

University of New Mexico



# **An Improved Framework for Diagnosing Confusable Diseases Using Neutrosophic Based Neural Network**

Okpako Abugor Ejaita

University of Port Harcourt, Choba Rivers State, Nigeria. okpako.ejaita@gmail.com Asagba P.O.

University of Port Harcourt, Choba Rivers State, Nigeria. lopito2013@yahoo.com

#### ABSTRACT

The two major motivations in medical science are to prevent and diagnose diseases. Diagnosis of disease 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 disastruous, and such, adequate care must be ensured. 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. In this research, we present the analysis of the existing systems and finally present a framework for the diagnosis of confusable disease using neutrosophic-based neural network.

### **General Terms**

Clinical Decision Support System, Medical Diagnosis, Machine Learning, Soft Computing.

### Keywords

Decision Support System, Medical Uncertainties, Neutrosophic Logic, Confusable Diseases

### 1. INTRODUCTION

Decision making in medical science is unique and is quite different from other science disciplines since it is a known fact that scientists tend to look for typical, normal phenomena while medical sciences look out for the atypical, abnormal, morbid phenomena. Medical decision making is a collaborative process between Physicians, Patients, and lab technologists typically through exchange of information that would ultimately guide the physician to make appropriate and proper therapeutic recommendations. There is an exponential amount of data generated daily in the medical domain thereby opening doors for all forms of uncertainties such as incompleteness of information, inconsistent description of disease symptoms, overlapping diseases symptoms, just to mention a few and has led to difficulties in properly diagnosing diseases in such situations. Medical uncertainty is an inherent phenomenon in medical science; it is what fuels medical research, prompts patients to seek medical attention

and stimulate medical intervention notwithstanding, it poses challenges in diagnostic decision making. In recent times, the negative effect of medical uncertainties has attracted attention due to the emerging realities of this period in medical sciences where evidence based, shared decision making and patient-centered care has brought to fore the limitation of scientific knowledge. The effect of uncertainties in the medical domain has been acknowledged by researchers since the 1950's when the sociologist Renee Fox conducted a seminal studies documenting how physician struggle with uncertainty during their trainings. Brause (2001) highlighted that almost all the physicians are confronted during their formative years by the task of learning to diagnose. Central to good diagnosis, is the ability of an experienced physician to know what symptoms or vitals to throw away and what to keep in the diagnostic process.

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. Sisson et al (2007), 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.

A pertinent question would be "how computers have helped in medical diagnosis?" and "how can we improve on the existing systems". Computers have been employed widely in the medical sector in recent time, from local and global patient and medicine databases to emergency networks, or as digital archives. Meanwhile, in the case of medical diagnosis, due to the complexity of the task, it has not been realistic to expect a

fully automatic, computer-based, medical diagnosis system. However, recent advances in the field of intelligent systems are materializing into a wider usage of computers, armed with Artificial Intelligence (AI) techniques. It is therefore imperative to have a decision support to assist in the diagnostic decision making. A decision support system in this context is a computer based information system that supports medical staff in diagnostic decision making. A properly designed medical decision support systems is interactive software whose intent is to help medical practitioners to semantically sieve through a deluge of raw data in order to identify and solve medical problems.

