

Decision Support System for Histopathological Diagnosis of Breast Diseases in Women

Aderonke A. Kayode¹, Babajide S. Afolabi², Bernard I. Akhigbe³, Ifiok J. Udo⁴ and A. Ominiya⁵
Information Storage and Retrieval Group
Department of Computer Science & Engineering,
Obafemi Awolowo University, Ile-Ife, Nigeria.

Abstract

This paper presents a representation of histological features for histopathological diagnosis of breast diseases in women. Hence, a Decision Support System (DSS) for histopathological interpretation and diagnosis of breast diseases was implemented and evaluated. The Expert knowledge used was elicited through interview and literature search. The needed diagnostic knowledge was represented using diseases' profile in the form of frame. UML, JAVA and MYSQL were used for the design and implementation of the system. 150 samples of retrospective cases were used for the system's implementation, while a Consultant Pathologist's interpretation was used to evaluate the system. Results for Sensitivity, Specificity, Positive Prediction Value and the Negative Prediction Value are 97.7%, 95.0%, 99.2% and 86.3% respectively. Thus, the result showed that the system is capable of assisting an inexperienced pathologist in making accurate, consistent and timely diagnoses, also in the study of diagnostic protocol, education, self-assessment, and quality control.

Keywords: *Histological features, Histopathological diagnosis, Expert knowledge, Diagnostic knowledge, Pathologist, Decision Support System*

1. INTRODUCTION

Histopathological interpretation of multiple types of tissue and cytological specimens do provide crucial information. The accuracy of this information is critical to the provision of good health care. Tissue evaluations are generally viewed as "gold standard" medical facts that provide the highest quality, most reliable, diagnostic evidence available. There is likely no investigative modality that can match the economic yield of information available to a

skilled pathologist, when interpreting a tissue section stained with basic hematoxylin and eosin [1], [2], [3].

However, the diagnosis of breast diseases using histopathological means requires both visual and logical skills to accurately interpret microscopic images. The cognitive heuristics involved in the recognition of pathologic visual patterns are clearly related to training and experience. But the decision-making processes involved in this realm are poorly defined. As in many situations, the quality of decisions is important; aiding the deficiencies of human judgment and decision making has been a major focus of science throughout history [4].

The concept of a Decision Support System (DSS) is extremely broad and its definitions vary depending on the author's point of view [4]. But first, decisions are often the choices made between alternatives and are based on estimates of the values of these alternatives. Hence, supporting a decision means helping people (who work alone or in a group), to gather intelligence, generate alternatives and make choices. The DSS is usually referred to as computer applications that perform this type of supporting role. They can take any different forms and can be used in many different ways [5]. [6] defined it as "a computer-based system that aids the process of decision making". More precisely, [7] defined it as an interactive, flexible, and adaptable computer-based information system. The system specially developed for providing solution to non-structured management problem for improved decision making, utilizes data, provides an easy-to-use interface, as well as allow the decision maker's own insights to be uninfluenced. For [8], it is an interactive computer based system that help decision makers utilize data and models to solve unstructured problems. As a

knowledge-based system, it is also a formalized knowledge domain that is amenable to computational reasoning [9].

Artificial Intelligence in Medicine (AIM) as a field emerged in the early 1970's in response to several simultaneous needs, opportunities, and interests. An increased demand for high-quality medical services coupled with the explosive growth of medical knowledge has led to the suggestion that computer programs could be used to assist physicians and other health care providers in discharging their clinical roles in diagnosis, therapy and prognosis [10]

The use of DSS in the diagnoses of breast cancers is not new. One of the earliest studies encountered was the one by [11]. In the study mammographic image analysis was investigated using a decision table to represent all the parameters and possibilities. About 41 rules were created, and all were centered upon masses and lesions. The other related works were mostly based on Artificial Neural Networks (ANN) for decision making in the diagnoses of breast cancer. Some of the diagnostic activities were also related to or influenced by breast mammographic decisions [12], [13], [14], [15], [16], and [17]. In like manner [18] proposed an artificial intelligent algorithm for tumor detection in screening mammogram. [19] devised a multiple circular path convolution neural network system for detection of mammographic masses. [20] classified mammographic breast density using a combined classifier paradigm. Also, [21] worked on the automatic classification of mammography parenchyma patterns in breast cancer and [22] presented an automatic mammographic diagnosis system for detecting breast cancer based on association rules (AR) and neural network (NN).

