International Journal of Advancements in Technology

Malgwi et al., Int J Adv Technol 2019, 10:1
DOI: 10.4172/0976-4860.1000224

Research Article Open Access

Multi-Agent-Based Performance Analysis of Classifiers for Breast Tumours

Malgwi YM*, Wajiga GM and Garba EJ

Department of Computer Science, Modibbo Adama University of Technology, Nigeria

***Corresponding author:** Malgwi YM, Department of Computer Science, Modibbo Adama University of Technology, Nigeria, Tel: (+234) 8059058056; E-mail: yumalgwi@mautech.edu.ng

Received date: January 11, 2019; **Accepted date:** February 01, 2019; **Published date:** February 08, 2019

Abstract

The challenging effect of selecting the best classifier among many classifier algorithms has been a big problem in data mining. Machine learning is widely used in bioinformatics and particularly in breast cancer diagnosis. This study is based on developing and evaluating different classifier algorithm (k-NN, J48, Decision table, Decision stump, and Naïve Bayes) in order to find the best among them using multi-agent platform and MYSQL for the diagnosis of breast tumors based on associated symptoms and risk factors of cancer diseases. Java Agent Development Environment (JADE) was used for the modeling and simulation. The results and the accuracy score were tested with a breast tumor clinical datasets which were gotten and formed from FMC Yola and FMC Gombe in Nigeria using 10-fold Cross-validation method. The results of the analysis reveal that k-NN classifier has a greater performance capability over other classification algorithms; hence, it is selected to be the best among the tested classifiers with higher accuracy score and lower false positive rate value.

Keywords: Diagnosis; Breast cancer; k-Nearest neighbors; J48; Decision table; Decision stump; Naïve bayes; JADE; Breast cancer

Introduction

Data mining is recognized to be a tool that can be used on a database for medical purposes. It improves the detection of diseases, sensitivity, and specificity. Accompanied costs are reduced drastically where an unwanted medical test is bypassed. The study on breast cancer prediction has been on for several years. The machine learning classifier systems in medical diagnosis are on the increase [1]. The classifier algorithms help experienced/inexperienced physicians to diagnosis accurately by minimizing possible errors.

The predictive method is a machine learning technique, supervised and assuming the existence of a group of labeled instances for each class of objects. The process in classification is characterized into three [2] (Figure 1):

- The Input is a set of attributes with instances, including a class attribute, predictable
- The Classifier is meant to predict the class of the instance
- Output, a pattern classifier that classifies the instance in a certain category based on the other attribute

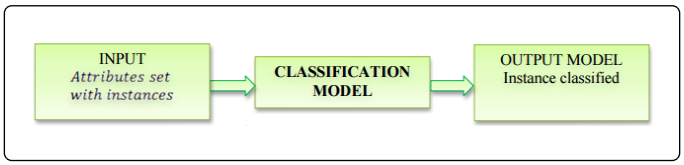

Figure 1: The process of classification.

Classification is a kind of complex optimization problem used by data mining. Many ML techniques have been applied by researchers in solving a classification problem. Data mining is the science used to discover knowledge from databases. It is concerned with choosing the most appropriate tools from the available techniques for summarization, classification, regression, association, clustering and searching for patterns or models of interest. Each instance used by machine learning and data mining algorithms is formatted using some set of fields (features, attributes, inputs, or variables). When the instances contain the correct output (class label) then the learning process is called supervised learning. On the other hand, the process of machine learning without knowing the class label of instances is called unsupervised learning [3]. In this study, the focus is on supervised machine learning.

Tumour is a mass tissue resulting from the unusual creation of cells in the body. Tumors can either be benign which will not develop into cancer or malignant which allow the spread of abnormal cellular growth to become uncontrollable. When a tumor is depicted in the body, the patient has to undergo a biopsy or mammography to determine whether it is malignant or benign.

In order to classify a tumor, more precisely tools are still needed to help oncologists to diagnose a breast tumor. Numerous research efforts have been conducted in the area of breast tumor detection and classification using various classification algorithms in order to develop an adaptive system that can classify and detect breast cancer without delay.

In this study, we intend to integrate both the features of the Agent and Machine learning to implement, simulate and undergo a comparative study on various classifiers (k-NN, J48. Decision table, Decision stump and Naïve Bayes) algorithms and to identify the best classifier for breast tumour classification using symptoms and risk factors of breast cancer.

Methods

Studies have been reported that have focused on breast cancer. These studies have applied different approaches to the given problem and achieved high classification accuracies. Details of some of the previous research works are given in the following: A comparative survey was carried out for the diagnosis and prediction of breast cancer and analysis of survivability of patients having breast cancer using data mining techniques [4]. Three data mining technique (Naïve Bayes, the back-propagated neural network and the C4.5 decision tree algorithms) were investigated and run on SEER public data.

