TELKOMNIKA, Vol.17, No.6, December 2019, pp.3145~3154
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v17i6.12715

A web/mobile decision support system to improve medical diagnosis using a combination of K-Mean and fuzzy logic

Zainab T. Al-Ars*[1], Abbas Al-Bakry[2]
[1]University of Information Technology and Communication,
Department of Remote Sensing and GIS, College of science, University of Baghdad, Iraq
[2]University of Information Technology and Communications Baghdad, Iraq
*Corresponding author, e-mail: zainabdrweesh@scbaghdad.edu.iq

Abstract

This research provides a system that integrates the work of data mining and expert system for different tasks in the process of medical diagnosis, and provides detailed steps to the process of reaching a diagnosis based on the described symptoms and mapping them with existing diagnosis available on the web or on a cloud of medical knowledge based, aggregate these data in a fuzzy manner and produce a satisfactory diagnosis of the persisting problem. The mobile phone interface would make the system user-friendly and provides mobility and accessibility to the user, while posting updates and reading in details the steps that led to the decision or diagnosis that is reached by the K-mean and the fuzzy logic inference engine. The achieved results indicate a promising diagnosis performance of the system as it achieved 90% accuracy and 92.9% F-Score.

Keywords*: AI, communication, expert system, fuzzy logic, K-Mean clustering*

1. Introduction

Health care fields have found their way among those fields to make use of computer sciences. The intersection between computer sciences and health care has been formulated into a new filed of studies called “medical informatics”, where medical data are collected, stored, processed, analyzed, retrieved, and used in various medical related operations. Different disciplines have emerged to cover the wide range of specialties required by the medical informatics field. These include [1]:

- Bioinformatics: workers in this field are concerned with collecting, storing, retrieving and analyzing medical data either for research purposes or to provide better patients’ care.
- Public health informatics: workers of this area define the way public can make use of medical records and what records are available to researchers and medical practitioners.
- Electronic Health Records: these records are exchanged between different healthcare providers to provide better care for patients, and workers of this field are responsible for securing this exchange, and allowing only authorized personnel have access.
- Health Data Analyst: this type of analysts uses medical data to define trends and relations between different health records to come up with predictions and recommendations to some medical situations.

The medical domain is one of the most important and critical domains that has direct interaction with the humans’ wellbeing. Some wrongly diagnosed illness or badly made decisions can have a serious effect on someone’s health or life. Errors in diagnosis can be related to different reasons, like the lack of experience in the healthcare provider, inaccurate description of the symptoms, or the lack of information available for the medical staff. Wrong decisions could originate from lack of cooperation between different healthcare providing department or hesitance in making the decision. Medical institutes are in constant need for computerized help to provide better diagnosis to certain illnesses, whether for collecting large amounts of data, analyzing complex input, organizing and classifying data, finding relations

Received March 21, 2019; Revised July 2, 2019; Accepted July 18, 2019

between these data and many other operations that when computerized would provide great help in the diagnosis process.

One of the areas where medical informatics show its great influence on health care is the help in providing diagnosis of diseases according to some rules that are put to work on huge amounts of medical records of patients and medical sciences. Sometimes, these diagnoses can help in accelerating providing required medical assistant faster, which would help improve the quality of life for human beings. Medical decision making process involves different actions to be taken before reaching a satisfactory decision that can help in improving patients' care outcomes, like diagnosis, prognosis, treatment, and therapeutic monitoring [2].

Fuzzy logic is an approach to computing and problem solving that provides "degrees of true" not the common binary true or false solutions. Decision making systems that are based on fuzzy logic resemble the way a human being makes decisions, by having levels of truthiness, according to [3]. Medical diagnosis system has a series of fuzzy data as an input to the diagnosis process, where a fuzzy logic based decision making system need to be adopted to deal with these "Shades" of correct and wrong decisions [4]. Classifying these data in fuzzy sets is necessary in the process of diagnosis to deal with different levels of uncertainty in the production of the final diagnosis (decision).

The widespread of mobile and wearable devices provides an opportunity and also concern in the process of collecting, analyzing, and transmission of medical data. Opportunities shown in different aspects, most important ones show in the time efficiency of data collection, the flexibility of patient's movement and personal time management, in addition to being easily adaptable to existing mobile application installed on smart devices (like smart mobile phones).

2. Related Work

Diagnosis of a disease or a medical care is looked at as a decision making process that is based on the collected data and the relations that could be conducted from the integration of these data. Since these data are mostly in natural language and can be explained with some degree of uncertainty, which affects the classification on which the rules used to make a decision depends, most researchers in Decision Support Systems (DSS) generally prefer the use of fuzzy logic techniques.