However, none of the aforementioned studies presented results that showed evidence of first, the inclusion of histological perspective; and secondly a system capable of assisting a Pathologist who is not specialized in the aspect of breast cancer diagnosis. Instead, their aim was at automating mammographic classification and diagnoses of breast cancers. Thus, this system proposes a decision support system for histopathological diagnosis of breast diseases as an improvement of earlier mammographic works. The next section presents a justification for this work, while the section that follows contains the architecture proposed for the system. The section after this presents the methodology of the work, and the results are discussed thereafter. Finally, the conclusion followed.

2. Justification of study

The abundance of histological patterns is often of complex and variable nature. When this occurs in breast diseases,

the pathologist is thus presented with several diagnostic problems. Therefore, the lofty challenge this portends, is that the Pathologist is required to be fully conversant with the diversity of possible patterns, recognize and diagnose them, timely and accurately. Hence, a Pathologist who is not a specialist in the pathology of the breast has to refer to textbooks and study past diagnosis before concrete diagnosis can be made and conclusion reached. This however can be painstaking, time consuming and the overall cost implication at the long run, both in terms of life and so on, is what considering. Hence, there is the need for a system, which can assist the Pathologist to reach timely and accurate decision, especially when life is at stake.

3. Architecture of Decision Support System

Three fundamental components of DSS were identified as, namely: the database management system (DBMS); the model-base management system (MBMS); and the dialogue generation and management system (DGMS) [24]. The DBMS stores information (which can be further subdivided into that derived from an organization's traditional data repositories, external sources such as the Internet or from the personal insights and experiences of individual users). The MBMS handles representations of events, facts, or situations, using various kinds of models. Thirdly, the DGMS, otherwise known as the User interface (UI), is the component that allows users to interact with the system. It was stated that the main product of an interaction with a DSS is insight. As a result, since users of the system are often managers who may not be computer-trained, there is the need to be equipped DSS with intuitive and easy-to-use interfaces [26]. The figure 1 below gives a precise view of a typical architecture of a DSS.

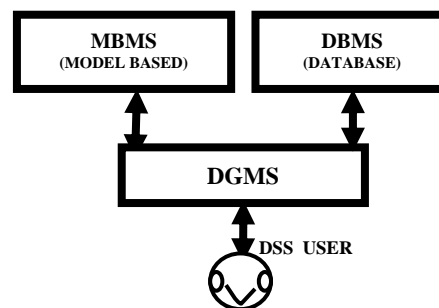


Figure 1: Decision Support System Architecture

4. Methodology of work

4.1 Knowledge Elicitation

Table 1 below shows the clinical and histological (microscopic) features that were used for the diagnosis of

breast diseases in the study. There were two major challenges with this endeavour: The acquisition of human expert knowledge and the provision of a bibliographical material on breast diseases that is broad by a knowledge engineer. The purpose for these; first was because expert knowledge is very important in the development of any knowledge based systems, such as a DSS. Secondly, a broad bibliographical material on breast diseases will provide sufficient background knowledge and obliterate the subjectivity of a single expert [27]. The knowledge used for the study was acquired by interviewing an expert (a pathologist), who is specialized in breast diseases. The study covers 21 breast diseases, which includes 13 benign lesions and 8 malignant lesions. One questionnaire per disease was employed. Figure 2 and figure 3 below are examples of the questionnaire. While figure 2 presents the Clinical information of the disease under diagnoses, figure 3 presents the microscopic (histological) features of the disease.

Table 1: Clinical and Microscopic Features used in the Histopathological diagnosis of breast diseases in Women.

