www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 8 August 2017, Page No. 22250-22258
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i8.06

Data Mining For Breast Cancer Classification

Akinsola Adeniyi F[1], Sokunbi M.A[2], Okikiola F.M[3], Onadokun I.O[4]

[1]Yaba College Of Technology, Yaba, Lagos Nigeria
akinsolafun@yahoo.com
[2] Yaba College Of Technology, Yaba, Lagos Nigeria
mikesokunbi@yahoo.com
[3] Yaba College Of Technology,Yaba, Lagos Nigeria
Sade.mercyl@yahoo.com
[4] Yaba College Of Technology, Yaba, Lagos Nigeria
onadokun@yahoo.com

Abstract: Breast cancer poses a serious threat in several developed as well as developing countries like Nigeria. The use of machine learning and data mining techniques has revolutionized the whole process of predicting the presence of breast cancer. In this work, we evaluate and investigate three selected classification algorithms using Waikato Environment for Knowledge Analysis (WEKA for short). WEKA is an open source data mining software mainly used for academic and research purpose. The algorithms tested are C4.5, multi-layer perceptron and Naive Bayes. Experimental results show that C4.5 proves to be the best algorithm with highest accuracy.

Keywords: Breast cancer, Data mining, Prediction, WEKA

1. Introduction

Data mining which is sometimes called data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information. It is the process of discovering new patterns from large data sets involving methods from statistics and artificial intelligence and also database management. Data mining concept is actually part of the knowledge discovery process. Data mining are used today by companies with a strong consumer focus in various domains such as medical, healthcare, higher education, telecommunication etc. There are several data mining functions such as Concept descriptions, Association Rules, Classification, Prediction, Clustering and Sequence discovery to find the useful patterns

Breast cancer is an uncontrolled growth of breast cells. It refers to a malignant tumor that has developed from cells in the breast. Breast Cancer constitutes a major public health issue globally with over 1 million new cases diagnosed annually, resulting in over 400,000 annual deaths and about 4.4 million women living with the disease. It also affects one in eight women during their lives. It is the commonest site specific malignancy affecting women and the most common cause of cancer mortality in women worldwide. It is also found in men but not very common. Statistics available in Nigeria are largely unreliable because of many factors that have not allowed adequate data collection and documentation; but according to Dr Chinyere Akpanika, [1] a Gynaecologist at the University of Calabar Teaching Hospital (UCTH) said Nigeria recorded over 100,000 new cases of cancer annually, added that early detection of the disease increased the chance of survival of an affected person. According to her, 85 per cent of women who have breast cancer do not have a family history of breast cancer. The Medical Director, Optimal Cancer Care Foundation, Dr Femi Olaleye, [2] said breast cancer killed one in every 25 Nigerian women.

Okobia [3] said "Late presentation of patients at advanced stages when little or no benefit can be derived from any form of therapy is the hallmark of breast cancer in Nigerian women." This is indeed a worrisome trend and it appears to be the norm in Nigeria.

Cancer occurs as a result of mutations, or abnormal changes, in the genes responsible for regulating the growth of cells and keeping them healthy. That changed cell gains the ability to keep dividing

without control or order, producing more cells just like it and forming a tumor. A tumor can be benign (not dangerous to health) or malignant (has the potential to be dangerous).

2. Related Work

In their work Giarratana, G *et.al* [4] they used different data mining technique for diagnosis & the prognosis of breast cancer with the main parameter of male and female gene behavior, they take gene expression data set of 311 instance to test and validate model and major the performance. They prove classification data mining algorithm provide more optimum outcome.

Abdelaal, M.M.A. *et.al* [5] the authors of this paper focuses on diagnosing of early Brest cancer in women using SVM, Tree Boost and Tree Forest data mining classification technique.

Qi Fan *et.al* [6] the authors of this paper focus on SEER public use-data to predict Brest Cancer. They use pre-classification method and find a possible solution to discover thein formation of Brest Cancer.