Fuzzy logic classification defines sets of data based on a threshold of provided values. If the developed defines this threshold ahead of running the fuzzy logic mechanism, then the sets are considered to be of type-1 fuzzy sets, where the intervention of the developer is needed. While in other systems the threshold is adapted from the fuzzificaiton process and can be dynamically changes throughout the course of the fuzzification process, so the resulting sets are considered to be of type-2 fuzzy sets [3]. Due to the huge amounts of data and the wide variations of their values and sources, recent researchers focus on adopting type-2 sets in the fuzzy logic process, as in the work in [5], where type-2 fuzzy sets were used in an automated decision making system for home health care of diabetes management of home treated patients.

Another research adopted fuzzy decision making technique in the selection of vaccination application of a healthcare system [6]. Choosing the suitable vaccine to admit to patients is challenging and the decision depends on linguistic variables provided by physicians to rate the alternatives. The inference engine was provided with Fuzzy distance measure for all alternatives of the training sets (varied between people, spatial and temporal data) along with crisp values of fuzzy weights. Testing the system showed that People and temporal data provided to the DSS were the most suitable vaccination method for protecting people from H1N1 influenza epidemic (the test case).

In home treated patients, the decision making efficiency and speed could be the divider between good life and serious medical issues. The authors of [7] adopted a fuzzy logic DSS in Personal Home Healthcare System for Cardiac Patients. Data are collected from the patients through sensors (could be fuzzy or crisp data), where are used as input to the fuzzifier and whereby the membership functions defined these data are applied to their actual values (membership function) to help set a degree of correctness in the rules base. With the wide spread of fuzzy data, other types of fuzzy sets were developed like intuitionistic fuzzy set, type-n fuzzy set, fuzzy multisets, and hesitant fuzzy set [4]. Researchers had attempts towards benefiting the integration between these types of sets (in addition to the mostly used type-2 sets) as in the work in [4] where hesitant decisions were supported through the integration of type-2 fuzzy set, hesitant sets and intuitionistic fuzzy when having doubts between different values to be classified.

Samuel et al introduced a Web-Based Decision Support Sys-tem (WBDSS) coupled with fuzzy logic (FL) for typhoid fe-ver (TF) diagnosis. The KB of the system contained a fuzzy inference system (FIS). The system was developed to aid in the provision of accurate and timely TF diagnosis. Studies on the proposed system were performed based on the medical records of TF patients. The efficiency of the proposed system was based on the standard statistical metrics, while the achieved results showed a 94% efficiency of the system in providing an accurate diagnosis. The authors suggested the integration of ANN into the FL-based medical diagnosis sys-tems for better performances [8].

Fatumo et al designed a diagnostic ES called XpertMalTyph for the diagnosis of different types of typhoid and malaria complications. The ES simulates the skills of the medical expert in disease diagnosis using computers. Hence, the ES can provide similar services in the absence of an expert, making it possible to treat patients even from their homes. They suggest the implementation of the ES with artificial neural networks (ANN). The XpertMalTyph was executed in Java Expert System Shell [9].

3. System Architecture and Work Flow

The system allows the user to register information and the system will diagnose the patient's condition and find out biological problem and its cause depending on a database that is stored online. System makes judges depending on AI program built on fuzzy logic. The database information is captured depending on data acquisition unit which gets new information only from users who input the diagnostics to database after visiting physician.

The novelty in this system will be in providing user with full description on how did it reach to the result, like writing down in English what steps it took to find out what's wrong. This is the major data acquisition unit job which includes filtering information and causes and providing description for each. Such an intelligent system and with enough database will be able to reply with meaningful description with humans normally via text.

System will include an android application that will be able to listen to humans describing problems, convert speech to text then skimming text for useful information that will be used to build diagnostics, then convert the reply from text to speech and say it to patient. Such an interactive system will be an amazing result of applying AI in useful biomedical apps. The hybrid system is implemented over 2 phases: phase 1 involved clustering of data collected through data acquisition unit, and phase 2 is the expert system that holds the AI module (fuzzy logic).

3.1. Data Clustering

An intelligent system requires huge amounts of data to build its knowledge and learn or make decisions according to these data and the learnt patterns or relationships between them. An efficient clustering mechanism is required to make sure that these huge amounts of data are usable and would enhance the learning and decision-making processes. These "big data", especially in the healthcare industry, are changing the way patients and doctors handle care. The bigger data involved, the more efficient healthcare services are, yet the harder it is to manage these data.