In the purview of computing, decision making in medical diagnosis is all about problem solving strategies which is done by taking potential candidate solutions from the possibilities of various solutions. But often times one is faced with the problem of how to choose from the abundant alternatives that have confusing or conflicting symptoms. If physician's premises are wrong, then the final decision is also wrong which ultimately leads to cases of misdiagnosis whose cost is obvious. It is pertinent to note that we can successfully select the numbers of features that would optimally help in the diagnostic process but as to what values this features can have, which areas needs further probing cannot be empirically ascertained. Medical uncertainties come in different flavors and shapes, but its impact which comes along the lines of class overlap or confusable symptoms is of interest to us. It has continually affected the diagnostic decision of diseases which have ultimately led to performance degradation amidst the supposedly high percentage of accuracy of some re-known classifiers mostly when considered in relation to practical implementation in medical domain. The complexity of the management of low prevalent diseases in the midst of high prevalent ones is to a larger extent attributed to the fact that other diseases have signs and symptoms that are similar to those presented by patients of low prevalent ones. For example Typhoid which is highly prevalent in the Niger Delta region of Nigeria and Hepatitis disease which is low prevalent have some common symptoms and sometimes could be very confusing to novice practitioners and patients in rural areas to diagnose correctly and as such as in most cases would overly conclude it for Typhoid. It should also be noted that in medical decision making, different types of misclassifications or misdiagnosis have different costs. For example, in Hepatitis diagnosis, a false positive decision translates into an unnecessary biomarkers test or liver biopsy which is associated with both emotional, financial cost and other inherent complications. False negative decision on the other hand, however, means a missed Hepatitis-positive which in turn can be deadly.

Medical diagnosis must therefore take into consideration issues of uncertainty and class imbalance which comes either in form of confusability or overlaps ,incomplete information, vagueness, inconsistency or indeterminacy, disease prevalence in order to make a reliable decision towards the prediction and eventual treatment of a disease. Neutrosophic logic is a new logic which is an extended and general

framework to measure the truth, indeterminacy, and falsehoodness of the information and as such suitable for handling issues of uncertainties thus giving fair estimate about the reliability of information. This research work proposes a framework that uses the tripartite membership power of Neutrosophic logic and combining it with the conventional Neural Networks in order to estimate a confusability measurement for two confusable diseases resulting from class overlap in lieu of providing an innovative approach that might be useful to support decisions about medical diagnoses for confusable diseases.

### 2.0 RELATED LITERATURE

Evans and Gadd [3], describe four different levels into which clinical knowledge is organized in a medical problem solving context. They stated that **Observations** are units of information that are recognized as potentially relevant in a problem solving context, however they do not constitute clinically useful facts. **Findings** are observations that have potential clinical significance (e.g. symptoms). **Facets** are clusters of findings that are suggestive of pre-diagnostic interpretations while **clinical diagnosis** is the level of classification that encompasses and explains all levels beneath it. The model is hierarchical with facets and diagnoses serving to establish a context in which observations and findings are interpreted, and to provide a basis for anticipating and searching for confirming or discriminating findings.

Oguntimelehin et al [17] opined that medical diagnosis is simply the task of categorization which allows physician to make predictions using clinical situations and to determine appropriate cause of action. They said it is a complex decision process that involves a lot of vagueness and uncertainty management especially when the disease has multiple symptoms. Diagnosis has been seen generally as the identification of the nature and cause of a certain phenomenon. Several disciplines make use of it but we are only considering it in the parlance of medical science and to put it in more simplistic form, it is the answer to the question of whether a system( in this case human body) is malfunctioning or not, and to the process of computing the answer. Expert diagnosis would not be trivialized in this regard, which is majorly based on experience with the system. Using this experience, a mapping is built that efficiently associates the observations to the corresponding diagnoses.

### 2.1: Medical Uncertainties

Mishel[13] defined uncertainty in illness as the inability to determine the meaning of illness-related events. McCormick [11] opines that uncertainty is a component of all illness experiences and it is believed to affect psychosocial adaptation and outcomes of disease and as such high levels of uncertainty are related to high emotional distress, anxiety and depression. Peter Szolovits [19] opines that "Uncertainty is the central, critical fact about medical reasoning. Patients cannot describe exactly what has happened to them or how they feel, doctors and nurses cannot tell exactly what they

