International Journal of Multidisciplinary Research and Growth Evaluation.

Cancer classification: A study of eight machine learning algorithms for optimal classification of the nature of cancer

Alaa Khouloud [1*], Atounti Mohamed [2], Bailoul Charaf Eddine [3]
[1, 2] Applied Mathematics and Information Systems Laboratory, Mohammed First University, Nador, Morocco
[3] Applied Mathematics and Computer Science Laboratory, Cadi Ayyad University, Marrakech, Morocco

* Corresponding Author: **Alaa Khouloud**

Article Info

ISSN (online): 2582-7138
Volume: 03
Issue: 01
January-February 2022
Received: 09-01-2022;
Accepted: 17-01-2022
Page No: 482-488
DOI:
https://doi.org/10.54660/anfo.2022.3.1.23

Abstract
Cancer mainly affects women and is the most dangerous disease in the world. Curing cancer through early detection of cancer and scientific research is our main goal because early detection helps to eliminate cancer completely. After reviewing several articles, we found that there are several techniques available for cancer detection. In this paper, we applied eight data mining techniques: Deep Neural Networks, Artificial Neural Networks, Naive Bayesian Classifier, Classification Trees, Fuzzy C-Means, Logistic Regression, Discriminant Analysis and K-Nearest Neighbor Classifier to this problem and show their significant results on real data. Therefore, among all data mining methods used, good results can be obtained by applying Deep Learning Neural Networks to cancer detection.

Keywords: Bayesian Classifier, Cancer detection, Classification, Classification Trees, Confusion Matrix

1. Introduction

Cancer is often seen as a painful and incurable disease. Most people think that there is no scarier diagnosis than cancer and this view of cancer is overly generalized and exaggerated. Although cancer is a potentially fatal and serious disease, it is wrong to believe that all forms of cancer are fatal and incurable. There are many types of cancer, many of which can now be effectively treated to reduce, slow or eliminate the impact of the disease on patients' lives. While a cancer diagnosis can still leave patients feeling out of control and helpless, in many cases there are now reasons for hope rather than despair.

Millions of cells make up our body, each of them being an autonomous living unit. Each cell coordinates with the other cells that make up the tissues and organs of our body. This coordination is reflected in the way our cells reproduce. Normal cells in the body divide and grow for a period of time and then stop dividing and growing. Thereafter, they reproduce only when necessary to replace dying or defective cells. When this process of cell reproduction gets out of control, cancer occurs. In other words, cancer is a disease characterized by unwanted, uncontrolled and uncoordinated cell division.

Cancer cells continue to divide and grow throughout their lives, replicating into more and more harmful cells. The abnormal division and growth seen in cancer cells is caused by damage to the cell's DNA. Cellular DNA can be defective and damaged in several ways. For example, environmental factors (such as exposure to tobacco smoke) can set off a chain of events that lead to defects in cellular DNA and thus to cancer. In addition, defective DNA can be inherited from our parents.

When cancer cells replicate and divide, they often form a cluster of cancer cells called a tumor. Tumors cause many symptoms of cancer by destroying and crushing surrounding noncancerous cells and tissues and by exerting pressure. Tumors come in two forms: benign and malignant. Benign tumors are usually not life-threatening. Benign tumors are not cancerous, they do not

spread as much as cancerous tumors and therefore do not grow. Malignant tumors, on the other hand, spread and grow to other parts of the body. The process by which cancer cells move from the original tumor site to other parts of the body is called metastasis.

To predict a cancer, the existing literature uses several classification techniques. This study aims to compare the accuracy of eight data mining methods: Deep Neural Networks (DNNs), Artificial Neural Networks (ANNs), Naive Bayesian Classifier (NB), Classification Trees (CTs), Fuzzy C-Means (FCM), Logistic Regression (LR), Discriminant Analysis (DA) and K-Nearest Neighbor Classifier (KNN) based on the biological information of a patient. The central idea is to provide a computer learner with a set of training data made up of certain characteristic values which are inherent in the system in which we want to make the cancer detection. After a learning process, the program is supposed to be able to correctly classify a biopsy not previously seen as benign or malignant, given certain characteristics of that patient.

