# An Improved Fuzzy-based Model for Diagnosing Confusable Diseases in Nigeria

Friday E. Onuodu[1], Ephraim N. Eyina[2]
[1]Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria
[2]Department of Computer Science, Ignatius Ajuru University of Education, Rivers State, Nigeria

***Abstract*—Decision-making in the medical field remains a difficult task for medical doctors as a result of confusable diseases. Indeed there is uncertainty when a medical doctor encounters confusability of symptomatic presentation of diseases due to the fact that they share common symptoms and as such becomes difficult for the medical doctor to correctly diagnose them. A major landmark of medical practitioners is the ability to carry out successful diagnoses and prevent diseases. Hence, it is very essential to implement a system that reduces cases of mis-diagnosis which could arise from confusable diseases. In this study, we developed an Improved Fuzzy-based Model for diagnosing Confusable Diseases in Nigeria. Software Development Lifecycle Methodology (SDLC) was adopted in this approach. Furthermore, we implemented with Hypertext Pre-processor and MySQL database as backend. The obtained results of the study showed that the proposed system outperforms the existing system in diagnostic speed and symptom processing. The evaluated parameters of the proposed system obtained values of 54 seconds and 45 seconds respectively for diagnostic speed and symptom processing when compared to the existing system which had values of 67 seconds and 84 seconds respectively. In addition, this study could be beneficial to medical doctors and their respective patients in the Nigerian Health Sector.**

***Keywords*—Confusable Diseases, Diagnosis, Fuzzy-based Model, Improved, Symptoms**

## I. INTRODUCTION

In medical science, diagnosis of a disease is very complicated, and many tests must be done on patients to obtain a near accurate diagnosis. This has given rise to computerized diagnostic tools, intended to aid the physician in making primary medical decisions and hence an early diagnosis. A major area for such computerized tools is in the domain of confusable diseases. In this study, we proposed an improved fuzzy-based diagnosis system for confusable diseases in health sector. Diagnosis of diseases must be done with care since it is the first stage of therapeutic actions towards eventual management of the disease; a mistake at this stage is disastrous and such adequate care must be ensured. Medical Diagnosis can be defined as the process of finding out which disease or condition explains a person's symptoms and signs. Confusable diseases are diseases that share common symptoms and as such become difficult for the medical doctor to correctly diagnose them. The process of diagnosing confusable diseases is usually complex especially when there is no automated system to back it up. Medical Diagnostic system is very crucial in generating accurate and rapid diagnosis of disease. Particularly, in today's era, there are several diseases whose symptoms are quite similar in initial stage, but at the same time, initial level diagnosis is also required to be accurate. In other words, there is need for automating diagnosis systems in order to have accurate diagnosis results [1]

In rural settings of Africa, clinical diagnosis (based on symptoms) remains the only option for diseases diagnosis and this makes accurate diagnosis unlikely. Several factors such as lack of good water supply, high prevalence of asymptomatic infections in rural communities, improper waste management system, inadequate healthcare facilities, and widespread practice of self-treatment for clinical diseases all contribute to the issues faced in rural settings. Asogbon et al [2] stated that "the administration of efficient healthcare services has been a major challenge in developing countries due to inadequate healthcare delivery personnel and the inappropriate diagnostic techniques often adopted". Medical Diagnosis can be defined as the process of finding out which disease or condition explains a person's symptoms and signs. Furthermore, the proposed algorithm to be adopted for the study is a Fuzzy-based Model.

Diagnosis becomes difficult in medical domain due to influence of medical uncertainties that arises from confusability in disease symptomatic presentation between two diseases. This confusability of these diseases stems from the overlaps in the disease symptomatic presentation and has led to misdiagnosis with various degrees of associated costs and in worst cases led to death. The ability of the physician(s) to thoroughly scan through the series of laboratory tests and symptoms of a patient which are time varying as the case may be and pick out meaningful and useful information that 'stand-out', for proper identification of a disease (amongst several diseases which would sometimes share common symptom ) makes a good physician. It is not overly out of place to say that perception plays a central role amidst skills and experiences garnered by an expert physician during his or her education pursuit, in order to perform a near accurate or accurate diagnosis of a disease. Most researchers opined that medical diagnosis is both science and arts; where the art is what separate between two well-trained medical personnel, thus is very necessary to talk of it if we are aiming at developing an application that would sieve through data and provide semantically relevant information amidst the wide range of uncertainties in a manner that simulate a human expert physician [3]