A cluster holds a collection of data items that are aggregated together based on some similarities between them. K-means clustering is one of the simplest and most popular learning algorithms, due to its simplicity and low computational costs. Clustering of data using k-means starts with a first group of randomly selected centroids, which are used as the starting points for every cluster, and then performs iterative calculations to optimize the positions of the centroids, and the data items that are similar to that centroid [10]. Clusters creation stops when:

- The centroids have stabilized i.e. (There is no change in their values because the clustering has been successful).
- The defined number of iterations has been achieved. This number is defined by the programmer or set based on certain threshold of the number of clusters.

The K-means clustering algorithm uses iterative refinement to produce a final result. The algorithm inputs are the number of clusters K and the data set. The data set is a collection of features for each data point. The algorithm starts with initial estimates for the K centroids, which can either be randomly generated or randomly selected from the data set. The algorithm then iterates between two steps [11]:

1. Data assignment step:

Each centroid defines one of the clusters. In this step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. More formally, if c_i is the collection of centroids in set C, then each data point x is assigned to a cluster based on

$$S_i = \overbrace{c_i \in C}^{agr\ min} \ dist(c_i.x)^2 \tag{1}$$

where *dist (·)* is the standard (L_2) Euclidean distance. And for each i^{th} cluster centroid the set of data point assignments is S_i.

2. Centroid update step:

In this step, the centroids are recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster.

$$c_i = \frac{1}{|S_i|}\sum_{x_i} \in \ S_i{}^{x_i} \tag{2}$$

the algorithm iterates between steps one and two until a stopping criterion is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).

This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome. Choosing *K* (denoting the number of clusters) is considered to be the backbone of the algorithm to run as desired. To find the number of clusters in the data, the user needs to run the K-means clustering algorithm for a range of *K* values and compare the results. In general, there is no method for determining exact value of *K*, an estimation is performed using some tested technique. One of the metrics that is commonly used to compare results across different values of *K* is the mean distance between data points and their cluster centroid. Since increasing the number of clusters will always reduce the distance to data points, increasing *K* will always decrease this metric [12-15]. Mean distance to the centroid as a function of *K* is plotted and the "elbow point" where the rate of decrease sharply shifts, can be used to roughly determine K [16].

A number of other techniques exist for validating *K*, including cross-validation, information criteria, the information theoretic jump method, the silhouette method, and the G-means algorithm. In addition, monitoring the distribution of data points across groups provides insight into how the algorithm is splitting the data for each *K*. The proposed hybrid system is dedicated to use the k-means algorithm in clustering of collected data that are used to get the proper medical diagnosis, based on symptoms submitted by the user. The clustering model building process is shown in Figure 1.

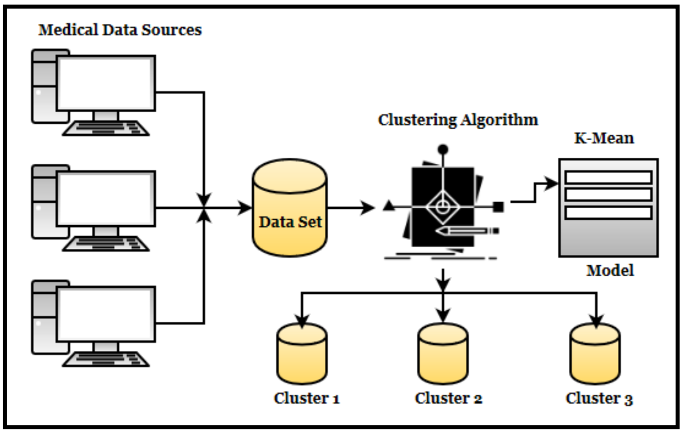

Figure 1. Phase 1 of the system’s building (Build a clustering Model)

3.2. The Expert System

An Expert System is one of the most common applications of artificial intelligence. It is a computer program that simulates the decision and actions of a person or an association that has specialist facts and experience in a particular field [17]. Normally, such a system contains a knowledge base containing accumulated experience and a set of rules for applying the knowledge base to each particular situation. The major features of expert system are user interface, data representation, inference, explanations etc. Advantages of expert system are increased reliability, reduced errors, reduced cost, multiple expertise, intelligent database, reduced danger etc. Disadvantages of expert system are absence of common sense and no change with changing environment [18, 19].