Age group – (pre menopause, menopause, post menopause, nursing mother)
Nipple discharge – (watery, cloudy, bloody)
Nipple inversion – (yes, no, not applicable)
Lesion class – (benign, malignant)
Growth type – (infiltrating, non -infiltrating)
Proliferation – (stromal, epithelial)
Cell layer – (double, multiple)
Cell spacing – (regular, irregular, not applicable)
Cell cohesion – (tightly, loosely)
Cell population – ((monomorphic – (complete, incomplete), pleomorphic)
Lesion feature – (necrosis, hemorrhage, not applicable)
Inflammatory cells – (present, absent, not applicable)
Cytological atypia – (yes, no, not applicable)
Papillary growth – (present, absent, not applicable)
Apocrine metaplasia – (present, absent, not applicable)

Myoepithelial cell – (present, absent, not applicable)

Please tick (✓) the one that applies

(a) Age group			(b) Nipple Discharge		
	No	Yes		No	Yes
Pre-Menopausal			Bloody		
Pre-Menopausal			Cloudy		
Menopausal			Watery		
Nursing mother					
(c) Nipple inversion					
Yes					
No					
Not Applicable					

Figure 2: Clinical Information

Please tick (✓) the one that applies

(a) Lesion Class			(b) Proliferation			
	No	Yes		No	Yes	
Malignant			Stromal			
Benign			Ephithelial			
(c) Growth type			(d) Cell layer			
	No	Yes		No	Yes	
Infiltrating			Double			
Non-infiltrating			Multiple			
(e) Cell Spacing			(f) Cell cohesion			
	No	Yes	NA		No	Yes
Regular				Loose		
Irregular				Tight		
(g) Cell Population				No	Yes	
Complete monomorphic						
Incomplete monomorphic						
Pleomorphic						
(h) Lesion Feature		No	Yes	NA		
Necrosis						
Hemorrhage						
		No	Yes	NA		
(i) Cytologic						
(ii) al atypia						

(j) Papillary growth			
(k) Myoepithelial cells			
(l) Apocrine metaplasia			
Inflammatory cells			

Figure 3: Microscopic Information

4.2 Knowledge Representation

Frame in the form of "disease profiles" was used to represents knowledge in the system. The choice of frame was informed by the fact that as a representational scheme, it is rich enough to capture pathological knowledge. With this a sufficient structure was provided to allow a very useful component coordination, which is often important and the case in histological domain. A total of 21 histological disease profiles were generated. Figure 4 below shows the representational scheme used for this study.

```

ENTITY complex sclerosing lesion BEGIN
CLASS benign lesions;
OCCURRENCE a common lesion occurring
primarily in menopausal women;
SET clinical features BEGIN
age group= menopausal (A);
nipple discharge = bloody (A);
nipple inversion = not applicable(A);
END
SET microscopic features BEGIN
lesion class = benign (A), malignant(N);
proliferation = stromal (N), epithelial (N);
myoepithelial cells = present (A), absent(N);
apocrine metaplasia = present (A), absent(N);
Inflammatory cells = not applicable (A);
cell population = not applicable (A);
cell spacing = not applicable(A);
cell cohesion = not applicable(A);
periductal fibrosis = not applicable (A);
lesion features = not applicable (A);
END
END
    
```

Figure 4: A Sample Frame

4.3 System Design Using Unified Modeling Language (UML)

The system was designed using the Unified Modeling Language (UML). As an object oriented programming (OOP) tool, it was chosen so that the need for object modeling and the relationship between objects and classes in the design phase of program would be met. Also, with its use there was no need for any flowchart representation, since UML suffices to allow the design to be viewed from different perspectives. In order to adequately represent the structure of the propose system; the activity, sequence, and use case diagrams were used.

Figure 5 below is the Activity diagram used to depict the sequential flow of activities, which was used to model the various actions that will be performed when an operation is being executed as well as the results of the corresponding action. The sequence diagram in figure 6 describes how the objects in the system interact over time. Objects identified for the system were the patients, user, security system and the DSS. They interact in the sequence shown in the sequence diagram by passing messages across the timelines. These messages are the actions carried out by the objects in the system in a chronological order. For the use case diagram it was meant to describe the system's behaviour from a user's viewpoint. It was valuable as a veritable aid during system analysis and helped with the understanding of user's requirements. In the use case diagram shown in figure 7; the actor is the user (expert), since he/she is the individual that interacts with the system. While resources were identified as system security, clinical information, microscopic information and the diagnosis; the clinical and microscopic information were put into the knowledge base.

