Advances In Multidisciplinary & Scientific Research
A Multidisciplinary & Interdisciplinary Journal

Vol. 2 No. 3, September 2016

Web-Based Decision Support System for Diagnosis of Ebola Virus Disease Using Bayesian Networks

[1]Dogo E.M., [2]Kolo J.G., [3]Aror O. & [4]Rahman A.T.
[1, 2, 3, 4]Department of Computer Engineering
Federal University of Technology
Minna, Nigeria
[1]eustace.dogo@futminna.edu.ng, [2]jgkolo@futminna.edu.ng, [3]odafe41@gmail.com, [4]rahmantaofeek@gmail.com

[1]Corresponding author

ABSTRACT

The recent epidemic of the Ebola Virus Disease (EVD) left many dead in West Africa and in other parts of the world. A major problem faced was late diagnosis or diagnostic error of the disease; this was due to largely unavailability of medical professionals familiar with the disease and low doctor to patient ratio. An accessible method for reliable diagnosis is required to offset the low ratio of doctors to population. This paper presents the application of Bayesian networks for diagnosis of EVD. A general procedure for implementing a Bayesian network model is proposed; thereafter we demonstrate how the resulting Bayesian network can be applied in a medical diagnostic decision support system. The system uses the questionnaire method to elicit symptoms and is accessible through web browsers over the internet and mobile phones to potential patients and medical practitioners. The system developed is able to provide diagnosis in the form of probabilities, for the presence or absence of EVD in an individual. The probability of an individual infected by the disease depends on present or absent of particular symptoms according to the gathered disease pathology. The system was successfully developed, and had a diagnostic accuracy of 77% when compared to the World Health Organization (WHO) algorithm, but refinements of the conditional probability distribution would provide the most accurate sensitivity to symptoms and also improve the accuracy of diagnosis. Finally, web functionality, performance and usability test on the developed web application is carried out by simulating various load patterns and the result was generally acceptable.

Keywords: Artificial Intelligence, Bayesian network, Expert systems, Diagnosis, Ebola Virus disease

Aims Research Journal Reference Format:
Dogo E.M., Kolo J.G., Aror , O. & Rahman A.T. (2016): Web-Based Decision Support System for Diagnosis of Ebola Virus Disease Using Bayesian Networks. Advances in Multidisciplinary Research Journal. Vol. 2. No. 3, Pp 213-232.

1. INTRODUCTION

EVD is a severe, often fatal illness in humans, with a fatality rate of 90% for all infected. Since it was first discovered and named after a river in the Democratic Republic of Congo (formerly Zaire) in 1976, there has been regular outbreaks of the various Ebola strains which were conferred to a particular geographical area, but the 2014 epidemic was the most devastating covering many West African countries, US and Europe (Centers for Disease Control and Prevention, 2015). This was mainly due to easier human movement within and across borders to other countries. The virus is the most virulent and belongs to the family Filoviridae which was responsible for the recent epidemic in West Africa (Centers for Disease Control and Prevention, 2015). The first outbreaks of Ebola is reported to have occurred in remote villages in Central Africa, in regions where tropical rainforests are present and it infected over 318 people in Zaire, with a mortality rate of 88%.

A few months later in that year, Ebola emerged in Sudan, infecting 284 with a mortality rate of 53% in both cases, the disease was spread by close personal contact, mainly within hospitals and many medical care personnel were infected (World Health Organisation, 2014), (Lorente, Blanch, & Esteban, 2015). However, researchers have put forward the theory that the first patient became infected through contact with an infected animal. Sources of human infection are mainly through contact with body fluids of an infected individual, this normally happens during the care of an infected individual either in clinical or residential settings, handling victims' corpses and consumption of an infected animal (Onwuakor, 2014). Time of infection to onset of symptoms ranges from 2-21 days, illness is usually abrupt and is characterized by the following signs and symptoms which are well documented in literatures: Headache, Joint Pain, Muscle aches, Sore throat, Weakness, Fever, Diarrhoea, Vomiting, Stomach Pain, Hiccups, Rash, Internal and external bleeding in many patients (Centers for Disease Control and Prevention, 2015).

Various modeling approach have been proposed by researchers for prediction of future Ebola epidemic (Siettos, Anastassopoulou, Russo, Grigoras, & Mylonakis, 2015), but the consensus among researchers is that there will be future Ebola outbreaks. With tremendous broadband and mobile communication penetration in West Africa were the Ebola epidemic was hard hit, the goal and contribution of this study is to develop a simple and intuitive medical decision support system for diagnosis of EVD leveraging on internet and mobile technology in order to enhance surveillance and awareness (Musa, et al., 2015) and reducing patient care misdiagnosis errors (Omole, Rawlings, & Festus, 2013), (Croskerry, 2003). As past Ebola outbreaks were mistaken for yellow fever, dysentery, typhoid fever (Kucharski & Piot, 2014), and sometimes malaria.

2. THEORETICAL BACKGROUND

2.1 Artificial Intelligence and Medical Diagnosis

Artificial Intelligence (AI) is defined as man-made systems that think, act and behave rationally given what they know (Russel & Norvig, 2003). Diagnosis is a fitting problem area for artificial intelligence where no efficient algorithmic solutions exist, because all the symptoms for all faults are not known in advance. The effectiveness of diagnostic reasoning lies in the ability to infer using a variety of information and knowledge sources, connecting or selecting between different structures to reach the appropriate conclusions (Angeli, 2010). Probabilistic reasoning is the use of the diagnostic value of specific symptoms, signs, or tests to rule in or rule out a diagnosis (McCormack, Hudson, & Hoare, 2012).

2.2 Expert systems (ES)

Expert systems are part of a general category of computer applications known as Artificial intelligence, and consist of dedicated embedded or software solutions tailored to provide analyses, specific to a problem domain that can function in the position of an expert in that problem domain. ES are beneficial to human because it's designed to act the place of a human expert or designed to guide them. ESs are designed for different functions, like medical diagnosis, financial forecasts, etc. (Ryan, 2014).

2.3 Decision Support Systems

Decision support systems are dedicated embedded or software solutions tailored to provide analyses, specific to a problem domain to augment the analyses of a practitioner in making better quality decisions. Clinical decision support systems are "active knowledge systems which use two or more items of patient data to generate case-specific advice." (Wyatt & Spiegelhalter, 1991). Programs can also support a clinician's therapy planning, improving the quality of care, as shown in a recent trial (Conejar & Kim, 2014). The goal of clinical decision support systems (CDSS) is to emulate the clinician's thought process during diagnosis (Algarin, 2011). Every clinical decision support system must provide the following to be useful in a real-world diagnostic environment: Patient data acquisition, Patient data validation, Medical knowledge modelling, Medical knowledge elicitation, Medical knowledge representation, Medical knowledge reasoning, System performance and Integration to workflow (Algarin, 2011).

Figure 1 shows the architecture of an expert system which consists of 3 major parts, the knowledge base, reasoning/inference engine, and user interface.