H.S.Hota [7] built a classification model using various intelligent techniques such as ANN(Artificial Neural Network),Unsupervised Artificial Neural Network, Statistical technique and decision tree. Experimental results show a testing accuracy of 97.73% from which the efficiency of the ensemble model was highlighted.

S.VijayaRani *et.al* [8] the authors analyzed the performance of c4.5, RIPPER and PART algorithm. Time and Number of rules generated were taken as the measures to analyze Breast cancer Wisconsin data and heart disease data. The author concludes that PART algorithm is best suited for the above said data.

Shukla *et.al* [9] examined a knowledge based system for early diagnosis of breast cancer using Artificial Neural Networks(ANN) and Neuro fuzzy system. The performance measures considered were accuracy of diagnosis, training time, Number of neurons and No. of epochs. Simulation results show that this knowledge based system enhanced the survival rates effectively.

In Nov 2012, a system which detects the cancer stage as benign or malignant using Adaptive Resonance Theory (ART2) neural network was proposed by Sonia Narang *et.al* [10]. Neural network approach was adopted to handle Wisconsin breast cancer data. As it is a continuous data clustering was applied to extract knowledge and performance measures such as precision, recall and accuracy was analyzed. From the results it has been observed that by using ANN model accuracy rate could be improved.

Andreeva,*et.al.* [11] made a comparative study of different learning models used in data mining and provided some practical guidelines to select an algorithm for a specific medical application. Many classification algorithms were applied for breast cancer, diabetes and iris data. Among various classification algorithms Bayesian classification and SMO served with highest accuracy.

Abdulah,*et.al* [12] conducted a research on comparison of four data mining tools namely weka, orange, tanagra, KNIME for classification purpose. In order to judge the toolkits nine different datasets were used by them. Results concluded Weka toolkit was the best one in terms of classifiers applicability issue.

In this paper, a comparative study of three supervised learning algorithms are made to best predict the breast cancer dataset using WEKA proved to be the best toolkit.

3 Causes of breast cancer

3.1 Genetic factors

1. **Gender**: Breast cancer occurs nearly 100 times more often in women than in men.
2. **Age**: Two out of three women with invasive cancer is diagnosed after age 55.
3. **Race:** Breast cancer is diagnosed more often in women and it occurs in every races.
4. **Family history and genetic factors:** If a relative has been diagnosed with breast or ovarian cancer, such person has a higher risk of being diagnosed before the age of 50.
5. **Personal health history:** if a person has been diagnosed with breast cancer in one breast, there is an increased risk of been diagnosed with breast cancer in the other breast in the future.
6. **Menstrual and reproductive history:** Early menstruation (before age 12), late menopause (after age 55), having your first child at an older age, or never having given birth can also increase the risk of breast cancer.
7. **Certain genome:** this is cause by mutation in certain genes and can be determined through genetic test as individual with certain gene mutation can pass it onto their children.
8. **Dense breast tissue:** this can increase the risk of having breast cancer and makes lumps harder to detect.

3.2 Enviromental and lifestyle risk factors

1. **Lack of physical activity:** A sedentary lifestyle with little physical activity can increase the risk for breast cancer.
2. **Poor diet:** a diet high in saturated fat and lacking fruits and vegetables can increase the risk for breast cancer.
3. **Being overweight or obese:** This can also increase the risk for breast cancer.
4. **Drinking alcohol:** frequent consumption of alcohol can increase the risk of breast cancer.
5. **Radiation chest:** having a radiation to the chest before the age of 30 can increase the risk of breast cancer.
6. **Combined hormone replacement**: Taking combined hormone replacement therapy, as prescribed for menopause can increase the risk of breast cancer.

4) **Types of breast cancer tumour**

There are various types of tumors which are; Benign, Premalignant and Malignant.

