International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-6, September 2015
ISSN: 2395-3470
www.ijseas.com

Data Mining: A Hybrid Approach on the Clinical Diagnosis of Breast Tumor Patients

Onuodu F. E.[1], Eke B. O.[2]
[2] bathoyol@gmail.com, University of Port Harcourt, Port Harcourt, Nigeria
[1] University of Port Harcourt, Port Harcourt, Nigeria

Abstract
There are huge numbers of patient's data that exist in the databases of some laboratories all over the world today, but clinical diagnosis is still being handled by manual computation. This paper presents a system that diagnoses breast tumors and categorizes them as either cancerous or not, using a hybrid data mining algorithm proposed in the paper. The system "learns" by clustering known patient data sets to find underlying classification patterns. The data set used in this work is obtained from the University of California Irvine (UCI) repository. Our design was built using the Object Oriented Analysis and Design methodology (OOADM) and implemented with Java programming language and MySQL as the database admin back end. The results obtained show that the system could assist clinicians diagnose easily by identifying tumors that are cancerous or non-cancerous, thereby suggesting early treatment for patients with breast cancer diseases.

Keywords: *Data mining, Hybrid Algorithm, Diagnosis, Breast, Cancer.*

1. Introduction

The widespread use of computer and information technology has made extensive data collection in business, manufacturing and medical organizations a routine task. This explosive growth in stored data has generated an urgent need for new techniques that can transform the vast amounts of data into useful knowledge. Data mining is, perhaps, most suitable for this need.

Clustering can be recognized as the unsupervised classification of patterns into groups. The goal of a clustering algorithm is to group the objects of a database into a set of meaningful subclasses[1]. Hospitals and healthcare centers are now capable of collecting more comprehensive data that allows the application of methods in order to improve the management of diagnostic tasks involving the coordination of laboratory equipment. Such technologies can provide accurate information on stages of disease conditions. The scientists have been using these databases for performing number of experiments.

In data mining, clustering is a widely used technique that partitions a data set consisting of n points embedded in an m-dimensional space into k distinct clusters such that the data points within the same cluster are more similar to each other than to data points in other clusters [2]. For the purpose of this research, we use partition cluster to enhance and give accurate diagnosis of clinical tumors. Most of these approaches are based on the iterative optimization of a criterion function depicting the agreement between the data and the partition. There are huge numbers of patient's data that exist in the databases of some laboratories all over the world today, but clinical diagnosis is still being handled by manual computation. This paper presents a system that diagnoses breast tumors and categorizes them as either cancerous or not, using the hybrid algorithm. The system "learns" by clustering known patient data sets to find underlying classification patterns. The data set used in this work is obtained from the University of California Irvine (UCI) repository [3]. Our design was built using the Object Oriented Analysis and Design methodology (OOADM) and implemented with Java programming language and MySQL as the database. The results obtained show that the system could assist clinicians diagnose easily by identifying tumors that are cancerous or non-cancerous, thereby suggesting early detection and treatment for patients with breast cancer diseases.

1.1 Data Mining

Past literatures have reviewed that Data Mining is an important research field in the area of clinical diagnosis. These days, the role of data generation and collection are producing data sets from variety of scientific disciplines. The most common type of breast cancer is ductal carcinoma, which begins in the lining of the milk ducts (thin tubes that carry milk from the lobules of the breast to the nipple). Another type of breast cancer is lobular carcinoma, which begins in the lobules (milk glands) of the breast. Invasive breast cancer is breast cancer that has spread from where it began in the breast ducts or lobules to surrounding normal tissue. Breast cancer occurs in both men and women, although male breast cancer is rare [4].

Breast cancer classification divides breast cancer into categories according to different schemes, each based on different criteria and serving a different purpose. The major categories are the histopathological type, the grade of the tumor, the stage of the tumor, and the expression of proteins and genes. As knowledge of cancer cell biology develops these classifications are updated. The purpose of classification is to select the best treatment. The effectiveness of a specific treatment is demonstrated for a specific breast cancer usually by randomized and controlled trials. That treatment may not be effective in a different breast cancer. Some breast cancers are aggressive and life-threatening, and must be treated with aggressive treatments that have major adverse effects. Other breast cancers are less aggressive and can be treated with less aggressive treatments, such as lumpectomy. Treatment algorithms rely on breast cancer classification to define specific subgroups that are each treated according to the best evidence available. Classification aspects must be carefully tested and validated, such that confounding effects are minimized, making them either true prognostic factors, which estimate disease outcomes such as disease-free or overall survival in the absence of therapy, or true predictive factors, which estimate the likelihood of response or lack of response to a specific treatment [5].