The structure of this paper is as follows : we first introduce the reader to the cancer detection domain. In section 2, we present and explain the eight techniques used in this work (DNNs, ANNs, NB, CTs, FCM, LR, DA and KNN). Finally, we discuss the numerical results obtained by the eight techniques, followed by a conclusion.

2. Literature Review

In the age of information explosion, companies are producing and collecting huge volumes of data on a daily basis. The major challenge for businesses is transforming the information into actionable results and discovering useful insights from the database. Data mining is the process of exploring and analyzing large amounts of data to discover meaningful rules and patterns, and it is an indispensable tool in the decision support system and plays a key role in cancer detection. The advantages and disadvantages of the eight data mining techniques used in our paper are reviewed.

2.1 Artificial Neural Networks and Deep Neural Networks

ANNs are composed of a large number of artificial neurons and are modeled by the human brain. But in ANNs, neurons have fewer connections and roles and they are less numerous compared to biological neurons. However, the functioning of both networks is the same. The first step, as in an ANN network, is to receive inputs from other neurons or perceptrons connected to the other end of the network. Then, applying an activation function to these received inputs, which leads to the activation level of the neurons which is in fact the output value of the neuron. There are different activation functions that can be applied to create non-linearity in the neural network, such as sigmoid function, step function, linear function. In the multilayer backpropagation network, the sigmoid function is used and its activation value is between -1 and 1. ANNs are self-adaptive frameworks preparing data whose functionality and structure are stimulated by cognitive processes.

A deep neural network (DNN) is a network that consists of more than three layers of neurons, the input layer, the hidden layers and the output layer. It is based on the idea of successive layers, these layers define the depth of the model. DNNs are a type of multilayer feedforward network (MLP) where data flows from the input layer to the output layer. In addition, it is a multi-step information purification process, where information passes through successive filters and emerges with a progressively clean solution for many tasks. Neural networks must develop complex models to obtain the best solutions to complex problems. By complex model, we usually mean the extra number of hidden layers or neurons in the hidden layers between the input and output layers, so that it can learn data representations with several levels of abstraction, which is why it is called a deep neural network. Although it increases computational power, it appears with an improved solution where the number of features is in the thousands.

Fig. 1 is a model of multilayer perceptron which consists of four layers: first layer is the input layer, two hidden layers and last layer is the output layer.

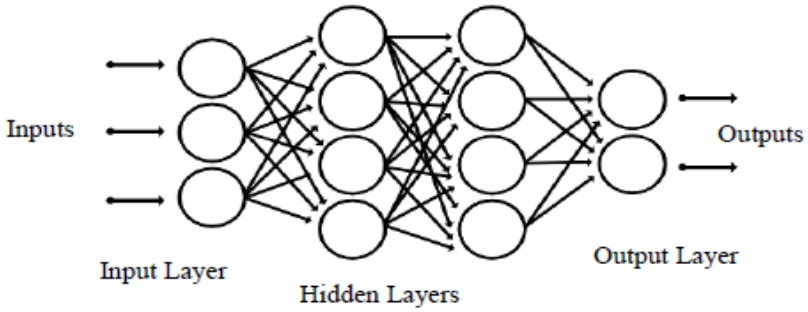

Fig 1: A multilayer perceptron networks

To develop meaningful relationships between inputs and outputs through a learning process, DNNs use nonlinear mathematical equations. Backpropagation appeared in 1970, but is considered the workhorse of neural network learning in the current environment. A central algorithm in a DNN to develop a meaningful relationship between inputs and outputs through the learning process.

Backpropagation is widely used in the multilayer perceptron for weight adjustment, because in the MLP neurons have weights assigned by random values associated with the inputs. Thus, when an error is made, a large number of weights must be adjusted with a sample of training data.

