decision Tree Model 4	CHAID	89.67%	89.00 - 90.30	80.14%	79.63 - 80.63	82.6%
	CRT	93.62%	93.05 - 94.14	79.86%	79.35 - 80.35	83.2%

Table 7. Decision trees testing results: sensitivity, specificity and accuracy

Decision Tree Testing		Sensitivity		Specificity		Accuracy
		Value	95% Confidence Interval	Value	95% Confidence Interval	
Decision Tree Model 1	CHAID	63.66%	62.07 - 65.22	71.22%	70.35 - 72.09	69.3%
	CRT	67.33%	65.66 - 68.97	69.69%	68.83 - 70.55	69.2%
Decision Tree Model 2	CHAID	68.63%	66.97 - 70.24	71.27%	70.40 - 72.12	70.7%
	CRT	67.24%	65.56 - 68.87	70.82%	69.95 - 71.67	70%
Decision Tree Model 3	CHAID	66.32%	64.75 - 67.85	72.19%	71.31 - 73.04	70.7%%
	CRT	65.11%	63.55 - 66.65	71.42%	70.54 - 72.28	69.8%
Decision Tree Model 4	CHAID	88.2%	87.10 - 89.21	80.38%	79.61 - 81.14	82.4%
	CRT	94.02%	93.19 - 94.76	79.93%	79.16 - 80.69	83.2%

The results showed that in a CHAID decision tree, the overall accuracy for the model-3 increased from 71.10% to 82.6% for the model-4. Similarly, the accuracy jumps from 70.9% for model-3 to 83.2% for model-4 for a CRT based decision tree.

The 95% survival sensitivity has returned a specificity of 22, 25, 29 and 62 percent respectively for the four models. CHAID model-4 decision tree results have an overall accuracy of 82.6% with a specificity of 80.14% and a sensitivity of 89.67%. These numbers suggest that model-4 of CHAID is a very well-performing model.

The ROC area under curve for the four models developed on CRT based tree are 71.9%, 72.8%, 72.7% and 87.4% respectively. The 95% survival sensitivity has returned a specificity of 24, 29, 28 and 62 percent respectively for the four models. CRT model-4 decision tree tests have an overall accuracy of 82.2%. The specificity and sensitivity for the model-4 are 79.86% and 93.62%. These numbers indicate that model-4 CRT is a very well-performing model. ROC values of the developed decision tree models using CHAID and CRT are summarized the Table 8.

Table 8. ROC of Decision tree using CHAID and CRT

	Model-1	Model-2	Model-3	Model-4
CHAID	72.0%	73.2%	73.6%	87.6%
CRT	71.9%	72.8%	72.7%	87.4%

Figure 3 has the ROC graphs for the decision trees developed using CHAID technique. From the results of Table 8, we notice that there is a significant increase in the area under curve for model-3 and model-4. Clearly, approximately there is an 88% chance that model-4 can distinguish between positive and negative cases.

Figure 4 has the ROC graphs for the decision trees developed using CRT technique. As observed in CHAID models, even in CRT based models, there is a significant increase in the area under curve for model-3 and model-4 with model-4 having an 87.4% chance of distinguishing between positive and negative cases.