4.1 **Benign Tumor:** A benign tumor (benign neoplasm) cannot metastasize - it cannot spread. Examples include uterine fibroids and moles. "Benign" means it is non-progressive, it remains as it is. Most benign tumors are not harmful to human health. Even though they are not cancerous, some may press against nerves or blood vessels and cause pain or other negative effects. Benign tumors of endocrine tissues may result in the excessive production of some hormones. Examples of benign tumors include:

a) **Adenomas:** Adenomas are tumors that arise from glandular epithelial tissue - epithelial tissue is the thin membrane that covers glands, organs and other structures in the body. A polyp in the colon is a type of adenoma. Although adenomas are not cancerous, they can change and become so; then they are called adenocarcinomas.

b) **Fibroids (Fibromas):** Fibroids (fibromas) are benign tumors that grow on fibrous or connective tissue of any organ in the body. Uterine fibroids are common. Uterine fibroids can cause vaginal bleeding, pelvic pain or discomfort, and urinary incontinence.

c) **A soft fibroma of the eyelid:** The fibroma durum (hard fibroma) is made up of many fibers and few cells. The fibroma molle (soft fibroma) is made up of several loosely connected cells and less fibroid tissue. Soft fibroma is usually found in the armpits, groin, neck and eyelids. There are many types of fibromas, such as angiofibroma, cystic fibroma (fibroma cysticum), myxofibroma (fibroma myxomatodes), nonossifying fibroma, ossifying fibroma, cemento-ossifying fibroma, pleomorphic fibroma, fibroma of tendon sheath nuchal fibroma, chondromyxoid fibroma, desmoplasmic fibroma, collagenous fibroma, and perifollicular fibroma. Some fibromas can cause symptoms and may require surgical removal. Rarely, fibroids can change and eventually become cancerous, they are then called fibrosarcomas.

d) **Hemangiomas:** A hemangiomas on the scalp of a child. Hemangiomas are benign tumors which consists of a collection of too many blood cells. They can sometimes be seen on the surface of the skin and are colloquially called strawberry marks. The majority of hemangiomas appear at birth and gradually go away after some months or years. Hemangiomas do not usually require any treatment. If they affect the patient's ability to eat, hear or see, the doctor may recommend treatment with corticosteroids. If the patient is over 10 years of age, they are more commonly removed today using laser surgery.

e) **Lipomas:** Lipomas are the most common form of soft-tissue tumor. Lipomas consist of adipose tissue (fat cells). Most of them are very small, painless, soft to the touch, and generally movable. They are more common among people aged 40+ years. Experts disagree on whether lipomas can change and become cancerous (malignant).

4.2) **Premalignant Tumor:** A premalignant or precancerous tumor is one that is not yet malignant, but is about to become so. Examples of premalignant growths include:

a) **Actinic keratosis** - also known as senile keratosis or solar keratosis is a premalignant growth consisting of crusty, scaly and thick patches of skin. Fair-skinned people are more susceptible to these types of growths, especially those who are exposed to sunlight (it is linked to solar damage).They are seen as potentially premalignant because a number of them progress to squamous cell carcinoma. Doctors usually recommend

treating them because of this. There is a 20% risk that untreated lesions eventually become cancerous. Continuous sun exposure increases the risk of malignancy.

b) **Dysplasia of the cervix** - the normal cells lining the cervix of the uterus change. The growth can be premalignant, a prelude to cervical cancer. Cervical dysplasia is diagnosed with a PAP smear. According to the National Institutes of Health, USA, about 5% of PAP smears detect the presence of cervical dysplasia. They are more common in women aged 25 to 35. They may be removed with Cryotherapy (freezing), or conization (the cone of tissue from the cervix is removed).

c) **Metaplasia of the lung** - the growths occur in the bronchi, tubes that carry air from the windpipe into the lung. The bronchi are lined with glandular cells, which can change and become squamous cells. Metaplasia of the lung is most commonly caused by smoking.

d) **Leukoplakia** - thick, white patches form on the gums, bottom of the mouth, insides of the cheeks, and less commonly on the tongue. They cannot be scraped off easily. Experts believe tobacco smoking and/or chewing is the main cause. Although Leukoplakia is rarely dangerous, a small percentage are premalignant and can eventually become cancerous. Many mouth cancers occur next to areas of leukoplakia. If smokers quit, the condition usually clears up. Quitting both alcohol and tobacco together has better results. The patches can be removed using laser, a scalpel or a cold probe that freezes the cancer cells (cryoprobe).