Backpropagation is an algorithm used to find the minimum value of the error function by adjusting the weights using the delta rule or gradient descent. The goal of backpropagation is to adjust and optimize the weights so that the neural network can learn to correctly map inputs to outputs. First, the error is calculated and propagated through the network, after the synaptic weights are adjusted according to the error correction rule. The error signal propagates backwards

through the network, while the function signal is an input signal and propagates forward through the network. The first step in backpropagation is to train a model by adjusting the model weights after adjusting the input weight, which is an iterative procedure. The next step is to find the minimum error. The model is considered the best when the error is minimal, otherwise the parameter update procedure continues until the minimal error is reached. In short, the goal of backpropagation is to reach global minima, so its value must be realized. Fig. 2 shows the working of backpropagation.

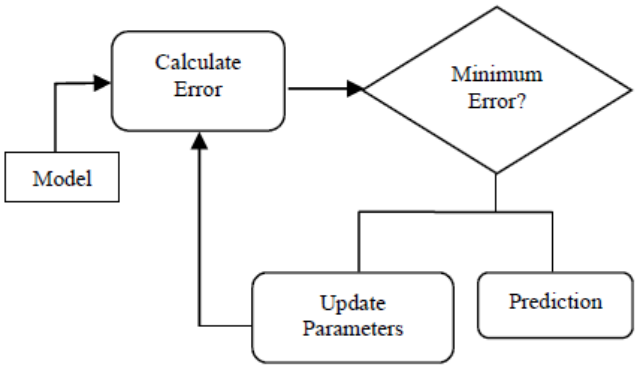

Fig 2: Backpropagation

DNNs can easily handle interactive and non-linear effects. The main drawback of DNNs is that they cannot provide a simple probabilistic classification formula.

2.2 Classification Trees

Each internal node in the tree represents a test on an attribute, each branch represents a test result, and the leaf nodes represent classes. The root node is the highest node in the tree. CTs are applied for discrete quantitative variables or qualitative variables. CTs rank observations based on all supervised and explanatory variables by the response variable.

Segmentation is usually done using a single explanatory variable. CTs are based on the minimization of impurities (the variability of the response values of the observations). TSTs can lead to simple classification rules and handle the non-linear effects of the explanatory variables.

CTs are considered the simplest of the machine learning methods. It is a completely transparent classification method that looks like a series of "if-then" instructions organized in a tree. The classification can be simply applied by answering a question in the decision tree. This algorithm divides the data into two or more sets.

The division is done on the basis of the most significant attributes in order to create groups as distinct as possible using information gain and entropy. Entropy measures the impurity of the result class in the subset with ps attributes in a data set D as shown in the formula :

$$H(p_1, p_2, \ldots, p_s) = \sum_{i=1}^{s} (p_i \log(\frac{1}{p_i}))$$

The information gain is calculated as the difference between the entropy of the data set and the entropy of the split attribute, as shown in the formula :

$$Gain(D, S) = H(D) - \sum_{t=1}^{s} p(D_i)\, H(D_i)$$

But their sequential nature and algorithmic complexity can make them dependent on the observed data, and even small changes can transform the tree structure. It is difficult to generalize a tree structure designed for one context to other contexts.

2.3 Fuzzy C-Means

Fuzzy clustering is a powerful unsupervised method for model building and data analysis. Fuzzy clustering is more natural than hard clustering in many situations. Objects located at the boundary of several classes are not required to belong entirely to one of them, but are assigned membership degrees between 0 and 1 indicating their partial membership. The FCM algorithm is the most widely used. FCM clustering was first introduced in the literature for a particular case (m = 2) by Joe Dunn in 1974. The general case (for any m greater than 1) was developed by Jim Bezdek in his PhD thesis at Cornell University in 1973. It can be improved by Bezdek in 1981. FCM uses fuzzy partitioning so that a data point can belong to all groups with different membership degrees between 0 and 1.

Algorithm

1. Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ with $U^{(k)}$

$$C_i = \frac{\sum_{j=1}^{n} {u_{ij}}^m x_j}{\sum_{j=1}^{n} {u_{ij}}^m}$$

3. Update$U^{(k)}$, $U^{(k+1)}$
4.

$$d_{ij} = \sqrt{\sum_{i=1}^{n} (x_i - c_i)}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{kj}})^{2/(m-1)}}$$

5. If$||U(k+1) - U(k)|| < \varepsilon$, then STOP; otherwise return to step 2.

Here *m:* is any real number greater than 1,
u_{ij}*:* is the degree of membership of x_i in the cluster *j*,
x_i*:* is the *i*th of d-dimensional measured data,
c_j*:* is the d-dimension center of the cluster.

