Journal of Mathematics and Statistics Studies
ISSN: 2709-4200
DOI: 10.32996/jmss
Journal Homepage: www.al-kindipublisher.com/index.php/jmss

JMSS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Forecasting Breast Cancer: A Study of Classifying Patients' Post-Surgical Survival Rates with Breast Cancer

Md Nurul Raihen[1] ✉ and Sultana Akter[2]

[1]*Visiting Assistant Professor, Department of Mathematics and Statistics, Stephen F. Austin State University, Nacogdoches, TX, USA*
[2]*MS Student in Statistics, Department of Statistics, Western Michigan University, Kalamazoo, MI, USA*

Corresponding Author: Md Nurul Raihen, **E-mail**: Nurul.Raihen@sfasu.edu

| ABSTRACT

Breast cancer is the most lethal form of cancer that can strike women anywhere in the world. The most complex and tough undertaking in order to lower the death rate is the process of predicting a patient's likelihood of survival following breast cancer surgery. Due to the fact that this survival prediction is linked to the life of a woman, effective algorithms are required for the purpose of making the prognosis. It is of the utmost importance to accurately predict the survival status of patients who will have breast cancer surgery since this shows whether or not doing surgery is the actual approach for the specific medical scenario. Given the gravity of the situation, it is impossible to overstate how important it is to investigate new and improved methods of prediction in order to guarantee an accurate assessment of the patient's chances of survival. In this paper, we collect data and examine some models based on the survival of patients who underwent breast cancer surgery. The goal of this research is to evaluate the forecasting performance of various classification models, including the Linear regression model, logistic regression analysis, LDA, QDA, KNN, ANN, and Decision Tree. The results of the experiment on this dataset demonstrate the better performance of the came up with ANN approach, with an accuracy of 82.98 percent.

| KEYWORDS

Prediction, Breast cancer, Classification, Regression analysis, KNN, ANN, Machine learning models

| ARTICLE INFORMATION

ACCEPTED: 13 May 2023 **PUBLISHED:** 15 May 2023 **DOI:** 10.32996/jmss.2023.4.2.8

1. Introduction

When ranked by mortality rate, breast cancer is second only to lung cancer in terms of severity. Approximately 30% of all newly diagnosed cancer patients are women with breast cancer (Jemal, 2005). Regular sentinel node biopsies are performed on patients with proven breast cancer to check for cancerous cells in the lymph nodes. The lymph nodes are small, bean-shaped organs that serve as filters for the lymph fluid pathways. The lymphatic system is similar to the circulation (blood) system in that it travels the entire body. It transports both liquids and cells. Axillary lymph nodes are the primary lymph node distribution site for breast cancer metastasis (Willard-Mack, 2006).

When it comes to evaluating and figuring out the prediction of models, there are a lot of different ways that traditional methods can be used (Raihen et al., 2023). On the other hand, these tend to be time-consuming and costly. Some methods can be inconsistent and inefficient. Because of these drawbacks, researchers have been working hard to come up with new techniques for evaluating the basic features of data, such as Haberman's survival. One such option is a machine learning system. Karimi et al., 2017; Mollazade et al., 2012 Stein and Raihen et al. 2023 have described machine learning allows for the automated extraction of visual features for use in classifying the predictor variable [Karimi, Friedman, 1997; Mollazade, 2012, Stein, 2023].

Polat (2007) and Shawarib (2020) have discussed that in the field of information technology (IT), an artificial neural network (ANN) is a computerized system designed to mimic the functioning of a real brain's neural networks. ANNs, or just neural networks, are one type of deep learning technology that is part of the larger field of AI [Polat, 2007, Karimi, 2017].

Tumors are often classified as benign or malignant using Machine Learning methods. Prediction and survival rates for individuals with breast cancer can be improved through early detection [Akay et al., 2009]. This will aid patients in receiving timely medical care (Karimi et al., 2017). Patients with benign tumors may prevent unnecessary tests and surgeries. When used in medicine, the Artificial Neural Networks method can improve healthcare value by predicting outcomes, reducing costs, and saving lives. Artificial Neural Networks are a useful tool for identifying tumors.

In this research, a deep learning and artificial intelligence system were created to sort Haberman survival breast cancer varietals like the patient survived 5 years or more and the patient died within 5 years. The second part of the research describes the process of acquiring the data; the manipulations applied to them, the feature extraction phase, the performance evaluation, and the cross validation. The study's classification models are discussed in detail in the third and fourth parts. The study's findings are broken down in depth in section four. Discussion questions are provided in the final portion.