Classification of breast cancer is usually, but not always, primarily based on the histological appearance of tissue in the tumor. A variant from this approach, defined on the basis of physical exam findings, is that inflammatory breast cancer (IBC), a form of ductal carcinoma or malignant cancer in the ducts, is distinguished from other carcinomas by the inflamed appearance of the affected breast, which correlates with increased cancer aggressively [4]. Risk factors for developing breast cancer include: female sex, obesity, lack of physical exercise, drinking much alcohol, hormone replacement therapy during menopause, ionizing radiation, early age at first menstruation, having children late or not at all, and older age [5]. A 2013 Cochrane review stated that it is unclear if mammographic screening does more good or harm. A 2009 review for the US Preventive Services Task Force found evidence of benefit in those 40 to 70 years of age. And the organization recommends screening every two years in women 50 to 74 years old [6]. The medications tamoxifen or raloxifene may be used in an effort to prevent breast cancer in those who are at high risk of developing it. Surgical removal of both breasts is another useful preventative measure in some high risk women. In those who have been diagnosed with cancer, a number of treatments may be used, including surgery, radiation therapy, chemotherapy, hormonal therapy and target therapy. Types of surgery vary from breast-conserving surgery to mastectomy. In those in whom the cancer has spread to other parts of the body, treatments are mostly aimed at improving quality of life and comfort.

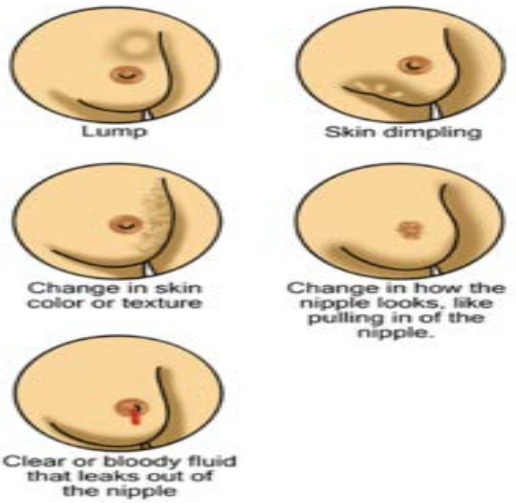


Figure 1: Breast Cancer, Source (Wiki, 2015)

Figure 1 gives an illustration of breast cancer and the various stages it goes through. In stage one, the lump formation begins, stage two shows the skin dimpling, in stage three, there is a change in the colour and texture of the skin, and in the final stage, there could be fluid discharge. Outcomes for breast cancer vary depending on the cancer type, extent of disease, and person's age. Survival rates in the developed world are high, with between 80 percent and 90% of those in England and the United States alive for at least 5 years. In developing countries survival rates are poorer. Worldwide, breast cancer is the leading type of cancer in women, accounting for 25 percent of all cases. In 2012 it resulted in 1.68 million cases and 522,000 deaths [6]. It is more common in developed countries and is more than 100 times more common in women than in men. Figure 2 shows a tumor spreading in a patient's chest [12].

1.2 Medical History and Physical Exam

In the case where a patient has any signs or symptoms that might mean breast tumor, the first thing is to consult with a doctor. The doctor asks questions about the symptoms as well as any other health problems, and possible risk factors for benign breast conditions or malign breast cancer. The patient's breasts will be thoroughly examined for any lumps or suspicious areas and to feel their texture, size, and relationship to the skin and chest muscles. Any changes in the nipples or the skin of the breasts will be noted. The lymph nodes in the armpit and above the collarbones may be palpated (felt), because enlargement or firmness of these lymph nodes might indicate spread of breast cancer. The doctor will also do a complete physical exam to judge the patient's general health and whether there is any evidence of cancer that may have spread.

1.3 Imaging Tests Used To Evaluate Breast Disease

Imaging tests use x-rays, magnetic fields, sound waves, or radioactive substances to create pictures of the inside of the body. Imaging tests may be done for a number of reasons, including to help find out whether a suspicious area might be cancerous, to learn how far cancer may have spread, and to help determine if treatment is working.

1.4 Mannograms

A mammogram is an x-ray of the breast. Screening mammograms are used to look for breast disease in women who have no signs or symptoms of a breast problem. Screening mammograms usually take two views (x-ray pictures taken from different angles) of each breast. For a mammogram, the breast is pressed between two plates to flatten and spread the tissue. This may be uncomfortable for a moment, but it is necessary to produce a good, readable mammogram. The compression only lasts a few seconds. If a patient has tumor symptoms (like a lump or nipple discharge) or an abnormal result on a screening mammogram, the patient will have a diagnostic mammogram. This will include more images of the area of concern. If the diagnostic mammogram shows that the abnormal area is more suspicious for cancer, a biopsy will be is needed to tell if it is cancer [7].