The algorithm works by assigning a membership to each data point corresponding to each cluster center based on the distance between the data point and the cluster center. The closer the data point is to the cluster center, the closer its membership to the particular cluster center. Obviously, the sum of the members of each data point must be equal to one. After each iteration, the members and cluster centers are updated according to the formula.

2.4 Naive Bayesian Classifier

NB is based on the assumption of conditional independence of classes and on Bayes' theory, which assumes that the effect of the value of an attribute on a given class is independent of the values of the other attributes. Bayesian classifiers provide a theoretical justification for other classifiers that do not use Bayes' theorem. Considering that another event has already occurred, we can calculate the probability of an event occurring. It can be written as follows:

$$\Pr(c/X) = \frac{\Pr(X/c).\Pr(c)}{\Pr(X)}$$

Where the posterior probability of the target class c Pr(c/X) is calculated from Pr(c), Pr(X/c) and Pr(X).

The main weakness of NB is that the predictive accuracy is strongly correlated to the assumption of conditional independence of the classes. This assumption simplifies the calculation. However, in practice, dependencies may exist between the variables.

2.5 K-Nearest Neighbor Classifier

The basis of KNN classifiers is learning by analogy. Given an unknown sample, a KNN classifier searches for the KNN closest to the unknown sample. The proximity is expressed in terms of distance. The unknown sample is mapped to the most common class among its KNNs.

The KNN classification follows the following four steps:

- Calculating the distance metric between the test data point and all labeled data points.
- The ranking of the labeled data points in ascending order by that distance metric.
- Selecting the first k labeled data points and looking at the class labels.
- Find the class label that own the majority of these k labeled data points and assign it to the test data point.

This algorithm consumes a lot of memory and works poorly for large data. It requires a large distance function to calculate the similarity.

It is not necessary to establish a predictive model before classification, which is the advantage of this approach. However, the predictive accuracy of KNN is strongly affected by the cardinality of the k-neighborhood and the distance measure, as well as KNN does not produce a simple classification probability formula, which are the drawbacks of this approach.

2.6 Logistic Regression

The LR method is considered a special case of linear regression models and is the gold standard for binary classification problems. However, the binary response variable violates the normality assumptions of general regression models.

This is a multiple regression with explanatory variables that can be categorical or continuous and a dependent variable that is a categorical dichotomy. Probability estimates for each instance of the variables are produced by the model. The logistic regression model specifies that the appropriate function of the adjusted probability of the event is a linear function of the observed values of the explanatory variables.

A threshold can be defined, to predict to which class an observation belongs. On the basis of this threshold, the estimated probability obtained is classified into classes.

To obtain a complex decision threshold, we can increase the polynomial order, the decision threshold can be linear or non linear (see Fig. 3).

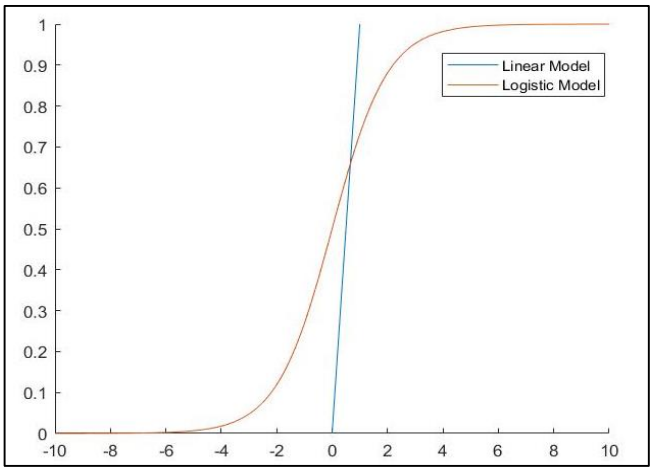

Fig 3: Linear and logistic models

This approach produces a simple probabilistic classification formula, which is the main advantage of LR. However, LR cannot adequately deal with the interactive problems of explanatory variables and nonlinear effects.