*A. Statement of the Problem*

This study addresses the problem of inability to diagnose and interpret confusable diseases in Nigerian Health Sector. Confusable diseases are diseases with incomprehensive symptoms. From an in-depth feasibility study of related issues carried out, the mentioned problem is a system-based problem that requires improvement in order to accept different formats of datasets on confusable diseases and further diagnose the diseases in order to obtain the required treatment action. Furthermore, the study will address the mentioned problem by providing an improved fuzzy-based system for diagnosis of confusable diseases using fuzzy cognitive engines. In addition, the study will be divided into two levels: the first level will recognize precise confusable diseases and provide fuzzy symptoms; while the second level will take the fuzzy symptoms and finally obtain fuzzy treatment actions.

*B. Aim and Objectives*

The aim of this study is to develop an improved fuzzy-based model for diagnosing confusable diseases in Nigeria. The specific objectives of the study include to:

i) design a Fuzzy-based Information System for Confusable Diseases.
ii) implement the system with Hypertext Pre-processor and MySQL database as backend.
iii) compare our results with the existing system for diagnosing confusable diseases.

*C. Overview of Confusable Diseases*

When two or more diseases have some overlapping symptoms which make it naturally difficult for a physician to establish the right diagnosis, it is referred to as confusable diseases in medical parlance. In order to diagnose confusable diseases properly, a diagnostic criterion for a particular disease is needed so as not to confuse it with other diseases because of shared symptoms. For a diagnosis to be effective in this regard, the target disease has to be recognized in a pool of confusable diseases and suggested two ways to handle this: by recognition of the combination of symptoms of the target disease or by exclusion of confusable disease as the cause of the symptoms.

Confusable disease is poised with the following problems outline herewith.

i) Confusable disease manifests the same symptoms thereby leading to imprecise or incomplete diagnosis by the physician.
ii) A disease at one stage can manifest similar symptoms with a different disease at another stage.
iii) Failure to correctly diagnose a confusable disease would lead to a physician giving the wrong treatment to the patient.
iv) Patients may be suffering from more than one confusable disease.

*D. Fuzzy Logic*

Fuzzy logic is a branch of science that is extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Fuzzy logic may be applied to many fields, including control systems, neural networks and artificial intelligence (AI). Fuzzy logic can be used to describe how information is processed inside human brains. For example, it can be argued that humans do not know the difference between fat and thin. Five people may be fat and not have the same severity of fatness. Or, one person may appear thin, compared to another, while both are actually fat. Using fuzzy logic, you can assign different logic values for fatness, ranging from 0 to 1, according to severity of fatness. Variables between the extremes of zero and one are closer to the concept of probability, which means there is a major correlation between the science of probability and fuzzy logic. However, fuzzy logic refers to intensity of truth, while probability refers to likelihood.

Fuzzy logic is an extension of Boolean logic by Lotfi Zadeh in 1965 based on the mathematical theory of fuzzy sets, which is a generalization of the classical set theory. By introducing the notion of degree in the verification of a condition, thus enabling a condition to be in a state other than true or false, fuzzy logic provides a very

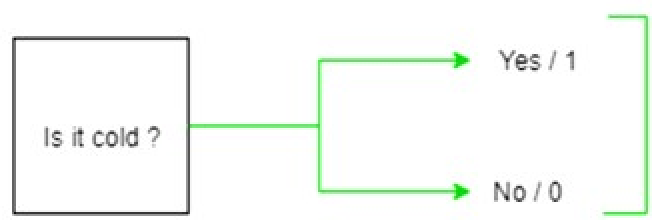


Fig. 1: Simple Illustration of Fuzzy Logic (Source: [2])

valuable flexibility for reasoning, which makes it possible to take into account inaccuracies and uncertainties. One advantage of fuzzy logic in order to formalize human reasoning is that the rules are set in natural language. Intuitively, it thus seems that the input variables like in this example are approximately appreciated by the brain, such as the degree of verification of a condition in fuzzy logic. Fuzzy logic can be conceptualized as a generalization of classical logic. Fuzzy logic can be used as an interpretation model for the properties of neural networks, as well as for giving a more precise description of their performance. We will show that fuzzy operators can be conceived as generalized output functions of computing units. Fuzzy logic can also be used to specify networks directly without having to apply a learning algorithm. An expert in a certain field can sometimes produce a simple set of control rules for a dynamical system with less effort than the work involved in training a neural network.