4.3) **Malignant Tumor**

Malignant tumors are cancerous tumors, they tend to become progressively worse, and can potentially result in death. Unlike benign tumors, malignant ones grow fast, they are ambitious, they seek out new territory, and they spread (metastasize).

The abnormal cells that form a malignant tumor multiply at a faster rate. Experts say that there is no clear dividing line between cancerous, precancerous and non-cancerous tumors - sometimes determining which is which may be arbitrary, especially if the tumor is in the middle of the spectrum. Some benign tumors eventually become premalignant, and then malignant. Metastasis - malignant tumors invade nearby cells, and then the cells near those, and spread. Some cells can break off from the tumor and spread to various parts of the body through the bloodstream or the lymphatic system, and establish themselves anywhere in the body, and form new malignant tumors. Metastasis is the process by which cancer cells spread from their primary site to distant locations in the human body. For example, a patient may have started off with melanoma (skin cancer) which metastasized in their brain. The cancer cells that metastasize are the same as the original ones. If a lung cancer spreads to the liver, those cancer cells that grow in the liver are lung cancer cells which have acquired the ability to invade other organs.

There are different types of tumors, which are made up of specific types of cancer cells:

a) Carcinoma - these tumors are derived from the skin or tissues that line body organs (epithelial cells). Carcinomas can be, for example, of the stomach, prostate, pancreas, lung, liver, colon or breast. Many of the most common tumors are of this type, especially among older patients.

b) Sarcoma - these are tumors that start off in connective tissue, such as cartilage, bones, fat and nerves. They originate in the mesenchymal cells outside the bone marrow. The majority of sarcoma tumors are malignant. They are called after the cell, tissue or structure they arise from, for example fibrosarcoma, liposarcoma, angiosarcoma, chondrosarcoma, and osteosarcoma.

c) Lymphoma/Leukemia - cancer arises from the blood forming (hematopoietic) cells that originate in the marrow and generally mature in the blood or lymph nodes. Leukemia accounts for 30% of childhood cancers. Leukemia is thought to be the only cancer where tumors are not formed.

d) Germ cell tumor - these are tumors that arise from a germ cell, pluripotent cells (cells than can turn into any kind of cell). Germ cell tumors most commonly present in the ovary (dysgerminoma) or testicle (seminoma). The majority of testicular tumors are germ cell ones. Less commonly, germ cell tumors may also appear in the brain, abdomen or chest.

e) Blastoma - tumors derived from embryonic tissue or immature "precursor" cells. These types of tumors are more common in children than adults. "Blastoma" is often the root word used in longer ones that describe

tumors, for example, medulloblastoma and glioblastoma are kinds of brain tumors, retinoblastoma is a tumor in the retina of the eye, osteoblastoma is a type of bone tumor, while a neuroblastoma is a tumor found in children of neural origin.

5) Methodology

The study utilized breast cancer data obtained from Federal government Hospital Lagos. Three supervised learning algorithms namely: C4.5, Multi-layer perceptron (MLP) and Naïve Bayes were investigated using WEKA toolkit. The simulation result thereafter compared.

A) Breast cancer data mart

The source dataset used was extracted from the database of Federal government hospital in Lagos. It has a total of 33,000 data and a dimension of 3000 rows and 11 columns. A number of transformations had to be performed before a suitable working dataset was built.

The initial dataset was obtained towards managing the health of people with breast cancer. These data were not designed with Data Mining in mind. Fields appeared more than once.

The final dataset contained over 1700 unique instances each being represented by 11 attributes (Sample Code Number, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion Fibrous, Epithelial Cell Size, Bare Nucleicx, Bland Chromatin, Normal Nucleoli, Mitoses, Diagnosis of tumors) including the predicted class.