2.7 Discriminant Analysis

DA, or Fisher's rule, is a commonly used technique for supervised classification problems that takes labels into account. The objective of DA is to project features from a higher dimensional space onto a lower dimensional space. DA is another technique applied to the binary outcome of the response variable.

DA is related to analysis of variance (ANOVA) and regression analysis, which also attempt to express a dependent variable as a linear combination of other measures. However, ANOVA uses a categorical dependent variable and continuous independent variables, whereas ANOVA has a continuous dependent variable and categorical variables.

DA is based on the assumption that, for each given class, the explanatory variables are distributed according to a multivariate normal distribution, and it is an alternative to logistic regression.

The objective of Fisher's rule is to minimize the distance within each group and to maximize the distance between the different groups. The advantages and disadvantages of the FD are similar to those of the LR.

3. Classification Accuracy Among Data Mining Techniques

3.1 Description of the Data

The dataset is used to design a classifier that classifies cancers as either benign or malignant depending on the characteristics of sample biopsies. The inputs matrix is a 9*699 matrix defining nine attributes of 699 biopsies.

- Clump thickness.
- Uniformity of cell size.
- Uniformity of cell shape.
- Marginal Adhesion
- Single epithelial cell size.
- Bare nuclei.
- Bland chomatin.
- Normal nucleoli.

- Mitoses.

The target matrix is a 2*699 matrix where each column indicates a correct category with a one in either element 1 or element 2.

- Benign.
- Malignant.

The data were randomly divided into two groups, one for model training and one for model validation. A review of the literature reveals that error rates have often been used as a measure of model classification accuracy. For the binary classification problem, the root mean square error (MSE) gives a better solution to compare the performance of different models. Therefore, our study used MSE to examine the classification accuracy among the eight data mining techniques. MSE is defined as follows :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Where

- Y : is the vector of observed values of the variable being predicted.
- $\hat{Y}$: being the predicted values.
- n : is the number of observations of the output vector Y.

We are especially interested in knowing how many malignant biopsies are classified correctly as malignant and at the same time how many benign biopsies are classified incorrectly as malignant. The first is called the true positive rate and the latter the false positive rate.

The accuracy, sensitivity, and specificity are also used to evaluate models' performance, as follows:

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

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

3.2 Results

The MSEs of the eight data mining techniques are shown in Table 1.

Table 1: Classification accuracy

Method	Error rate	
	Training	Validation
Classification Trees	0.042	0.066
Logistic Regression	0.026	0.057
NB Classifier	0.034	0.066
Discriminant Analysis	0.040	0.047
K-nearest neighbor	0.042	0.057
Deep Neural Networks	0.020	0.033
ANNs	0.023	0.027
Fuzzy C-Means ($m = 2$)	0.636	0.510

In both training data and validation data, based on error rate (MSE), DNNs achieve the best performance with the lowest error rate (0.02 and 0.03 respectively, see Table 1) in 14 iterations. Since the efficient dataset used to measure the generalization classification accuracy of the models is the validation data, we can therefore conclude that the DNNs are the best model among the eight methods.

It is often thought that DNNs are a fast, reliable and easy technique to obtain good results in different areas. In practice, it can be seen that the great difficulty in the application of DNN lies in a good compromise between the different parameters to be chosen and in the choice of a good set of pretreatment operations.

The deep network architecture used here is given in Fig. 4, we used a deep network with three hidden layers, the first one contains 30 neurons, the second one 10 neurons and the last one 10 neurons.

Fig. 4. Deep network architecture

To better visualize our results obtained by the best model (DNNs), we used the confusion matrix is a very popular measure used while solving classification problems. It can be applied to binary classification as well as for multiclass classification problems (see Fig. 5).