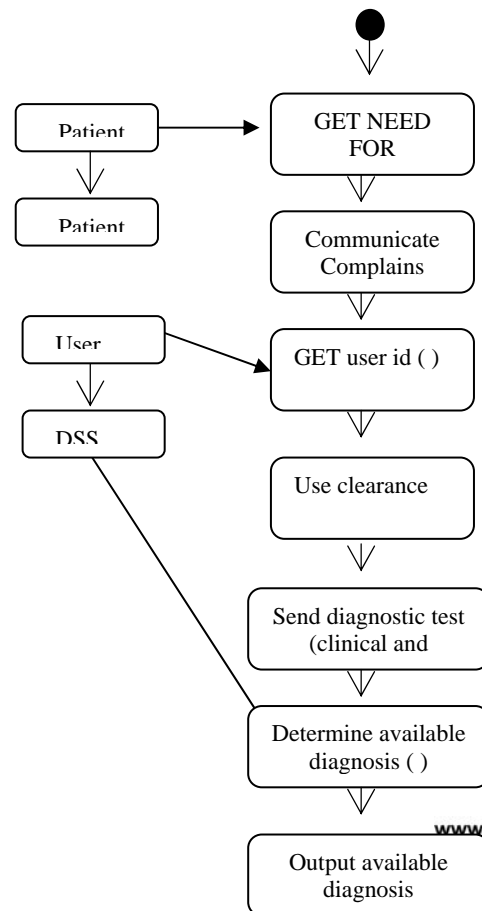


Figure 5: Activity diagram

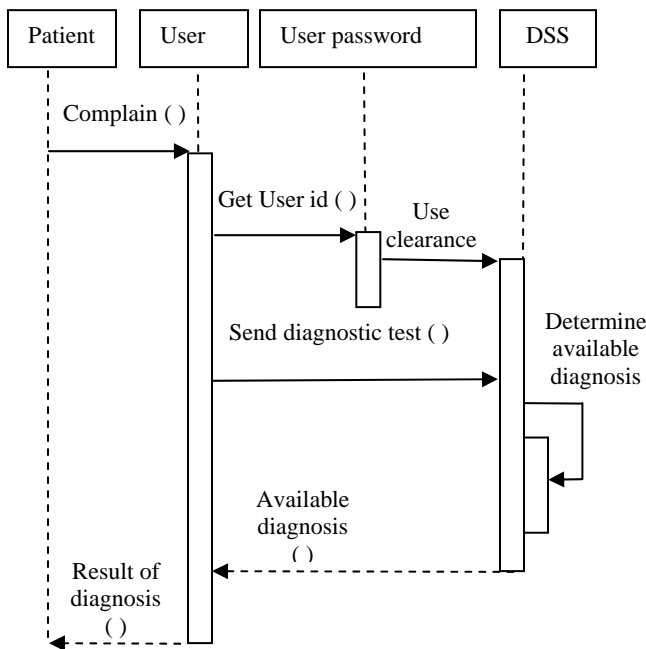
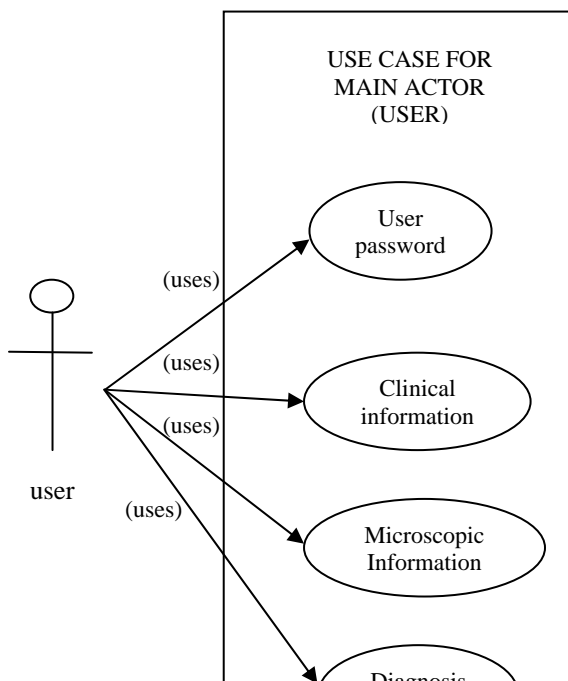


Figure 6: Sequence diagram



4.4 User Interface Design and System Integration

In order to build a user-friendly and intelligent environment, JAVA was used to implement the various user interface forms (UIFs) for the DSS. They contain questions necessary for determining abnormalities. The user of the system is requested to answer questions by choosing answers from a given lists. By this the user assists the DSS to reach a valid diagnosis. Figure 8 below shows a login UIF, which displays after the initial launching of system by a user. Using the login UIF, the user enters his/her username and the password. If the password is correct an introductory screen, as shown in figure 9 is displayed; then the user is prompt to characterize the clinical and the microscopic (histological) abnormalities through the UIF shown in figure 10 below.