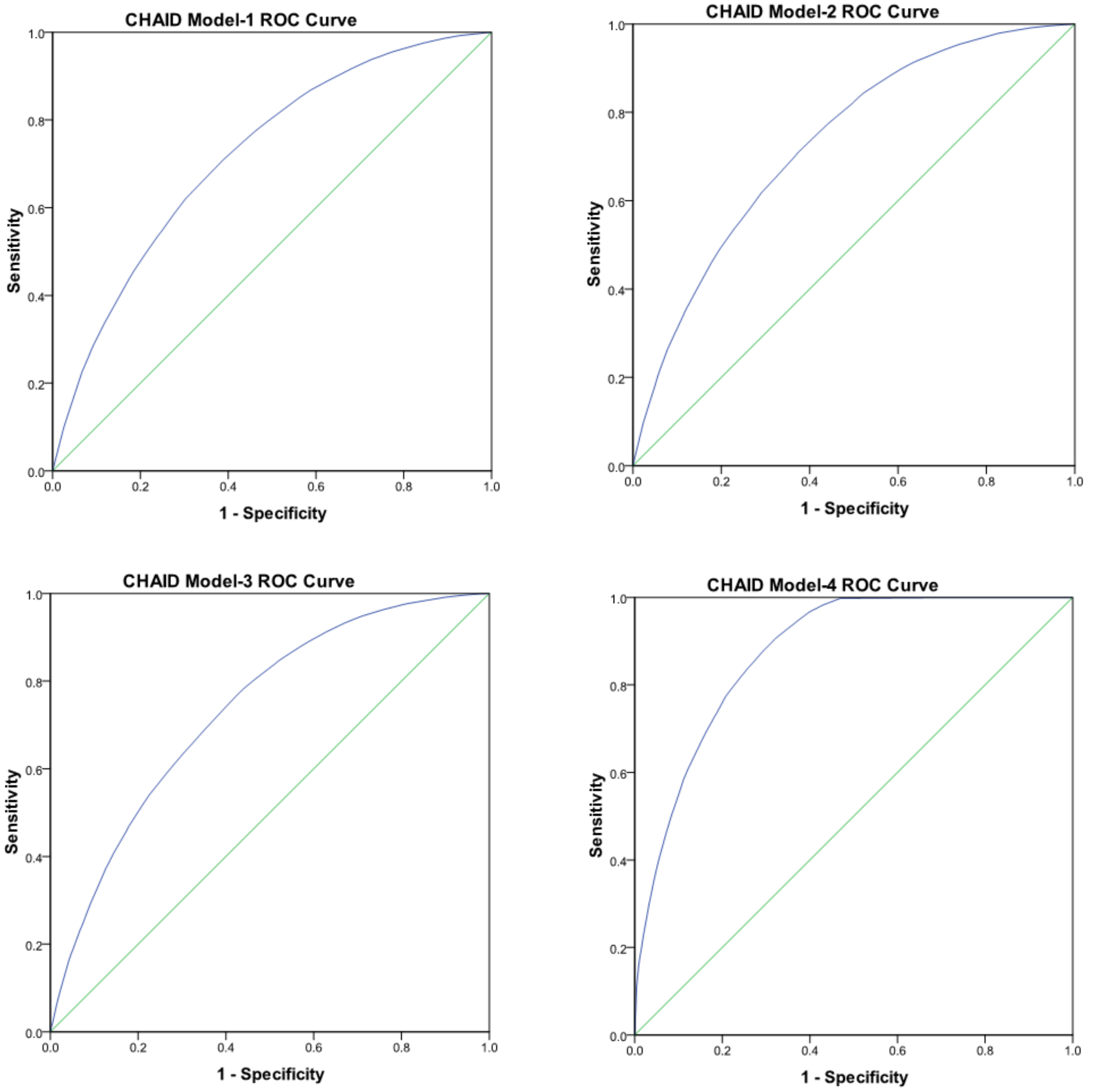


Figure 3. ROC areas for CHAID based Decision trees

By default, CHAID uses multiway splits while CRT uses binary splits. This disparity between the CRT and the CHAID affects even the decision tree structures. Segmentation or classification has various significant applications.

- When the dependent variable is categorical, CHAID, Chi-square Automatic Interaction Detector, splits the tree based on chi-square test. On the other hand, CRT uses impurity reduction as a measure to split.
- In order to produce a smaller tree rather than an exhaustive one, CHAID uses the forward stepwise stopping rule. CRT intentionally overfits and uses validation data to prune back to identify the best and smallest tree.

Finally, if the aim is to identify or explain the relationship between a response variable and a collection of explanatory variables, one may prefer CHAID, while CRT is best suited for constructing a regression model. At this viewpoint, in this paper, we conclude that for the survival classification of breast cancer in women, CHAID decision tree performed marginally better than CRT.

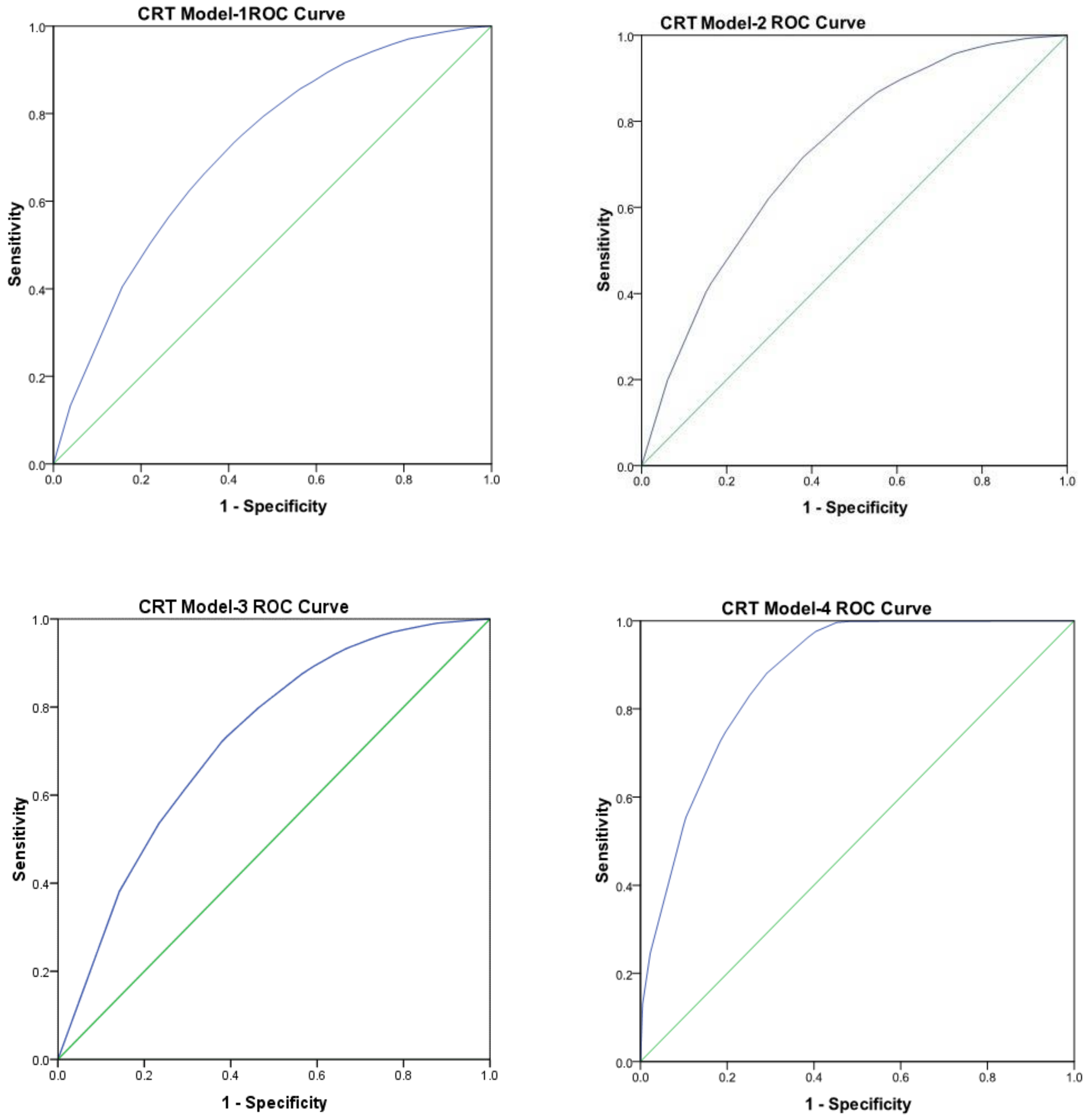


Figure 4. ROC areas for CRT based Decision trees

5. Performance Evaluation of Models

One of four outcomes are possible when evaluating a predictive binary classification model:
(a) a true positive (TP) (b) a false positive (FP) (c) a true negative (TN) (d) a false negative (FN).