Confusion matrix		Predicted class	
		0	1
Actual class	0	True positive (TP)	False negative (FN)
	1	False positive (FP)	True negative (TN)

Fig 5. Confusion matrix

On the confusion matrix graph, the columns correspond to the actual class (target class) and the rows correspond to the predicted class (output class). The off-diagonal cells correspond to observations that are incorrectly classified. The diagonal cells correspond to observations that are correctly classified. The number of observations and the percentage of the total number of observations are shown in each cell.

The bottom line of the graph shows the percentages of all examples in each class that are correctly and incorrectly classified. These measures are often referred to as the recall (or true positive rate) and false negative rate, respectively. The column on the far right of the graph shows the percentages of all examples predicted to belong to each class that are correctly and incorrectly classified. These measures are often referred to as accuracy (or positive predictive value) and false discovery rate, respectively. The cell at the bottom right of the graph indicates the overall accuracy.

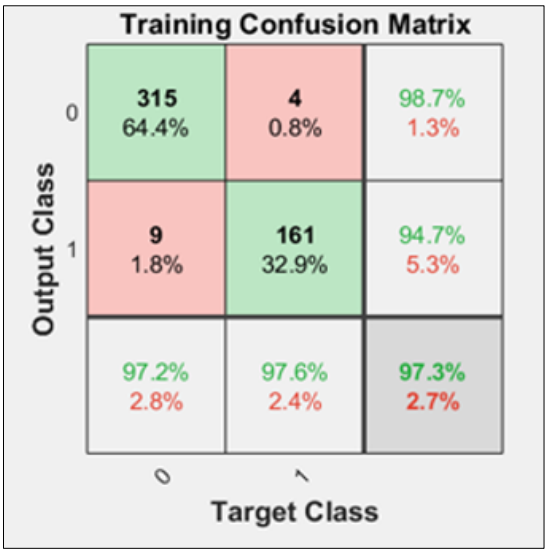

Fig 6: Training confusion matrix

The first two diagonal cells in the training confusion matrix (see Fig. 6) show the number and percentage of correct classifications obtained by the trained network. For example, 315 biopsies are correctly classified as benign, this corresponds to 64.4%$of all the data. Similarly, 161 biopsies are correctly classified as malignant and this corresponds to 32.9% of our data. 4 biopsies are incorrectly classified as malignant and this corresponds to 0.8% of the data. Similarly, 9 biopsies are incorrectly classified as benign and this corresponds to 1.8% of our data.

Out of 319 benign biopsies, 98.7% are correct and 1.3% are wrong. Out of 170 malignant biopsies, 94.7% are correct and 5.3% are wrong. Out of 324 benign cases, 97.2% are correctly predicted as benign and 2.8% are predicted as malignant. And out of 165 malignant cases, 97.6% are correctly classified as malignant and 2.4% are classified as benign.

Overall, 97.3% of the predictions are correct and 2.7% are wrong.

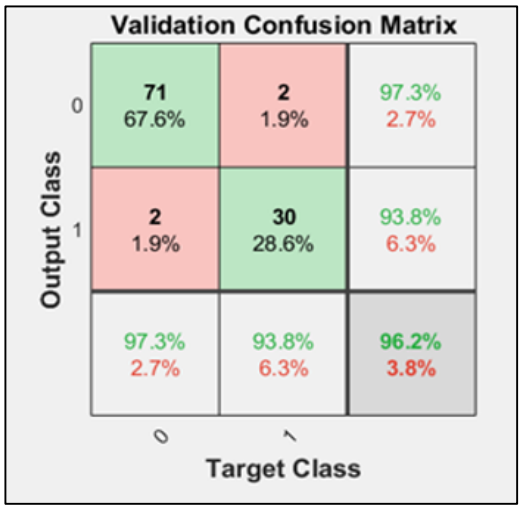

Fig 7: Validation confusion matrix

In the validation confusion matrix (see Fig. 7), 71 biopsies are correctly classified as benign, this corresponds to 67.6% of all the data. Similarly, 30 biopsies are correctly classified as malignant and this corresponds to 28.6% of our data. 2 biopsies are incorrectly classified as malignant and this corresponds to 1.9% of the data. Similarly, 2 biopsies are incorrectly classified as benign and this corresponds to 1.9% of our data.