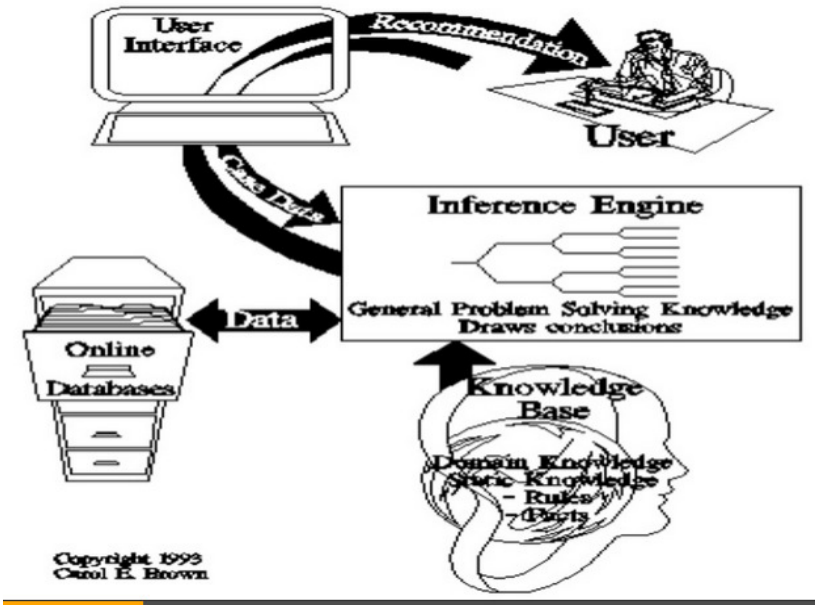

Figure 1: Architecture of an Expert System (Brown, 1993)

i. **Knowledge Base**: It contains the information gotten from experts in the particular field of operation and also governs by the logic rules on how that information is applied. The knowledge based is also divided into factual and heuristic knowledge. Factual knowledge is that knowledge of task domain that is widely shared. They are gotten from journals, textbooks, websites etc. while the heuristic knowledge is less exhaustive, more experiential and more judgmental knowledge of performance.
ii. **Inference Engine or Reasoning**: It actually serves as the problem solver and it acts as the engine of the software by interpreting the submitted problem against the set rules /logic of information stored in the Knowledge base to be used or followed.
iii. **Interface**: It allows the user to be able to access the software, as intuitively as possible. Making it user friendly implies simplicity to the users and also allows the user to interact with the expert system through the use of human language e.g. English.

2.4 Bayesian Network & Theory

Bayesian or belief networks are a construct to coherently model domain knowledge, particularly causality, independence/dependence and uncertainty. It performs inference against a body of knowledge by ascertaining the degrees of belief by observation within that body of knowledge. Bayesian networks have been applied successfully in the field of medical diagnosis (Onisko, Druzdzel, & Wasyluk, 1999); (Lin & Haug, 2008); (Sun, Tang, Ding, Lv, & Cui, 2011). Bayesian networks are apt for the development of diagnostic problems because they cope very well with uncertainty, inconsistent data and still provide accurate diagnosis. In addition, they are, in some cases, not as computationally expensive as other machine learning techniques like neural networks (Prasadl, Krishna Prasad, & Sagar, 2011). They use joint distributions of data on a particular problem domain, as a “knowledge base” (Russel & Norvig, 2003) to make inference on queries about that problem domain.

Therefore, the proposed system with the use of Bayesian networks would provide fast and accurate results as will be required with the diagnosis of EVD. Bayesian network is built on Bayes theorem, and conditional probability that a disease is present in a finding, $\mathbf{P(disease|finding)}$, using the probability of finding a given disease, $\mathbf{P(finding|disease)}$, the probability of disease, P(disease), and the probability of a finding, P(finding) given by equation 1:

$$\mathbf{P(disease|finding)} = \frac{\mathbf{P(finding|disease)P(disease)}}{\mathbf{P(finding)}} \quad (1)$$

Bayesian network, any node is conditionally independent always of its all non-descendants given that node's parents (Horny, 2014). There currently exist two techniques for building Bayesian networks; parameter learning and structural learning. In parameter learning, a pre-existing network structure is used by the developers and then, the values of the network structure are estimated and then optimized (Wong & Guan, 2008) while in structural learning, a Bayesian network is specified by an expert and used and its then used to perform inference (Webb & Andrew, 2011). Bayesian networks have seen widespread application, particularly for their ability to represent the conditional probabilistic relationships with incomplete data (e.g. expert opinion, field/survey data, historical data). It is also applied in multi-class classification problems and handwriting recognition. A joint probability distribution can be calculated that describes the Bayesian network using the chain rule (Pearl, 1988).

Bayesian network represented by a directed acyclic graph (DAG) ***G = (V, E)***, and each node $i \in V$ corresponds to a random variable X_i with a finite set $\mathbf{X}_i$ of mutually exclusive states. The relation ***pa(i)*** provides the set of all parents for a node ***i*** DAG ***G*** (Vomlel, 2005) . The conditional probability table for each node $i \in V$ is populated by equation 2:

$$(X_i)_{i \in V} = P(X_i \mid (X_j)_{j \in pa(i)}) \quad (2)$$

And, the joint probability distribution for the whole network is given by equation 3:

$$P((X_i)_{i \in V}) = \prod_{i \in V} P(X_i \mid (X_j)_{j \in pa(i)} \quad (3)$$

As such Bayesian networks model causal relationships amongst entities. Causality occurs when a characteristic or behaviour of a particular entity has direct influence on the behaviour and or characteristic of another entity within some domain. Causal networks attempts to formally represent the relationships among entities and the quantitative and qualitative influences they exert on one another (Ben-Gal, Ruggeri, & Kennett, 2007).

3. RELATED WORKS

Considerable literatures on expert systems for medical diagnosis support system have been carried out focusing on one or many diseases. Software systems that possess domain specific knowledge that assist users make better decisions and adopt better strategies are important for the trend of innovation to continue. There are also commercially available Clinical Decision Support Systems (CDSS) such as DXplain (Computer Science Laboratory, Massachusetts General Hospital, US), Isabel (Isabel Healthcare Inc., USA) and VisualDX (logical Images, Rochester, USA), but they are not readily available for use in Ebola epidemic prone West Africa region and they present a clinician with numerous diagnostic suggestions which could be a distraction and burden on the medical personnel especially in an epidemic scenario. Past research on ES for medical diagnosis can be broadly classified into the following rule-base, probabilistic network (Bayesian network) and Machine learning approach (Onisko, Lucas, & Druzdzel, 2001), (Kononenko, 2001):

3.1 Rule based method

(Fred, Filipe, Partinen, & Paiva, 2000), developed an auxiliary diagnosis system that analysed polysomnographic (PSG) data and provided a medical doctor with diagnosis consistent with the PSG data. The system featured a knowledge-based built with CLIPS (C Language Integrated Production System) and all its features, also a GUI (graphical user interface). Also, it provided the functionality of updating the rules in the knowledge-base after the system has been deployed. The system is heavily platform dependent and this can limit its usability, since it was built for Windows. (Wiriyasuttiwong & Narkbuakaew, 2009), proposed a knowledge-based diagnosis expert system that performed inference on signs and symptoms based on production rules and forward chaining that are used in performing inference on a patient's symptoms. The system showed a high confidence level (99%) on diagnosis of test and validation data. Nevertheless, the system using a fact list of twenty (20) signs and symptoms could be seen as too few a number for diagnosis.

(Abu-Naser, El-Hissi, Abu-Rass, & El-khozondar, 2010), developed a rule based expert system for the diagnosis of diseases of the endocrine. Knowledge engineering was carried out, and the facts, rules and procedures were built with Java Expert System Shell (JESS) and questionnaire method to collect the symptoms of patients was used. It demonstrated admirable accuracy. Nevertheless, it had limited accessibility and a limited disease diagnosis as defined in the knowledge base. (Hassan, Sher-E-Alam, & Chowdhury, 2010), developed a fuzzy medical expert system for the diagnosis of human diseases. It had a rule base that consisted of symbolic rules that were gotten from fuzzification of input data. It was able to capture the uncertainty and ambiguity inherent in diagnosis of diseases and it was also accessible over the internet. Nevertheless, the system had a limited detection rate because of the number of pathological test it could accommodate per diagnosis of disease in real time.

(Borgohain & Sanyal, 2012) developed a rule based diagnostic expert system that provided diagnosis for cerebral palsy. The system was built with JESS, it provided the user with a questionnaire which the user answered and the answers were weighted. Diagnosis was performed from rules collected during knowledge engineering process. The output of the system classified according to the severity i.e. mild, moderate or severe. The system didn't support remote access (i.e. web or internet). (Tunmibi, Adeniji, Aregbesola, & Dasylva, 2013), built an expert system for the diagnosis of fever. It was implemented with a rule-base, the rules defined after interviews with medical doctors. The system provided an easily accessible and usable interface and could diagnose a number of fever variations. The system was unable to present a diagnosis with incomplete information, because of the rule-based technique implemented.