A Fuzzy Expert System, that is a decision-making system based on Fuzzy Logic inference, is a group of membership functions and rules. These functions and rules are used to reason about data. Fuzzy expert systems are oriented toward numerical processing. It takes numbers as input, and then translates the input numbers into linguistic terms like Small, Medium and large. Then the task of Rules is to map the input linguistic terms onto similar linguistic terms describing the output. Finally, the translation of output linguistic terms into an output number is done [20]. These rules are built in an if-then manner, and are evaluated in parallel in the inference engine, which indicates that the orders of these rules are not important. The terms used to explain the rules are defined ahead of initializing the inference engine. These terms are nouns and adjectives that are used to describe the input data (like high, normal, big, low, etc.) [21, 22].

Fuzzy logic relies on having learning rules, according to which an inference engine makes decisions regarding the problem at hand. The input to these fuzzy sets is in natural language (mostly) and the output can either be crisp or in natural language [3]. A fuzzy classifier is a procedure of labeling sets in the decision making or machine learning algorithm that embeds uncertainty measures, aka; fuzzy logic, within its workflow. A fuzzy classifier uses a rule base that can be expressed as a fuzzy knowledge base to Convert crisp input into a linguistic variable in a process named fuzzififcation, then an inference engine makes fuzzy decisions based on pre-set rules, and the fuzzy output is defuzzified by converting it back into crisp output using membership functions analogous to the ones used in the fuzzification phase [22]. The analysis of fuzzy input in order to produce a fuzzy decision is based on three main operations:

– Receive the fuzzy input: that could be from a single source, multiple heterogonous sources or even distributed data sources.
– Processing of these fuzzy inputs in a “fuzzification” technique that relies on a set of rules that are set according to human thinking “if-then” procedures in simple natural language, in addition to traditional processing methods.
– Reaching weighted results after passing the fuzzy rules and assembling them into a single decision related to the problem that guides other parts of the system or the human user what to do after de-fuzzification of the results to make them understandable.

The expert system that is built using a fuzzy logic inference module uses the clustered data produced by the data collection unit using the k-means algorithm to build the fuzzy knowledge base. Once the user enters his/her data then the symptoms he/she suffers from, the system matches these symptoms to a suitable cluster. This cluster is determined by a fuzzy logic system. Another fuzzy logic module uses patient’s attributes (age, environment, gender) along with the cluster’s attributes and features extracted from the symptoms, to make a decision on which diagnosis is the most relevant. Figure 2 shows the main structure of the expert system’s functionality.

The fuzzy logic module produces a ranked list of matching diagnosis. Each diagnosis is given a confidence level in percentage. Based on this level, the system chooses the one with the highest level and sends it to the user’s device. If this confidence level is low (below 60%) then the system prompts the user either to enter more symptoms or to suggest adding the diagnosis (assuming that a physician is using the system or the patient has consulted a doctor to make a judgment). The user sends this suggested diagnosis back to the expert system, which uses this diagnosis to alter the rules, or add a new rule to cover this diagnosis and learn from it to produce a more reliable diagnosis for similar cases in the future.

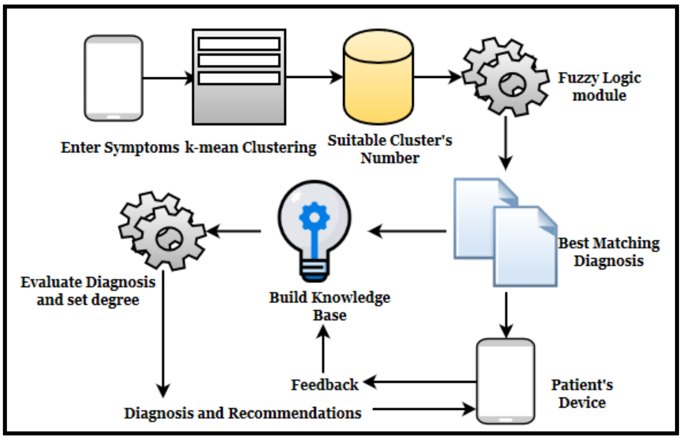

Figure 2. Expert System's main structure

4. System Evaluation

The system was tested by a group of experts in the field of medical diseases diagnosis and found to give a high performance and accurate results together with the ease of use. Central Pediatric Teaching Hospital, World Health Organization (Factsheet), WebMD, Mayoclinic, healthline, and other certificated website were used for data collection. The medical dataset of 350 records (diseases) and 300 features (symptoms, disease description and the possible advice and treatment according to the disease degree where collected, analyzed, and preprocessed to the required format [23].