2. Literature Review

Big Data is a very powerful tool for fighting breast cancer. The rise of data mining in healthcare, paired with powerful machine learning, is set to make advanced predictive analysis a game changer in terms of lowering risk, diagnosing the disease early, and lowering mortality rates from breast cancer. According to the American Cancer Society, there will be around 276,480 new cases of invasive breast cancer in 2020 (Breast Cancer Statistics, Boston College). In addition to that, another 48,530 women have been diagnosed with the non-invasive form of breast cancer known as "carcinoma in situ" (CIS). CIS is the first stage of breast cancer. It is estimated that breast cancer will claim the lives of around 42,170 women in the United States in the year 2023.

That's why it's important to consider the research carried out in recent years using machine learning systems logistic regression, SVM, KNN, and decision tree techniques on cancer disease, in particular breast cancer. Okamura et al. 1993 used a Bayesian classifier for classification. The results indicated that the classification was more accurate than those obtained through manual human effort [Okamura, 1993].

The principal component analysis (PCA) method was used by Karimi (2017) to identify the best aspect of features from data. For instance, ANN, KNN, and SVM were utilized in the food classification process. They used the SVM technique to get more efficient and accurate classification results. MATLAB was used by Angadi and Hiregoudar (2016) to classify a set of data. In their study, they achieved an average of 85% accuracy using color and size features for raisin data.

The author Abbas Hussein (2002) provided a novel method for the identification of breast cancer. This method makes use of statistical methods in conjunction with swarm optimization, and the author reported that it had an accuracy of 88.71%. The Hidden Markov model was proposed by Bahrampour et al. [2015] in order to predict the mortality from breast cancer. The percentage of correct classifications achieved by using this method is 0.939. In reference to [Delen, 2005], three different data mining strategies were offered to estimate breast cancer patients' chances of survival. Artificial neural networks, decision trees, and logistic regression are the three methods that they offer for data mining. These methods have an accuracy of 81.2%, 83.6%, and 88.2%, respectively.

3. Methodology

We compare the predictive abilities of the multinomial logistic regression, LDA, QDA, KNN, ANN, and decision tree with a cross-validated K to determine the error rate of the Bayes classifier. The error rate of the tests was measured using a cross-validation procedure with a 10-fold sampling size. To create a fair comparison between methods, we use 60% for the training set and 40% for the test set, and sometimes 50% for each, and then see which one yields the best results for this classification. We plan on calculating the RMSE with PCR and PLS regression and using lasso regression for regularization. Given that the dataset is manageable in size. So, we delve into the numbers to investigate each group and determine why they fit into this categorization the way they do.

3.1 Data Source

Patients who had breast cancer surgery are included in this data collection that tracks their outcomes after treatment. For this study, we made use of the Census Income Data Set from the UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival). Patients who had breast cancer surgery at the University of Chicago's Billings Hospital between 1958 and 1970 are represented in this collection.

3.2 Data Structure

In this article, I detailed the various data analytic operations I performed on this dataset, as well as the methods I used to draw conclusions about the survival rates of surgical patients and to predict those rates. In order to properly conduct any data analysis or data operation, we must first possess extensive background knowledge in the field in question. Therefore, I will discuss data set characteristics and how they relate to one another. Three of the attributes in this data collection are features, and one is classification. There are also 306 data occurrences, where data represents.

Lymph Node: Bean-shaped lymph nodes cleanse lymph fluid as they travel through the body. Lymph nodes work to prevent cancer cells from spreading throughout the body by filtering lymph fluid as it drains from the breast and into the circulatory system. Lymph nodes under the elbow that contain cancer cells indicate a more aggressive form of the disease. Axillary cells are what we see in our data (0–52).

Age: An individual's chronological age at the time of operation. (Age from 30 to 83).

Operation year: The year that the patient had their operation (1958–1969).

Survival Status: It indicates whether or not the patient lived for more than 5 years after treatment. 1 indicates that the patient lived for 5 years or more, and a status of 2 indicates that the patient did not make it to year 5. The data is clean and suitable for analysis because all but the classification variables are numerical. Age is also a key factor in this case, which I will consider in my mortality analysis. Table 1 displays the range of the input factors that were considered in this analysis.