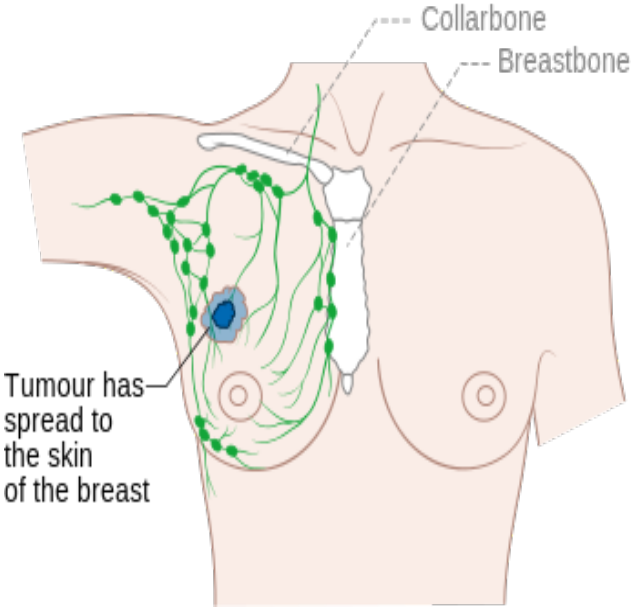


Figure 2: Tumor spreading

1.5 Breast Ultrasound

Ultrasound, also known as sonography, uses sound waves to outline a part of the body. For this test, a small, microphone-like instrument called a transducer is placed on the skin (which is often first lubricated with ultrasound gel) [14]. It emits sound waves and

picks up the echoes as they bounce off body tissues. The echoes are converted by a computer into a black and white image that is displayed on a computer screen. This test is painless and does not expose you to radiation. The use of ultrasound instead of mammograms for breast cancer screening is not recommended. However, clinical trials are now looking at the benefits and risks of adding breast ultrasound to screening mammograms in women with dense breasts and a higher risk of breast cancer [7].

1.6 Magnetic Resonance Imaging (MRI) Of The Breast

MRI scans use radio waves and strong magnets instead of x-rays. The energy from the radio waves is absorbed and then released in a pattern formed by the type of body tissue and by certain diseases. A computer translates the pattern into a very detailed image. For breast MRI to look for cancer, a contrast liquid called gadolinium is injected into a vein before or during the scan to show details better. MRI scans can take a long time − often up to an hour. For a breast MRI, the patient has to lie inside a narrow tube, face down on a platform specially designed for the procedure. The platform has openings for each breast that allow them to be imaged without compression. The platform contains the sensors needed to capture the MRI image [6]. It is important to remain very still throughout the exam.

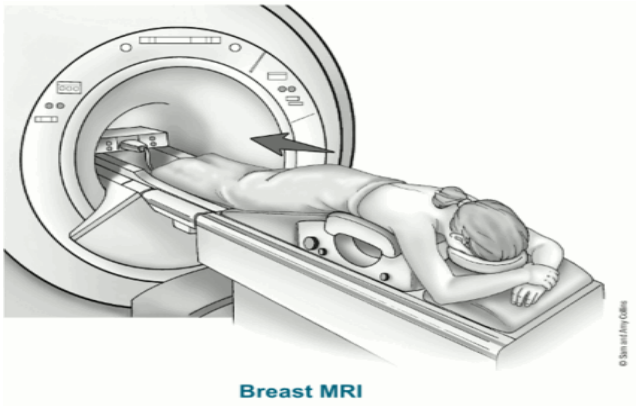


Figure 3: MRI machine used in diagnosis

2. Data mining

There are challenges in traditional data analysis techniques as new types of datasets emerge. In order to cope with these new challenges, researchers have been developing more efficient and scalable tools that can more easily handle diverse types of data. In particular, data mining draws upon ideas such as sampling, estimating and hypothesis testing from statistic, search algorithms, modeling techniques and learning theories from artificial intelligence, pattern recognition and machine learning[8]. Data mining has been adopting from other areas such as optimization, evolutionary computing, information theory, signal processing, visualization and information retrieval and the use of databases in order to provide support for storage, index and query processing. Figure 4 represents relationship of data mining with other areas [9].

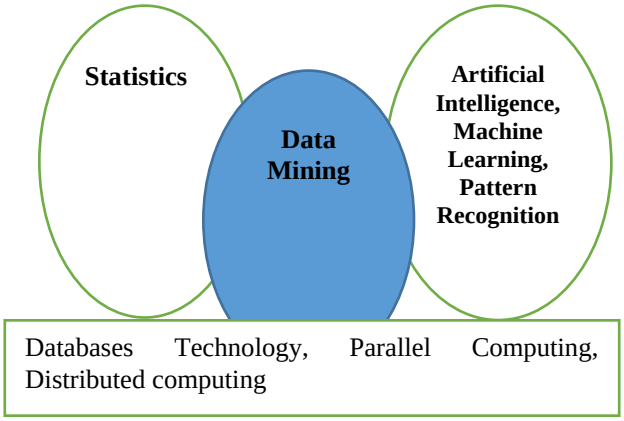


Figure 4: Data mining as a confluence of many disciplines

International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-6, September 2015
ISSN: 2395-3470
www.ijseas.com

2.1Analysis

The existing system of clinical diagnosis involves the use of tested software to analyze results of laboratory experiments on breast tumors. Patient tumor sample is obtained, taken to the laboratory for testing, the result is then analyzed using special software tools [9]. The disadvantages of the existing system has no way to learn from patient data, even though large amounts of data exists in the databases of laboratories [11].