observe, laboratories report results only with some degree of error, physiologists do not understand precisely how the human body works, medical researchers cannot precisely characterize how diseases alter the normal functioning of the body, pharmacologists do not fully understand the mechanisms accounting for the effectiveness of drugs, and no one can precisely determine one's prognosis". Paul et al(2011) opine that irrespective of the visible negative effect of uncertainty in various domain and most importantly to the medical domain, there is limited comprehensible way of addressing the problems it poses in relation to layperson, physicians and patients and health policy makers. According to Smithson [26] this knowledge gaps reflect limitations in empirical evidence; however, a more fundamental problem is the absence of a shared concept of uncertainty, and a lack of integration of insights from different disciplines. Uncertainty is not a monolithic phenomenon and such in considering it, the varied meanings and synonyms should also be considered. Bammer et al [1] opined that there are multiple varieties of uncertainty, which may have distinct psychological effects and thus warrant different courses of action, thus there is, need to have an organized conceptual framework that categorizes these multiple varieties of uncertainty in a coherent, useful way.

### 2.2: Confusable Diseases

This research work pointed out the serious effect of uncertainty, yet how it affect medical diagnosis needs to be elucidated. 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. Fries et al.[5] opined that 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. Joop [8] opined that 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.

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

### 2.3 Clinical Decision and Support Systems

In literature, many researchers have given their definitions of Clinical Decision Support Systems (CDDS). Musen [15] defined a CDSS as any piece of software that takes information about a clinical situation as inputs and that produces inferences as outputs that can assist practitioners in their decision making and that would be judged as "intelligent" by the program's users. Miller and Geissbuhler [12] defined a CDSS as a computer-based algorithm that assists a clinician with one or more component steps of the diagnostic process. Sim et al [22] defined CDSS as a software that is designed to be a direct aid to clinical decision-making, in which the characteristics of an individual patient are matched to a computerized clinical knowledge base and patient specific assessments or recommendations are then presented to the clinician or the patient for a decision. In more recent studies, researchers have been trying to classify CDSSs in the literature so as to provide a holistic picture of CDSSs. For example, Berlin et al [2] did research on CDSS taxonomy to describe the technical, workflow, and contextual characteristics of CDSSs, and the research results are very useful for researchers to have a comprehensive understanding of various designs and functions of CDSSs.

A general model of all clinical and decision support system is shown in Fig 2.1. the interaction is simple: A patient clinical signs and symptoms or lab tests is fed into the system having the inference mechanism component which in turn in consultation with the knowledge base proffer a diagnostic and therapeutic recommendation to the doctor who in turn advise the patient accordingly.

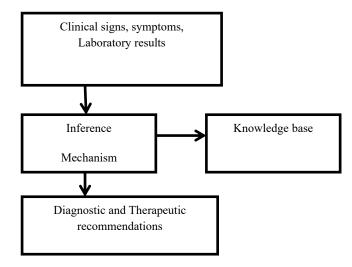


Fig. 2.1: A general model of CDSSs (Source: Lincoln 1999, Reggia 1983)

### 2.4: Neutrosophic Logic

Neutrosophic Logic represents an alternative to the existing logics as a mathematical model of uncertainty, vagueness, ambiguity, imprecision, undefined, unknown, incompleteness, inconsistency, redundancy, contradiction. It is a non-classical logic. It is a logic in which each proposition is estimated to have the percentage of truth in a subset T, the percentage of indeterminacy in a subset I, and the percentage of falsity in a subset F, where T, I, F are defined above, is called Neutrosophic Logic.

A neutrosophic set A in X is characterized by a truth membership function TA, a indeterminacy- membership function IA and a falsity-membership function FA. TA(x), IA(x) and FA(x) are real standard or non-standard subsets of]-0,  $1+\lceil$ .

That is

TA:  $X \varepsilon$  ]-0, 1+ [

*IA*:  $X \varepsilon$  ]-0, 1+ [

*FA*:  $X \varepsilon$  ]-0, 1+ [

There is no restriction on the sum of TA(x), IA(x) and FA(x),

$$-0 \le \sup TA(x) + \sup IA(x) + \sup FA(x) \le 3+\dots 2.1$$

### 2.5: Conditional Probabilities

In medical diagnosis, there are many variables that contribute to the diagnostic process of arriving at a particular disease with varied values of the variables which ultimately in most cases leads to some forgivable errors. As good as this may sound, there is a level of tolerable errors that would be associated with every instance of diagnosis of such disease but it is very unrealistic to quantify the errors for all instances of the disease owing to the fact that we would have just a handful of sample data (due to low prevalence) and as such there is going to be many evaluation of the decision variables. In order to accomplish this feat in less time and space, conditional probabilities become handy.