Table 1. Attribute name and classification

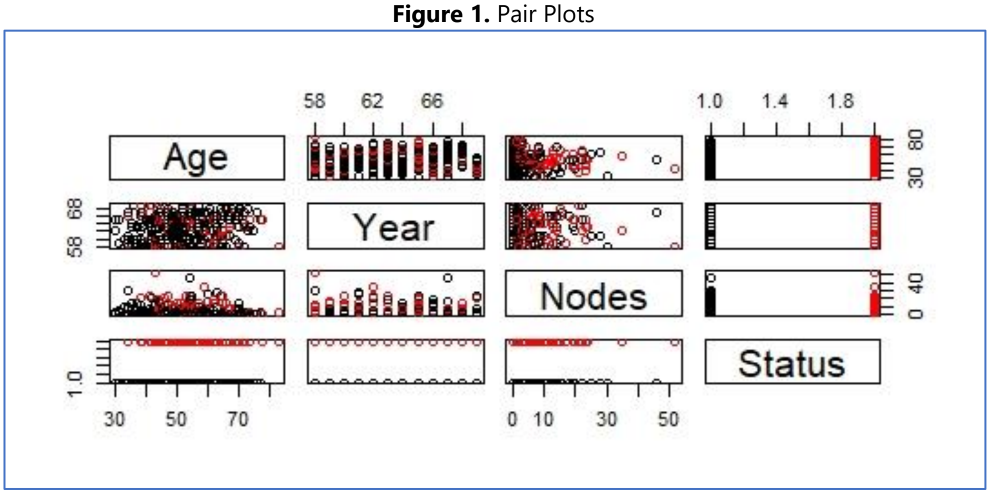
Attribute Name	Category
Patient's age at the time of surgery	Numerical
The patient's surgery years (1958 and 1970)	Year-1900, numerical
The number of positive axillary nodes discovered	Numerical
Survival Status	Class: 1 = The patient survived 5 years or more 2 = The patient died within 5 years

There are four variables and a total of 306 data points in the collection; perhaps I can get by with the first 200. where a survivor status of 1 indicates that the patient lived for 5 years or more and a status of 2 indicates that the patient did not make it to year 5. The data is clean and suitable for analysis because all but the classification variables are numerical. Age is also a key factor in this case, which I will consider in my mortality analysis.

3.3 Data Plot

Using the given data, Figure 1 generates an Axes grid where each variable in the y-axis will be common to a single row, and each variable in the x-axis will be common to a single column.

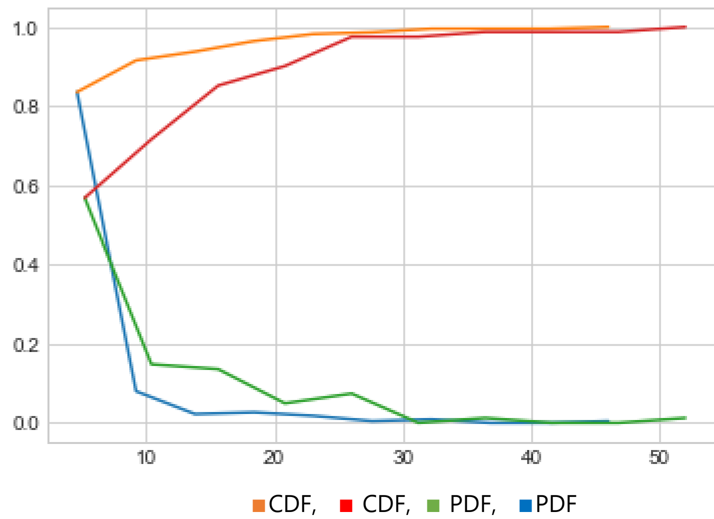
Figure 1. Pair Plots



- ■ The color black indicates that the patient has survived for at least five years.
- ■ The color red indicates that the patient died within 5 years.

The cumulative distribution function (CDF) of each independent variable against the class attribute gives us the distribution across the range of its value. The graph that depicts the CDF together with its PDF can be found in Figure 2.