Out of 73 benign biopsies, 97.3% are correct and 2.7% are wrong. Out of 32 malignant biopsies, 93.8% are correct and 6.3% are wrong. Out of 73 benign cases, 97.3% are correctly predicted as benign and 2.7% are predicted as malignant. And out of 32 malignant cases, 93.8% are correctly classified as malignant and 6.3% are classified as benign.

Overall, 96.2% of the predictions are correct and 3.8% are wrong.

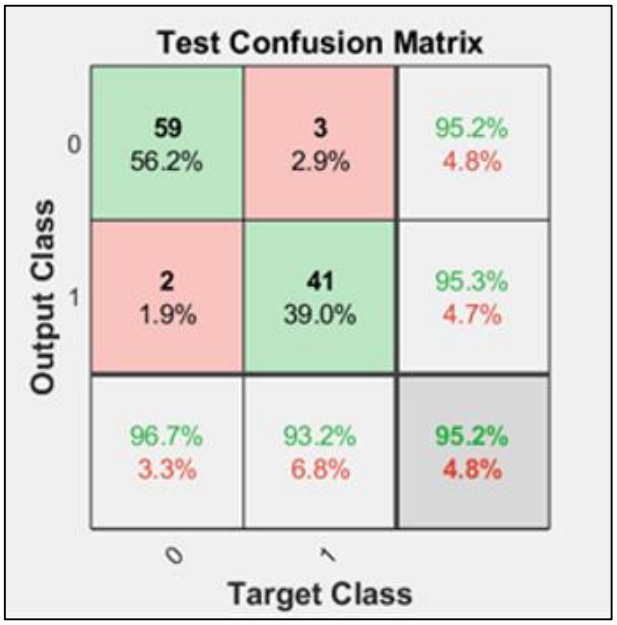

Fig 8: Test confusion matrix

Same for the test confusion matrix (see Fig. 8), 59 biopsies are correctly classified as benign, this corresponds to 65.2% of all the data. Similarly, 41 biopsies are correctly classified as malignant and this corresponds to 39\% of our data. 3 biopsies are incorrectly classified as malignant and this corresponds to 2.9% of the data. Similarly, 2 biopsies are incorrectly classified as benign and this corresponds to 1.9% of our data. \\.

Out of 62 benign biopsies, 95.2% are correct and 4.8% are wrong. Out of 43 malignant biopsies, 95.3% are correct and 4.7% are wrong. Out of 61 benign cases, 96.7% are correctly predicted as benign and 3.3% are predicted as malignant. And out of 44 malignant cases, 93.2% are correctly classified as malignant and 6.8% are classified as benign.

Overall, 95.2% of the predictions are correct and 4.8% are wrong.

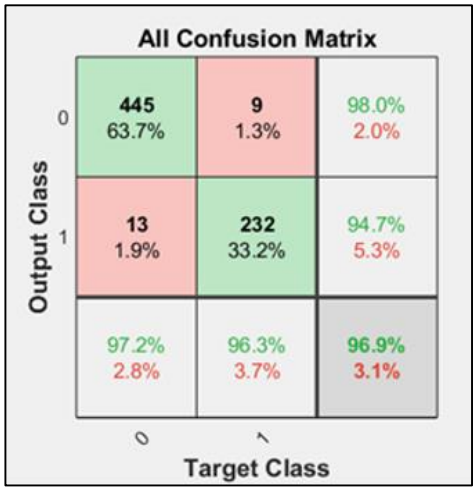

Fig 9: All confusion matrix

All these matrices are summarized in Fig. 9, where 445 biopsies are correctly classified as benign, this corresponds to 63.7% of all the data. Similarly, 232 biopsies are correctly classified as malignant and this corresponds to 33.2% of our data. 9 biopsies are incorrectly classified as malignant and this corresponds to 1.3% of the data. Similarly, 13 biopsies are incorrectly classified as benign and this corresponds to 1.9% of our data.