Conditional distributions are one of the key tools in probability theory for reasoning about uncertainty. They specify the distribution of a random variable when the value of another random variable is known (or more generally, when some event is known to be true).

Formally, conditional probability of X = e given Y = d is defined as

Note that this is not defined when the probability of Y = d is 0. The idea of conditional probability extends naturally to the case when the distribution of a random variable is conditioned on several variables.

As for notations, we write P(X|Y = d) to denote the distribution of random variable X when Y = d. We may also write P(X|Y) to denote a set of distributions of X, one for each of the different values that Y can take.

### 3.0: Analysis of Existing Systems

Proper diagnoses and prevention is the major concerns in medical science, it is there imperative to have systems that assist in medical diagnosis with such an accuracy comparable to human physicians. Many existing system have employed different approaches in ameliorating the effect of uncertainties yet there is still room for improvement so as to handle the diagnosis of confusable diseases.

A detailed review and analysis of existing system was carried out in order to bring to fore areas to improve on, in order to tackle the embarrassing effect of confusable disease diagnosis. We reviewed the following:-

- i. The approaches and methods used in the existing system in knowledge construction
- ii. The inference mechanism in handling uncertainties
- iii. Support for diagnostic criteria for reliability of prediction of disease in a two class of diseases diagnosis with confusable symptoms

### **3.1:** Architecture of the Existing systems Using Neural Network

A typical architecture for diagnosis of disease used in existing system using an Artificial Neural Network is shown in Fig 3.1.

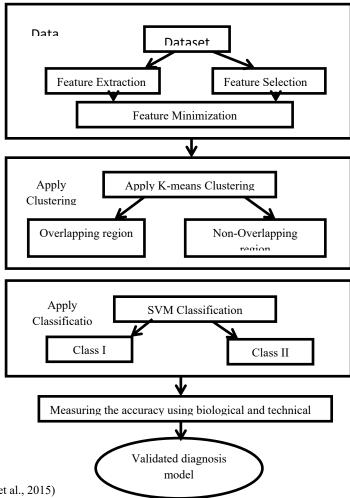


Fig. 3.1: Architecture of the Existing System (Source: Mohammed et al., 2015)

### 3.2 Limitations of the Existing System

The existing system has some limitations which prevent it from having a practically good performance as needed. The salient findings include

- 1. Though some of the existing system ensures multiple belongingness of a particular element to multiple classes with varied degree but capturing the neutralities due to class overlap or confusability which could degrade the prediction performance is missing.
- 2. The existing system is mute or unable to classify instances that falls under overlapping region and as such refers them for further medical probe. This clearly defeats timeliness and quality of service delivery we are seeking for in clinical diagnosis and as such not suitable to handle confusable diseases whose features are overlapped.
- 3. In diagnosing confusability in disease classes, some of the existing system used only unsupervised statistical approach such as k-means to separate the overlapping region from the non-overlapping region. K-means is very poor when it comes to data with serious overlapping; is unable to handle noisy data and outliers as well as not suitable for non-linear data sets. Supervised machine learning using neural network is more suitable for complex nature of biological systems and non-linear data sets.
- 4. There is no reliability or justification metric for the decision of the classification which serves as a diagnostic criterion that allows a disease to be definitely diagnosed or definitely excluded in cases of non-linear decision boundary