(Oguntimilehin, Adetunmbi, & Abiola, 2013), developed an expert system that will be used for the clinical diagnosis of the various levels of typhoid fever. The LEM2 rule induction algorithm was used for the classification of the patient data. The algorithm generated eighteen (18) unique rules that will classify a symptom into the five (5) categories of typhoid fever.

3.2 Bayesian Network Method

(Onisko, Druzdzel, & Wasyluk, 1999), developed a Bayesian network for determining liver disorders that was deployed in clinical practice and for medical training. The network structure was built on data of patients with the disease in a database and from taking surveys from specialists on liver diseases**).** (Milho & Fred, 2000), built a Bayesian network for the diagnosis of sleep disorders. It provided an intuitive web user-interface, particularly directed towards the lay person. The system was limited in that it didn't possess any learning algorithms.

(Lucas P. , 2001), developed a medical decision support system for the optimal prescription of antibiotics to patients with pneumonia in an ICU (Intensive Care Unit). Knowledge engineering was carried out on the nature of pneumonia per patient and a corresponding treatment taking into account the symptoms that patient presented. The work did not describe remote access and also, the nature of symptom collection.

(Van der Gaag, Renooij, Witterman, Aleman, & Taal, 2002), designed a probabilistic network for the appropriate therapy selection for patients with different stages of oesophageal cancer. They generated the network structure and parameters of the network nodes in conjunction with oncologists in the area. The work demonstrated an 85% diagnostic accuracy on real patients with different stages of oesophageal cancer. (Lin & Haug, 2008), developed a Bayesian network structure with the consultation of medical experts. The parameters were estimated using a machine learning algorithm. Data was stored in a database overtime and this is what the algorithm learnt from. The developed systems' performance was analysed with sample data and it proved to be quite accurate, because it allowed for the *"missingness"* value of the dataset to be represented.

(Wiegerinck, Kappen, & Burgers, 2010), developed a Bayesian network for medical diagnosis. It contained an extensive variable list (nodes), greater than one thousand (1000). Also, the knowledge base consisted of the conditional probability tables (CPT) of the variable relationships. The work proposes a trade-off between diagnostic accuracy and computational feasibility by making all the risk factors a default value and do not contribute to inference. Also, a GUI is designed to enable easy use. (Thirumuruganathan, 2010), presented a novel procedure in developing expert systems. Knowledge engineering was carried out and rules generated, these rules were given certainty factors that were then used to bootstrap a Bayesian network, which is then optimized with a learning algorithm. This provided for a more robust network, due to the ease with which experts better represent knowledge qualitatively as opposed to quantitatively. The diagnostic system was not web accessible.

(Kosarzycki, 2011), developed a machinery fault diagnostic tool using Bayesian networks. The tool was SOAP (Simple Object Access Protocol) and REST communication protocols. It featured a sensory network that provided input to the Bayesian network in near real time. It could diagnose the causes of failures after they had occurred and also predict the occurrence of failures in the future. The system is heavily domain dependent, requiring several Microsoft Windows only packages to run i.e. Microsoft Silverlight and .NET Framework 4.0. The work by (Sun, Tang, Ding, Lv, & Cui, 2011) augmented the incomplete data with expert knowledge. The network therefore incorporated logical information and hence more accurate diagnostic results. The system developed wasn't web accessible.

3.3 Machine Learning Approach

(Levine, et al., 2015) used logic regression a supervised machine learning algorithm to develop a novel and simple Ebola prediction scoring system for triaging Ebola risk in patients with suspected EVD using EVD testing result of 382 patients data collected at Ebola treatment unit in Liberia over a period of 16 weeks of which 160 patients tested positive for EVD. For the Ebola prediction score, 6 symptom-variables independently predictive of labratory-confirmed Ebola virus disease was used as against 14 symptoms by WHO, and the result showed that patients with higher Ebola prediction scores had higher probability of laboratory confirmed EVD. The study was able to empirically derive and internally validate a clinical prediction model for laboratory confirmed EVD, but was a paper rather than computer-based template.

(Oguntimilehin, Adetunmbi, & Olatunji, 2014), developed a decision support system that implemented the C4.5 algorithms to generate a decision tree classification model based on clinical records normalized into attributes and subsequent classes. The system was quite accurate; having a detection rate of 95% on validation data and it could handle categorical and numeric data values. The decision tree was used to generate 17(seventeen), as such the system classification model would be tedious to update, when new information is provided. Williams & Olatunji (2013) proposed hybrid architecture for the diagnosis of typhoid fever using MATLAB. The system used artificial neural networks in tandem with fuzzy logic. The artificial neural network was used to estimate the membership function for the typhoid symptom severity. This provided for a better fit for diagnosis from data. Nevertheless, the system was computationally expensive, requiring superior hardware, the system was also not accessible remotely (i.e. over the web or internet). Machine learning for medical diagnosis is also well documented in (Kononenko, 2001); (Sajda, 2006) and (Shalev-Shwartz & Ben-David, 2014).

3.4 Problem Statement

The EVD outbreak in 2014 showed a fundamental lacking in easily accessible and accurate diagnostic tools that can reduce the spread of the infection most especially in the affected regions of Africa and also to reduce the likelihood of exposing an individual who has similar symptoms to that of EVD but does not have the disease. This diagnostic tool would help medical professionals, relevant officers and lay individuals alike, by providing an easy and intuitive tool that is easily accessible over the web and through mobile phones.

This could, in tandem with other available systems reduce the death toll of the infectious Ebola disease. This work attempts to provide a computer-based diagnostic tool that could help offset the high patient to medical professional's ratio by providing an easily accessible and intuitive diagnostic tool for timely diagnosis of the disease, as early treatment commencement is increases the probability of survival.

3.5 Justification of Bayesian Networks over Rule-based Systems

Bayesian networks are able to handle uncertainty better than rule-based systems. Even if all the required variables are not presented, it is able to make accurate inferences. Rule-based approach to a remote heart monitor was compared to a Bayesian network approach; the rule-based approach proffered more false alarms than the Bayesian network approach (Nee & Hein, 2010). Bayesian approach is more appropriate in this case because of its ability to handle complexity and more importantly uncertainty in causal relationships. The relationships between the symptoms can be more accurately represented with Bayesian networks than defining each individual rule in a rule-base (Murphy, 1998).

4. METHODOLOGY

The system consists of the several interconnected components (modules), these modules which were developed incrementally contributed to the overall functionality of the system. The block diagram in Figure 3 shows the interconnection between these modules. Knowledge engineering was carried out to determine qualitative and quantitative relationship amongst the EVD causative symptoms. This involved review of literature and consultation with a medical doctor. Each of the symptoms corresponding to the EVD was represented as a node, and their relationships with one another or the lack thereof was represented by a node of the graph to the other nodes.

The designed network was implemented in Java 8 SE. This was then debugged using an open source tool - UnBBayes as shown in Figure 2. UnBBayes is a Java GUI tool for designing and testing Bayesian networks, also the API that included libraries for performing inference with these networks. The network uses the Junction Tree algorithm for inference (Matsumoto, et al., 2011). In Figure 2 the topmost nodes are describes as the root nodes, since they have no parents or ancestors. They are: *infected_resident, travel_risk, meat_consumption and profession_risk*. The node which has both parents and children is the *ebola node*. It is called an intermediate node. Finally, the nodes which have a parent *(ebola node)* and ancestors *(infected_resident, travel_risk, meat_consumption* and *profession_risk)* are low most nodes referred to as the children nodes.

The network is arranged such that the causes are at the top and the effects of those causes are at the bottom. The intermediate node describes a logical mid-point to arrive at an effect from all the possible causes. Each node can have either of two (2) possible states. The root and leaf nodes have the states of *Yes* or *No*, while the intermediate node has the states of *True* or *False.*

Infected_resident (IR): It is the first of the root nodes. This node describes an individual who has lived in the same premises with another individual who is confirmed to have been infected with EVD.