The central concern of implementing various modelling applications in this paper is to determine which of the techniques proposed would enhance predictive accuracy. Even a small fraction of a percent change will turn into substantial savings or increased revenues.

The performance of logistic regression models, artificial neural network models and decision tree models in this work is evaluated based on sensitivity, specificity, accuracy and ROC area under curve values. Sensitivity is the proportion of true positive elements that the model correctly classifies. Specificity is the proportion of true negatives that the model correctly classifies.

In general, we compare the area under the ROC curve, which is an easy way to test predictive binary classification models when the analyst or decision-maker does not have any knowledge about the cost or extent of classification errors. We compared the results of the four logistic models with the results of four artificial neural network models and four decision tree models, respectively.

Figure 5 and Figure 6 depicts the comparison of specificity and overall accuracy of the models developed. In comparing the specificity and overall accuracy of all models, model-4 clearly stands out as a better model. Models 1, 2 and 3 resulted with similar results of accuracy, specificity and ROC values.

Table 9 includes the performance assessment of logistic, artificial neural network, and decision tree techniques. Compared with logistic models, neural network and decision tree techniques reported almost the same overall accuracy for correct classification of breast cancer survival in women. Nonetheless, the specificity of the logistic model results is higher than the neural network and decision tree models. Models developed are ranked, given in Table 9, based on their performance and accuracy in classification. For all the four models, the area under the ROC curves of decision tree methods is comparatively higher than the other two techniques.

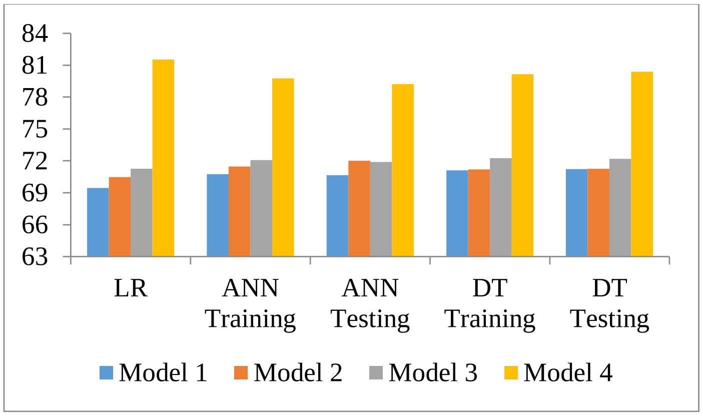


Figure 5. Specificity comparison of Logistic Regression, Artificial Neural Networks and Decision Tree models

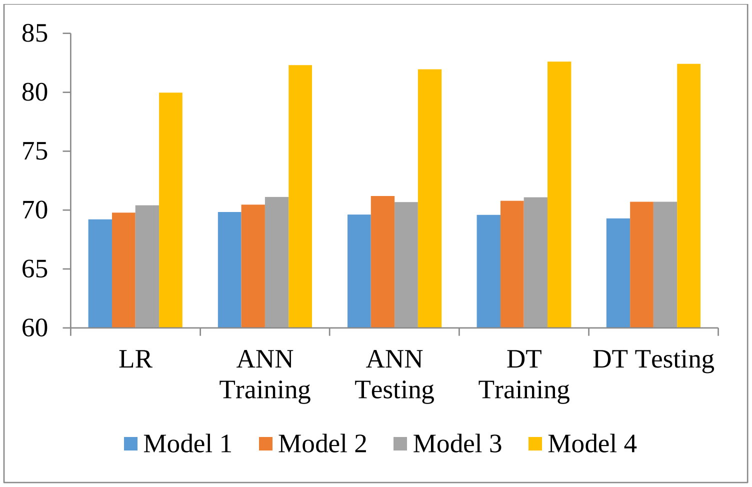


Figure 6. Comparison of overall accuracy of Logistic Regression, Artificial Neural Networks and Decision Tree models

Table 9. Performance evaluation of Logistic, ANN and decision tree models

	Overall Accuracy					Specificity				
Model	LR	ANN		CHAID		LR	ANN		CHAID	
		Train	Test	Train	Test		Train	Test	Train	Test
1	69.2	69.85	69.63	69.6	69.3	69.45	70.76	70.66	71.17	71.22
2	69.78	70.46	71.2	70.8	70.7	70.47	71.48	72	68.63	71.27
3	70.42	71.12	70.67	71.1	70.7	71.26	72.08	71.89	72.25	72.19
4	79.98	82.31	81.95	82.6	82.4	81.54	79.76	79.22	80.14	80.4
Rank	III	II		I		I	III		II	

Table 10 provides the specifics of comparing ROCs of the three different approaches along with their ranking.

Table 10. ROC values of Logistic, ANN and Decision Tree models

Models	LR	ANN	DT
Model-1	68.8%	72.1%	72.0%
Model-2	71.0%	73.1%	73.2%
Model-3	71.8%	73.8%	73.6%
Model-4	85.5%	87.4%	87.6%
Rank	**III**	**II**	**I**

6. Conclusion and Discussion

This research aimed to compare breast cancer survival predictions of the artificial neural networks (ANN), decision trees, and logistic models. In the present research, we established models with the attributes of a woman with breast cancer, including age, tumor size and cancer stage, treatment administered, and duration. Four models with different attributable variables are developed and compared using logistic, ANN and decision trees (using CHAID and CRT algorithms). The degree of generalization (or the precision of predictive ability) was determined and the corresponding predictive abilities of the developed models are assigned a rank in this order. Model-4, Model-3, Model-2, Model-1.