The first technique assumes mutually independent attributes which are achieved by pre-processing the data to remove the dependent categories. The method is used to represent, utilize, and learn the probabilistic knowledge. Results have been achieved. In the study, a multi-layer network with backpropagation is used and C4.5 decision-tree generating an algorithm which was based on the ID3 algorithm used as the third technique. Weka toolkit was also used to experiment these three data mining algorithms. The toolkit is developed in Java and is open source software issued under the GNU General Public License.

Elsalamony [5] compared and evaluated classification performance using classification prediction, accuracy, sensitivity and specificity running on a bank dataset using 17 features and 45,211 instances. He uses four different data mining techniques' models which includes Multilayer Perceptron Neural Network (MPLNN), Tree Augmented Naïve-Bayes (TAN), Logistic Regression (LR) and C5.0 Decision Tree Classifier. The new model achieved slightly better performance.

In [6] investigated and analyzed the classification performance of Best First Tree, K-nearest neighbor and Sequential Minimal Optimization classification techniques on breast cancer data using Weka data mining tool. Their experiment has used time, correctly classified instances and accuracy as the criteria for assessing the superiority of each algorithm. They found out that the performance of Sequential Minimal Optimization algorithms has better classification performance than the other two algorithms considering accuracy and low error rate.

Nachev [7] carried out a comparative analysis of Neural Networks, Logistic Regression, Naïve Bayes, Linear and Quadratic Discriminant Analysis Considering their performance at different levels of data saturation. He applied cross-validation on training and testing the datasets on a direct marketing response task. He went further to simplify the two hidden layers of architecture proposed by Elsalamony into a single layer structure. His results revealed that Neural Networks is found to be the best performing classifier in nearly all levels of saturation.

Four classification algorithms (J48, Classification and Regression Trees, Alternating Decision Tree, and Best First Tree) were analyzed running on the breast cancer data [8]. They have conducted the experiment with Weka tool and Cross-validation technique was employed using 10 folds, 9 folds were used for training each classifier and 1 fold for testing. The percentage split uses 2/3 of the dataset for training and 1/3 of the dataset for testing. The authors reported that J48 classifier has the highest accuracy with 99%.

The goal of the work is to develop a predictive model that will improve the efficiency of the directed campaign for a long term deposit subscription thereby reducing the number of customers to be contacted by phone [9]. The data mining technique used includes Naïve Bayes, Decision Tree and Support Vector Machine classifiers. The methodology involves using a Portuguese bank and a dataset (Cross Industry Standard Process for Data Mining). The results from the analysis revealed the SVM is found to be the most reliable predictive algorithm.

A decision tree classifier model was developed to classify a tumor to either benign or malignant using the breast cancer dataset [10]. Weka was used for the experimentation in order to simplify the prediction task. They have also taken the decision tree model using the if-then rules in order to enhance the performance of decision trees earlier used.

Prusty [11] compare results when he applied the Naïve Bayes and Decision tree algorithms to the datasets. The results were obtained when the unbalanced data were compared with results obtained when the data was balanced with equal selections of "yes" and "no" in the response class using the same classification algorithm. The Results showed that the Area under Curve value improved after balancing the response class.

Senturk and Resul [12] carry out an analysis of performance on seven classification models. The classification models include Discriminant Analysis, Artificial Neural Networks, Decision Trees, Logistic Regression, Support Vector Machines, Naïve Bayes and the K-nearest neighbor to improve early diagnosis of breast cancer using RapidMiner Tool. They came to a conclusion that the SVM classifier outperforms the other six classifiers with 96% accuracy score.

Williams K et al. [13] Predicted and determined the most effective and efficient model using the Naïve Bayes and J48 decision trees to predict breast cancer risks in Nigeria. The experiment was run on datasets from cancer registry of LASUTH, Ikeja in Lagos containing 69 instances with 17 attributes with their class label. The experiment was conducted using Weka. The authors reported that the J48 decision tree performs credibly better prediction for breast cancer risks with an accuracy value of 94.2%.

Model

In our proposed work we proposed a multi-agent system that would diagnose breast tumors using five different classifiers (k-NN, J48, Decision table, Decision stump, and Naïve Bayes). We introduced three agents namely the Medical practitioner Agent, Classifier Agent and the Database Agent where each agent perform an own task under the coordination of the Medical practitioner Agent; the Medical practitioner Agent enables the user to input his or her symptoms/risk factors in order for the classifier agent to classify a tumor. The classify agent (i.e. either k-NN, J48. Decision table, Decision stump or Naïve Bayes) is responsible for classifying the symptoms presented by the medical practitioner agent into either malignant or benign using the data mining and the classifier algorithm. The database agent stores and retrieves the information presented to it by the medical practitioner agent.