4.1. The Fuzzy Expert System

The fuzzy logic toolbox is one of most powerful tools that MATLAB provides, to help its users build, analyze, design and simulate a fuzzy logic system, through a set of applications that encapsulate functions easily used through its user-friendly interface and code integration module. The functions fuzzy logic toolbox provides cover the basic methods used by fuzzy logic, like fuzzy clustering, neuro-fuzzy learning, inference engine and rule base building. The rules are implemented using simple logic rules and integrated within the fuzzy inference system, which can be later used to simulate the fuzzy system as a while, or simulate it within Simulink (a part of Matlalb's environment) to simulate the fuzzy system within a comprehensive model of the entire dynamic system. The data acquisition unit collected and clustered big data regarding common diseases and related symptoms. These data are used to build rules to be integrated into the rule base of the fuzzy logic to decide on what is the most relevant disease the submitted symptoms show. An example of such a rule is:

- Disease (Patient, tuberculosis),
- Symptom (Patient, persistent_cough),
- Symptom (Patient, constant_fatigue),
- Symptom (Patient, weight_loss),
- Symptom (Patient, loss_of_appetite),
- Symptom (Patient, fever),
- Symptom (Patient, coughing_up_blood),
- Symptom (Patient, night_sweats).

where Tuberculosis is a lung disease whose symptoms are persistent cough, constant fatigue, weight loss, loss of appetite, fever, coughing up blood, night sweats.

The expert system takes into consideration other factors when building a diagnosis. As it is known to almost every healthcare practitioner; factors like age, environment, and previous health issues do affect the decision made about diagnosis. The fuzzy logic based expert systems, uses as input, in addition to the submitted symptoms, user's age, medication history, Body Mass Indicator, gender. These attributes are submitted by the user through the diagnosis unit, and used by the diagnosis unit to reach a decision. The fuzzy logic engine accepts non-crisp input of normalized data as a “Fuzzy” input for the inference engine to make a “fuzzy” decision. In other words; the fuzzy logic evaluation system takes input values between 0-1 by mapping each of the “fuzzy set” members through a membership function so that all values are between 0 and 1. The performance of our proposed system was evaluated using a confusion matrix which contains the information on the actual case of the patient diagnosed by the expert and the diagnosis predicted by our hybrid system. A two-class classifier confusion matrix is shown in Table 1 and the result is presented in Table 2.

Table 1. Confusion Matrix

Confusion Matrix		Predicted	
		Negative	Positive
Actual	Negative	TN: is the number of diagnosed cases (No, they don't have a disease) predicted correctly but they don't have the disease	FP: is the number of diagnosed cases (Yes, they do have a disease) predicted incorrectly but they do not have the disease.
	Positive	FN: is the number of diagnosed cases (No, they don't have a disease) predicted incorrectly and they do have the disease.	TP: is the number of diagnosed cases (Yes, they do have a disease) predicted correctly and they do have the disease.

Table 2. The Proposed System Confusion Matrix

Proposed System Confusion Matrix		Predicted		Total
		Incorrect by the FRS	Correct by the FRS	
Actual	Incorrectly diagnosis cases	6	1	7
	Correctly diagnosis cases	2	21	23

The performance of our proposed system was generally rated using the data in the matrix. Some metrics, including accuracy, precision, sensitivity (recall), F-measure (F1 score), and specificity were applied as the criteria to implement this evaluation. In (3) to (7) show the formulas for these metrics [24]:

$$Accuracy = \frac{TN + TP}{TN+FP+FN+TP} \times 100\% \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \tag{6}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

The F1-Score is the harmonic average of the precision and recall. The best value realized is one and the worst is zero [25]. As shown in Table 3, the F1 score of our system was 0.929 which indicates a very good performance.

Table 3. The Accuracy of the Proposed System

Results	
Accuracy	90%
Precision	95%
Sensitivity (Recall)	91%
Specificity	85%
F1-Score	0.929

4.2. User Interface

The system was implemented to be compatible with a website browser where the user can input the symptoms in the front end as a textual expression; these expressions are matched against the closest symptom collected by the data collection unit. The user can also choose the symptom from a list of pre-defined symptoms. The screenshot in Figure 3 shows the system’s start up page, where the user enters personal attributes to be used as input (in addition to the symptoms) to the fuzzy logic expert system.

The user enters his/her information and moves to the next page to start adding the sympotoms he/she has to help the system provide proper diagnosis. Figure 4 shows the symptoms entry web page. The symptoms entered by the user can be given a level of effect. How annoying or uncomfortable these symptoms are making the user. These levels are: Low (somewhat annoying), Moderate (annoying and causes pain sometimes), and High (this symptom is obstructing the user from performing daily chores). When the user finishes adding the symptoms and clicks the "finish" button, the system produces the diagnosis that produced the highest "confidence level" in the fuzzy logic expert system. This diagnosis is displayed with how sever it is (low, moderate, or high) along with proposed procedure to follow or a treatment. Figure 5 shows the diagnosis web page.