The class distribution is framed as Benign and malignant. There are 1 dependent variable and 9 independent variables. The values for the independent variables ranges from 1 - 10 and for class variable 2 for Benign and 4 for malignant tumor. The minimum possibilities for a person to get breast cancer are 1 and the ‘maximum possibilities are represented by the value 10.

B) Supervised learning algorithms

Three supervised learning algorithms: C4.5, multi-layer perceptron and Naive Bayes were used to classify the breast cancer data.

i) C4.5

C4.5 was developed by Quinlan Ross which is an extension to ID3 [13]. It is mainly used for generating decision tree. The splitting area defined here is gain ratio. C4.5 classification uses entropy and information gain for tree splitting. It is suitable for handling both categorical as well as continuous data. A threshold value is fixed such that all the values above the threshold are not taken into consideration. The initial step is to calculate information gain for each attribute. The attribute with the maximum gain will be preferred as the root node for the decision tree. Given a set S of cases, C4.5 first grows an initial tree using the divide-and-conquer algorithm as follows:

- If all the cases in S belong to the same class or S is small, the tree is a leaf labeled with the most frequent class in S.
- Otherwise, choose a test based on a single attribute with two or more outcomes. Make this test the root of the tree with one branch for each outcome of the test, partition S into corresponding subsets S1, S2,... according to the outcome for each case, and apply the same procedure recursively to each subset.

ii) Naive Bayes

Naive Bayes Classifier is a probabilistic model based on Baye's theorem. It is defined as a statistical classifier. It is one of the frequently used methods for supervised learning. It provides an efficient way of handling any number of attributes or classes which is purely based on probabilistic theory. Bayesian classification provides practical learning algorithms and prior knowledge on observed data [14].

- Let **X** be a data sample : class label is unknown
- Let H be a *hypothesis* that X belongs to class C
- Classification is to determine P(H|**X**), (i.e., *posteriori probability):* the probability that the hypothesis holds given the observed data sample **X**
- P(H) (*prior probability*): the initial probability
- P(**X**): probability that sample data is observed
- P(**X**|H) (likelihood): the probability of observing the sample **X**, given that the hypothesis holds
- training data **X**, *posteriori probability of a hypothesis* H, P(H|**X**), follows the Baye's theorem

$$P(X|H) = \frac{P(X|H)\ P(H)}{P(X)} = P(X|H) \times P(H)/P(X)$$

iii) **Multi-Layer Perceptron (MLP)**

It is a type of classifier that is based on neural networks. A multilayer perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs.

DATASETS USED

Dataset is a collection of data or a single statistical data where every attribute of data represents variable and each instance has its own description. For prediction of breast cancer we used breast cancer set for prediction and classification of algorithms in order to compare their accuracy using wekas explorer interface based on Naïve Bayes, C 4.5 and Multilayer perceptron (MLP).

WEKA accepts the data in ARFF format (attribute relation file format), CSV format that is comma separated values.

Table 1 Training dataset tuples

Attribute Name	Description	Category	Range Values
SCNo	Sample Code Number	Id	-
CT	Clump Thickness	Ordinal	1-10
UCSIZE	Uniformity of Cell Size	Ordinal	1-10
UCSHAPE	Uniformity of Cell Shape	Ordinal	1-10
MAF	Marginal Adhesion Fibrous: Fibrous bands tissue that form between two surfaces	Ordinal	1-10
ECSIZE	Epithelial Cell Size : Size of a single cell that forms tissues that lines the outside of the body and the passageways that lead to or from the surface	Ordinal	1-10
BN	Bare Nuclei	Ordinal	1-10
BC	Bland Chromatin: Evaluates for the presence of Bare bodies	Ordinal	1-10
NN	Normal Nucleoli	Ordinal	1-10
M	Mitoses: Cell growth	Ordinal	1-10
Dia	Diagnosis of tumours	Class	2,4

From Table 1 it is understood that UCSIZE (Uniformity of Cell Size), UCSHAPE (Uniformity of Cell Shape), CT (Clump Thickness), BN(Bland Chromatin) and MA(Marginal Adhesion) are the best attributes.