Fuzzy logic is now being used in many products of industrial and consumer electronics for which a good control system is sufficient and where the question of optimal control does not necessarily arises. The difference between crisp (i.e., classical) and fuzzy sets is established by introducing a membership function. Consider a finite set

$$X = \{x1, x2, ..., xn\} \quad (1.1)$$

which will be considered the universal set in what follows. The subset A of X consisting of the single element x1 can be described by then-dimensional membership vector

$$Z(A) = (1,0,0,...,0), \quad (1.2)$$

where the convention has been adopted that a 1 at the ith position indicates that xi belongs to A. The set B composed of the elements x1 and xn is described by the vector

$$Z(B) = (1,0,0, ...,1). \quad (1.3)$$

Any other crisp subset of X can be represented in the same way by an n-dimensional binary vector. But what happens if we lift the restriction to binary vectors? In that case we can define the fuzzy set C with the following vector description:

$$Z(C) = (0.5,0,0, ...,0) \quad (1.4)$$

In classical set theory such a set cannot be defined. An element belongs to a subset or it does not. In the theory of fuzzy sets we make a generalization and allow descriptions of this type. In our example the element x1 belongs to the set C only to some extent. The degree of membership is expressed by areal number in the interval [0, 1], in this case 0.5. This interpretation of the degree of membership is similar to the meaning we assign to statements such as “person x1 is an adult”.

## II. Related works

Alom et al [4] researched on A State-of-the-Art Survey on deep learning theory and architectures. The work presented a brief survey on the advances that have occurred in the area of Deep Learning (DL) which includes: Convolution Neural Network and Recurrent Neural Network. The result of the work showed that a statistical approach to Deep Learning theories is an investment for the performance of medical diagnosis. However, the work could not further discuss individual advanced techniques for training large-scale deep learning models and recently developed method of generative models.

Amir [5] looked at State-of-the-Art of Machine Learning (ML) Models in Energy Systems, a Systematic Review. The work presented the state of the art of Machine Learning (ML) models used in energy systems along with a novel taxonomy of models and applications. The work further concluded that there is an outstanding rise in the accuracy, robustness, and precision and generalization ability of the ML models in energy systems using hybrid ML models.

Feng [6] researched on Deep Learning Models for Bankruptcy prediction using Textual Disclosures. The work introduced deep learning models for corporate bankruptcy forecasting using textual disclosures. The result of the work was the construction of a comprehensive bankruptcy database of 11, 827 U.S. public companies and showed that deep learning models yield superior prediction performance in forecasting bankruptcy using textual disclosures. However, there was no adequate comparative analysis between the textual disclosure model and other existing deep learning models.

Tomasz [7] looked at An Evaluation of Effectiveness of Fuzzy Logic Model in predicting the business bankruptcy. The work presented a fuzzy-based system that predicts bankruptcy for one, two and three years before the possible failure of companies. The result of the work showed that the developed fuzzy model uses financial ratios that are dynamic. However, there was no adequate comparative analysis of the system with other existing ones.

Vaidehi [8] researched on a prediction system based on fuzzy logic. The work centered on the application of fuzzy logic in building a prediction system to predict the future occurrence of an event. The developed model consists of IF-THEN rules which formed vague predicates on their antecedent part while the consequent part is a linear or quadratic combination of the antecedent variables. However, there was no adequate comparative analysis of the system with other existing ones.

Sedat [9] researched on Fuzzy Logic Model for the prediction of cement compressive strength. The work developed a fuzzy logic prediction model for the 28-day compressive strength of cement mortar under standard curing conditions was created. The results of the work indicated that through the application of fuzzy logic algorithm, a more user-friendly and more explicit model than the ANNs could be produced within successfully low error margins. There were quite some percentage error levels in the developed fuzzy level.