Agents are considered to be trained intelligent systems capable of setting up the platform for diagnosing breast cancer. The agents themselves communicate with each other in the decision making process. If a Medical practitioner wants to diagnose a breast tumor patient, the medical practitioner can be invoked to which he/she has to specify the symptoms of the patient, the medical practitioner agent, in turn, will communicate to other agents and provide the corresponding information. The expert feedback will be displayed on the user side as

well as it will be stored in a database by the database agent for further references. A use case diagram for this system is produced as shown in Figure 2.

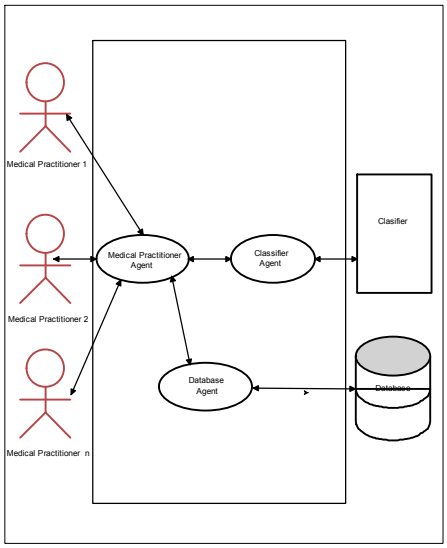

Figure 2: Use case diagram for MAS.

The model is intended to classify tumors based on organized parameters of cancer symptoms and risk factors, the proposed method accepts the information through the system's user interface by the medical practitioner. Symptoms T are entered into the system and the class Z of the patient is presented as a result. The Individual weight value associated with the symptom for a tumor disease D (XD) corresponding to a breast tumor type is given by the equations.

$$X_D = V_T D^* Z_T \qquad (1)$$

Where $V_T D$represents the weight of symptom T in disease D,Z_T Represents class of symptom T with which a patient reports

System Validation

We use cross-validation technique and metrics to evaluate the predictive performance of our classification model.

The pseudo code

To validate the model, cross-validation was used. The pseudo code for the procedure includes using a single parameter k that representing a group which a data sample can be split into. A value for k is chosen (Figure 3). The pseudo code for the validation is outlined as follows:

- To randomly Shuffle the dataset
- The dataset is split into k groups
- Consider each group as a unique group and testing the dataset
- The remaining groups are used as a dataset for the training
- Run the training dataset on the model and assessing it using the test datasets
- Hold the score and assess the scores from the model

Implementation

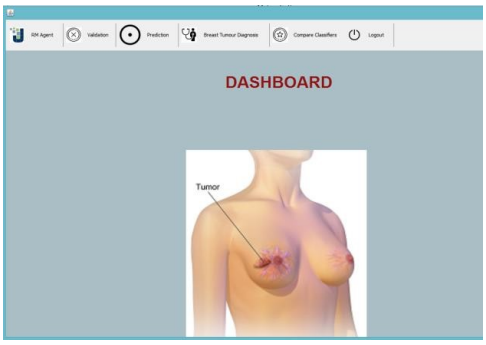

Figure 3: Dashboard interface.

Figure 4 shows the breast tumor diagnosis system interface. Patient data are inputted to run the diagnosis; the diagnostic result is then presented.

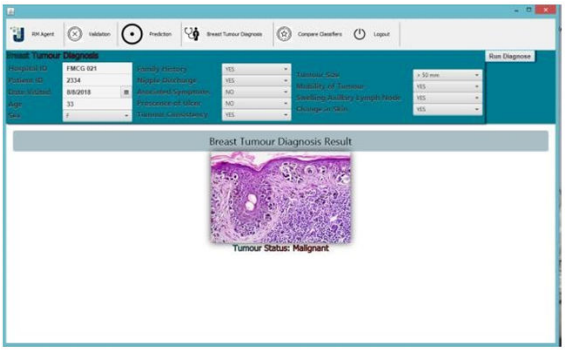

Figure 4: Tumour prediction interface.

Experimental Results and Analysis

This section gives the detailed results obtained by our proposed model for diagnosing breast tumor. The experimental result and the accuracy score were tested with a breast tumor clinical datasets which were gotten and formed from Federal Medical Centers (FMC) located in Yola (Adamawa state) and Gombe (Gombe state) in Nigeria. The attributes are selected based on the opinions of expert (Medical Practitioners) in the hospitals. The data was collected from 3rd May 2018 to 24th July 2018 at FMC Yola and Gombe. These datasets consist of 2,127 instances including 79 records with missing values, Noise and inconsistent data using 11 attributes. The breast tumor dataset collected was used to classify the malignant from benign. The dataset has a total of 2,048 rows and 11 columns indicating the attributes. Waikato Environment for Knowledge Analysis (WEKA) classes were used to analyze the datasets and evaluate the performance of the predictive result (Tables 1-12).