To maintain consistency in comparing these implemented models, we considered the accuracy of three techniques when classifying breast cancer survival results. The receiver operating characteristic (ROC) curves are plotted and estimated the area under the curves. The results indicated that Model-4 yielded better performance using logistic regression, ANN, and decision tree methods. The model classification accuracy is obtained as 79.98%, 82.31% and 82.6% respectively for logistic regression, ANN and decision tree models. In the models developed using ANN and decision tree methods, we find no significant difference in the results obtained. However, the ANN method provided a better specificity when compared to the logistic and decision tree models at a 95% sensitivity. In this study, the artificial neural networks and decision tree methods reported a modest improvement in the outcomes compared with logistic regression. Figure 7 is a comparison of the ROC curves of model-4 using all three techniques.

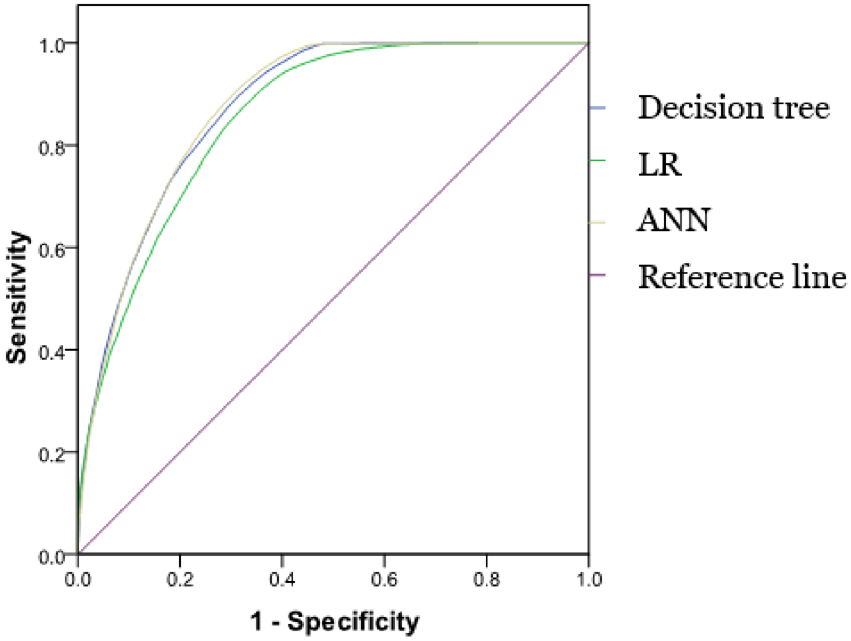


Figure 7. ROC comparison of the logistic regression, artificial neural network and decision tree models

We conclude in this analysis that ANN and decision tree models have a higher predictive probability compared to the logistic model.

Conflict of Interest

The authors confirm that there is no potential conflict of interest to publish the paper in the journal.

Acknowledgements

This research paper is the revised and extended version of the chapter 6 of the first author Ph.D. dissertation. We would like to thank Dr. Gangaram Ladde for his valuable suggestions in completing this research work. We would also like to thank Ms. Saisumasri Komarraju, for her unconditional support in formatting this article for publication.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*. John Wiley & Sons. New Jersey.

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). *Classification and regression trees*. Chapman & Hall/CRC press.

Duda, R.O., Hart, P.E., & Stork, D.G. (2012). *Pattern classification*. John Wiley & Sons. New York.

Friedman, J.H. (1977). A recursive partitioning decision rule for nonparametric classification. *IEEE Transactions on Computers*, *26*(4), 404-408.

Friedman, J.H., Kohavi, R., & Yun, Y. (1996, August). Lazy decision trees. In *AAAI-96 Conference Proceedings* (Vol. 1, pp. 717-724).

Gellar, J.E., Colantuoni, E., Needham, D.M., & Crainiceanu, C.M. (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association*, *109*(508), 1425-1439.

Jabri, M., & Flower, B. (1992). Weight perturbation: an optimal architecture and learning technique for analog VLSI feedforward and recurrent multilayer networks. *IEEE Transactions on Neural Networks*, *3*(1), 154-157.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed) *Frontiers in Econometrics*. Academic Press, New York, pp. 105-142.

Mohr, A.E., Riskin, E.A., & Ladner, R.E. (2000). Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction. *IEEE Journal on Selected Areas in Communications*, *18*(6), 819-828.

Mudunuru, V.R. (2016). Comparison of activation functions in multilayer neural networks for stage classification in breast cancer. *Neural, Parallel, and Scientific Computations*, *24*, 83-96.

Quinlan, J.R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, *27*(3), 221-234.

Widrow, B., & Lehr, M.A. (1990). 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, *78*(9), 1415-1442.

Zantema, H., & Bodlaender, H.L. (2000). Finding small equivalent decision trees is hard. *International Journal of Foundations of Computer Science*, *11*(02), 343-354.

Appendix

1. Data Source and Block Diagram

- Breast Cancer Data has been requested from https://seer.cancer.gov/seertrack/data/request/
- Pre-process the data using excel (and macros) for missing values, removing the unwanted columns and prepare the columns identifying attributable variables, and by merging columns (if needed).
- Our Final dataset included the following variables: age, tumor size, stage of cancer, treatment, duration, and censor.
- Sample of our data is available upon request.