Berkin [10] looked at Machine Learning Models to Predict Performance of Computer System Design Alternatives. The work concentrated on both the system design and the architectural design processes for parallel computers and developed methods to expedite them. The result of the work showed that by using 1% of the design space (i.e. cycle-accurate simulations), we can predict the performance of the whole design space within 3.4% error rate.

## III. Materials and methods

### A. Methodology.

The Methodology adopted for the Improved Fuzzy-based Model for diagnosing confusable diseases is Software Development Lifecycle Methodology (SDLC). The Software Development Lifecycle Methodology (SDLC) involves a standardized set of task carried out in order to improve an Existing Project. It is also likened to the framework that is required to structure, plan and control the process of developing a new project. Furthermore, the Software Development Lifecycle Methodology (SDLC) comprises of models such as the waterfall model, the prototyping model, the incremental model and the spiral model.

### B. Analysis of the Existing System

The Present Diagnostics System applied by most Health Institutions is Collaborative Filtering Diagnosis System (CF) as developed by [4]. Many existing Diagnosis systems rely on the Collaborative Filtering (CF) and have been extensively used in Health. They have proven to be very effective with powerful techniques in many famous Health companies (see figure 2). The Existing System presents an overview of the field of Diagnosis systems with current generation of recommendation methods and examines comprehensively CF systems with its algorithms. Collaborative Filter Diagnosis systems try to suggest services that are similar to the ones that the user liked in the past. The likeness of items is determined depending on the traits associated with the compared items. For example, if an individual user has favorably inquired from a Diagnosis system about certain symptoms, he or she should likely obtain swift response from the system.

Furthermore, Collaborative Filter Diagnosis treat suggestions as a user-specific category problem and learn a classifier for the patient's preferences depending on services traits. Techniques applying a Collaborative Filter-based recommendation strategy evaluate a set of documents and/or details of services previously ranked by a user and develop a model or user profile of user passions depending on the features of the things rated by that user. Based on the genuine and ordinary of this strategy the items that other users with similar tastes liked in the past are recommended to the target user. The likeness in taste of two users is computed with regards to the likeness in the past ratings of the users.

### C. Explanation of the Existing System Components

The following components of the Existing System are:

i) System User:

The system user is the individual who launches the system in search of recommended data.

ii) Server Request:

This is the web platform that is powered by a web server and also enables the user to access the backend of system.

iii) User Registration, Validation and Login:

This component enables the user to register with important details in order to be allocated a unique username and password for login.

iv) Medical Diagnosis Platform:

This is a Software tool and technique that provide suggestions for items to be of use to a user.

v) Collaborative Filter Technique:

This component interfaces with the database by trying to suggest products that are similar to the ones the user liked in the past.

v) Database:

This is an organized collection of related files that are usually in digital form.

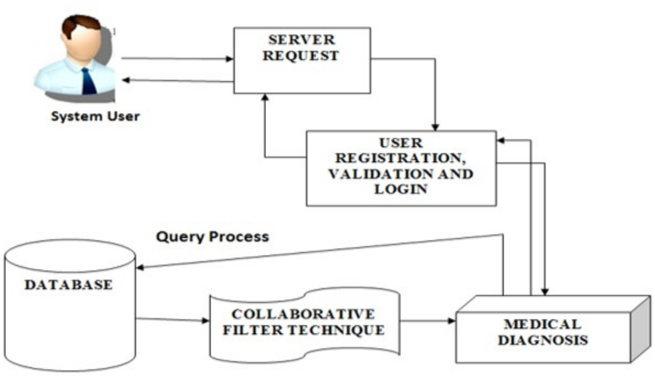


Fig. 2: Existing System Structure of a Collaborative Filter Medical Diagnosis System