2.2 The Proposed Algorithm

The proposed system employs the use of a hybrid algorithm in clustering patient data into a number of clusters (K) so as to find underlying patterns and similarities amongst the data. Patient breast tumors can thus be classified as either benign or malign [10]. The clustering process of the *k*-prototypes algorithm is similar to the *k*-means algorithm [16] except that it uses the *k*-modes approach to updating the categorical attribute values of cluster prototypes. The algorithm is as follows:

Step 1: Using Euclidean distance as a dissimilarity measure, compute the distance between every pair of all the objects as follow.

$$d_{ij} = \sqrt{\sum_{a=}^{p}(X_{ia} - X_{ja})^2} \qquad i\&j = 0,1,2 \ldots\ldots\ldots\ldots \text{n} \qquad (1)$$

Step 2: Calculate Mij to make an initial guess at the centers of the clusters

$$M_{ij} = \frac{d_{ij}}{\sum_{i=1}^{n} d_{ij}} \qquad (2)$$

Step 3: Calculate $\sum_{i=1}^{n} M_{ij}{}^{2}$ …… (3)

At each object and sort them in ascending order.

Step 4: Select K objects having the minimum value as initial cluster centroids which are determined by the above equation. Arbitrarily choose *k* data points from *D* as initial centroids.

Step 5: Assign each point *d*i to the cluster which has the closest centroid.

Step 6: Calculate the new mean for each cluster.

Step 7: Repeat step 5 and step 6 until convergence criteria is met.

3. Design

The design of a generic hybrid method is shown in figure 4, which simply explains the processes involved in clustering breast tumor diseases.

(a) The problem is formulated. This means that the variables or features on which clustering is done are selected. The set of variables selected should show the similarity between objects so that this similarity or dissimilarity is relevant for marketing problems. Anyhow the ideal features should be useful to distinguish the patterns belonging to different clusters, immune to noise, and easy to obtain and interpret.

(b) The data is prepared as often some changes are needed before conducting the analysis.

The process is designed and selected. The key questions here are how cluster should be formed. To answer this question, distance measure is selected. The most common approach is to measure similarity in terms of distance between pairs of objects. An often used measure of similarity is Euclidean distance or its square. This is the square root of the sum of the squared differences in values for each variable. As the distance measure is decided, clustering can be constructed as an optimization problem with a specific criterion function. We adopted the Object Oriented Analysis and Design Method (OOADM) in the analysis and design system of hybrid clustering algorithm to predict breast cancer problems [13]. We implemented with JAVA and MySQL as database. The nature of real world objects has a big impact in terms of their classes, attributes and operations.

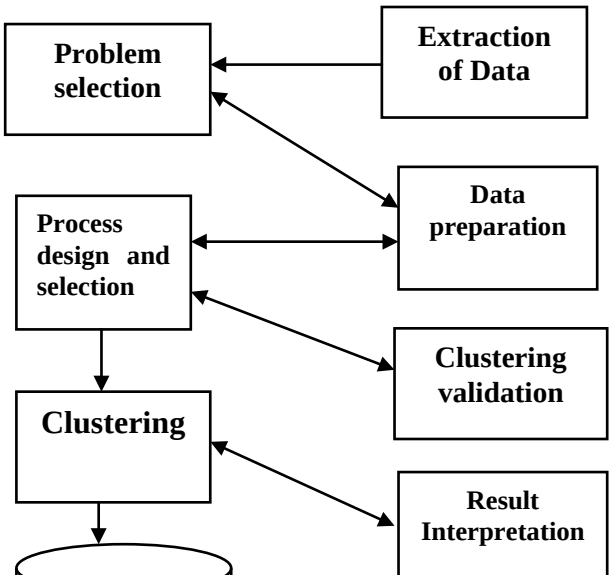


Figure 5: Clustering procedure with basic steps

4. Results and Discussion

Our application can be distributed by extracting the zipped folder called "BREAST TUMOR CLASSISIFIER". The Zipped folder contains the java archive file for running our application. It also contains a copy of our database, which is bundled together in the zipped folder.