Figure 8: Login Screen shot

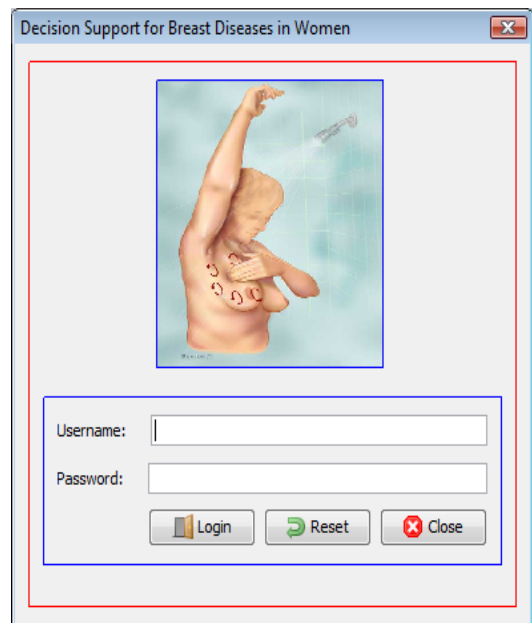




Figure 9: The Welcome screen shot

with the input data. When this is done, the inference engine takes over and produces a final diagnosis as shown in figure 11. The user may right-click on the diagnosed disease to see the evidence in favour of the disease this is displayed by the user interface in figure 12. On the contrary, whenever there is an error, the user is informed by a relative message. Also, the user is warned by a beeping sound whenever the request for continuity is made without any selected answer from the UIF. But ultimately the appropriate form, which contains the name of an expected diagnosis and its histological features displays to the user through the UIFs (figure 11 and 12) meant for the purpose. If no diagnosis is accomplished the user is informed by a corresponding message.