Results from validation using 10 fold cross-validation

Invalidating the percentage score of the model using the 10-fold cross validation method to evaluate the classification model, using the k-NN classifier, the model achieved a higher optimal value when the

value of k=12 with 100% score (which is the nearest neighbor). Figure 5 presents a graph for the optimal value of k.

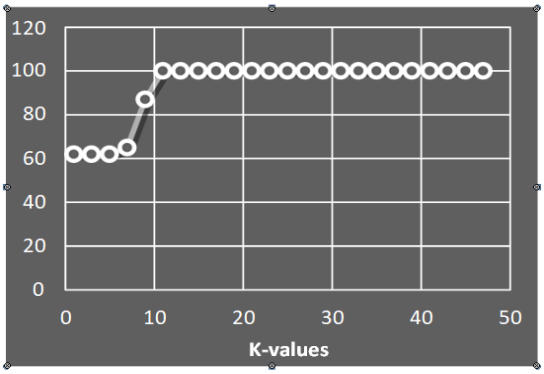

Figure 5: Result of cross-validation on k-NN classifier indicating the optimal value of k using 10 folds.

In Figure 6, 10 fold cross-validation methods were used on J48 classifier. The result indicates that 63% score on 1fold (k=1).

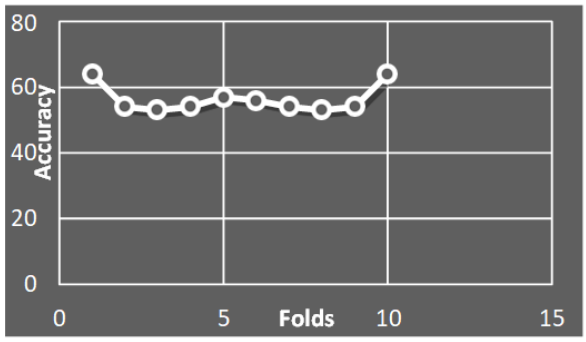

Figure 6: Result of cross-validation on J48 classifier using 10 folds.

Figure 7 shows a line graph of a decision table classifier using 10 folds. The result indicates 80% score when the number of the fold is 5 (K=5).

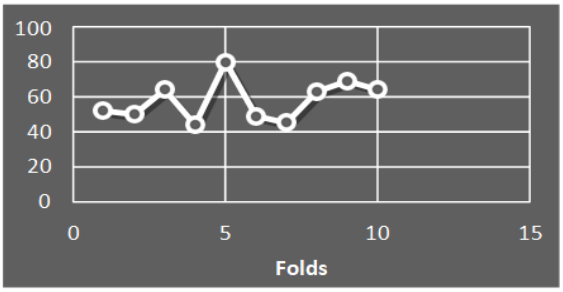

Figure 7: Result of cross-validation on decision table classifier using 10 folds.

Figure 8 presents a 10 fold cross validation line graph of decision stump achieving a score of 56% when the number of the fold is 1 (K=1).

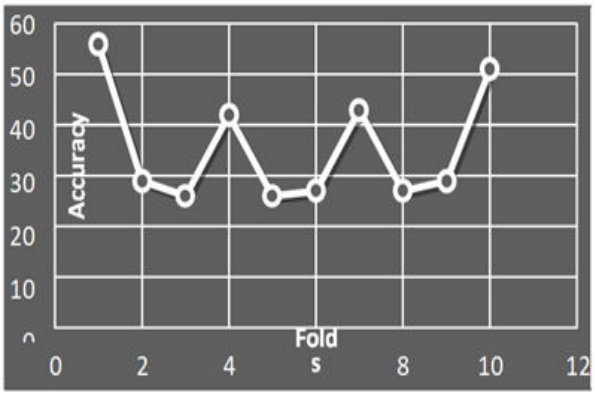

Figure 8: Result of cross-validation on decision stump classifier using 10 folds.

Figure 9 Presents a 10 fold cross-validation method was used on Naïve Bayes classifier. The result indicates a 100% score when the number of the fold is 2 (k=2).

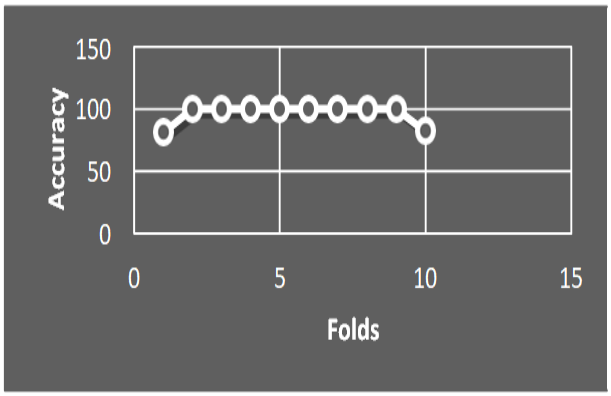

Figure 9: Result of cross-validation on Naïve Bayes classifier using 10 folds.