1. Double click on the jar file
2. On the Display window, select the number of clusters k.
3. Click "perform clustering".

Figure 6 shows the Breast tumor Classifier application. The application starts up, the user is then required to select a value of k, which represents the number of clusters[17]. On selecting the number of clusters (k), and clicking the "cluster" button, the system begins to iterate over the patient data in the database using the k-prototype algorithm. The data used for the application is retrieved from the University of California (Irvine) repository. Figure 7 shows the "winconson.txt" file which contains the data retrieved from the UCI repository. Each data set contains Eleven (11) attributes described in the "Winconson_names" file. The field attributes are as follows- sample code number, clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, Normal nucleoli, and mitosis. The last is the categorical attribute, which has a value of either four (4) or (2). A 2 denotes a benign tumor, while a 4 denotes a malign tumor.

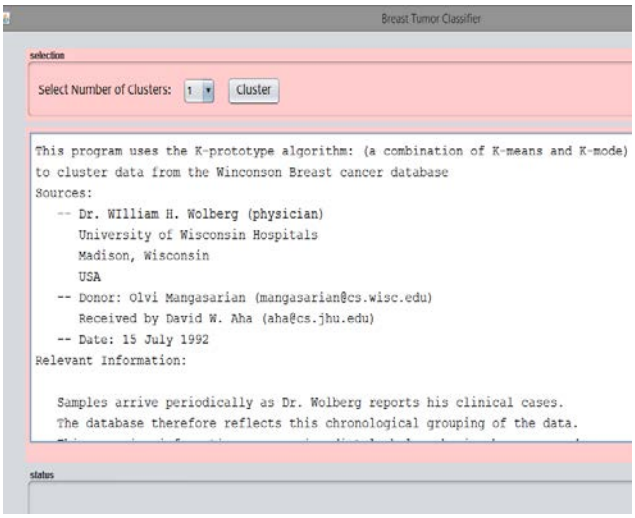


Figure 6: The Breast tumor application

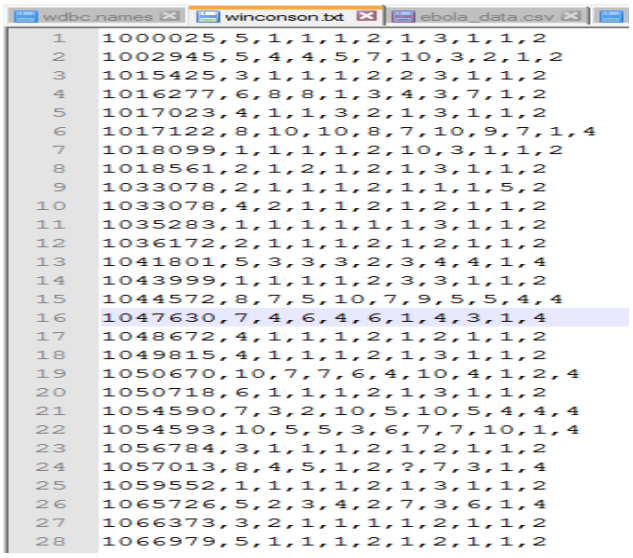


Figure 7: Patient data set retrieved from the UCI repository

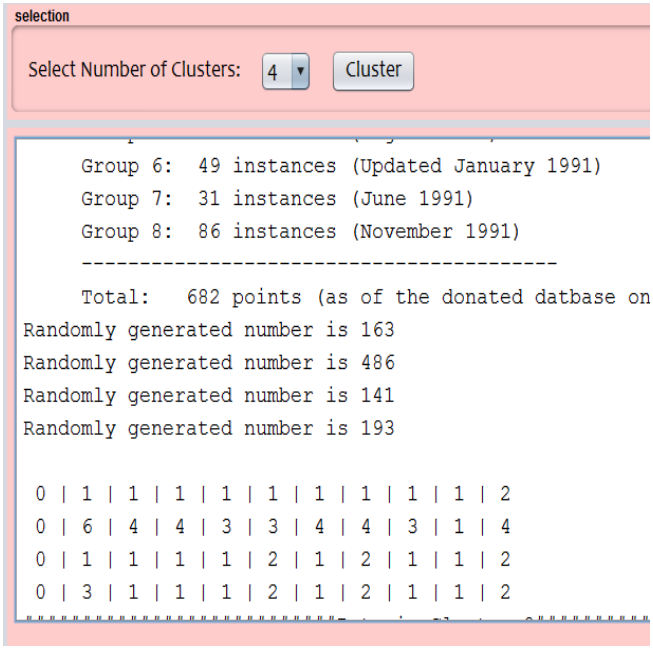


Figure 8: Centroid Initialization

Figure 8 shows the centroid initialization phase of the application. The system begins by retrieving all patient data available in the database. User selects four (4) clusters. Value of k is thus 4. The system proceeds to generate four random numbers, which it uses in picking four breast cancer data set from the data set available In the database. The initial centroids are shown in Figure 8.