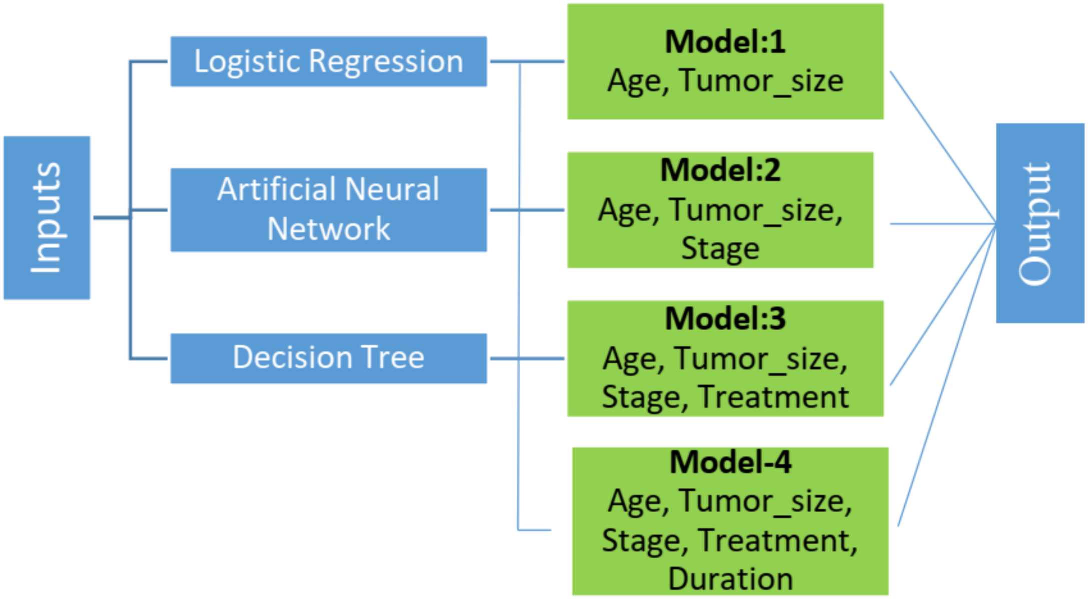


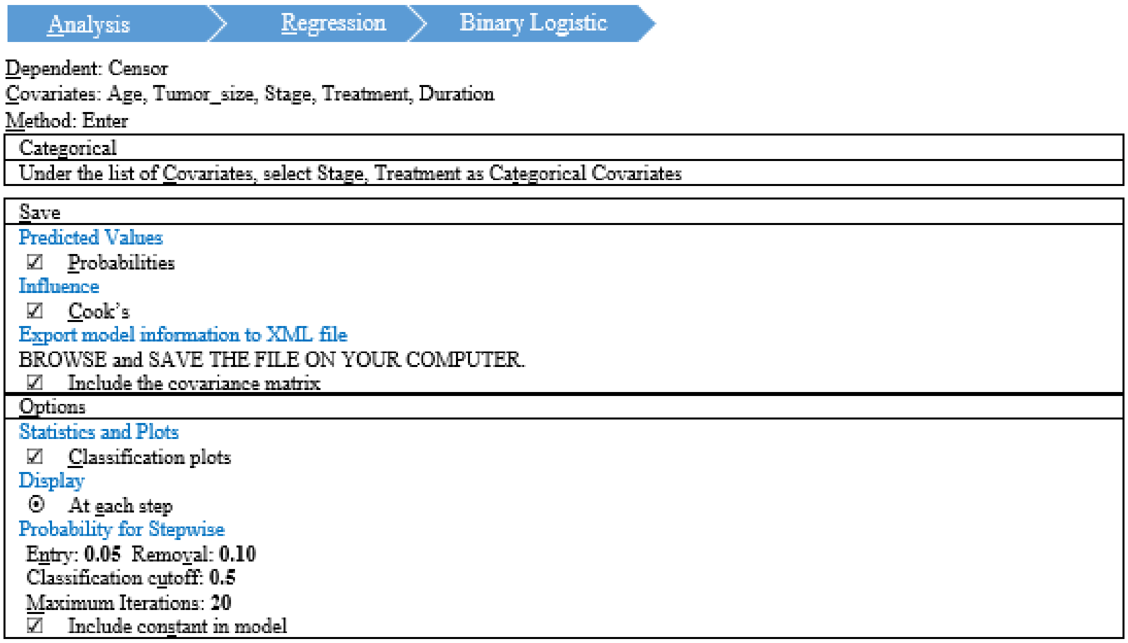
Figure 8. Block Diagram

Figure 8 is the complete block diagram of the research conducted in this paper.

2. Software

IBM SPSS Statistics, Version 22.

3. SPSS Set Up for Logistic Regression (LR) Models



Dependent: Censor
Covariates: Age, Tumor_size, Stage, Treatment, Duration
Method: Enter

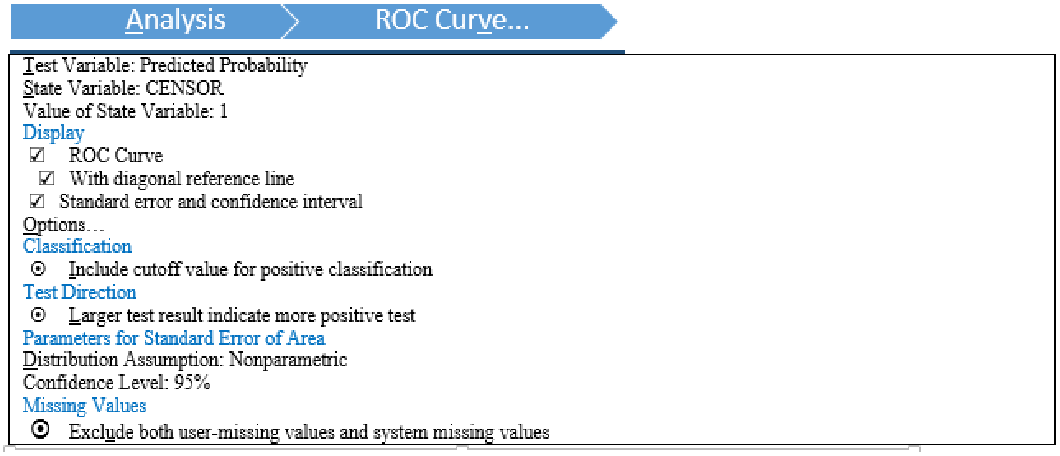
Categorical
Under the list of Covariates, select Stage, Treatment as Categorical Covariates