The user is able to enter symptoms in a similar manner to that through a web browser adding symptoms is easy through the drop-down list or by simply filling in the text box. Choosing the severity of the symptoms is done through radio buttons for easy and clear input. Once the symptoms are entered, the diagnosis window is displayed.This information is sent to the base station to the main diagnosis unit to add to the confidence level produced by the expert system to make future diagnosis even more accurate, and provide future users with a more reliable diagnosis system, and automatically modify the rules in the knowledge base of the system, and modify the learning and inference rules accordingly too.

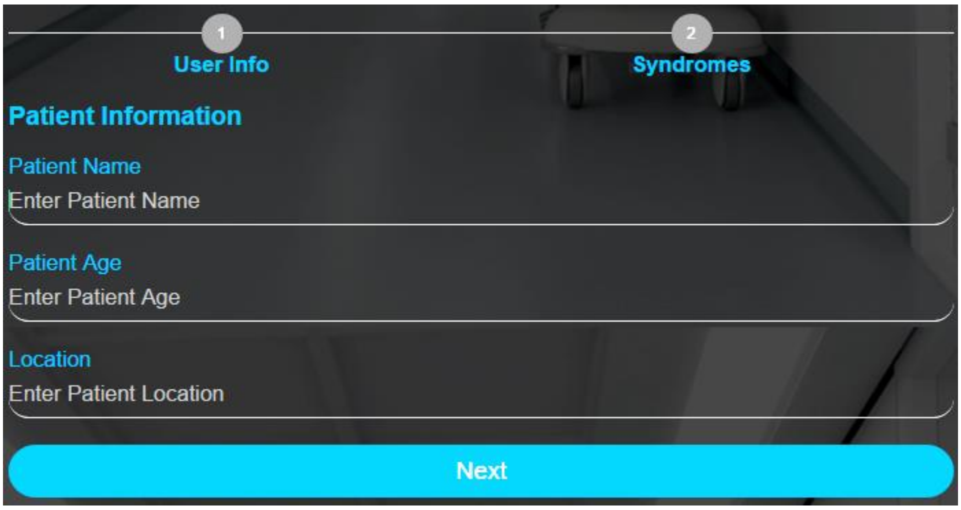

Figure 3. Start up web page for the proposed medical diagnosis system

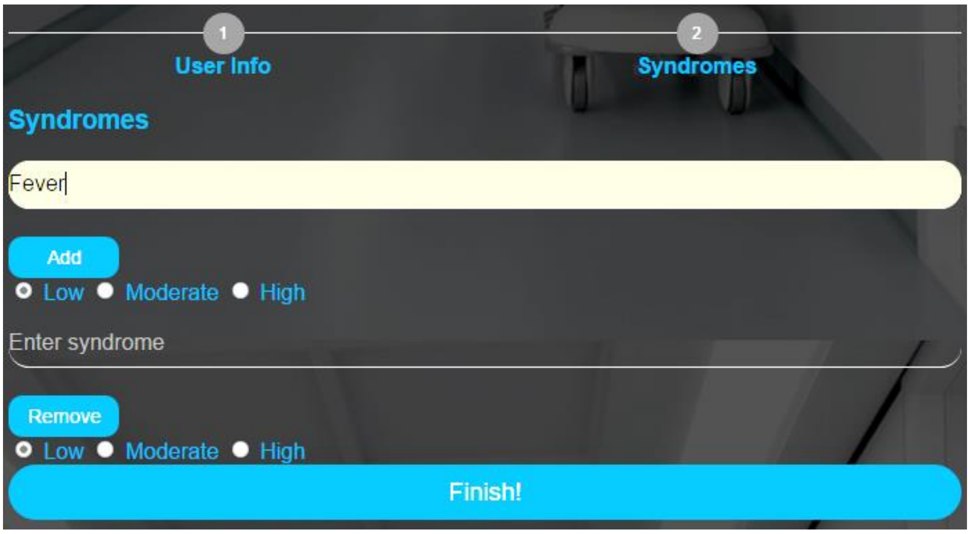

Figure 4. Entering symptoms

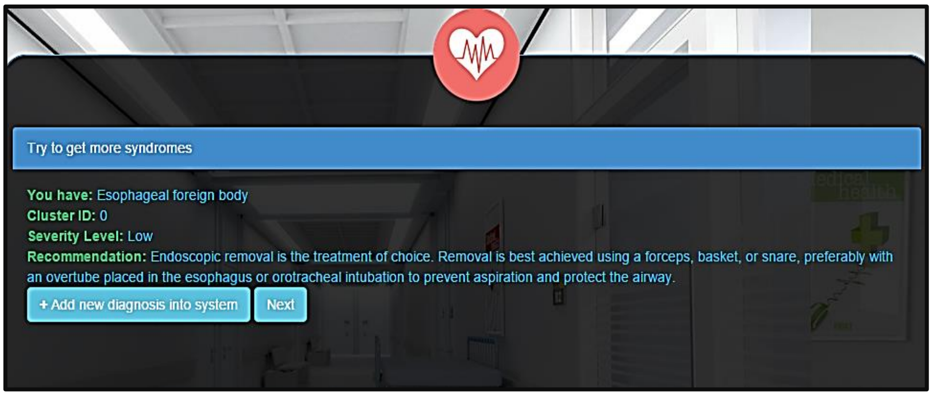

Figure 5. Diagnosis based on submitted symptoms