Travel_risk (TR): Also a root node. This node describes an individual who has travelled to a region (or country) which has been reported to have an EVD outbreak.

Meat_consumption (MC): This is node describes an individual who is in the habit of consuming meats from animals that are carriers of EVD. Also, it will report positive or a yes for an individual who has consumed a carrier within a short period after the onset of symptoms.

Profession_risk (PR): This node describes an individual whose profession puts him/her in a position of being infected with EVD. They include but are not limited to healthcare practitioners, especially those in affected regions.

Ebola: This is the node of interest; it is the one that is queried on the presence of EVD or not. The first state gives the probability of the user having EVD (a True state) given the evidences presented from observations. The second state gives the probability of the user not having EVD (a False state) given the evidences presented from observation.

Abdominal_pain (Ab): This is a leaf node; it exhibits the effect of its parent node being either *True* or *False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It checks whether the user has been experiencing abdominal pain or cramps within a few days of the diagnostic session.

Diarrhea (Di): This is a leaf node; it exhibits the effect of its parent node being either *True* or *False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It checks whether the user has been experiencing loose stooling especially after meals, and also rapid dehydration.

Anorexia (An): This is a leaf node; it exhibits the effect of its parent node being either *True* or *False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It checks whether the user has lost appetite for food, or is having difficulty with feeding.

Muscle_pain (Mu): This is a leaf node; it exhibits the effect of its parent node being either *True* or *False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It checks whether the user has been experiencing body pain or cramps, especially at regions around the joints.

Bleached_tongue (Bl): This is a leaf node; it exhibits the effect of its parent node being either *True* or *False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It checks whether the user has a tongue devoid of colour, that is a colour that is white, dry and having a flaky look.

Difficulty_swallowing (Df): This is a leaf node; it exhibits the effect of its parent node being either *True* or *False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It checks whether the user has been experiencing difficulty swallowing during meals within a few days of the diagnostic session.

Time_factor (Ti): This is a leaf node; it exhibits the effect of its parent node being either *True or False*. Also, it is one of the nodes from which evidences are collated as observations from the user being diagnosed. It ascertains the time frame of the onset of the first symptoms. If the user has been experiencing the symptoms for greater than twenty-one (21) days then it is assigned *Yes*, otherwise, it is assigned the value of *No*.

The designed Ebola Bayesian network uses the Junction Tree algorithm to perform inference. It is an improvement on the Variable Elimination algorithm. The Junction Tree algorithm queries, takes into cognizance all the nodes and the observation or the lack thereof that they possess.

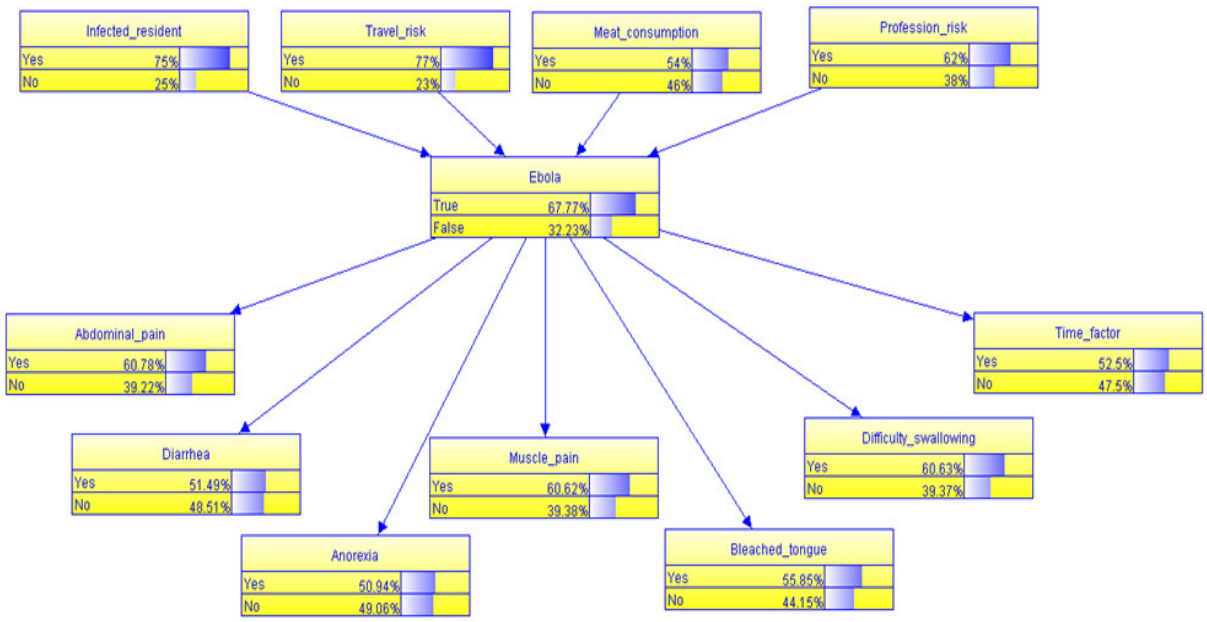

Figure 2: Bayesian Network Structure modelling of EVD diagnosis on UnBBayes

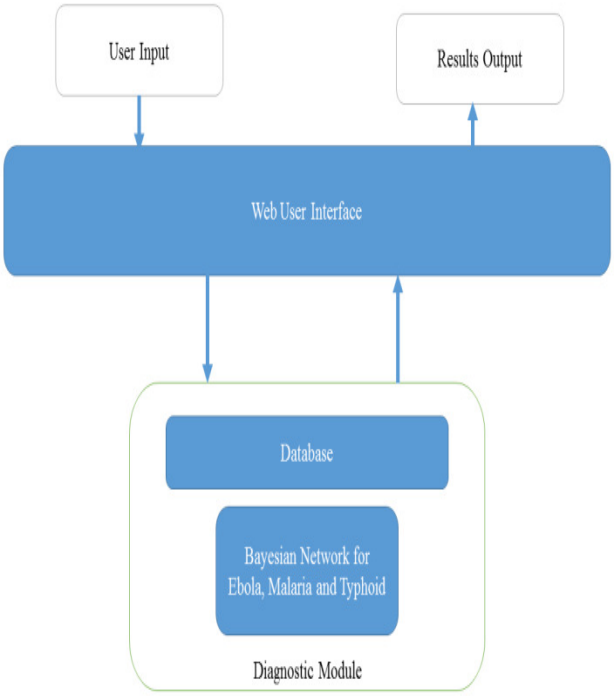

Figure 3: Block Diagram of Ebola Decision Support System.

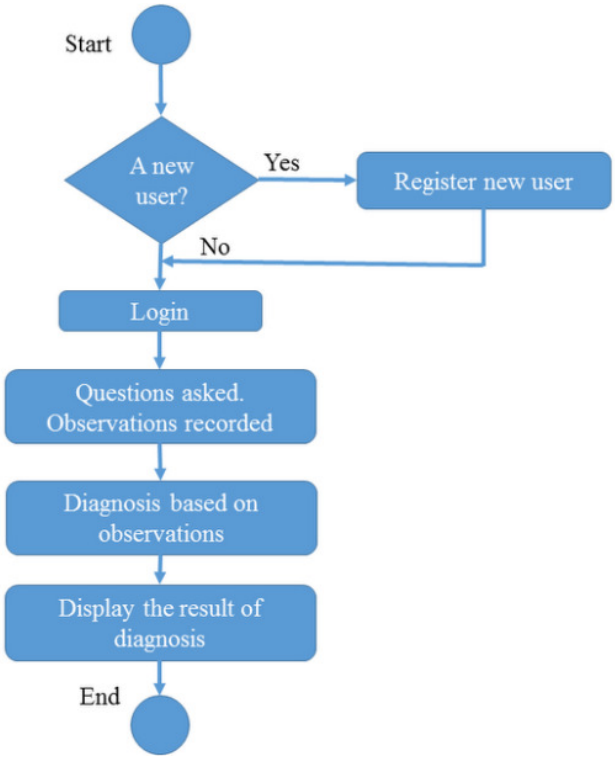

Figure 4: Activity Diagram of the Software System.