Prediction Results

In this section, also the tumor prediction depends on the inputted patient data based on the selection of a particular classifier in order to depict the tumor status. The results obtained are shown in Figures 10-14 using the same data with different classifiers and having the same tumor status (results).

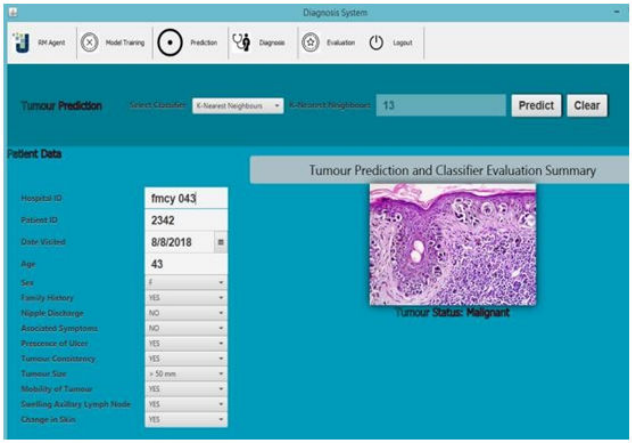

Figure 10: Result using k-NN classifier.

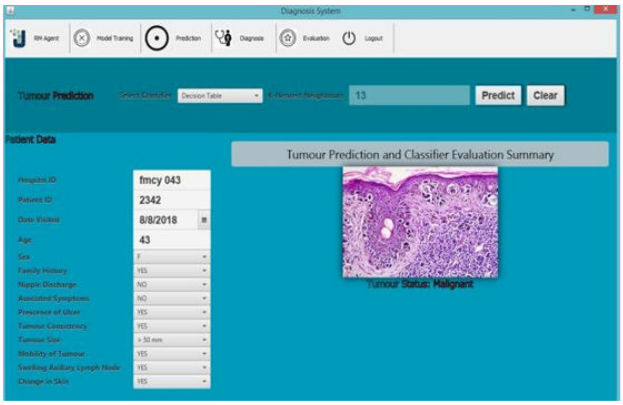

Figure 11: Result using decision table classifier.

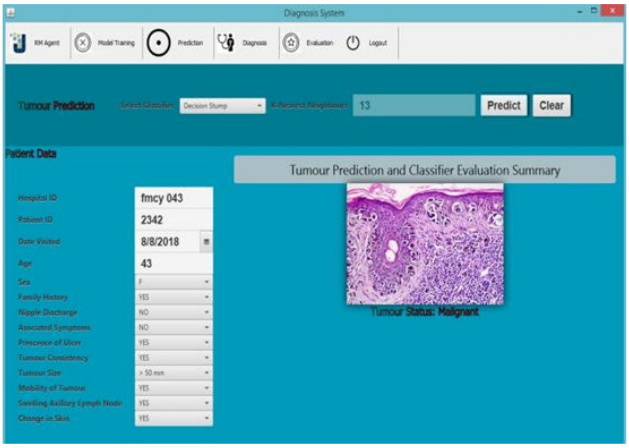

Figure 12: Result using decision stump classifier.

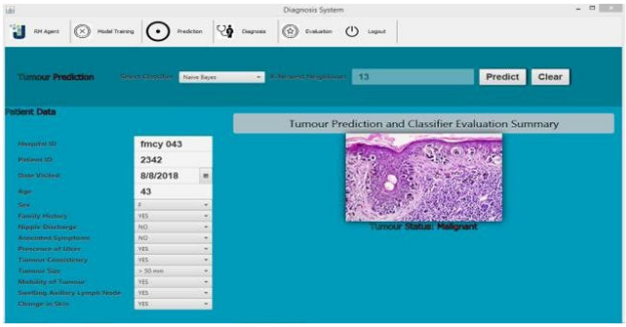

Figure 13: Result using naïve bayes classifier.

N=1,022	Predicted: No	Predicted: Yes	Total
Actual: No	TN=33	FP=03	36
Actual: Yes	FN=08	TP=978	986
Total	41	981	

Table 1: Row and column totals of the confusion matrix for the k-NN classifier.