C) WEKA

In order to carry out this experimentations and implementations, WEKA was used as the data mining tool. WEKA (Waikato Environment for Knowledge Analysis) is a data mining tool written in java developed at Waikato [15]. WEKA is a very good data mining tool for modeling breast cancer data. These are open source data mining software mainly used for academic and research purposes. Explorer, Experimenter and Knowledge flow are the interface available in WEKA that has been used by us. The data file normally used by WEKA is in ARFF file format, all data is considered as instances and features in the data are known as attributes.

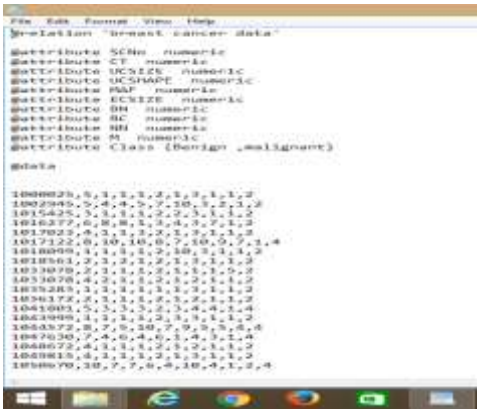
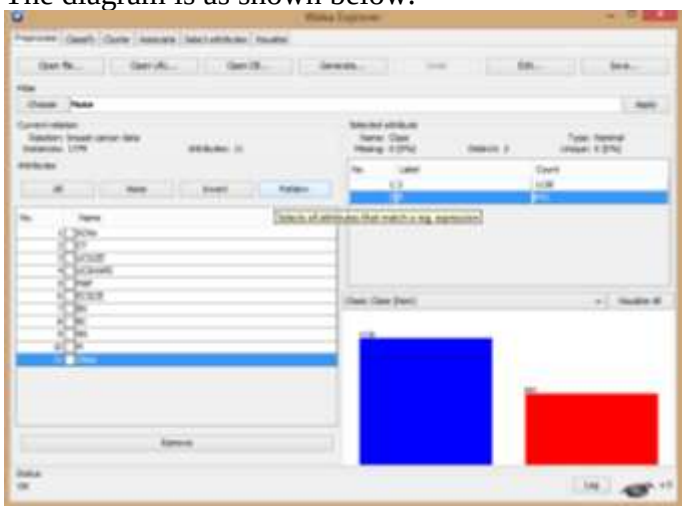


Fig. 1: sample ARFF (Attribute relation file format) breast cancer data in WEKA.

Preprocess Interface

This is the interface where the preprocess dataset will be loaded into the machine learning tool. The interface will show the number of instances, attributes and also calculate the mean and standard deviation by getting the minimum and maximum number from each attribute values from the dataset. The diagram is as shown below.



The performance of the three algorithms using WEKA is shown below in screenshot. They were evaluated based on two criteria (The Prediction Accuracy and Time Taken to build the model)