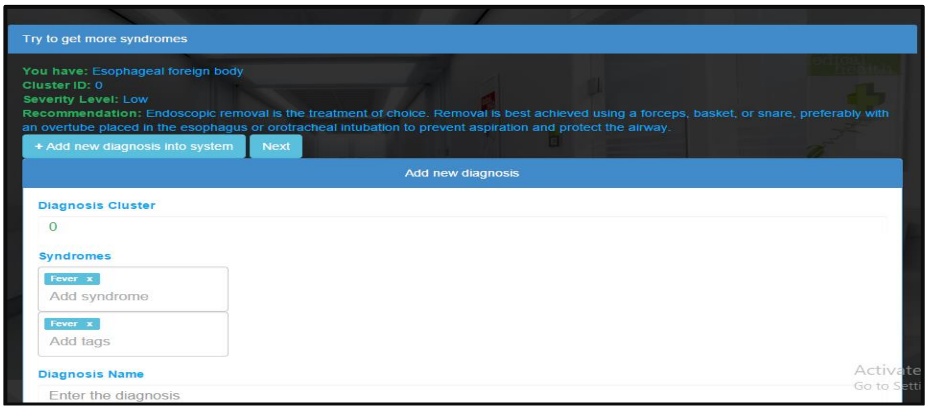

Figure 6. Part of add diagnosis web page

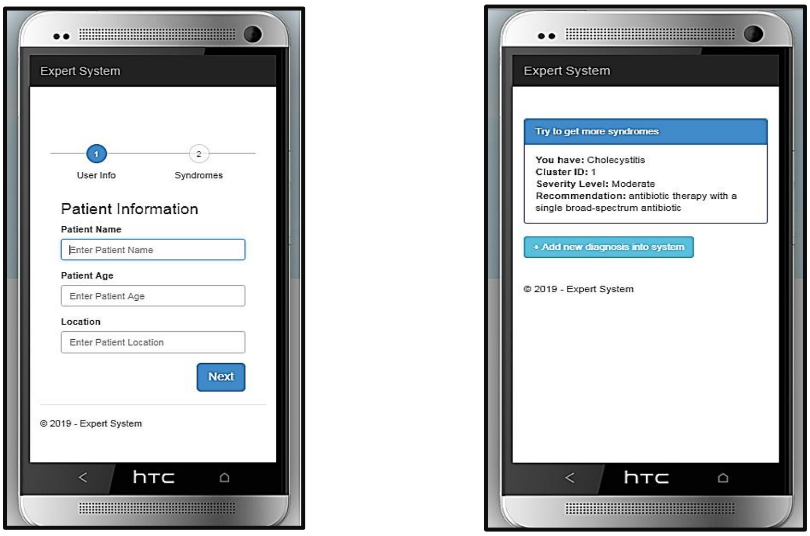

Figure 7. Android emulator operating the diagnosis system

5. Conclusion

An expert system that is built with a Fuzzy Logic decision making procedure is a big step towards having and adaptive and constantly developing system that builds its own knowledge base and provide users with an adequate diagnosis procedure that supports levels of severity of detected diseases, based on several information, among which are the symptoms provided by the system's users themselves. The use of clustering technique increases the system performance by groubing the most similer diseases which allows the fuzzy logic to look up for the right diagnosis in the nearest cluster according to Euclidean distance. The achieved results indicate a promising diagnosis performance of the system as it achieved90% accuracy and 92.9% F-Score.