Figure 4 is an activity diagram showing in a stepwise fashion the interaction of a user with the diagnostic software system. The web user interface shown in Figure 5 was developed using software tools specifically for intuitive user experience and optimal performance. It was built with HTML5, CSS and JavaScript. PHP was used to provide a dynamic web interface and connectivity with the database. CS6 Dreamweaver was the software tool used for designing the web.

Figure 5: Web User Interface

The database management system used was MYSQL 5.6.17. Tables were created for logical organization and classification of patient data. The demographic information was stored as variables assigned to it were names, gender, age, relationship status, occupation etc. A separate table was assigned to store answers gotten from the questions and the questions were divided into five (5) sections with 3 questions on each section thereabout making it fifteen answers in all. The table was then constructed according to the classification/ groupings of the question in order to make it easier and faster when variables are being assigned to each answer gotten from the user during the phase of integration.

4.1 Building Bayesian networks Model

The development of a prototype for medical diagnosis is carried based on the following phases: Knowledge acquisition, Knowledge representation, Coding and Testing.

4.2 Application Scenario

A diagnostic system usually starts with the patients' complaint and the doctor tries to learn more about the patient situation interactively through an interview. Typically, a patient visits the doctor with having a particular complaint; the doctor then begins by compiling the patients' symptoms, then, using the diagnostic software, tries to find the possible diseases. He then presents to the patient the possible outcome before the commencement of treatment. Also, the patient can access the decision support system through his mobile phone or personal computer himself, the system then diagnoses the patient regarding his complaint. The patient can then take the result to a medical practitioner for treatment.

5. TESTING, RESULTS AND PERFORMANCE EVALUATION

The results obtained from simulating patients who had symptoms consistent with a patient not having EVD and that having EVD was considered. The diagnosis obtained from the software system was compared to that of the WHO first contact triage algorithm for the same patient symptoms (World Health Organisation, 2014), (Levine, et al., 2015). The disparity was calculated to obtain the accuracy of the system's diagnosis**.** The developed Bayesian network was designed and tested using UnBBayes 4.21.14 (Stable) on a system with specifications: Intel® Core™ i5-2430M Processor, 2.40GHz, 6GB RAM

The Bayesian network was tested by passing inputs of each of the patient's, symptoms through the diagnostic software system and the *ebola node* of the Ebola Bayesian network was queried. When queried, the greater of the two (2) probabilities corresponding to the states of *True or False* was observed and recorded. The dataset used in assessing the Ebola Bayesian network was developer generated shown in Table 1; this was due to the unavailability of EVD patient records and data. It was generated in a pseudo-random fashion that included the definitive presence or absence of a symptom, also its complete unavailability (those cases where the user is uncertain). It contained a distribution of fifty-two (52) hypothetical users and their symptoms, these symptoms were passed to the nodes in the Bayesian network. In the dataset some symptoms are left as blank fields; this signifies a user who could not answer definitively whether he/she had that symptom.

The results obtained from querying each patient's data with the Ebola Bayesian network model is shown in Table 2, it indicates the presence of EVD. The WHO algorithm (World Health Organization, 2014) for an EVD suspect was used to establish the correct or true value of diagnosis. Based on the WHO algorithm twenty-nine (29) of the patients were positive suspects for EVD, they constituted 55.77% of the data set

Table 1: Data Testing

PATIENT	IR	TR	MC	PR	Ab	Di	An	Mu	Bl	Df	Ti
1	Y	Y	Y	Y	N	N	Y	--	N	Y	Y
2	Y	Y	Y	Y	N	N	Y	Y	N	Y	Y
3	Y	Y	Y	Y	--	N	Y	Y	--	Y	Y
4	Y	Y	Y	Y	N	Y	Y	Y	--	Y	Y
5	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N
6	N	Y	Y	Y	Y	Y	N	N	Y	Y	N
7	N	Y	N	Y	Y	Y	N	N	Y	Y	N
8	N	Y	N	N	Y	Y	N	--	N	Y	Y
9	N	N	N	N	Y	--	N	Y	N	Y	Y
10	N	N	N	N	--	N	N	Y	N	N	N
11	Y	Y	N	N	Y	N	N	Y	Y	N	N
12	Y	Y	Y	N	Y	N	Y	Y	Y	N	N
13	Y	Y	N	N	--	--	Y	N	Y	N	N
14	Y	N	Y	Y	N	--	Y	N	Y	N	N
15	Y	N	Y	Y	N	--	Y	N	Y	Y	N
16	N	N	Y	Y	N	Y	Y	N	N	Y	N
17	N	N	Y	Y	N	Y	Y	N	--	Y	N
18	N	N	Y	N	N	Y	Y	Y	--	N	Y
19	N	N	Y	N	N	Y	N	Y	N	Y	Y
20	N	N	N	N	N	Y	N	Y	N	Y	Y
21	Y	N	Y	Y	N	N	N	Y	Y	Y	Y
22	Y	N	Y	Y	N	N	--	--	Y	Y	Y
23	Y	N	N	N	N	N	--	--	N	Y	Y
24	Y	N	N	N	--	N	--	--	N	Y	Y
25	Y	N	N	N	Y	N	Y	N	--	Y	Y
26	Y	N	N	N	Y	--	Y	N	--	N	Y
27	N	Y	N	N	Y	--	Y	N	Y	N	Y
28	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y
29	Y	Y	N	Y	N	Y	N	Y	Y	--	Y
30	Y	Y	Y	Y	N	Y	N	Y	Y	--	Y
31	Y	Y	Y	Y	--	N	N	Y	Y	N	N
32	Y	Y	Y	Y	Y	N	Y	N	N	N	N
33	Y	Y	Y	Y	Y	N	Y	N	--	N	N
34	Y	Y	Y	Y	Y	Y	N	N	Y	N	N
35	N	Y	Y	Y	Y	Y	N	--	Y	N	N
36	N	Y	Y	Y	N	Y	N	Y	N	Y	Y
37	N	Y	Y	Y	N	Y	N	Y	--	N	Y
38	N	Y	N	N	N	Y	N	Y	--	N	Y
39	N	Y	N	N	N	N	N	Y	Y	--	Y
40	N	N	N	N	N	N	Y	N	Y	--	N
41	N	N	Y	N	Y	N	Y	N	Y	N	N
42	N	N	Y	N	Y	N	Y	--	N	Y	N
43	N	N	Y	Y	Y	N	Y	Y	N	Y	N
44	N	N	N	Y	Y	Y	Y	Y	N	Y	N
45	N	N	N	Y	--	Y	Y	Y	Y	Y	Y
46	N	N	N	N	--	N	Y	Y	Y	--	Y
47	N	N	N	N	--	N	Y	N	Y	--	N
48	Y	N	N	N	N	Y	N	Y	N	--	N
49	Y	N	Y	Y	Y	Y	N	Y	N	Y	Y
50	Y	N	Y	N	Y	Y	N	Y	N	Y	Y
51	Y	N	Y	Y	N	Y	N	Y	Y	Y	N
52	N	N	Y	N	N	N	Y	--	Y	N	Y