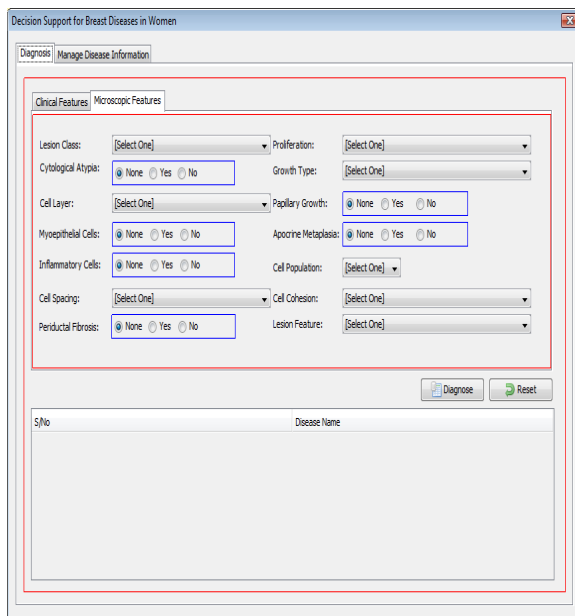


Figure 10: The Clinical and microscopic abnormalities form

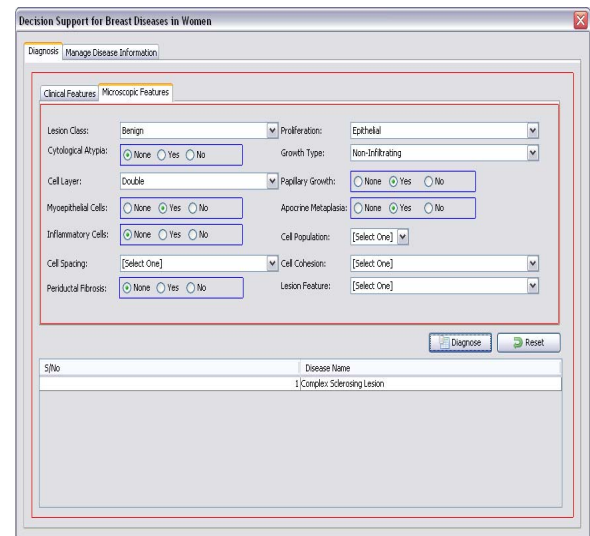


Figure 11: The Diagnostic form

After the characterization of the clinical and microscopic (histological) abnormalities, using the UIF mentioned above, user's answer as user's input data is feed into the system. As a result, the knowledge base is called and fed

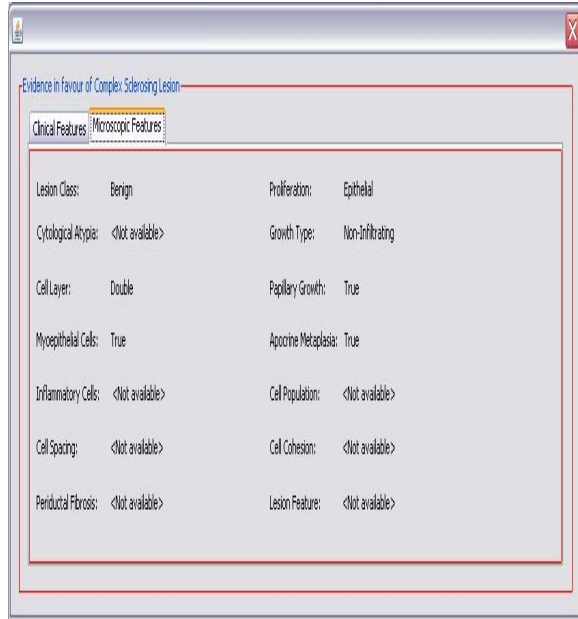


Figure 12: The Evidence Form

5. Result and discussion

The system was tested using 150 samples of breast disease in the form of slide preparations from retrospective cases, obtained from the Department of Morbid Anatomy, Obafemi Awolowo University Teaching Hospital, Ile-Ife, Nigeria. The evaluation was meant to determine if the system can assist a pathologist in making accurate, consistent and timely diagnoses. Using a Consultant Pathologist's interpretation as a "gold standard" (reference test), the system's sensitivity and specificity levels for detecting and localizing breast diseases were calculated.

The choice of the "gold standard" test was informed since it is the standard diagnostic test or benchmark that is regarded as definitive in medical sciences. For instance, with a given sample, a diagnostic system could lead to one of the four possible categories of results:

- (i) True positive (TP):
The diagnostic system yields positive test result for the sample and thus the sample actually has the disease;
- (ii) False positive (FP):
The diagnostic system yields positive test result for the sample but the sample does not actually have the disease;
- (iii) True negative (TN):
The diagnostic system yields negative test result for the sample and the sample does not actually have the disease; and

- (iv) False negative (FN):
The diagnostic system yields negative test result for the sample but the sample actually has the disease.

Figure 13 below shows the system classification of results by disease status. It is also a pictorial representation of the "Gold Standard (Reference) Test" results.

		Disease Present (+) Disease Absent (-)		
		TP	FP	
System Test	-	127	1	128
	+	3	19	22
		130	20	

Figure 13: Classification of the system results by disease status

The formulas for used for calculating sensitivity, specificity, PPV and NPV are:

$$Sensitivity = \frac{TP}{TP + FN} * (100\%) \dots\dots\dots (1)$$

$$Specificity = \frac{TN}{TN + FP} * (100\%) \dots\dots\dots (2)$$

$$PPV = \frac{TP}{TP + FP} * (100\%) \dots\dots\dots (3)$$

$$NPV = \frac{TN}{TN + FN} * (100\%) \dots\dots\dots (4)$$

Therefore, from table 2, using equations (i), (ii), (iii) and (iv), respectively, the Sensitivity, Specificity, Positive Prediction Value (PPV) and the Negative Prediction Value of the system are:

Sensitivity = 97.7%;
 Specificity = 95.0%;
 PPV = 99.2% and
 NPV = 86.3%.