Figure 2. PDF and CDF Plots

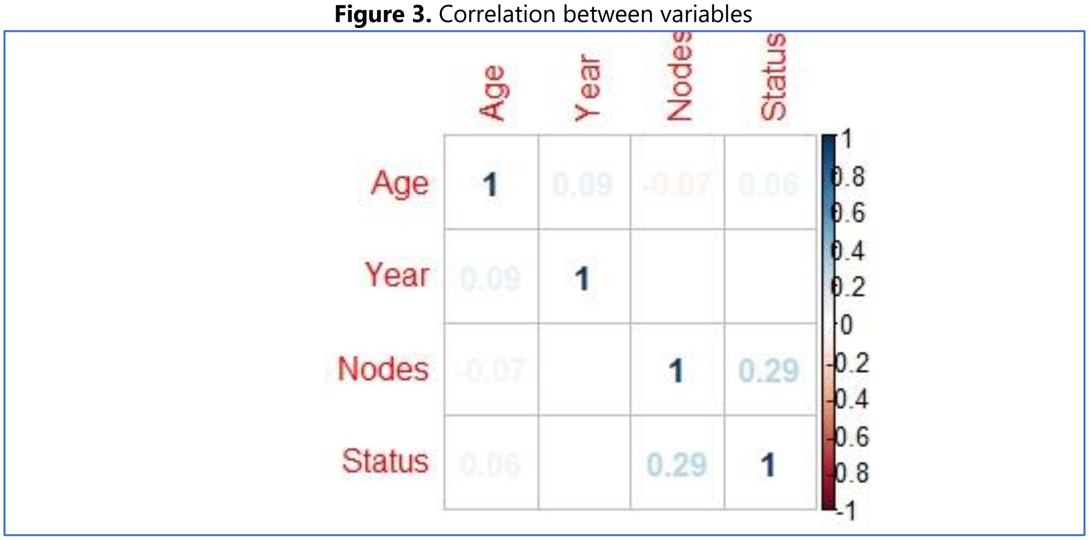


Cumulative density function (orange curve) for survival years plotted as a function of node count, where node count is independent of class 2. The CDF provides age-related data, such as the fact that sixty percent of patients with a five-year or longer survival rate following surgery had ten or fewer lymph nodes and that the median number of lymph nodes was fifty-five. The blue line in the following PDF figure depicts the distribution of patients' nodes across the node count. According to Blueline (PDF), no one with a diagnosis of more than 30 nodes has lived for more than 5 years after surgery.

3.4 Correlation

The intensity and direction of a relationship between two variables can be determined by the correlation coefficient, which is a value between -1 and 1. When conducting correlational research, you look into whether or not the changes that occur in one variable are connected with the changes that occur in other variables. The color blue draws attention to the highest degree of correlation between the variables, in which the nodes' status and each of the other variables is significantly associated with one another.

Figure 3. Correlation between variables



From Figure 3, we observe that nodes and status are highly correlated. The cancerous cells that have broken free from the original tumor are what cause the disease to spread. These cells have the greatest potential to spread to the lymph nodes that are located closest to the breast, which is damaged when breast cancer is developed. The involvement of lymph nodes is a crucial component in both the staging process and the treatment of breast cancer.

3.5 Regression Analysis

We performed regression analysis for predicting the relationships between a dependent variable (often called the "outcome" or "response" variable, or a "label" in machine learning jargon) and one or more independent variables (often called "predictors," "covariates," "explanatory variables," or "features"). In our case, we have

- Predicted variables: Status.
- Predictor variables: Age, Year, and Nodes.

Table 2. Predicted and Predicted variables analysis

	Linear Regression	**Logistic Regression**
Predicted Variables	Status	Status, family=binomial
Predictor Variables	Age, Year, Nodes	Age, Year, Nodes
Residuals	Min: -1.0169 1Q: -0.2390 Median: -0.1956 3Q: 0.3925 Max: 0.8657	Min: -2.3186 1Q: -0.7296 Median: -0.6581 3Q: 0.9320 Max: 1.9542
Coefficients	Pr (>|t|) Intercept 0.0214 Age 0.1268 Year 0.8427 Nodes 2.33e-0.7	Pr (>|z|) Intercept 0.487 Age 0.131 Year 0.824 Nodes 8.88e-06
Significance codes	Residual Std. error: 0.4243, Multiple R-squared: 0.08901 P-value: 3.45e-0.6	AIC: 335.95 Number of Fisher scoring iterations: 4

After carrying out a logistic regression using the backward deletion method, we find that the Nodes variable has a statistically significant impact when compared to the other variables, and the AIC value comes in at 335.95.

Combinations of two category variables can be seen as frequencies in a contingency table. Rows in a contingency table represent categories for one variable, while columns represent categories for another. We perform an analysis of the error rate produced by the logistic regression model and obtain the contingency table below.