DOI: 10.18535/ijecs/v6i8.06

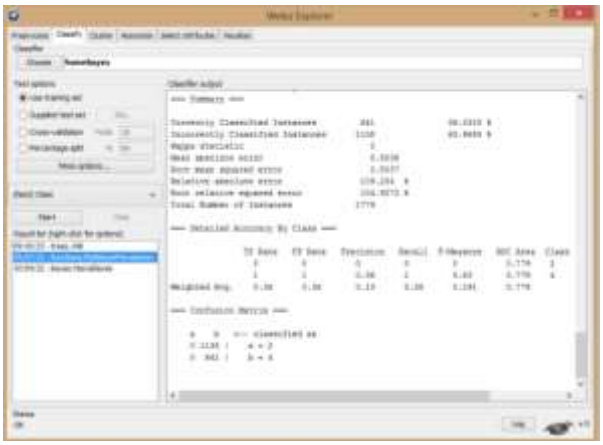


Fig. 2 Dataset training output multilayer perceptron screenshot

Fig. 3 Dataset training output for Naïve Bayes screenshot

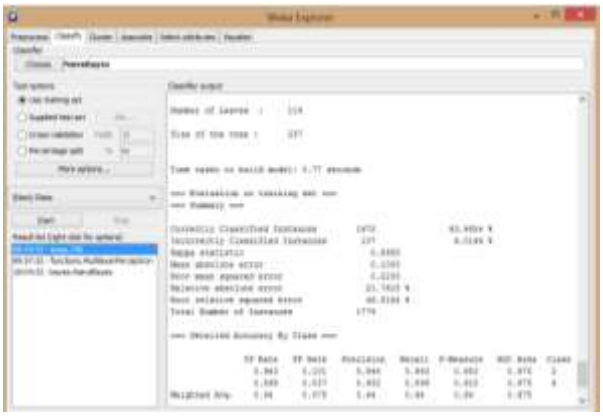


Fig. 4 Dataset training output for C4.5 screenshot

6) **Result**

The simulation results are partitioned into several sub items for easier analysis and evaluation. The algorithm will be compared by analyzing their performance; execution time, mean absolute time and Root mean squared error as shown in Tables 1 and 2 below. Table 1 mainly summarizes the result based on accuracy and time taken for each simulation, while Table 2 shows the result based on error during simulation. Figure 4 shows the graphical representation of the simulation result.

Table 2: Accuracy of various classifiers compared

Algorithm (Total Instances, 1779)	Correctly Classified Instances % (value)	Incorrectly Classified Instances % (Value)	Time Taken (seconds)	Kappa Statistic
C4.5	**93.9854 (1672)**	**6.0146 (107)**	**0.28**	**0.8685**
Naïve Bayes	**76.5037 (1361)**	**23.4963 (418)**	**0.07**	**0.5035**
Multilayer perceptron	**83.8673 (1492)**	**0.0562 (16.1327)**	**12.68**	**0.6416**

Table 3: Training and simulation errors

Algorithm (Total Instances, 1779)	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error (%)	Root Relative Squared Error (%)
C4.5	**0.1003**	**0.2233**	**21.7615**	**46.5164**
Naïve Bayes	**0.2503**	**0.4167**	**54.2951**	**86.7976**
Multilayer Perceptron	**0.2179**	**0.3354**	**47.2562**	**69.8599**

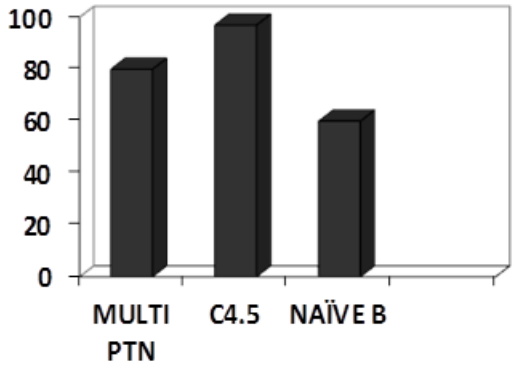


Fig. 5: Accuracy of Basic Classifiers

7) Conclusion

As a conclusion, we have met our objective which is to evaluate and investigate three selected classification algorithms. The best algorithm based on the breast cancer data is C4.5 with an accuracy of **93.9854%** and the total time taken to build the model is at **0.28 seconds,** followed by the Multi-layer perceptron with an accuracy of **83.8673%** and the total time taken to build the model is at **12.68 seconds** and Bayes network classifier with an accuracy of **76.5037%** and the total time taken to build the model is at **0.03 seconds.** These results suggest that among the machine learning algorithm tested, C4.5 classifier has the potential to

significantly improve the conventional classification methods for use in medical or in general, bioinformatics field.

8) **Recommendation**

This system is to create the flexible and dynamic system that helps the hospital management to control or maximize the prediction of breast cancer. From the technology view, it is suggested that this system add another available function not only to be used by the doctor or nurses but should be able to relate to patient on its own just like the ATM machine's technology. The additional features must give benefit to either the doctor or the patient. Nevertheless the development of the system must be according to current demand which the medical environment requires.