0 | 1 | 1 | 1| 1 | 1 | 1 | 1 | 1 | 1 | 2

0 | 6 | 4 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 4

0 | 1 | 1 | 1| 1 | 2 | 1 | 2 | 1 | 1 | 2

0 | 3 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2

The explanation is that, centroid(0) contains patient breast tumor data that are categorized as bening (non – cancerous). Centroid(1) contains patient breast tumor data that are categorized as malign (cancerous). Centroid(2) contains tumor data that are categorized as malign, and Centorid(3) contains tumor data that are categorized as bening. The system proceeds to the assignment phase, where it assigns each patient tumor data to the centroid having the minimum distance from the tumor data. Figure 9 shows the result of the assignment phase. Patient tumor data have been assigned to Cluster(0). The system uses K-means algorithm to generate the new attribute value for each attribute. It computes the Euclidean distance d_i between each Cluster C_i and each tumor data point p_i. It uses a function to find the minimum distance $d_{\min}$. Thus, a tumor data point p_i is allocated to the centroid that produces $d_{\min}$. This process is repeated for each cluster.

After each iteration, the system has to determine the value for the category-class of each cluster, which is a categorical value; this is unlike the numerical values that can easily be derived using K-means. To determine the new categoy-class, the system uses the K-modes algorithm [11]. It computes a mode m_i for each category-class c_i in each cluster. The mode refers to the number of occurrences. The new category-class for the centroid will therefore be the most occurring m_i in the cluster. For figure 10, the new category-class will thus be 2, since 2 occurs more times than 4. The system has to re-calculate centroid values for each cluster on every iteration. Figure 9 shows the result of the computation. The system is done with the entire clustering process, which runs for a total of 206 seconds. The attribute values for centroid(0) are shown in Figure 10. Figure 11 shows the result discussion. The system not only classifies the patient data, it also interprets the result in human-readable form, which is missing in most clustering applications.

########Data in Cluster 0########
1070935 | 1 | 1 | 3 |1 | 2 | 1 | 1 | 1| 1 | 2
1165297 | 2 | 1 | 1 |2 | 2 | 1 | 1 | 1| 1 | 2
1167439 | 2 | 3 | 4 |4 | 2 | 5 | 2 | 5| 1 | 4
1173347 | 1 | 1 | 1 |1 | 2 | 5 | 1 | 1| 1 | 2
1184184 | 1 | 1 | 1 |1 | 2 | 5 | 1 | 1| 1 | 2
1190485 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
1197440 | 1 | 1 | 1 |2 | 1 | 3 | 1 | 1| 7 | 2
1199219 | 1 | 1 | 1 |2 | 1 | 1 | 1 | 1| 1 | 2
1204242 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
1214092 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
183913 | 1 | 2 | 2 |1 | 2 | 1 | 1 | 1| 1 | 2
560680 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
688033 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
693702 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
704097 | 1 | 1 | 1 |1 | 1 | 1 | 2 | 1| 1 | 2
792744 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
704097 | 1 | 1 | 1 |1 | 1 | 1 | 2 | 1| 1 | 2
814911 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
826923 | 1 | 1 | 1 |1 | 2 | 1 | 1 | 1| 1 | 2
831268 | 1 | 1 | 1 |1 | 1 | 1 | 1 | 3| 1 | 2

Figure 9: Patient tumor data in cluster(0) after first iteration

5. Conclusion

The manual system of diagnosis has the drawback of time and resource wastage. It leads to loss of lives that could have been saved, since early detection helps save lives. This could be achieved by automating the manual system and the using a reliable and tested application to manage these data. Therefore, the implementation of this design will assist professionals in the Clinical diagnosis of Breast Tumor in carrying out their service effectively and efficiently. The work has been able to review the traditional diagnostic process and how a system that learns from patient data will better improve diagnostic process. The research documentation presented here includes the design and implementation of breast tumor diagnosis system, which implements and improves diagnosis process to provide an output that is consistent with laboratory needs. An intelligent system, such as the proposed system will result in an overall reduction of problem areas that have become evident in the existing system. We hereby recommend that the system should be deployed online so that patients can make use of it.