Table 3. Contingency table for regression analysis

Status	**False**	**TRUE**
1	214	10
2	68	13

The contingency table reveals that the error rate for survival status after surgery is 0.2557377, which is equivalent to 25.57%. Therefore, logistic regression accurately classifies 74% of the data.

3.6 Other Model Analysis

Data can be better understood when visualized with the help of data analysis. The use of data in the form of models for data analysis allows for the examination of correlations between variables, the prediction of outcomes, and the informing of decision making.

3.6.1 LDA, QDA and KNN Analysis

The LDA methods and logistic regression are closely related to one another. Logistic regression and LDA both generate linear decision boundaries. In a non-linear setting, KNN is better than LDA and Logistic. QDA is a middle ground between KNN, LDA, and logistic regression. The LDA and logistic regression techniques tend to perform well when the true decision boundaries are linear. QDA may produce better outcomes when the boundaries are moderately non-linear. A non-parametric technique, such as KNN, can be superior for much more intricate decision boundaries. However, the level of smoothness for a non-parametric approach must be carefully chosen.

Table 4. LDA, QDA and KNN analysis

Model	Contingency Table	Error Rate
Linear Discriminant Analysis (LDA)	0 1 1 214 10 2 66 15	0.2491803
Quadratic Discriminant Analysis (QDA)	0 1 1 213 11 2 64 17	0.2459016
K Nearest Neighborhood (KNN)	0 1 1 107 9 2 26 10	0.2302632

■ Status = 1 ■ Status =2, ■ True=1, ■ False=0

3.6.2 Neural Network

The human brain serves as inspiration for a subfield of machine learning known as neural networks or simulated neural networks. They function similarly to biological neurons when it comes to coordinating to reach a conclusion. There are three layers to a neural network: the input, the hidden, and the output. In a deep neural network, data is sent into the first layer, where it is processed by several hidden layers before being output by the last layer (Charro, Fernando, Alaa Haj Ali, Nurul Raihen, Monica Torres, and Peiyong Wang, 2023).

An input layer, several hidden layers, and an output layer make up a feedforward neural network. Since no backpropagation occurs, this type of learning is known as "feedforward." The fields of classification, speech recognition, face recognition, and pattern recognition make extensive use of them (Raihen et a., 2022, Verma et al., 2006).

For complex machine-learning tasks, a wide variety of neural network types are used. There is no universally applicable model architecture in our current toolkit. Frank Rosenblatt developed the first successful neural network architecture, the Perceptron, in 1958. The five most used neural network architectures in Computing.

Feedforward Neural Networks (FNNs): An input layer, several hidden layers, and an output layer make up a feedforward neural network. Since no backpropagation occurs, this type of learning is known as "feedforward." The fields of classification, speech recognition, face recognition, and pattern recognition make extensive use of them (Zeng, 2014).

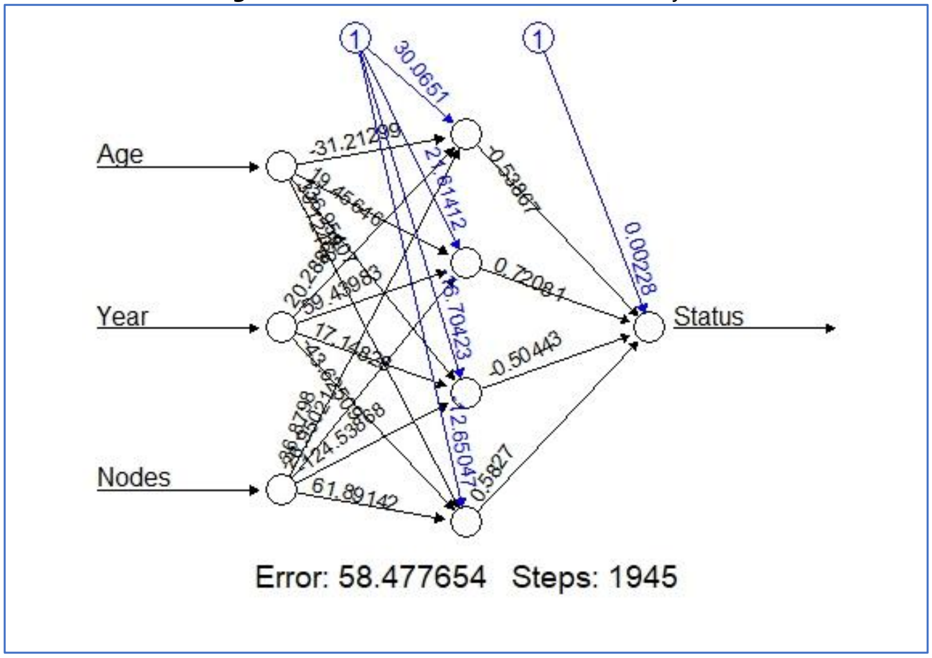
Multi-Layer Perceptrons (MLPs): The inability of a feedforward neural network to learn by backpropagation is addressed by Multi-Layer Perceptrons (MLPs). It's two-way, with many secret levels and activation functions. In MLPs, inputs are propagated forward, and weights are updated by backpropagation. They are the backbone of modern artificial intelligence, enabling everything from computer vision to language technology (Montazeri, 2014).