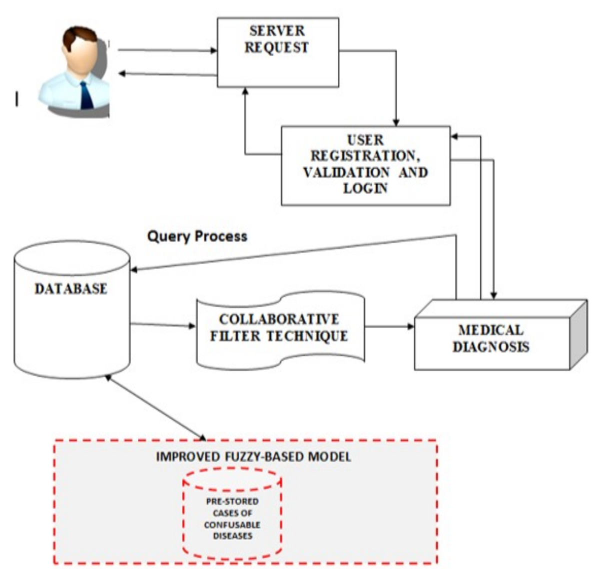


Fig. 3: Proposed System Structure of an Improved Fuzzy-based System for Confusable Diseases

*D. Disadvantages of the Existing System*

The following disadvantages of the Existing System are:

i) it does not implement a Fuzzy-based Concept for diagnosing any diseases to potential patients
ii) The Existing Data also needs a lot of Big Data in order to make efficient and reliable recommendations to its users.

*E. Explanation of the Proposed System Components*

The following component of the Proposed System is:

i) Improved Fuzzy-based Model:

This component enables the system to recommend vital information to the user based on the referencing of similar cases that relates to the user's request.

*F. Advantages of the Proposed System*

The following advantages of the Proposed System are:

i) Efficient User Graphical Interface for Patient-Physician Communication
ii) Speed in acceptance and processing of diagnostic information from potential patients.
iii) The ability of the Improved Fuzzy-based System to update itself in the long-run

*G. Existing System Algorithm*

Step 1:

START

Step 2:

DECLARE ALL VARIABLES

Step 3:

SR, UV, U, P, L, UR, QP, DSO. WHERE SR IS SERVER REQUEST, UV IS

USER VALIDATION, U IS USERNAME, P IS PASSWORD, L IS LOGIN, UR

USER REGISTRATION, QP IS QUERY PROCESS, DSO IS DIAGNOSIS SYSTEM OUTPUT

Step 4:

INITIATE SR

Step 5:

INITIATE UR

Step 6:

PROCESS UR

Step 7:

INITIATE UV

Step 8:

UV = L * (U + P)

Step 9:

INITIATE QP

Step 10:

OBTAIN DSO

Step 11:

STOP

*H. Proposed System Algorithm*

Step 1:

START

Step 2:

DECLARE ALL VARIABLES

Step 3:

SR, UV, U, P, L, UR, QP, DSO, IFBM. WHERE SR IS SERVER REQUEST, UV IS

USER VALIDATION, U IS USERNAME, P IS PASSWORD, L IS LOGIN, UR

USER REGISTRATION, QP IS QUERY PROCESS, DSO IS DIAGNOSIS

SYSTEM OUTPUT AND IFBM IS IMPROVED FUZZY-BASED MODEL

Step 4:

INITIATE SR

Step 5:

INITIATE UR

Step 6:

PROCESS UR

Step 7:

INITIATE UV

Step 8:

UV = L * (U + P)

Step 9:

INITIATE QP

Step 10:

ACTIVATE IFBM

Step 11:

INTERFACE IFBM-DATABASE

Step 12

SEARCH FOR SIMILAR CASES RELATED TO THE USER'S REQUEST

Step 13

ALERT MATCH FOUND

Step 14

UPDATE SYSTEM DATABASE

Step 15

OBTAIN DSO

Step 16

STOP

## IV. Result and Discussions

### *A. Choice and Justification of Programming Language used*

We implemented the Proposed System design with PHP, JavaScript Programming Language, Hypertext Markup Language, Cascading Style Sheet and MySQL Relational Database Management System. JavaScript is a server-side scripting language that is used for making web pages interactive. It is supported by all major web browsers. This is a programming language that is used by web developers for the creation of contents that communicate with databases. Secondly, PHP can be used for the development of web-based applications, system function performance; HTML is an acronym for Hypertext Markup Language and is used for structuring web pages. It consists of tags and is also supported by all major web browsers. Cascading Style Sheet (CSS) is a web development content that is used for styling and beautifying web pages. MySQL is the world's most popular open source database. With its proven performance, reliability and ease-of-use, MySQL has become the leading database choice for web-based applications, used by high profile web properties including Facebook, Twitter, YouTube, Yahoo and many more. Oracle drives MySQL innovation, delivering new capabilities to power next generation web, cloud, mobile and embedded applications.