Out of 454 benign biopsies, 98% are correct and 2% are wrong. Out of 245 malignant biopsies, 94.7% are correct and 5.3% are wrong. Out of 458 benign cases, 97.2% are correctly predicted as benign and 2.8% are predicted as malignant. And out of 241 malignant cases, 96.3% are correctly classified as malignant and 3.7% are classified as benign.

Overall, 96.9% of the predictions are correct and 3.1% are wrong. It is to note that DNNs is more accurate and efficient in comparison with the seven other techniques.

The values of the parameters used in the performance evaluation of our best and accurate model are presented in Table 2.

Table 2: Classification results

	Accuracy	Sensitivity	Specificity
DNNs	0.968	0.980	0.946

4. Conclusion

An early cancer detection plan is a key component of a comprehensive cancer control plan. It allows cases to be detected at an earlier stage, when treatment is more effective and the chances of a cure are greater. This paper examines eight major classification techniques in data mining : Deep Neural Networks, Artificial Neural Networks, Naive Bayesian Classifier, Classification Trees, Fuzzy C-Means, Logistic Regression, Discriminant Analysis and K-Nearest Neighbor Classifier and compares the performance of classification and the predictive accuracy among them. The results show that Deep Neural Networks perform classification more accurately than the other methods (an error rate equal to 0.02 and 0.03 respectively) concerning cancer detection.

5. References

1. Jo H, Han I. Integration of case - based forecasting, neural network, and discriminant analysis for bankruptcy prediction. Expert Systems with Applications. 1996; 11(4):415-422.
2. Muller GH, Steyn-Bruwer BW, Hamman WD. Predicting financial distress of companies listed on the JSE – comparison of techniques. South African Journal of Business Management. 2009; 40(1):21-32.
3. Hunter WC, *et al*. The Asian financial crisis : origins, implications, and solutions. Springer, 1999.
4. Basel Committee on Banking Supervision, International convergence of capital measurement and capital standards: A revised framework comprehensive version. Bank of International Settlements, 2006.
5. Thanh TN, Binh AN, Manh H, Tung VN, Giao NPA method for speech Vietnamese recognition based on deep learning. International Journal of Multidisciplinary Research and Growth Evaluation. 2021, 2(4):152-156.
6. Xiong, Tengke, *et al*. Personal bankruptcy prediction by mining credit card data. Expert systems with applications. 2013; 40(2):665-676.
7. Berry M, Linoff G. Mastering data mining: The art and science of customer relationship management. New York: John Willey & Sons, Inc, 2000.
8. Paolo G, Linoff G. Bayesian data mining, with application to benchmarking and credit scoring. Applied Stochastic Models in Business and Society. 2001; 17:69-81.
9. Han J, Kamber M. Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann, 2001.
10. Hand DJ, Mannila H, Smyth P. Data mining: Practical machine learning tools and techniques. Cambridge: MIT Press, 2001.
11. Paolo G. Applied data mining: Statistical methods for business and industry. New York: John Wiley & Sons, Inc, 2003.
12. Witten IH, Frank E. Data mining: Practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufman, 1991.
13. Islam MJ, Jonathan Wu QM, Ahmadi M, Sid-Ahmed MA. Investigating the performance of Naïve-Bayes Classifiers and K-nearest neighbor classifiers. IEEE, International Conference on Convergence Information Technology, 2007, 1541-1546.
14. Chandola V, Arindam B, Kumar V. Anomaly detection: a survey. ACM Computing Surveys. 2009; 41(3):15-72.
15. Patil S., Nemade V., Soni P.K. Predictive Modelling for credit card fraud detection using data analytics. Procedia Computer Science, 132, 385-395.
16. Asha RB, Suresh Kumar KR. Credit card fraud detection using artificial neural network. Global Transitions Proceedings, 2021.
17. Prajal S, Pranali T, Ketan N, Mahyavanshi N. A novel idea for credit card fraud detection using decision tree. International Journal of Computer Applications, 2017; 161(13):6-9.
18. Dunn JC. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics. 1973; 3:32-57.
19. Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.
20. Giao NP. Short solution proposal for stroke detection based on deep learning and 3D point cloud. International Journal of Multidisciplinary Research and Growth Evaluation. 2021; 2(5):334-336.