References

[1] Parker KR, Srinivasan SS, Houghton RF, Kordzadeh N, Bozan K, Ottaway T, Davey B. Health informatics program design and outcomes: Learning from an early offering at a mid-level university. *Education and Information Technologies*. 2017; 22(4): 1497-1513.

[2] Shen Y, Colloc J, Jacquet-Andrieu A, Lei K. Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system. *Journal of Biomedical Informatics*. 2015; 56: 307-317.

[3] Bellman RE, Zadeh LA. Decision-making in a fuzzy environment. *Management Science*.1970; 17(4): B-141.

[4] Xia M, Xu Z. Hesitant fuzzy information aggregation in decision making. *International Journal of Approximate Reasoning*. 2011; 52(3): 395-407.

[5] Gatton TM, Lee M. *Fuzzy logic decision making for an intelligent home healthcare system*. In 2010 5th International Conference on Future Information Technology.2010.

[6] Lopez D, Gunasekaran M. *Assessment of vaccination strategies using fuzzy multi-criteria decision making*. In Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015). Springer, Cham. 2015; 195-208.

[7] Hussain A, Wenbi R, Xiaosong Z, Hongyang W, da Silva AL. Personal home healthcare system for the cardiac patient of smart city using fuzzy logic. *Journal of Advances in Information Technology*.2016; 7(1): 58-64.

[8] Samuel O, Omisore M, Ojokoh B. A Web Based Decision Support System driven by Fuzzy Logic for the diagnosis of typhoid fever. *Expert Systems with Applications*. 2015; 40:4164-417.

[9] Fatumo S, Adetiba E, Onolapo J. Implementation of XpertMalTyph: An Expert System for Medical Diagnosis of the Complications of Malaria and Typhoid. *IOSR Journal of Computer Engineering (IOSR-JCE).* 2013; 8(5): 34–40.

[10] Page JT, Liechty ZS, Huynh MD, Udall JA. BamBam: genome sequence analysis tools for biologists. *BMC Research Notes*.2014; 7(1): 829.

[11] Hartigan JA, MA Wong. A K-means clustering algorithm. *Applied Statistics.* 1979; 28:100–108.

[12] F Yuan Z, H Meng, HX Zhang, CR Dong. *A New Algorithm to Get the Initial Centroids*. *Proc. of the 3rd International Conference on Machine Learning and Cybernetics,* 2004; 26-29.

[13] Z Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*. 1998; 2: 283-304.

[14] D Birant A. Kut. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*. 2007; 60(1): 208-221.

[15] ALN Fred, JMN Leitão. *Partitional vs hierarchical clustering using a minimum grammar complexity approach*. Proc. of the SSPR & SPR 2000. LNCS 1876. 2000; 193-202.

[16] K-means algorithm in Python tutorial, available through: https://pythonprogramminglanguage.com/kmeans-elbow-method/ accessed, Nov. 2018.

[17] Bassem S. Medical Expert Systems Survey. *International Journal of Engineering and Information Systems (IJEAIS*). 2017; 1(7): 218-224.

[18] De Kock E. Decentralising the Codification of Rules in A Decision Support Expert Knowledge Base, M.Sc. Thesis. Faculty of Engineering. Built Environment and Information Technology, University of Pretoria; 2003.

[19] Jimmy S, Dinesh G, Abhinav B. Medical Expert Systems for Diagnosis of Various Diseases. *International Journal of Computer Applications*. 2014; 93(7): 36-43.

[20] Imène B, Noria T. A multi-agent framework for a web-based decision support system applied to manufacturing system. *CIIA*. 2009; 9.

[21] William S, James J. Fuzzy Expert Systems and Fuzzy Reasoning, John Wiley & Sons, Inc., 2005.

[22] Rimpy N. Medical Expert System-A Comprehensive Review. *International Journal of Computer Applications*. 2015; 130(7): 44-50.

[23] Centers for Medicare & Medicaid Services. *ICD-10-CM Official Guidelines for Coding and Reporting*. 2012 ICD-10-CM and GEMs. Retrieved May 17, 2012.

[24] Zainab T. Al-Ars and Abbass Al-Bakry. Iraq's Major Infectious Disease Diagnosis Using A Fuzzy Rule-Based System. *International Journal of Engineering & Technology* .2018; 7(4): 4943-4948.

[25] Chala D, Million M, Debela T. Developing a Knowledge-Based System for Diagnosis and Treatment of Malaria. *Journal of Information & Knowledge Management*. 2016; 15(4): 108-112.