### *B. Discussion of Results*

Figure 4, shows the welcome page, figure 5 is the registration page of the proposed system, figure 6 shows the database confirmation of the registered user, figure 7 is the user login page, figure 8 is the disease diagnosis platform and figure 9 is the confusable disease symptom diagnosis. Figure 10 is the system diagnosis output. Furthermore, the Proposed System also supports offline evaluation for its users. When evaluating diagnosis systems with offline experiments, there are no actual users. Instead, a collection of data based on earlier user choices is used to simulate real user queries. The results of the queries can be used to evaluate the prediction power of the diagnosis. The queries should, to produce proper results, be similar to the queries that users will make in a production version of the system. Since an offline experiment requires no user interaction, it is a cheap and efficient approach of evaluating a recommendation system. It does however only evaluate a small set of possible user queries, and does not provide any information on user behavior of the system. An offline experiment can be used on different recommendation algorithms to evaluate them against each other. The data sets used to simulate user queries can either be designed by hand or by using a random selector on stored queries. Different types of biases can be introduced when generating the user queries, but methods such as reweighting or re-sampling can lower the influence of these biases. In a user study the diagnosis system is tested on real users who are given tasks to complete. Their behavior and their satisfaction with the recommendations are then measured, and the feedback used to improve the system. User studies can use questionnaires to get feedback from users on their experience. The main advantages of a user study are that a broad range of questions can be answered and that a large amount of data, both qualitative and quantitative can be collected. The test users should represent the actual population of desired users, or else the results might be more vulnerable to biases such as non-response bias. By evaluating a diagnosis system with the online evaluation method, data on real behavior and user interaction can be collected from a live system.

Since an offline experiment requires no user interaction, it is a cheap and efficient approach of evaluating a recommendation system. It does however only evaluate a small set of possible user queries, and does not provide any information on user behaviour of the system. An offline experiment can be used on different recommendation algorithms to evaluate them against each other. The data sets used to simulate user queries can either be designed by hand or by using a random selector on stored queries. Different types of biases can be introduced when generating the user queries, but methods such as reweighting or re-sampling can lower the influence of these biases. In a user study the diagnosis system is tested on real users who are given tasks to complete.