Save
Predicted Values
☑ Probabilities
Influence
☑ Cook's
Export model information to XML file
BROWSE and SAVE THE FILE ON YOUR COMPUTER.
☑ Include the covariance matrix

Options
Statistics and Plots
☑ Classification plots
Display
⦿ At each step
Probability for Stepwise
Entry: **0.05** Removal: **0.10**
Classification cutoff: **0.5**
Maximum Iterations: **20**
☑ Include constant in model

Figure 9. SPSS set up for Logistic Regression Models

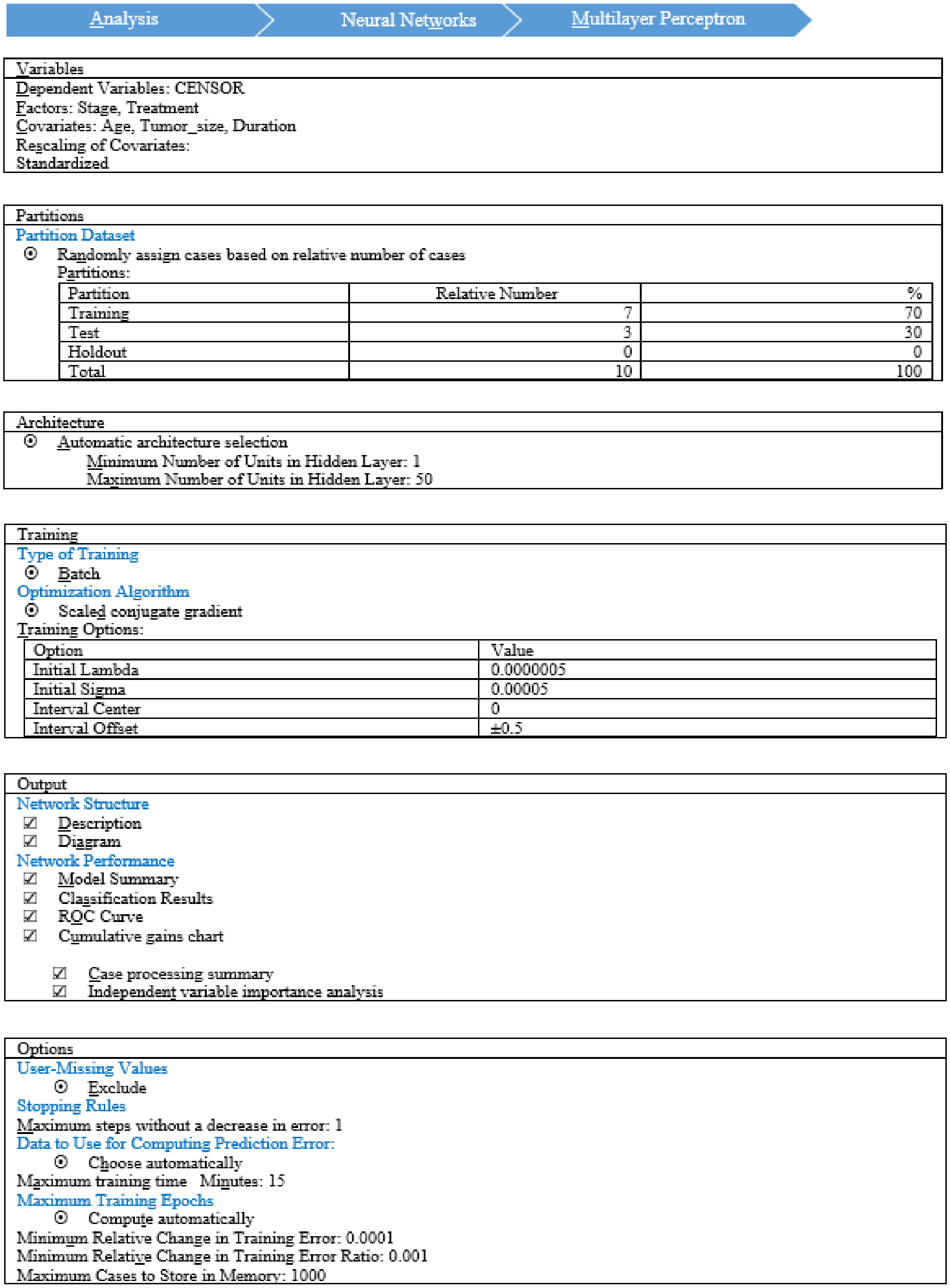
4. ROC Curve for Logistic Regression



Test Variable: Predicted Probability
State Variable: CENSOR
Value of State Variable: 1
Display
☑ ROC Curve
☑ With diagonal reference line
☑ Standard error and confidence interval
Options…
Classification
⦿ Include cutoff value for positive classification
Test Direction
⦿ Larger test result indicate more positive test
Parameters for Standard Error of Area
Distribution Assumption: Nonparametric
Confidence Level: 95%
Missing Values
⦿ Exclude both user-missing values and system missing values

Figure 10. ROC Curve for Logistic Regression

5. SPSS Set Up for Artificial Neural Network (ANN) Models



Variables

Dependent Variables: CENSOR
Factors: Stage, Treatment
Covariates: Age, Tumor_size, Duration
Rescaling of Covariates:
Standardized