Convolutional Neural Networks (CNNs): Computer vision, image identification, and pattern recognition are typical applications of Convolution Neural Networks (CNN). Important picture features can be extracted with its help, thanks to its many convolutional layers. CNN's convolutional layer generates a map by convolving over pictures with a custom-made matrix (filter). Convolutional neural networks often include several layers: input, convolution, pooling, fully linked, and output (Sivachitra et al., 2015).

Recurrent Neural Networks (RNNs): Sequential data, such as texts, visual sequences, and time series, are ideal for Recurrent Neural Networks (RNNs). They function similarly to feed-forward networks, except instead of receiving input from a single sequence, they get input from multiple sequences. Non-Linguistic Processing, Sales Forecasting, and Climate Prediction All Make Use of RNNs (Friedman, 2005).

As can be seen in Figure 4, the Multilayer Perceptron architecture was used to construct the neural network, which consists of an input layer, two hidden layers with four 1x1 nodes, and an output layer.

Figure 4. Neural Network for two hidden layers



3.6.3 Decision Tree

A decision tree is a type of hierarchical model used to help make important choices by visually representing the alternative outcomes, costs, and benefits of those choices in a tree structure [Willard-Mack, 2006].

The nodes in a decision tree can be one of three varieties [Yeh, 2009]

- Squares are commonly used to denote decision nodes.
- Often depicted as a series of concentric circles, nodes of chance.
- Points of termination are usually denoted by triangles.

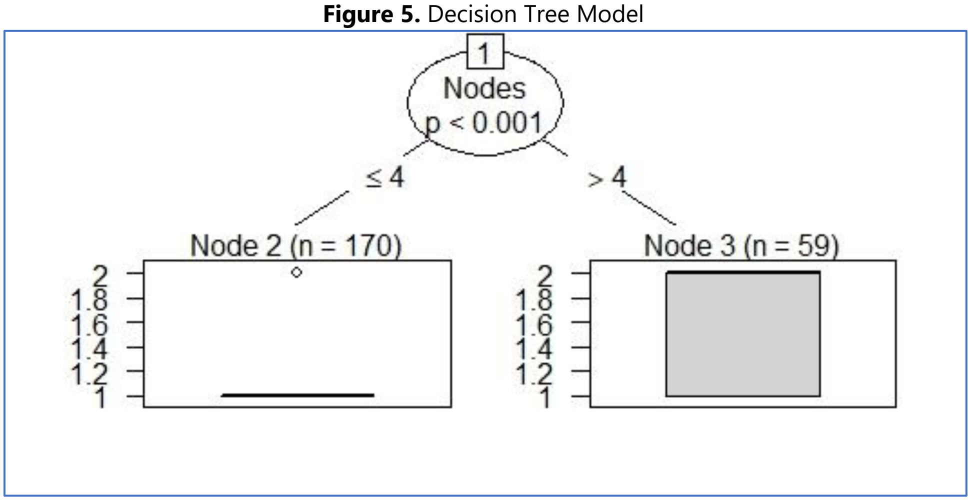
We calculate the values accuracy by using the following formula:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)},$$

Where TP = True positive, TN = True negative, FP = False Positive, FN = False Negative.

Figure 5. presents a decision tree model for analyzing formula and trees.

Figure 5. Decision Tree Model



Our data was divided into a training set (80%) and a testing set (20%). In order to make a survival prediction, we used a decision tree model. After that, we made use of a contingency table to determine the error rate, and we had an error rate of 33.76%, which means that 66.24% of patients were accurately classified in terms of whether or not they lived longer than 5 years or passed away within 5 years.