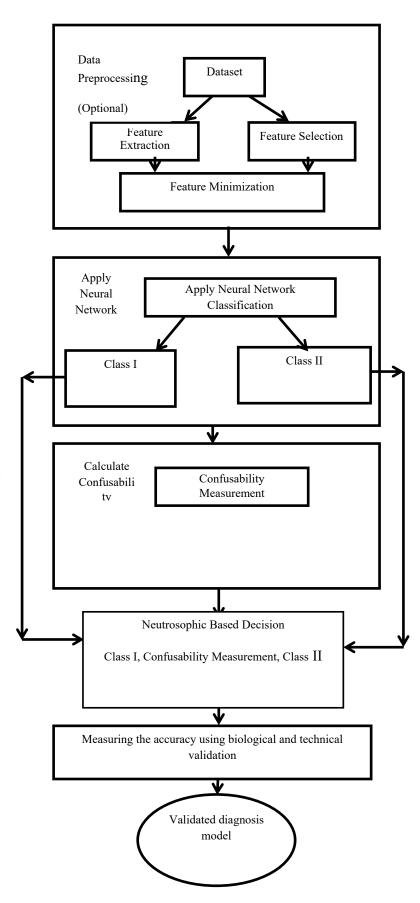


Fig. 3.2: Proposed System Architecture

## **4.1: Brief Description of the Components of the Proposed System**

This section talks about the brief description of the various components in the proposed system architecture.

# 1. Patient Symptoms and Signs Subcomponent

Disease symptoms are the biological indicators which are associated with the clinical presentation of disease as learnt from medical literature and expert physicians. George et al (2000) opine that a symptom is a visible or even a measurable condition indicating the presence of a disease and thus can be regard as an aid towards diagnosis. It is based on this clinical presentation that a doctor or physician makes a tentative judgment about the state of the patient and consequently a test for confirmation.

### 2. Feature Selection Sub Component

It is important to note that the essence of feature selection in this research is to help reduce the dimension of a dataset of features potentially relevant with respect to the diagnosis of the diseases, finding the best minimum subset without transforming the data into a new set. The feature selection process points out all the input features relevant for the diagnosis of the diseases, and it is an indispensable data preprocessing step. The difficulty of extracting the most relevant variables is due mainly to the large dimension of the original feature set, the correlations between inputs which cause redundancy and finally the presence of variables which do not affect the diseases. In this research, we will employ feature selection using genetic algorithm for the feature searching techniques. The genetic algorithm was originally used to select binary string but it has been used been used in recent times to explore the inter-dependencies between the bits in the string, hence the choice of its usage. Singh et al (2016) have successfully used it for feature selection and its performance was superlative.

### 3. Confusability Measurement

There are two components that make up this component-Vagueness and multidimensional interpolation of the errors. The confusability measurement provides information on amount of uncertainty associated with such a classification that would have degraded the performance and is on this basis that final diagnosis is made. Confusability Measurement is 1-|Tm(class I) -Tm(Class II)|, where Tm means the truth membership.

# 4. Inference Engine/Decision support Component

The decision made by the inference system through the neural work is optimized by this component by taking the result of the inference sub-component as input and with the aid of the result confusability measurement, a decision is ultimately made. The supporting components for the confusability measurement are multidimensional interpolation and vagueness calculated from the two networks which objectively influences the result of the proposed system thereby optimizing the practical implementation of the system in regards to sensitivity and specificity in an environment poised with class overlaps.

### 5.0: Conclusion

To make proper, reasonable and appropriate medical decision in the diagnosis of confusable diseases, the knowledge base and the inference mechanism play an indispensable role as they are the heart of clinical decision support systems. Once such clinical decision and support systems are built, we are faced in most times with a large feature set of symptoms which needs to be pruned to improve the performance of the system with regards to accuracy of classification. The key quality in this study is to achieve a better and proper diagnosis of confusable diseases. A genetic algorithm is applied in the feature selection phase. In quantifying the confusability, a multidimensional interpolation of error is plotted in the multidimensional feature space while the vagueness is calculated from the two class Neural Network as |1-(class Aclass B)|, both vagueness and the errors form the confusability measurement. The inference mechanism is also improved by employing the concept of neutrosophic logic thereby having a tripartite membership (Degree of class A, Confusability Measurement, Degree of class B) rather than just two in order to make therapeutic recommendations. With these consideration, it is hope that there is going to be an obvious improvement in the system performance in terms of handling confusability in disease symptomatic presentations and eventually renders a proper diagnosis. Therefore, in this study, the architecture for diagnosing confusable disease was developed using the concept of neutrosophic logic in combination with neural network. This will be able to capture and quantify the confusability in this situation and ultimately being used in the decision making process.