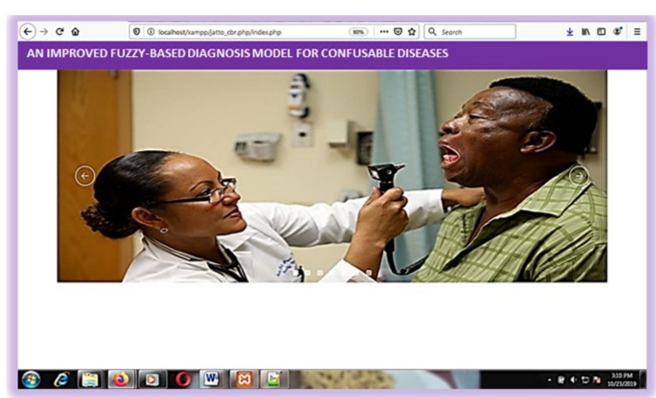


Fig. 4: Welcome Page

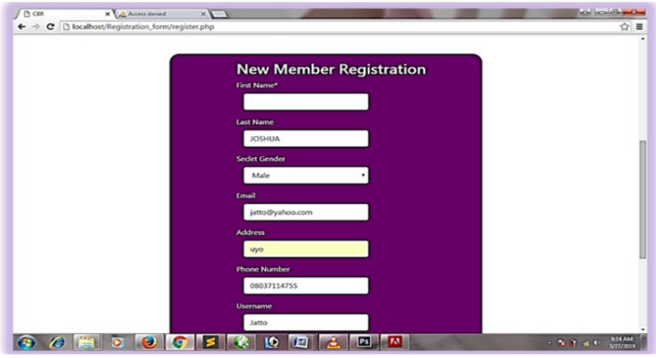


Fig. 5: Registration Page

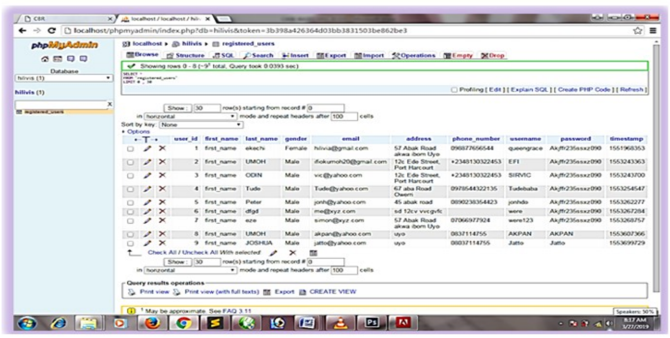

Fig. 6: Database Confirmation of Registration

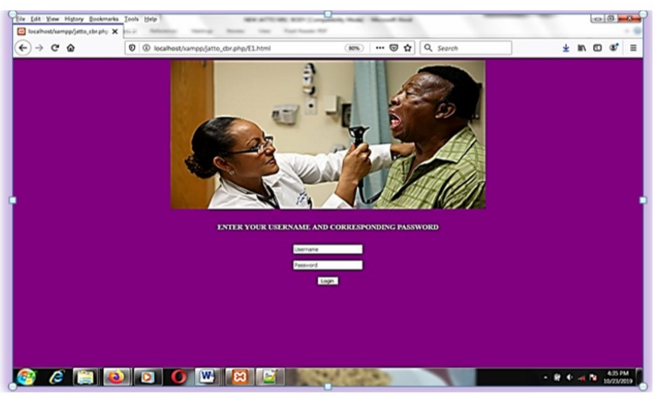


Fig. 7: User Login Page

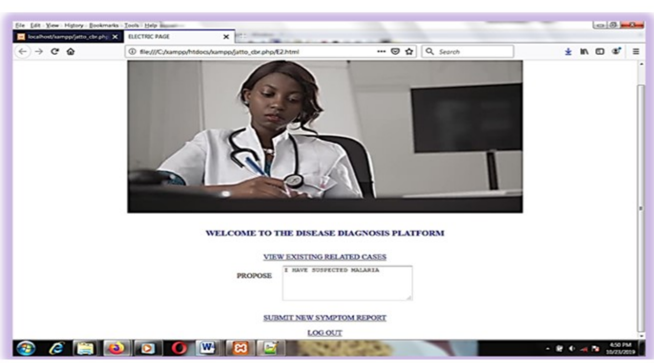


Fig. 8: Disease Diagnosis Platform

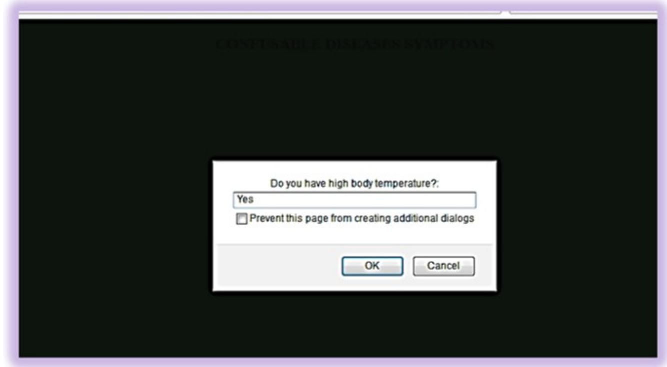


Fig. 9: Confusable Diseases Symptoms

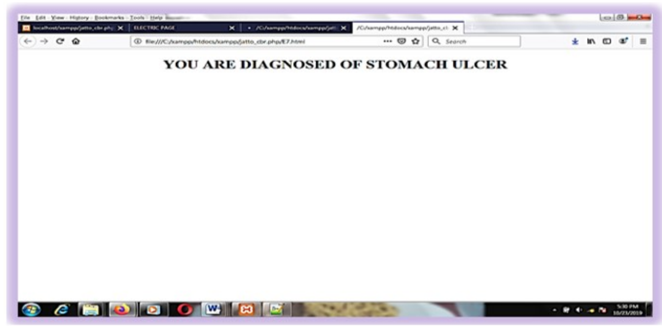


Fig. 10: System Diagnosis Output

TABLE I

COMPARATIVE ANALYSIS

| SN | EXISTING SYSTEM | Time in seconds | Time in Seconds | PROPOSED SYSTEM |
|---|---|---|---|---|
| 1. | Diagnostic Speed | 54 | 67 | Diagnostic Speed |
| 2. | Symptoms Processing Speed | 45 | 84 | Symptoms Processing Speed |