4. Results and Discussion

Our research was to forecast the outcome using only the three input factors and the single output variable. In order to determine something, we carried out a series of models. In addition to that, we used k-fold cross validation to evaluate the performance of each model. It divides the dataset into k roughly equal-sized parts/folds. Each fold is tested in turn, and the remaining portions are trained on. Following the completion of this process k times, the overall performance is evaluated by calculating the mean score across all of the tests.

Our neural network experiments show that the optimal number of hidden layers is two. We used a total of 306 records, including 152 validation samples and 153 training samples. The network topology was uncovered through a process of trial and error. (as seen in Figure 4). We tested a prototype system with a limited network and expanded it over time. Finally, we discovered that the best outcomes are achieved with the following network architecture: The architecture consists of three input neurons, two hidden layers of one neuron each, and a single output neuron. On a standard machine with 16 GB of RAM memory running Windows 11, we trained the network for 31452 epochs (as indicated in Figure 4). A precision of 82.98% was achieved.

We found the following result after compiling data from each model:

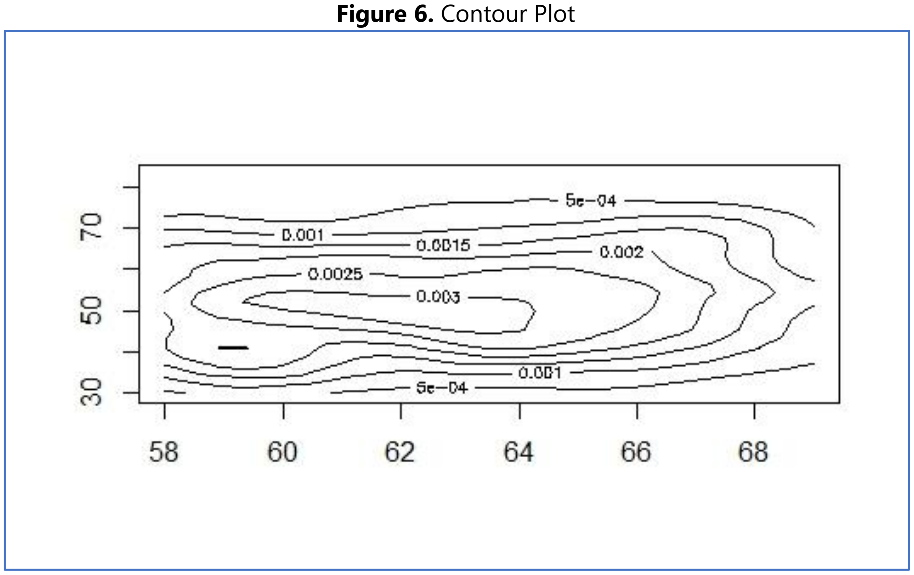
Table 5. Comparing the error rate for each model

Model	Linear Regression	Logistic Regression	Linear Discriminant Analysis	Quadratic Discriminant Analysis	K-Nearest Neighborhood	Neural Network	Decision Tree
The average Accuracy	0.5757	0.7443	0.7509	0.7541	0.7698	0.1702	0.3376

Cancer survival rates, often known as survival statistics, are estimates of the proportion of patients who are able to beat a particular form of the disease after a given period of treatment. Patients who have more than one lymph node affected are not expected to live. The greater the number of nodes, the lower the probability of survival. From Table 5, we show that ANN is the optimal approach for our data.

Moreover, observe that between the years 1960 and 1964 in Figure 6, a greater number of surgeries were performed on patients in the age range of 45 to 55.

Figure 6. Contour Plot



■ x-axis = Year, ■ y-axis=Age

5. Conclusion

We are able to classify the patient's survival status more accurately using the feature of auxillary nodes than we are using any of the other features. Yes, it is possible to diagnose cancer using the above data set by employing a variety of data analysis strategies and making use of a wide range of R modules.

In this article, we categorized patients into those who lived for at least 5 years and those who did not based on the predictive power of a neural network. The overall precision of our network was 82.98%. How to access and navigate the dataset in order to brainstorm approaches to cleaning the data and choose the best models. Methods for assessing a set of probabilistic models and optimizing their output through careful data cleaning are provided. What fitting a final model entails and how it might be used to make probabilistic predictions.