Table 1: Result and Discussion table

CENTROID	RESULT
[0]	A Breast tumor with clump thickness 1, uniformity of cell size 1, uniformity of cell shape 1, marginal adhesion 1, single Epithelial cell size 1, bare nuclei 1, bland chromatin 1, normal nuceoli 1 and mitosis 1 will most likely be benign (Not cancerous)
[1]	A Breast tumor with clump thickness 6, uniformity of cell size 4, uniformity of cell shape 4, marginal adhesion 3, single Epithelial cell size 3, bare nuclei 4, bland chromatin 4, normal nuceoli 3 and mitosis 1 will most likely be Malignant (cancerous)
[2]	A Breast tumor with clump thickness 1, uniformity of cell size 1, uniformity of cell shape 1, marginal adhesion 1, single Epithelial cell size 2, bare nuclei 1, bland chromatin 2, normal nuceoli 1 and mitosis 1 will most likely be benign (Not cancerous)
[3]	A Breast tumor with clump thickness 3, uniformity of cell size 1, uniformity of cell shape 1, marginal adhesion 1, single Epithelial cell size 2, bare nuclei 1, bland chromatin 2, normal nuceoli 1 and mitosis 1 will most likely be benign (Not cancerous)

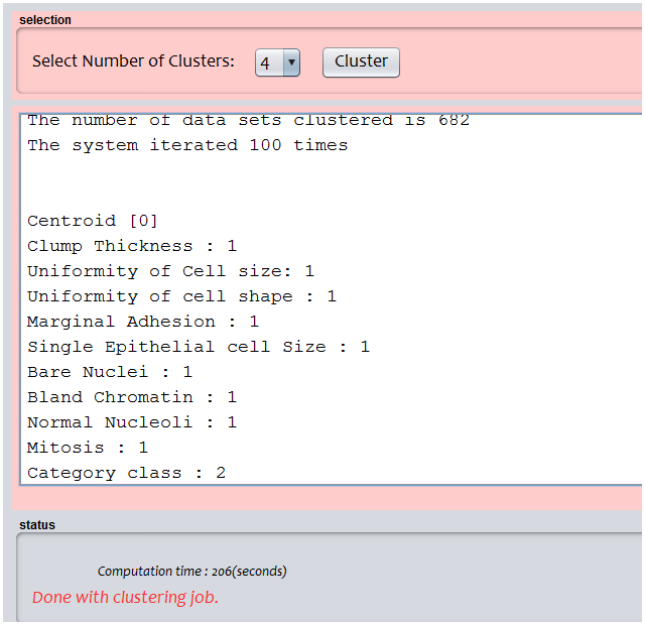


Figure 10: Final centroid value for centroid [0]

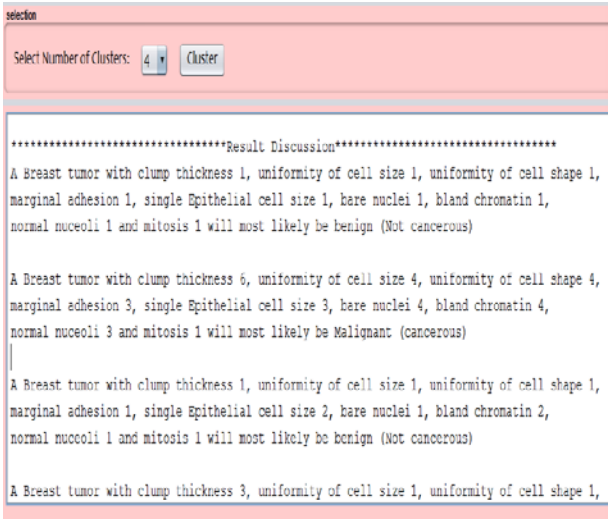
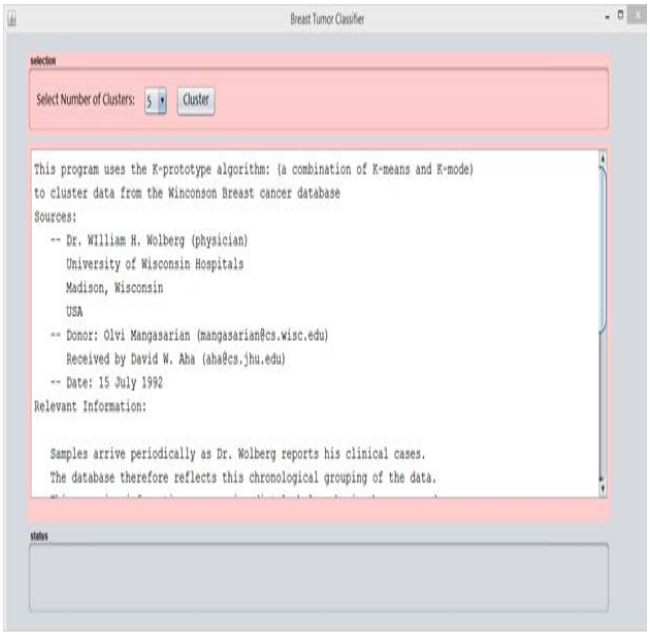
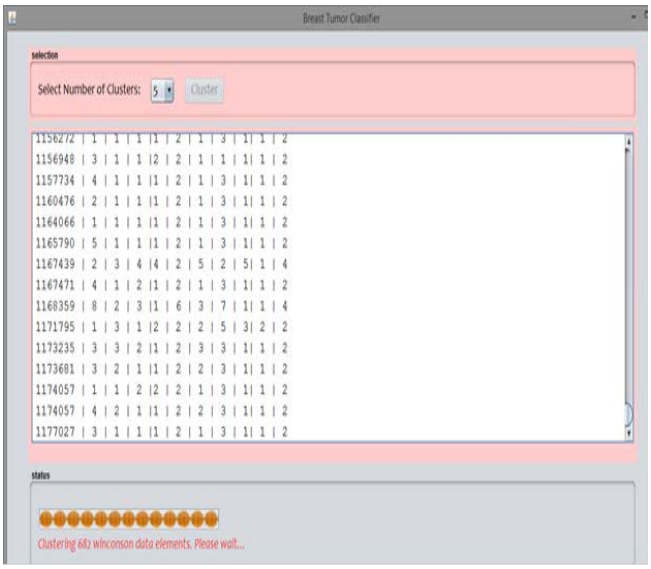