(Source of data: Program implementation and Research Findings)

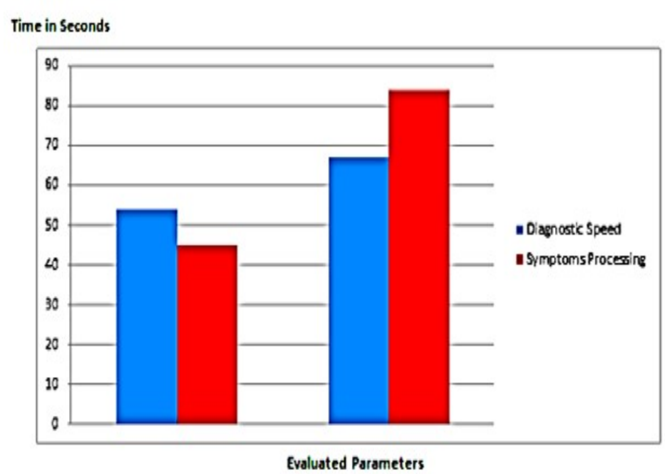


Fig 11: Performance Evaluation Chart

## V. Conclusions

In conclusion, the proposed system is a problem-solving methodology that tries to solve new problems by re-using specific past experiences stored in example cases. A case models a past experience, storing both the problem description and the solution applied in that context. All the cases are stored in the case base. When the system is presented with a new problem to solve, it searches for the most similar case(s) in the case base and reuses an adapted version of the retrieved solution to solve the new problem. In the revise step the system adapts the solution to fit the specific constraint of the new problem. Whereas in the review step the constructed solution is evaluated by applying it to the new problem, understanding where it fails and making the necessary corrections; in a diagnosis task, for instance, the system acquires the patient symptoms (new problem) and tries to give the final diagnosis based on past patient.

## REFERENCES

[1] Andrienn, K. (2018), A fuzzy-based system for decision support, International Journal of Computer Applications (IJCA), 2(7), 42 – 47

[2] Asogbon P., Aristo G., and Asharfa, F. (2017), Fuzzy Logic Fundamentals, International Journal of Engineering Technology (IJET), 4(8), 61 – 64

[3] Felix N. (2014), DeepQB: Deep Learning with Player Tracking to Quantify Quarter-back Decision Making and Performance, MIT SLOAN; Sports, Analytic Conference, March 1 – 2

[4] Alom O., Ipswitch L., and Mainz M. (2019), Collaborative Filtering Techniques for Medical Diagnosis, International Journal of Computer Applications (IJCA), 9(4), 22 – 29

[5] Amir V. (2019), State-of-the-Art of Machine Learning (ML) Models in Energy Systems, a Systematic Review, Energies Journal, MDPI, received: 12th February 2019; Accepted; 28th March 2019; Published: 4th April 2019

[6] Feng G. (2019), Deep Learning Models for Bankruptcy prediction using Textual Disclosures, European Journal of Operational Research 274(2019), 743 – 758

[7] Tomasz A. (2011), Can you trust this prediction? Auditing Point-wise Reliability after learning, Proceedings of the 22nd International Conference on AI and Statistics (AISTATS), Naha, Japan, Volume 89

[8] Vaidehi C. (2011), An Evaluation of Effectiveness of Fuzzy Logic Model in predicting the business bankruptcy, Institute for Economic Forecasting, Romanian Journal of Economic Forecasting.

[9] Sedat N. (2014), Fuzzy Logic Model for the prediction of cement compressive strength, Cement and Concrete Research, 34(2014), 1429 – 1433

[10] Berkin H. (2010), Machine Learning Models to Predict Performance of Computer System Design Alternatives, 37th International Conference on Parallel Processing, DOI 10.1109/ICPP.2008.36