From table 1

- Two predictable classes are used: "Yes" and "No" as the presence of the disease is predicted
- "Yes" indicates the presence of the disease and "No" indicates they don't have the disease
- Out of the total number of 1,022 predictions by the classifier 1,022 have been tested for the presence of the disease
- The classifier predicted "Yes" 988 times and "No" 34 times out of the 1,022 cases
- 986 patients have the disease and 36 patients do not have the disease

Definition of the basic terms used:

- ***True positives (TP):*** Predicted "Yes having the disease but in a real sense, they have the disease
- ***True negatives (TN):*** Predicted "No" but in a real sense, the patient does not have the disease
- ***False Positive rate:*** Predicted "Yes" but actually the patients have the disease
- ***False negatives (FN):*** Predicted "No, but actually they have the disease

Computation for the confusion matrix:

- ***Accuracy:*** In our sample how frequent is the classifier correct? (TP+TN)/total=(978+33)/1.022=0.9892
- ***Misclassification (Error) rate:*** How frequent it is not correct? (FP+FN)/total=(33+08)/1.022=0.0401
- ***True positive rate (Sensitivity):*** How frequently does it predict "Yes" when actually it is "Yes"?

TP/Actual yes=978/986=0.9919

- ***False positive rate:*** How frequently does it predict "No" when actually it is "Yes"?
  FP/Actual No=03/36=0.08333
- ***Specificity:*** How frequently does it predict "No" when actually it is "No"?
  TN/Actual No=33/36=0.9167
- ***Precision:*** How frequent it is correct when it predicts "Yes"? TP/Predicted yes TP/Predicted yes=978/981=0.9969
- ***Prevalence:*** How frequently does the "Yes" condition occur?
  Actual yes/total=986/1.022=0.9648

Terms	Score	Percentage score
Accuracy	0.9892	98.9%
Error rate	0.0401	4%
True positive rate	0.9919	99.2%
False positive rate	0.0833	8.3%
Specificity	0.9167	91.7%
Precision	0.9969	99.7%
Prevalence	0.9648	96.5%

Table 2: Terms totals of the confusion matrix for k-NN classifier.

The percentage accuracy score of the model is estimated at 98.9% with a reasonable low false positive rate of 8.3%.

N=1,022	Predicted: No	Predicted: Yes	Total
Actual: No	TN=26	FP=06	32
Actual: Yes	FN=10	TP=980	990
Total	36	986	

Table 3: Row and column totals of the confusion matrix for J48 classifier.

From table 3

Computation for the confusion matrix:

- ***Accuracy:*** In our sample how frequent is the classifier correct?
  (TP+TN)/total=(980+26)/1,022=0.9843
- ***Misclassification (Error) rate:*** How frequent it is not correct?
  (FP+FN)/total=(06+10)/1,022=0.0157 (Equivalent to 1 minus Accuracy).
- ***True positive rate (Sensitivity):*** How frequently does it predict "Yes" when actually it is "Yes"? TP/Actual yes=980/990=0.9898
- ***False positive rate:*** How frequently does it predict "No" when actually it is "Yes"?
  FP/Actual No=06/32=0.1875
- ***Specificity:*** How frequently does it predict "No" when actually it is "No"?
  TN/Actual No=26/32=0.8125 (equivalent to 1 minus False Positive Rate)
- ***Precision:*** How frequent it is correct when it predicts "Yes"?
  TP/Predected yes=980/986=0.9939
- ***Prevalence:*** How frequently does the "Yes" condition occur?
  Actual yes/total=990/1,022=0.9687

Terms	Score	Percentage Score
Accuracy	0.9843	98.4%
Error Rate	0.0157	2%
True Positive Rate	0.9898	99%
False Positive rate	0.1875	18.8%
Specificity	0.8125	81.3%
Precision	0.9939	99.4%
Prevalence	0.9687	96.9%

Table 4: Terms totals of the confusion matrix for J48.

The percentage accuracy score of the model using J48 classifier is estimated at 98% with a false positive rate of 18.8%.

N=1,022	Predicted: No	Predicted: Yes	Total
Actual: No	TN=22	FP=10	32
Actual: Yes	FN=11	TP=979	990
Total	33	989	

Table 5: Row and column totals of the confusion matrix for decision table classifier.

From table 5

Computation rates from the confusion matrix:

- ***Accuracy:*** In our sample how frequent is the classifier correct?
  (TP+TN)/total=(979+22)/1022=0.9795
- ***Misclassification (error) rate:*** How frequent it is not correct?
  (FP+FN)/total=(10+11)/1022=0.0206
  (Equivalen to 1 minus Accuracy)
- ***True positive rate (Sensitivity):*** How frequently does it predict "Yes" when actually it is "Yes"?
  TP/Actual yes=979/990=0.9889
- ***False positive rate:*** How frequently does it predict "No" when actually it is "Yes"?
  FP/Actual No=10/32=0.3125
- ***Specificity:*** How frequently does it predict "No" when actually it is "No"?
  TN/Actual No=22/32=0.6875 (equivalent to 1 minus False Positive Rate)
- ***Precision:*** How frequent it is correct when it predicts "Yes"?
  TP/Predicted yes=979/989=0.9899
- ***Prevalence:*** How frequently does the "Yes" condition occur?
  Actualal yes/total=990/1,022=0.9687

Terms	Score	Percentage score
Accuracy	0.9795	98%
Error Rate	0.0206	2%
True Positive Rate	0.9889	98.9%
False Positive rate	0.3125	31.3%
Specificity	0.6875	68.8%
Precision	0.9899	99%
Prevalence	0.9687	96.9%

Table 6: Terms totals of the confusion matrix for the decision table.