Partitions

Partition Dataset

⦿ Randomly assign cases based on relative number of cases

Partitions:

Partition	Relative Number	%
Training	7	70
Test	3	30
Holdout	0	0
Total	10	100

Architecture

⦿ Automatic architecture selection
Minimum Number of Units in Hidden Layer: 1
Maximum Number of Units in Hidden Layer: 50

Training

Type of Training
⦿ Batch
Optimization Algorithm
⦿ Scaled conjugate gradient
Training Options:

Option	Value
Initial Lambda	0.0000005
Initial Sigma	0.00005
Interval Center	0
Interval Offset	±0.5

Output

Network Structure
☑ Description
☑ Diagram
Network Performance
☑ Model Summary
☑ Classification Results
☑ ROC Curve
☑ Cumulative gains chart

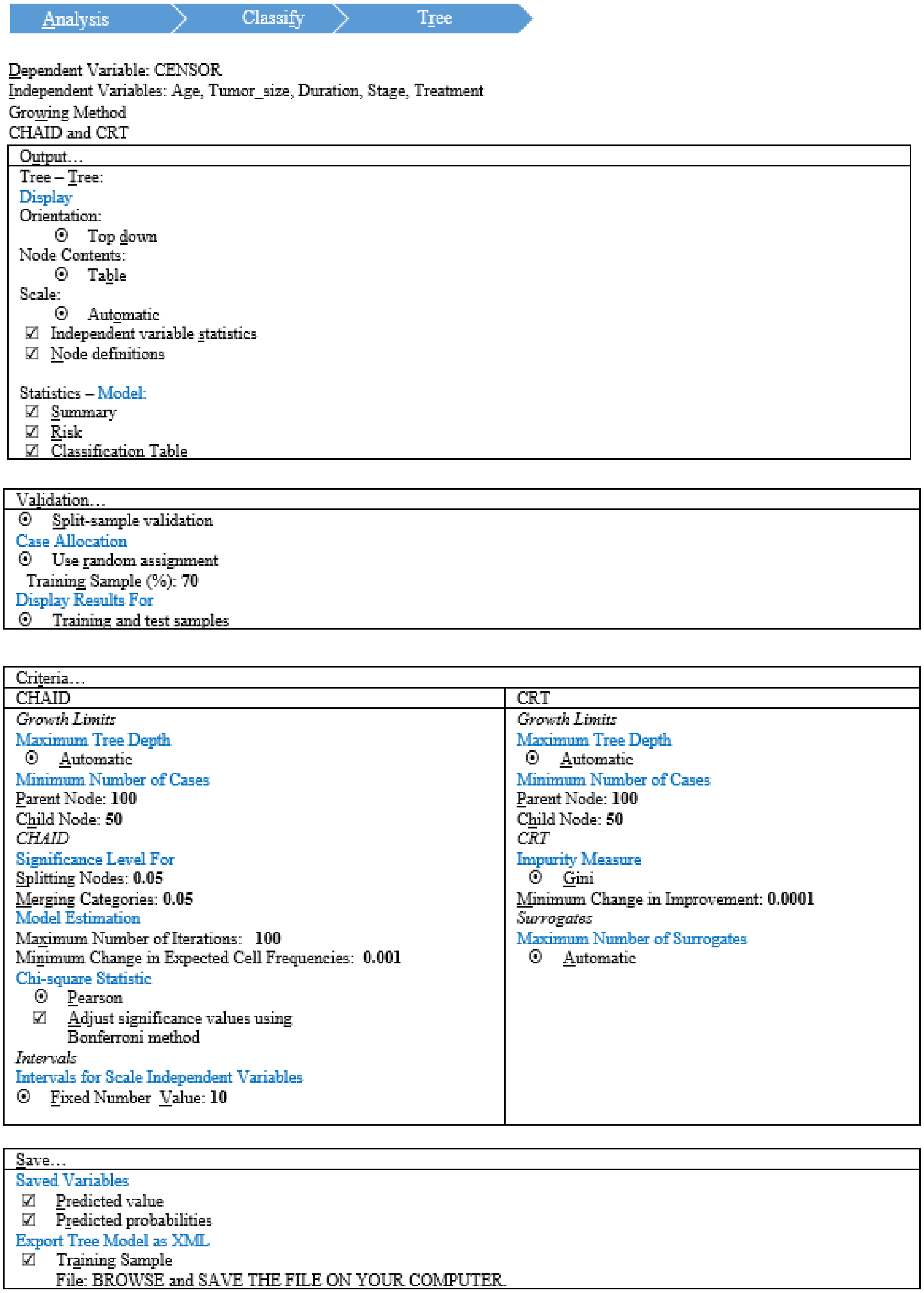
☑ Case processing summary
☑ Independent variable importance analysis

Options

User-Missing Values
⦿ Exclude
Stopping Rules
Maximum steps without a decrease in error: 1
Data to Use for Computing Prediction Error:
⦿ Choose automatically
Maximum training time Minutes: 15
Maximum Training Epochs
⦿ Compute automatically
Minimum Relative Change in Training Error: 0.0001
Minimum Relative Change in Training Error Ratio: 0.001
Maximum Cases to Store in Memory: 1000

Figure 11. SPSS set up for Artificial Neural Networks

6. SPSS Set Up for Decision Tree (DT) Models



Analysis > Classify > Tree

Dependent Variable: CENSOR
Independent Variables: Age, Tumor_size, Duration, Stage, Treatment
Growing Method
CHAID and CRT

Output…
Tree – Tree:
Display
Orientation:
⊙ Top down
Node Contents:
⊙ Table
Scale:
⊙ Automatic
☑ Independent variable statistics
☑ Node definitions