References

1. Dr Chinyere Akpanika, (2016, October 29). Nigeria records 100,000 new breast cancer cases yearly — Gynaecologist. *Punch Newspaper*. Retrieved from (http://punchng.com)
2. Dr Femi Olaleye, (2013, Nov. 22). One in every 25 Nigerian women dies of cancer – Expert. *Vanguard Newspaper* Retrieved from *http://www.vanguardngr.com.*
3. Okobia Michael N, Clareann H Bunker, Friday E Okonofua, & Usifo Osime (2006, Feb. 21) "Knowledge, attitude and practice of Nigerian women towards breast cancer: A cross-sectional study" Retrieved from (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1397833/)
4. Giarratana, G. ;Dipt.diElettron. e Inf., Politec. di Milano, Milan, Italy ; Pizzera, M. ; Masseroli, M. ; Medico, E. (2009). "Data Mining Techniques for the Identification of Genes with Expression Levels Related to Breast Cancer Prognosis" . DOI:10.1109/BIBE.2009.37
5. Abdelaal, M.M.A. ; Stat. & Math. Dept., Ain Shams Univ., Cairo, Egypt ; Farouq, M.W. ; Sena, H.A. ; Salem, A.-B.M. "Using data mining for assessing diagnosis of breast cancer a literature review" Retrieved from (http:/./ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=Brest+cancer+papre+in+data+mining) on 15/6/17
6. Qi Fan ; Dept. of Comput. Sci., Huaibei Coal Ind. Teacher Coll., Huaibei, China ; Chang-Jie Zhu ; Liu Yin "Predicting breast cancer recurrence using data mining techniques a literature review" Retrieved from (http://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=Brest+cancer+papre+in+data+mining) 15/6/17
7. Hota, H.S. (2013) "Diagnosis of Breast Cancer Using Intelligent Techniques", *International Journal of Emerging Science and Engineering*, ISSN:2319-6378,Vol.1,Issue 3.
8. Vijayarani, S. and Divya, M. (2011) "An Efficient Algorithm for Generating Classification Rules", *International Journal of Computer Science and Technology*,Vol.2,Issue 4,.
9. Shukla, A. Tiwari R. & Kapur, P. "Knowledge based approach for diagnosis of breast cancer", (2009) *Advance Computing Conference ,2009. IACC, 6 & 7March 2009*. E-ISBN:978-1-4244-2928-8, published in IEEE xplore digital library.
10. Sonia Narang,Harsh K Verma & UdaySachdev,(2012) "Breast Cancer Detection using ART2 Model of Neural Networks", *International Journal of Computer Applications*, ISSN:0975-8887, Vol.57-No.5,
11. Andreeva, P., Dimitrova, M., & Pertia, R. (n.d.). *Data Mining Learning Models and Algorithms for Medical Application. https://pdfs.semanticscholar.org/1d23/ed260d90f41bae16c5974e3ea369a1cfd83d.pdf*
12. Abdulah, W. H., Al-Radaideh, Q. A., Al-Kabi, M. N., & Al-Shawakfa, E. M. (2011). A Comparison Study between Data Mining Tools over some Classification Methods. *International Journal of Advanced Computer Science and Applications*. (DOI) : 10.14569/SpecialIssue.2011.010304
13. Xindong Wu ,Vipin Kumar ,J. Ross Quinlan, et al.,‖ Top 10 algorithms in data mining‖, *Knowledge Information System*, Vol.14, pp.1–37.
14. Han J .and Kamber, M.,―Data mining: concepts and techniques, Academic Press, ISBN 1- 55860-489-8.
15. WEKA Retrieved from http://www.WEKA.com/

DOI: 10.18535/ijecs/v6i8.06

Author Profile

Akinsola Adeniyi Foluso obtained B.Sc in Computer Science/Accounting from the Ogun State University Ago Iwoye in 1994. He also obtained M.Sc in Computer Science from the University of Lagos Nigeria in 2008. He worked briefly in some Corporation and now a full time lecturer at Yaba College of Technology Lagos Nigeria

.