Table 2: Result Obtained from Diagnosis of Patients

EVD Bayesian (True)	EVD Bayesian (False)	WHO Algorithm
T=99.56	F=0.44	T
T=99.72	F=0.38	T
T=99.96	F=0.04	T
T=99.99	F=0.01	T
T=93.5	F=6.5	T
T=93.5	F=6.5	T
T=79.66	F=21.34	T
T=99.68	F=0.32	T
T=96.98	F=3.02	F
T=0.34	F=99.86	F
T=64.13	F=35.87	T
T=30.18	F=69.82	T
T=28.95	F=71.05	F
T=16.9	F=83.1	F
T=73.06	F=26.94	T
T=30.93	F=69.07	T
T=53.99	F=46.01	T
T=9.09	F=90.91	F
T=98.37	F=7.63	F
T=96.57	F=3.43	F
T=99.76	F=0.24	T
T=99.76	F=0.24	T
T=95.56	F=4.44	F
T=98.23	F=1.77	F
T=99.68	F=0.32	T
T=98.17	F=1.83	T
T=98.86	F=1.14	F
T=99.96	F=0.04	Y
T=99.84	F=0.16	T
T=36.8	F=63.2	T
T=36.8	F=63.2	F
T=13.75	F=86.25	F
T=29.47	F=70.53	F
T=76.36	F=23.64	T
T=66.48	F=33.52	T
T=99.68	F=0.32	T
T=98.41	F=1.59	T
T=94.38	F=5.62	F
T=96.2	F=3.8	F
T=1.46	F=98.54	F
T=7.45	F=92.65	F
T=20.06	F=79.94	T
T = 51.11	F=40.89	T
T=24.38	F=85.62	F
T=0.01	F=99.99	F
T=0.62	F = 97.38	F
T=0.62	F = 97.38	F
T=1.38	F=85.62	F
T=85.62	F=1.38	T
T=85.62	F=1.38	T
T=85.62	F=1.38	T
T=1.38	F=85.62	F

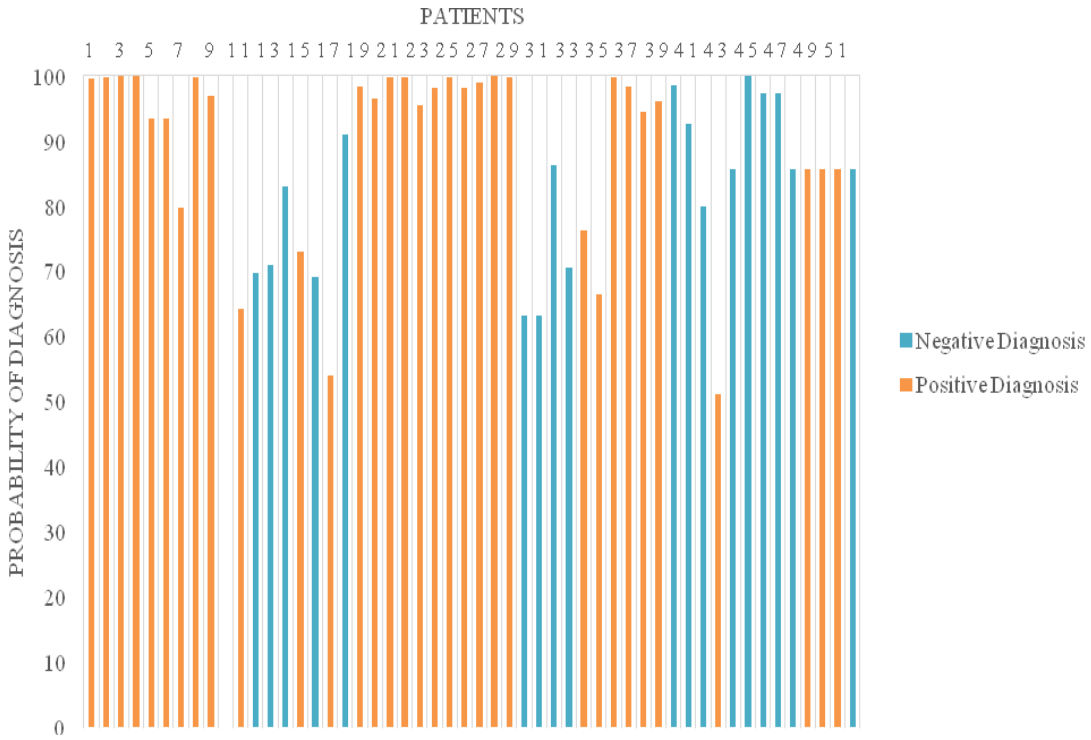

Figure 6: Positive and Negative Diagnosis Distribution

Figure 6 shows the distribution of positive to negative diagnosis using the Ebola Bayesian network on the test data, and the strength of each diagnosis. The vertical axis represented the probability of either a positive or negative diagnosis and the horizontal axis represented the patient data being diagnosed. The results were compared to the results of the WHO algorithm for each patient and tallied, the results of the Ebola Bayesian network differed from the WHO algorithm in twelve (12) users. Accuracy was determined by using the relationship:

$$\text{Accuracy} = \frac{\text{Total results} - \text{Errors}}{\text{Total results}} \times 100 \qquad (4)$$

$$\text{Accuracy} = \frac{52-12}{52} \times 100 \approx 77\%$$

The developed Bayesian network performed with admirable accuracy considering the difficulties encountered with the unavailability of clinical data required for building the Bayesian network structure and estimating the conditional probabilities of the network. On analysis of the results obtained from the Ebola Bayesian network, 64% of the users were positively diagnosed for EVD and 36% were negatively diagnosed. Out of the users positively diagnosed, 24 of them had probabilities greater than 90% and for those negatively diagnosed, only 6 had probabilities greater than 90%. The Ebola Bayesian network displayed a tendency toward positive diagnosis; this was for two major reasons: the estimated conditional probability distribution was not correctly representing the qualitative truth of the problem domain and the dataset was biased towards the positive diagnosis of EVD. Figure 7 shows login for all users and contain fields for email and password to be inserted. This password will give access to the questionnaire, which is the main page. It only gives access to people who already have their email registered initially from the sign up. If the user doesn't have a registered email, then the user has the option to click on sign up to begin registration.

The home page is where the questions are being displayed for the user to answer by clicking in the boxes. The page comprises of five (5) sections of questions which are to be answered before the result is being computed and displayed to the user. Figure 8 is the fifth phase of questionnaire; it allows the clinical question and answer section to continue. It must be completed before submission by clicking the finish button; it is then directed to result page.

Testing on the web responsiveness was carried out using third party software (Webserver Stress Tool version 8.0) and the result is shown in Figure 9. Server response time was measured in milliseconds against steadily increasing user clicks. Response time increased greatly at 34 users, with response a time of 800ms. This showed the expected degradation in performance with increase in user activity.