Figure 11: Result Discussion

APPENDIX : SAMPLE PROGRAM OUTPUT



OUTPUT FOR BREAST TUMOR CLASSIFIER



OUTPUT FOR CLUSTERING OF SAMPLES

Result of classification

References

[1] Chetna S. and Garima M.. (2013). A Linear PCA based hybrid K-Means PSO algorithm for clustering large dataset. International Journal of Scientific & Computer Science and Information Systems, Vol12 142-158, London.

[2] Guha, S., Rastogi, R., Shim K. (1998) CURE: An Efficient Clustering Algorithm for Large Databases, Published in the Proceedings of the ACM SIGMOD Conference.

[3] Asuncion A., and Newman D. J. (2007), UCI machine learning repository: http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science.

[4] NCI (2014) Breast Cancer Treatment PDQ, National Cancer Institute. 2014-05-23. Retrieved 29 June 2014.

[5] WCR (2011) World Cancer Report. International Agency for Research on Cancer 2008" Retrieved 2011-02-26.

[6] WCR (2014) World Cancer Report 2014. World Health Organization. 2014. pp. Chapter 5.2.ISBN92-832-0429-8.

[7] Srimani P. K. and Manjula S. K. (2013), Medical Diagnosis Using Ensemble Classifiers-A Pattern Recognition Letters, 25 (11), 1293-1302. Proceedings of the IEEE International Conf. Data Mining, USA, 331-338.

[8] Bradley, P.S, Fayyad, U.M. (1998): Refining Initial Points for K-Means Clustering, Proceedings Conference Proceedings, 10th International Conference on Modeling and Applied Conference Proceedings, Advances in Data Mining–Applications and Theoretical.

[9] Gumjan V. and Vineeta V, (2012): An Introduction to Cluster Analysis for Data Mining International Journal Of Computer Application, Vol. 48, 17-20.

[10] Gower, J., Diday, E. (1991): Symbolic Clustering Using a New Dissimilarity Measure, Pattern Recognition Letters , 24(6), 567-578.

[11] Cookey R., and Lerner S. (2009) Advances over K-means". International Journal of knowledge-based and Intelligent and anomalous cluster initializing in K-Means clustering," *Pattern Recognition*, vol. 45,

[12] Fayyad, Usama; Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996), from Data mining to knowledge discovery in data base.

[13] Irny, S.I and Rose, A.A (2005), Designing a Strategic Information System Planning

[14] Jinchao Ji, Wei Pang, Chunguang Zhou, Xiao Han, and Zhe Wang (2012), "A fuzzy k-prototype clustering algorithm for mixed

numeric and categorical data," Knowledge-Based Systems, vol. 30, pp. 129-135.

[15] Pang-Ning-Tan M. S. (2006) “Data mining,” in Introduction to data mining, Pearson

[16] Xu R. and Wunsch D. (2005), “Survey of Clustering Algorithms,” IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678.

[17] Wesan, Barbakh and Colin Fyfe. (2008) “Local vs. global interactions in clustering algorithms: with Categorical Values, Data Mining and Knowledge Discovery 2, Netherlands ,283–304.

First Author

Dr. Onuodu, Friday E. is a Lecturer at the University of Port Harcourt. His research interest is in Data Mining and Data Extraction using both Mobile Devices and Desktops. He also has interest in Soft Computing and data analysis. He has many publication in learned journals.

Second Author

Eke Bartholomew PhD, MCPN, MACM, FIPMD is a Software Engineering and Computer Science Lecturer at the University of Port Harcourt and Java Application Developer in Oyol Computer Consult Inc. His research interest is in Software Engineering Methodologies and Embedded systems.