6. Conclusion

In this paper we presented a DSS, which could be used by stakeholders for arriving at very vital decisions regarding the diagnosis of breast diseases (DBD) in women. The focus was on the development of a DSS that can assist Pathologist, especially those who may not be specialist in the area of breast cancer treatment. Thus, the system attempts to improve the effectiveness of diagnosis (in relation to accuracy, timeliness and quality) that is performed by a human pathologist, rather than improve their efficiency with respect to decision making. Therefore, the diagnoses made by the system are at least as good as those made by a human expert. Nevertheless, the knowledge for interpreting slides of diseased breast tissues could be very valuable. It allows an experienced Pathologist, which is specialized in breast pathology to follow a systematic, orderly and logical line of reasoning, and arrive at a proper DBD disease. As a result, the system presented in this paper was developed to simulate the above logical pattern of reasoning that Pathologist often used for DBD. The aim of developing the DSS was to find a way of assisting inexperienced pathologists in the histological interpretation of the slides of diseased breasts.

The value of the system presented in this paper could be quantified by the fact that Information and communication technology is ubiquitous. As such the need for the system is crucial as a response to the support of AIM. The system's diagnostic efficiency is based on the presence of the graphical user interface. This adequately assists the user through a cooperative style of interaction, so that an analytical, systematic reasoning can be followed by the physician without being bored. The UIFs are also presented in such a way that a good oversight is obtained without an increase in its complexity in terms of appearance. The system's knowledge base was evaluated by an experienced pathologist who is an expert in histopathological interpretation of breast diseases. It is however, recommended that the extension of the system would be to include larger portions of textual descriptions that could be viewed, under the microscope, by the user to clarify terms and assist in the recognition of features. The system can be further enhanced to become very sophisticated. This will include the provision of a database of examples of images which may be used for comparison. Also, the user-interface may be improved, so that the physician can answer questions in a more flexible way. That is, instead of answering a NO/YES question, the user could have a scaled range of possible answers between two extremes. Of course, this requires that the answers be combined with the knowledge base using fuzzy logic concepts.

With the introduction of fuzzy concepts, more accurate reasoning and results will be obtained. This would require, of course, certain changes in the user-interface to account for fuzzy-linguistic user answers. Finally, the knowledge

base at present contains 21 disease profiles. There are many more types of breast diseases that have not yet been included. This could also be done, so that the system is expanded to accommodate more breast diseases.

7. References

- [1] Hicks, D. G., Kulkarni, S., and Hammond, M. E. (2008). The role of the indispensable surgical pathologist in treatment planning for breast cancer. *Arch.Pathol.Lab Med.*, 132, 1226-1227
- [2] Rosai, J. (2007). Why microscopy will remain a cornerstone of surgical pathology. *Lab Invest.*, 87, 403-408.
- [3] Rosai, J. (2001). The continuing role of morphology in the molecular age. *Mod.Pathol.*, 14, 258-260
- [4] Druzdzel, M. J. and Flynn R. R. (2002). Decision Support Systems. Encyclopedia of Library and Information Science. Allen Kent (ed.), New York: Marcel Dekker, Inc., 2002. Retrieved from <http://www.pitt.edu/~druzdzel/abstracts/dss.html> on 24/01/2011 @ 7:50pm.
- [5] Alter, S. L. (1990). Decision support systems: current practice and continuing challenges. Reading, Mass., Addison-Wesley Publishers.
- [6] Finlay, P. N. (1994). Introducing decision support systems. Oxford, UK Cambridge, Mass., NCC Blackwell; Blackwell Publishers.
- [7] Turban, E. (1995). Decision support and expert systems: management support systems. Englewood Cliffs, N.J., Prentice Hall. ISBN 0-024-21702-6
- [8] Sprague, R. H. and E. D. Carlson (1982). Building effective decision support systems. Englewood Cliffs, N.J., Prentice-Hall. ISBN 0-130-86215-0
- [9] Marakas, G. M. (1999). Decision support systems in the twenty-first century. Upper Saddle River, N.J., Prentice Hall.
- [10] Choua, S. M., Leeb, T. S., Shaoc, Y. E., and Chenb, I. F. (2004). Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 27, 133-142.
- [11] Cook H. M. and Fox, M. D. (1987). Artificial Intelligence Applied to Mammographic Image Analysis. *Electronic Imaging*. Pp 1154-1158.
- [12] Wu Y, Giger, M.L., Doi, K., Vyborny, C. J., Schmidt, R. A., and Metz, C. E. (1993). Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* 1993; 187:81-87.
- [13] Floyd, C. E., Lo, J. Y., Yun, A. J., Sullivan, D. C. and Kornguth, P. J. (1994). Prediction of Breast Cancer Malignancy using an Artificial Neural Network. *Cancer*. 74: No. 11.