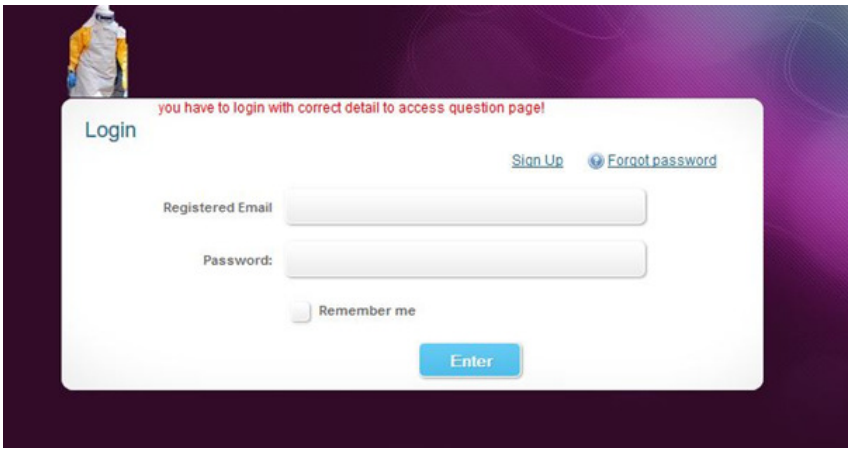

Figure 7: Web Login Interface

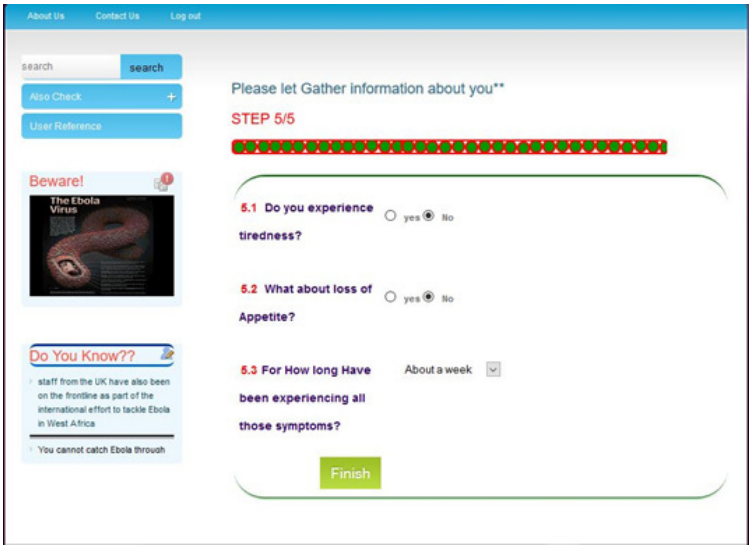

Figure 8: Fifth Phase of Questionnaire

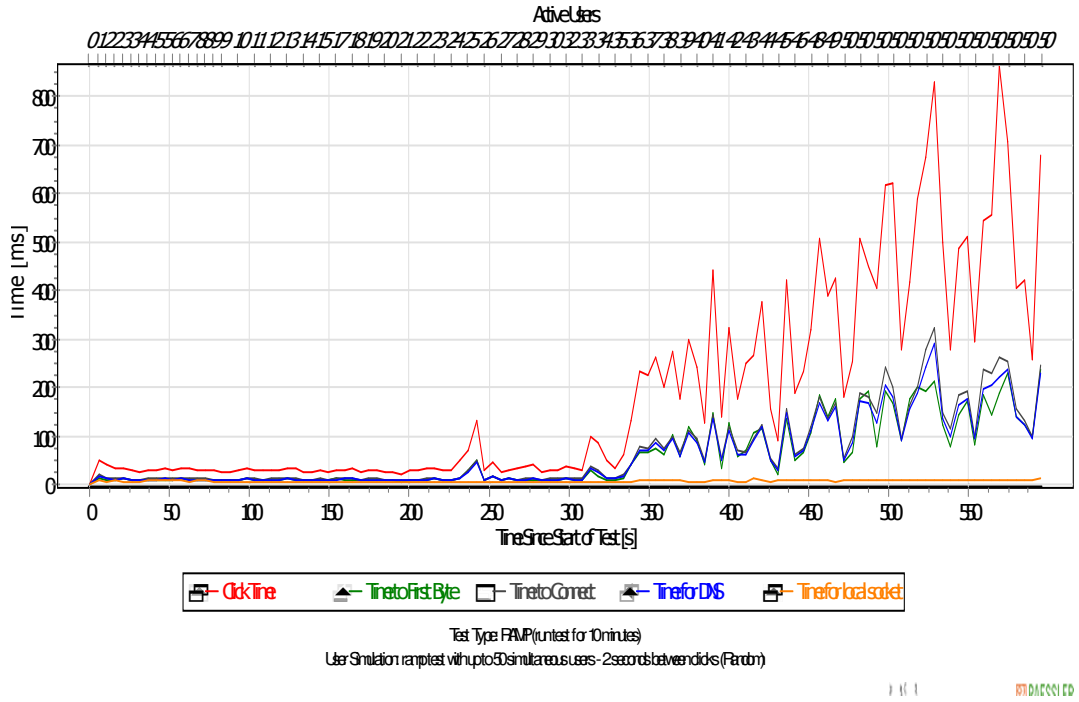

Figure 9: Time of Response for the web Interface to User Clicks

6. CONCLUSION AND FUTURE WORK

In this work, the development of a web-based clinical decision support for diagnosis of EVD was presented. Bayesian networks provide a framework for the representation of medical diagnostic procedure. The expert knowledge and literature helped build the network structure and also the parameters for the network. The network was then deployed on a web-server and was accessible via the web (http) protocol on personal computer systems and mobile phones. The developed Bayesian network was tested with 52 cases and had a diagnostic accuracy of 76.92% when compared to the WHO algorithm for EVD detection.

It is our opinion that the use of this tool in tandem with rapid diagnostic kit, and improved health care infrastructure will go a long way at containment of future Ebola outbreaks.

The following are future work to be carried out:

a) Evaluation and testing with real patient data in a clinical setting for further verification and optimization of system.
b) Addition of other commonly diagnosed diseases like Typhoid and Malaria, to improve the usefulness of the tool.

Acknowledgement
The authors appreciate the generous assistance of Dr Abdulsalam Abdulfatai of the University Health Service, Federal University of Technology, Minna Nigeria for his medical expert advice and guidance.

REFERENCES

1. Abu-Naser, S., El-Hissi, H., Abu-Rass, M., & El-khozondar, N. (2010). An Expert System for Endocrine Diagnosis and Treatments Using JESS. *Journal of Artificial Intelligence, 3*(4), 239-251.
2. Algarin, A. D. (2011). *Clinical Decision Support Systems in Biomedical Informatics and their Limitations.*
3. Angeli, C. (2010). Diagnostic Expert Systems: From Expert's Knowledge to Real-Time Systems. (Sajjar, & Akerkar, Eds.) *Advanced Knowledge Based Systems: Model, Applications & Research, 1*, 50-73.
4. Ben-Gal, I., Ruggeri, F., & Kennett, R. (2007). Bayesian Networks. In *Encyclopedia of Statistics in Quality & Reliability.* Wiley & Sons.
5. Borgohain, R., & Sanyal, S. (2012). Rule Based Expert System for Cerebral Palsy Diagnosis. 1-4.
6. Brown, C. E. (1993). Introduction to Artificial Intelligence and Expert Systems. Oregon State University.
7. Centers for Disease Control and Prevention. (2015). Identify, Isolate, Inform: Emergency Department Evaluation and Management of Patients Under Investigation for Ebola Virus Disease.
8. Conejar, R. J., & Kim, H.-K. (2014). A Medical Decision Support System (DSS) for Ubiquitous Healthcare Diagnosis System. *International Journal of Software Engineering and Its Applications*, 237-244.
9. Croskerry, P. (2003). The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them. *Academic Medicine, 78*(8), 775-780.
10. Fred, A., Filipe, J., Partinen, M., & Paiva, T. (2000). PSG-Expert - An Expert System for the Diagnosis of Sleep Disorder. *Studies in Health Technology and Informatics*, 127-147.
11. Hassan, M. A., Sher-E-Alam, K. M., & Chowdhury, A. R. (2010, June). Human Disease Diagnosis Using a Fuzzy Expert System. *Journal of Computing*, 66-70.
12. Horny, M. (2014). *Bayesian Networks.* Boston University School of Public Health, Department of Health Policy & Management.
13. Kononenko, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. *Artificial Intelligence in Medicine, 23*(1), 89-109.
14. Kosarzycki, B. (2011). *Design and Implementation of a Computer Systems Diagnosis Tool Using Bayesian Networks.* Poznan University of Technology, Computer Science. Poznan: Poznan University of Technology. Retrieved December 18, 2015, from http://sirius.cs.put.poznan.pl/~inf80080/MastersThesisKosarzyckiINF80080_DiagnosisToolUsing BayesianNetworks.pdf
15. Kucharski, A., & Piot, P. (2014, Semptember 11). Cointaining Ebola virus infection in West Africa. *Europe's journal on infectious disease epidemiology, prevention and control, 19*(36). Retrieved December 9, 2015, from http://www.eurosurveillance.org/ ViewArticle.aspx?ArticleId=20899
16. Levine, A. C., Shetty, P. P., Burbach, R., Cheemalapati, S., Glavis-Bloom, J., Wiskel, T., . . . Kessely, T. (2015). Derivation and Internal Validation of the Ebola Prediction Score for Risk Stratification of Patients With Suspected Ebola Virus Disease. *Ànnals of Emergency Medicine*.
17. Lin, J.-H., & Haug, P. J. (2008). Exploiting missing clinical data in bayesian network modelling for predicting medical problems. *Journal of Biomedical Informatics*, 1-14.
18. Lorente, J. A., Blanch, L., & Esteban, A. (2015). Ebola Virus: Understanding the 2014 Outbreak. *Archivos de Bronconeumologia*, 59-60.
19. Lucas, P. (2001). Bayesian Networks in Medicine: a Model-based Approach to Medical Decision Making.
20. Matsumoto, S., Carvalho, R. N., Ladeira, M., Costa, P., Santos, L., Silva, D., . . . Machado, E. (2011, January). *UnBBayes: a java Framework for Probabilistic Models in AI.* Retrieved November 11, 2015, from http://www.researchgate.net/publication/229150811
21. McCormack, J., Hudson, B., & Hoare, K. (2012). Diagnosis in primary care: probabilistic reasoning. *Of Primary Health Care*, (p. 166).
22. Milho, I., & Fred, A. (2000). A User-Friendly Development Tool for Medical Diagnosis Based on Bayesian Network. 1-5.