The percentage accuracy score of the model using the decision table classifier is estimated at 98 % with a false positive rate of 31.3%.

N=1,022	Predicted: No	Predicted: Yes	Total
Actual: No	TN=20	FP=04	24
Actual: Yes	FN=10	TP=998	998
Total	30	992	

Table 7: Row and column totals of the confusion matrix for decision stump classifier.

From table 7

Computation rates from the confusion matrix:

- ***Accuracy:*** In our sample how frequent is the classifier correct?
  (TP+TN)/total=(988+20)/1,022=0.9863
- ***Misclassification (Error) rate:*** How frequent it is not correct?
  (FP+FN)/total=(04+10)/1,022=0.0137 (Equivalent to 1 minus Accuracy)
- ***True positive rate (Sensitivity):*** How frequent does it predict "Yes" when actually it is "Yes"? TP/Actual yes=988/998=0.9899
- ***False Positive rate:*** How frequent does it predict "No" when actually it is "Yes"?
  FP/Actual No=04/24=0.1667
- ***Specificity:*** How frequently does it predict "No" when actually it is "No"?
  TN/Actual No=20/24=0.8333 (equivalent to 1 minus False Positive Rate).
- ***Precision:*** How frequent it is correct when it predicts "Yes"?
  TP/Predicted yes=988/992=0.9959
- ***Prevalence:*** How frequently does the "Yes" condition occur?
  Actual yes/total=998/1,022=0.9765

Terms	Score	Percentage Score
Accuracy	0.9863	98.6%
Error Rate	0.0137	1.4%
True Positive Rate	0.9899	98.9%
False Positive rate	0.1667	16.7%
Specificity	0.8333	83.3%
Precision	0.9959	99.6%
Prevalence	0.9765	97.7%

Table 8: Terms totals of the confusion matrix for decision stump.

The percentage accuracy score of the model using decision stump classifier is estimated at 98.6% with a false positive rate of 16.7%.

N=1,022	Predicted: No	Predicted: Yes	Total
Actual: No	TN=16	FP=02	18
Actual: Yes	FN=09	TP=995	1,044
Total	25	997	

Table 9: Row and column totals of the confusion matrix for naïve bayes classifier.

From table 9

Computation rates from the confusion matrix:

- ***Accuracy:*** In our sample how frequent is the classifier correct?
  (TP+TN)/total=(995+16)/1,022=0.9892
- ***Misclassification (Error) Rate:*** How frequent it is not correct?
  (FP+FN)/total=(02+09)/1,022=0.0108
- ***True positive rate (Sensitivity):*** How frequent does it predict "Yes" when actually it is "Yes"?
  TP/Actual yes=995/1,004=0.9910
- ***False positive rate:*** How frequently does it predict "No" when actually it is "Yes"?
  FP/Actual No=02/18=0.1111
- ***Specificity:*** How frequently does it predict "No" when actually it is "No"? TN/Actual
  No=6/18 = 0.8889 (equivalent to 1 minus false positive rate)
- ***Precision:*** How frequent it is correct when it predicts "Yes"?
  TP/Predicted yes=995/997=0.9979
- ***Prevalence:*** How frequently does the "Yes" condition occur?

Terms	Score	Percentage Score
Accuracy	0.9892	98.9%
Error Rate	0.0108	1.1%
True Positive Rate	0.9910	99.1%
False Positive rate	0.1111	11.1%
Specificity	0.8889	88.9%
Precision	0.9979	99.8%
Prevalence	0.9824	98.2%

Table 10: Terms totals of the confusion matrix for Naïve Bayes.

Classifiers	Accuracy score (%)	False positive score (%)
k-NN	98.9	8.3
J48	98.4	18.8
Decision Table	98.0	31.3
Decision Stump	98.6	16.7
Naïve Bayes	98.9	11.1

Table 11: Comparison chart for classifiers.

The percentage accuracy score of the model using Naïve Bayes classifier is estimated at 98.9% with a false positive rate of 11.1%.