Statistics – Model:
☑ Summary
☑ Risk
☑ Classification Table

Validation…
⊙ Split-sample validation
Case Allocation
⊙ Use random assignment
Training Sample (%): **70**
Display Results For
⊙ Training and test samples

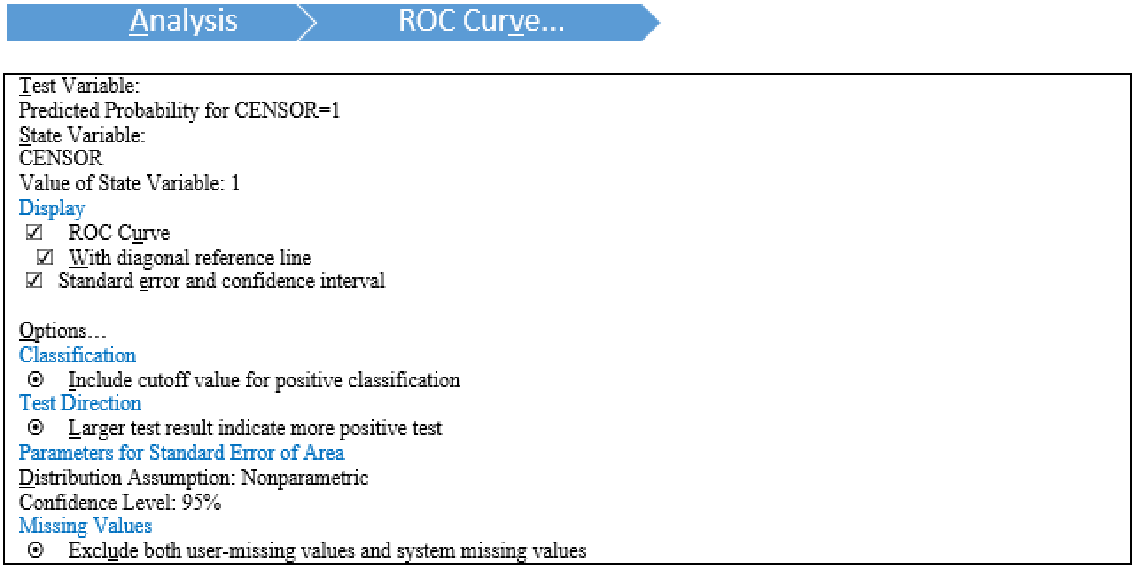
Criteria…

CHAID	CRT
Growth Limits	*Growth Limits*
Maximum Tree Depth	Maximum Tree Depth
⊙ Automatic	⊙ Automatic
Minimum Number of Cases	Minimum Number of Cases
Parent Node: **100**	Parent Node: **100**
Child Node: **50**	Child Node: **50**
CHAID	*CRT*
Significance Level For	Impurity Measure
Splitting Nodes: **0.05**	⊙ Gini
Merging Categories: **0.05**	Minimum Change in Improvement: **0.0001**
Model Estimation	*Surrogates*
Maximum Number of Iterations: **100**	Maximum Number of Surrogates
Minimum Change in Expected Cell Frequencies: **0.001**	⊙ Automatic
Chi-square Statistic	
⊙ Pearson	
☑ Adjust significance values using Bonferroni method	
Intervals	
Intervals for Scale Independent Variables	
⊙ Fixed Number Value: **10**	

Save…
Saved Variables
☑ Predicted value
☑ Predicted probabilities
Export Tree Model as XML
☑ Training Sample
File: BROWSE and SAVE THE FILE ON YOUR COMPUTER.

Figure 12. SPSS set up for Decision Tree Models

7. ROC Curve for Decision Tree



Test Variable:
Predicted Probability for CENSOR=1
State Variable:
CENSOR
Value of State Variable: 1
Display
☑ ROC Curve
☑ With diagonal reference line
☑ Standard error and confidence interval

Options…
Classification
⦿ Include cutoff value for positive classification
Test Direction
⦿ Larger test result indicate more positive test
Parameters for Standard Error of Area
Distribution Assumption: Nonparametric
Confidence Level: 95%
Missing Values
⦿ Exclude both user-missing values and system missing values

Figure 13. ROC curve for Decision Tree

Figure 9 explains the steps involved in modelling a logistic regression. Figure 10 gives the steps for generating ROC curves for a logistic regression model.

Figure 11 is a step-by-step break down for performing a neural network analysis along with generating ROC curves.

Figure 12 has the steps involved in performing a decision tree analysis by both CHAID and CRT techniques. Figure 13 gives the steps for generating ROC curves for a decision tree model.

8. Models Comparison

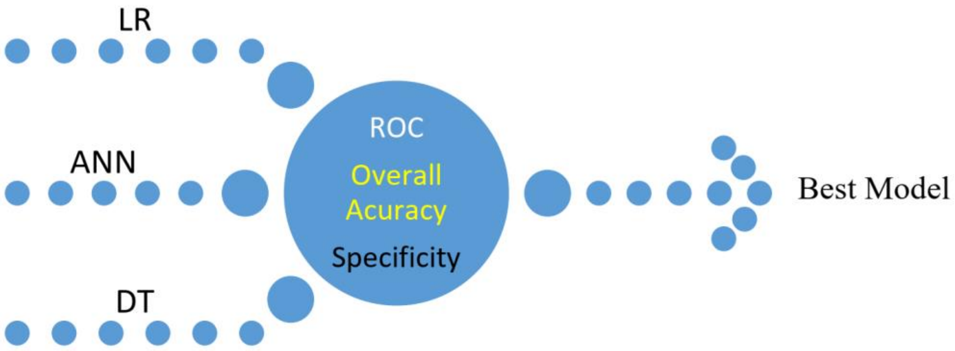


Figure 14. Models Comparison

Figure 14 above is the comparison of the models with evaluating ROC, overall accuracy and specificity factors in order to identify the best classification model.