There is more to a patient's prognosis than just their age and the year of their surgery. However, those under the age of 35 have a better chance of survival. The more positive axillary nodes there are, the less likely it is that patients will survive. We also learned that survival is not always ensured by the lack of positive axillary nodes. Patients of any age who have no lymph nodes involved have a better chance of survival. Patients rarely have more than 20 lymph nodes. Patients older than 45/50 with a lymph node count of 8 or more have a poor prognosis. So, it is clear that the survival status of a person deteriorates as the number of detected axillary nodes rises, so that a person with a low number of detected axillary nodes has an 80%–85% chance of survival, but a person with a high number of detected axillary nodes has a 0% chance of survival.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD: 0000-0003-2680-0658
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Akay, M F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert systems with applications* 36, 3240-3247.
[2] Angadi, S.A. and Hiregoudar, N. (2016). A Cost Effective Algorithm for Grading Raisins Using Image Processing*, International Journal of Recent Trends in Engineering Research* 2455-1457
[3] Abbass, H A. (2002) An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial intelligence in Medicine* 25: 265-281.
[4] Bahrampour, A., and Montazeri. M. (2015). Prediction of breast cancer mortality by hidden markov model. *School of Health, Kerman University of Medical Sciences, Kerman, Iran*.
[5] Charro, F, Alaa H, Nurul R and Peiyong W (2023). A bifurcation phenomenon in a singularly perturbed two-phase free boundary problem of phase transition. *Nonlinear Analysis: Real World Applications* 73 103911.
[6] Delen, D and Amit K ((2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine* 34 113-127.
[7] Friedman, N, and Moises G (1997). Bayesian network classifiers. *Machine learning* 29 131-163.
[8] Jemal, A, Taylor M, and Michael J. T. (2005). Cancer statistics, 2005. 10-30.
[9] Karimi, N., Kondrood, R.R and Alizadeh, T. (2017). An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms, Measurement, 68- 76, September
[10] Kamiński, B and Przemysław S. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of operations research* 26 135-159.
[11] Liu, Y and Lu Z (2012). Decision tree based predictive models for breast cancer survivability on imbalanced data. In *2009 3rd international conference on Bioinformatics and biomedical engineering*. 1-4. IEEE
[12] Mollazade, K and Arman A. (2012). Comparing data mining classifiers for grading raisins based on visual features. *Computers and electronics in agriculture* 84 124-131.
[13] Montazeri, M, Mahdieh M and Mohammad J. (2014). Identifying efficient clinical parameters in diagnosis of liver disease. *Health MED* 1115.
[14] Okamura, N and Thompson. (1993). Raisin grading by machine vision." *Transactions of the ASAE (USA)*
[15] Polat, K and Salih G. (2007). Breast cancer diagnosis using least square support vector machine." *Digital signal processing* 17, no. 4 694-701.
[16] Raihen, Md N and Nazmul S (2023). Food Satisfaction among Students: A Study of Present Public University Students in Bangladesh. *Journal of Mathematics and Statistics Studies*. 1 01-18.
[17] Raihen, Md N. (2022). A Bifurcation Phenomenon of Regularized Free Boundary Problems of Two-Phase Elliptic-Parabolic Type. PhD diss., Wayne State University.
[18] Stein J and Nurul R (2023). Convergence Rates for Hestenes' Gram–Schmidt Conjugate Direction Method without Derivatives in Numerical Optimization. *AppliedMath* 3, 268-285.
[19] Shawarib, M, and Samy, S. A (2020). Breast cancer diagnosis and survival prediction using JNN.
[20] Sivachitra, M., and Vijayachitra. S. (2015). Classification of post operative breast cancer patient information using complex valued neural classifiers. In *2015 International Conference on Cognitive Computing and Information Processing (CCIP)*, pp. 1-4. IEEE, 2015.
[21] Verma, D and Nidhi M (2017). Analysis and prediction of breast cancer and diabetes disease datasets using data mining classification techniques. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. 533-538. IEEE, 2017.
[22] Willard-Mack, C L (2006). Normal structure, function, and histology of lymph nodes. *Toxicologic pathology* 34: 409-424.
[23] Yeh, W, Wei-Wen C, and Yuk Y C (2009). A new hybrid approach for mining breast cancer patterns using discrete particle swarm optimization and statistical method. *Expert Systems with Applications*. 8204-8211.
[24] Zheng, B, Sang, W Y, and Sarah, S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications* 41, 1476-1482.