23. Murphy, K. (1998). A Brief Introduction to Graphical Models and Bayesian Networks.
24. Musa, E. O., Adedire, E., Adeoye, O., Adewuyi, P., Waziri, N., Nguku, P., . . . Olayinka, A. (2015). Epidemiological Profile of the Ebola virus disease outbreak in Nigeria, July-September 2014. *The PanAfrican Medical Journal, 22*. doi:10.11604/pamj.2015.21.331.5834
25. Nee, O., & Hein, A. (2010). Clinical Decision Support with Guidelines and Bayesian Networks. In G. Devlin (Ed.), *Decision Support Systems Advances in.* InTech.
26. Oguntimilehin, A., Adetunmbi, A. O., & Olatunji, K. A. (2014, June). A Machine Learning Based Clinical Decision Support System for Diagnosis and Treatment of Typhoid Fever. *International Journal of Advanced Research in Computer Science and Software Engineering, 4*(6), 961-969.
27. Oguntimilehin, A., Adetunmbi, A., & Abiola. (2013, July). Machine Learning Approach to Clinical Diagnosis of Typhoid Fever. *International Journal of Computer and Infromation Technology, 02*(04), 671-676.
28. Omole, O. I., Rawlings, O. I., & Festus, I. (2013). Classification of Mistakes in Patient Care in Nigerian Hospital. *Global Journal of Medical Research Interdsciplinary, 13*(3), 34-42.
29. Onisko, A., Druzdzel, M. J., & Wasyluk, H. (1999). A Bayesian Network Model for Diagnosis of Liver Disorders. *Eleventh Conference on Biocybernetics and Biomedical Engineering*, (pp. 842-846). Warsaw.
30. Onisko, A., Lucas, P., & Druzdzel, M. J. (2001). Comparison of Rule-Based and Bayesian Network Approaches in Medical Diagnostic Systems. 283-292.
31. Onwuakor, C. E. (2014). Ebola Outbreak in Africa: Current issues - A Mini Review. *American Journal of Medical Case Reports*, 184-186.
32. Pearl, J. (1988). *Probabilistic reasoning systems in intelligent systems: Networks of plausible inference.* San Francisco: Morgan Kaufmann Pulbishers Incorporated.
33. Prasadl, B. C., Krishna Prasad, P. E., & Sagar, Y. (2011). An approach to develop expert systems in medical diagnosis using machine learning algorithms(asthma) and a performance study. *International journal on soft computing*, 26-33.
34. Russel, S. J., & Norvig, P. (2003). *Artificial Intelligence - A Modern Approach.* Pearson Education Inc.
35. Sajda, P. (2006, April 7). Machine Learning for Detection and Diagnosis of Disease. *The Annual Review of Biomedical Engineering*, 8.1-8.29. doi:doi: 10.1146/ annurev.bioeng.8.061505.095802
36. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning from Theory to Algorithms.* Cambridge University Press.
37. Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., & Mylonakis, E. (2015, March 9). Modeling the 2014 Ebola Virus Epidemic - Agent-Based Simulations, Temporal Analysis and Future Predictions for Liberia and Sierra Leone. *PLOS Currents Outbreaks, I*. doi:10.1371/currents.outbreaks.8d5984114855fc425e699e1a18cdc6c9
38. Sun, Y., Tang, Y., Ding, S., Lv, S., & Cui, Y. (2011). Diagnose the mild cognitive impairment by constructing Bayesian network with missing data. *Expert Systems with Applications*, 442-449.
39. Thirumuruganathan, S. (2010). *Building Bayesian Network Based Expert Systems from Rules.* Master Science Degree in Computer Dissertation, University of Texas at Arlington, Graduate School, Arlington. Retrieved 2015
40. Tunmibi, S., Adeniji, O., Aregbesola, A., & Dasylva, A. (2013). A Rule Based Expert System for Diagnosis of Fever. *International Journal of Advanced Research, 1*(7), 343-348.
41. Van der Gaag, L. C., Renooij, S., Witterman, C. L., Aleman, B. M., & Taal, B. G. (2002). Probabilities for probabilistic network: a case study in oesophageal cancer. *Artificial Intelligence in Medicine, 25*, 123-148. Retrieved December 18, 2015, from http://www.socsci.ru.nl/ciliaw/Site%20Cilia/aimedicine.pdf
42. Vomlel, J. (2005). *Some applications of Bayesian networks.* Retrieved July 7, 2015, from www.utia.cas.cz/vomlel
43. Webb, K. D., & Andrew, R. (2011). *Statistical Pattern Recognition Third Edition.* John Wiley and Sons.
44. Wiegerinck, W., Kappen, B., & Burgers, W. (2010). Bayesian Networks for Expert Systems, Theory and Practical Applications. 1-33.

45. Williams, S. O., & Olatunji, O. M. (2013). Hybrid Intelligent System for the Diagnosis of Typhoid Fever. *Journal of Computer Engineering & Information Technology, 2*(2), 2-9.
46. Wiriyasuttiwong, W., & Narkbuakaew, W. (2009). Medical Knowledge-Based System for Diagnosis from Symptons and Signs. *International Journal of Applied Biomedical Engineering, 2*(1), 54-59.
47. Wong, M. L., & Guan, Y. Y. (2008). Learning bayesian networks from incomplete databases using a novel evolutionary algorithm. *Decision support systems*, 368-383.
48. World Health Organisation. (2014, March). *Clinical management of patients with viral haemorrhagic fever.*
49. World Health Organization. (2014). *Clinical management of patients with viral haemorrhagic fever: a pocket guide for the front-line health worker.* Retrieved from World Health Organization: http://apps.who.int/iris/bitstream/10665/130883/2/WHO_HSE_PED_AIP_14.05.pdf.
50. Wyatt, J., & Spiegelhalter, D. (1991). Field trials of medical decision-aids: potential problems. *Proceedings of the 15th Symposium on Computer Applications in Medical*, (pp. 3-8). Washington.

Biography of Authors

Eustace M. Dogo had his BSc and MEng in Electrical Engineering in 1996 and 1998 respectively from Saint Petersburg State Polytechnical University, Russia. He has working experience as an Information Technology expert in areas of Network Administration and Network Security Development & Deployment in Financial Institutions in Nigeria, as well as a Field Engineer with Schlumberger in Nigeria and Europe. He is now with the Department of Computer Engineering, Federal University of Technology, Minna, Nigeria. His research interest includes Wireless Sensor Networks, Artificial Intelligence, Computing Network and Cloud Computing.

Jonathan G. Kolo received his PhD from the University of Nottingham (Malaysia Campus) in 2013. He is currently a Senior Lecturer at Federal University of Technology Minna, Nigeria. His research interests are in the fields of signal processing, data compression, embedded systems, intelligent systems, wireless sensor networks and engineering education. He has published several journal and conference papers as well as chapters of books in these areas.

Odafe Aror is an undergraduate in the Department of Computer Engineering of the School of Engineering & Engineering Technology, Federal University of Technology Minna, Niger State, Nigeria. His area of interest is in Artificial Intelligence and Java programing.

Ayo T. Rahman is an undergraduate in the Department of Computer Engineering of the School of Engineering & Engineering Technology, Federal University of Technology Minna, Niger State, Nigeria. His area of interest is in Web design, DBMS and Computer networking.