- [14] Baker J. A., Kornguth, P. J., Lo, J. Y., Williford, M. E. and Floyd, C. E. (1995). Breast Cancer Prediction with Artificial Neural Network Based on B-RADS Standardized Lexicon. *Radiology*. 196: 817-822.
- [15] Lo, J. Y., Baker, J. A., Kornguth, P. J., Iglehart, J. D., and Floyd, C. E (1999). Predicting Breast Cancer Invasion with Artificial Neural Networks On The Basis of Mammographic Features. *Radiology*. 203: 159-163.
- [16] Chen, D. R., Chang, R. F., and Huang, Y. L. (1999). Computer-aided Diagnosis Applied to US of Solid Breast Nodules by Using Neural Networks. *Radiology*. 213: 407-412
- [17] Floyd, C. E., Lo, J. Y. and Tourassi, G. D. (2000). Case- Based Reasoning Computer Algorithm that Uses Mammographic Findings for Breast Biopsy Decisions, *Am. J. of Roentgenol*. 175: 1347-1352.
- [18] Lei Z. and Andrew K. C. (2001), "An artificial intelligent algorithm for tumor detection in screening mammogram," *IEEE transactions on medical imaging*, vol. 20, no. 7, July, 2001.
- [19] Huai Li, Shih-Chung B. L. Yue W., Lisa K. and Matthew T. F (2002), "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE transactions on medical imaging*, vol. 21, no, 2, February, 2002.
- [20] Bovis, K. and Singh, S. (2002). Classification of mammographic breast density using a combined classifier paradigm. In *Med. Image Underst. Anal.*, pp. 177–180.
- [21] Petroudi, S., Kadir, T., and Brady, M. (2003) "Automatic classification of mammography parenchyma patterns: A statistical approach." In *IEEE Conf. Eng. Med. Biol. Soc.*, volume 1, pp. 798–801. 2003.
- [22] Karabatak, M. and Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications* 36 (2009) 3465–3469. Available at <http://www.elsevier.com/locate/eswa>
- [23] Varsha, H. P., Dattatraya, S. B. and Vaishali S. P. (2006). An Automated Computer Aided Breast Cancer Detection System. *GVIP Journal*, Volume 6, Issue 1, July 2006
- [24] Sprague, R. H. and. Carlson, E. D (1982). *Building effective decision support systems*. Englewood Cliffs, N.J., Prentice-Hall. ISBN 0-130-86215-0
- [25] Haag, S., Cummings, M., McCubbrey D. J., Pinsonneault, A. and Donovan, R. (2000). *Management information systems for the information age*. McGraw-Hill Ryerson Limited: 136-140.

- [26] Sage, A. P. (1991). *Decision Support Systems Engineering*. John Wiley & Sons, Inc., New York.
- [27] Hayes- Roth, F., Waterman, D. A. and Lenat, D. B. (1983). *Building Expert Systems* (Reading, MA: Addison – Wesley)

8.0 Bibliography of Authors

Aderonke A. Kayode

She holds a B.Sc degree in Computer Science and is currently on her M.Sc Programme. She is a member of Information Storage and Retrieval Group in the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.

Afolabi, B.S. (Ph.D)

He is a Senior Lecturer in the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria and head of the Information Storage and Retrieval Group research team.

Akhigbe, B.I

He is a member of Information Storage and Retrieval Group in the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria. He has both B.Sc and M.Sc in Computer Science.

UDO Ifiok James

He is a member of Information Storage and Retrieval Group (ISRG), Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.

Ominiyi Adedayo

He is a member of Information Storage and Retrieval Group (ISRG), Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.