Classifier	Classification Score				Total
	Malignant		Benign		
K-NN	0.9892	98.9%	0.11	1.1%	1.00 (100%)
J48	0.9843	98.4%	0.16	1.6%	1.00 (100%)
Decision table	0.9795	98.0%	0.2	2%	1.00 (100%)
Decision stump	0.9863	98.6%	0.14	1.4%	1.00 (100%)
Naïve bayes	0.9892	98.9%	0.11	1.1%	1.00 (100%)

Table 12: Comparative analysis of different classifiers.

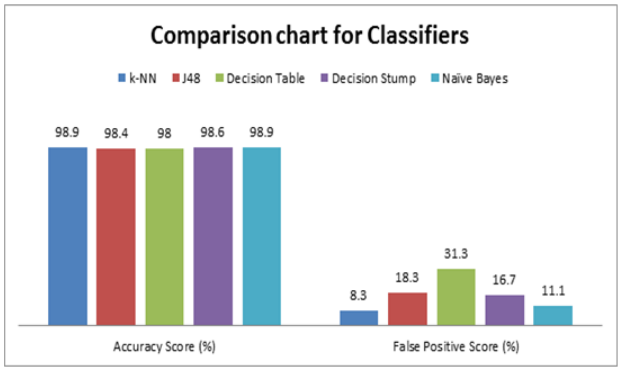

Figure 14: Comparison chart for classifiers.

Figure 14 shows a comparison chart for classifiers indicating accuracy scores versus Sensitivity rate of each classifier. It was observed that k-NN has an accuracy score of 98.9% with false positive rate is 8.3%, J48 has an accuracy score of 98.4% and the false positive rate is 11%, Decision Table has an accuracy score of 98% with false positive rate is 31.3%, Decision Stump has an accuracy score of 98.6% and the False positive rate is 16.7% while the Naïve Bayes has an accuracy score of 98.4% with false positive rate at 11.1%.

Figure 15 shows a chart indicating a comparative analysis for malignant and benign. k- NN has 98.9% for malignant and 1.1% for benign, J48 has 98.4% while benign has 1.6 %, Decision table has 98% for malignant and 2% for benign, Decision stump has 98.6% for malignant while benign has 1.4% and Naïve Bayes 98.9% for malignant and 1.1% for benign. This indicates that the classification score of the k-NN classifier is higher with 98.9% and 1.1% for malignant and benign respectively.

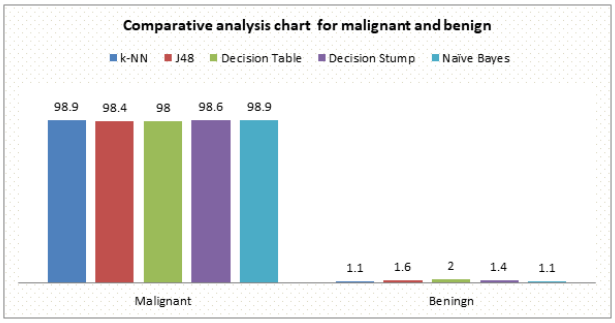

Figure 15: Comparative analysis chart for malignant and benign.

Conclusion

The findings of our results indicate that k-NN is selected to be the best classifier with an accuracy score of 98.9% and having lower false positive rate score of 8.3 which is assumed to make a reliable, confident and accurate diagnostic system.

References

1. Mitchell TM (1997) Machine learning. McGraw Hill.
2. Gorunesu F (2006) Data mining-concepts, models and techniques.
3. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifier. Mach learn 29: 131-163.
4. Hamid KKZ (2015) A comparative survey on data mining techniques for breast cancer diagnosis and prediction. Indian J Fundam Appl Sci 5: 4330-4339.
5. Elsalamony HA (2014) Bank direct marketing analysis of data mining techniques. Int J Comput App 85: 12-22.
6. Chaurasia V, Saurabh P (2014) A novel approach for breast cancer detection using data mining techniques. Int J Innov Res Comput Comm Eng 2: 2456-2465.
7. Nachev A (2015) Application of data mining techniques for direct marketing. Computational Models for Business and Engineering Domains.
8. Venkatesan E, Velmurugan T (2015) Performance analysis of decision tree algorithms for breast cancer classification. Indian J Sci Technol 8: 1-8.
9. Moro S, Laureano R, Cortez P (2011) Using data mining for bank direct marketing: an application of the CRISPDM methodology.
10. Shrivastava SS, Anjali S, Ramesh PA (2013) An overview of data mining approach on breast cancer data. Int J Adv Comput Res 3: 256-262.
11. Prusty S (2013) Data mining applications to direct marketing: identifying hot prospects for banking product. DePaul University, USA.
12. Senturk Z, Resul K (2014) Breast cancer diagnosis via data mining: performance analysis of seven different algorithms. Comput Sci Eng 4: 35-46.
13. Williams K, Peter AI, Jeremiah AB, Adeniran IO (2015) Breast cancer risk prediction using data mining classification techniques. Trans Net Commun 3: 1-11.