### 6.0: Future Work

In the paper, analysis of the existing systems was carried out and some limitations were highlighted for consideration. The proposed architecture provides an interface where a patient's symptom is captured by the system, the confusability measure is calculated and in consultation with the knowledge base, the inference mechanism makes its therapeutic recommendation to the doctors who in turn advise the patient accordingly. Future work will delve into the implementation procedure of the framework for the diagnosis of confusable diseases using two confusable diseases and the result from the implementation and evaluation will be provided. The interface for the system based on patients' symptoms will also be presented.

### References

- Bammer G, Smithson M and Group G. (2008). The nature of uncertainty. Uncertainty and Risk: Multidisciplinary Perspectives. London: Earthscan; 289–304.
- Berlin A., Sorani M., Sim I.(2006), Journal of Biomedical Informatics 39, 656.
- Evans, D.A., and Gadd, C.S. (1989). Managing coherence and context in medical problem-solving discourse. In D. Evans & V. Patel (Eds.), Cognitive Science in Medicine: Biomedical Modelling, 11-255. MIT Press, Cambridge, Massachusetts
- Fatumo S.A, Emmanuel Adetiba, J.O Onolapo (2013), "Implementation of XpertMaltyph: An Expert System for Medical Diagnosis of the Complication of Malaria and Typhoid ", IOSR Journal of Computer Engineering, 8(5), s34-40, www.iosrjournals.org.
- Fries J.F, Hochberg MC, Medsger TA, et al.(1194)
   Criteria for rheumatic disease. Different types and different functions. The American College of Rheumatology Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 37:454-62.
- George Becks, M. Dotoli, Diechrich craf keyserlink, Jan Jantzen (2000): fuzzy clustering. A versatile means to Explore medical Database, ESIT Aachen Germany.
- 7. John, R.I and Innocent, P.R (2005). Modeling uncertainty in clinical diagnosis using fuzzy logic: Systems, Man and Cybernetics Part B; Cybernetic, IEEE transaction on 35(6),1346-1358
- Joop P van de Merwe (2004)- Design of Criteria for Diagnosis - ESSIC Meeting - Copenhagen 4 June 2004Joop P van de Merwe - Design of Criteria for Diagnosis - ESSIC Meeting - Copenhagen
- Joop, V. M. (2005): Diagnosis and differential diagnosis of Alcoholic liver Diseases.
- Lincoln, M.J (1999), in clinical decision support systems. E.S Berner, Ed. (Springer-Verlag, Newyork) 169-198.
- 11. McCormick, K. M. (2002). A concept analysis of uncertainty in illness. Journal of Nursing Scholarship, 32 (2), 27-131.
- 12. Miller R. A. and Geissbuhler .A (1999): In Clinical Decision Support Systems E. S. Berner, Ed. (Springer-Verlag, New York), 3(34.)
- 13. Mishel, M. H. (1988). Uncertainty in illness. Image: Journal of Nursing Scholarship, 20(4), 225
- 14. Mohammed Abdullah Alghamdi, Sunil G Bhirud and Afshar M. Alam.(2015).Physician's Decision Process for Disease Diagnosis of Overlapping Syndrome in Liver Disease using Soft Computing Model. International Journal of Soft Computing and Engineering (IJSCE) 4(6).

- Musen M. A.(1997) , in Handbook of medical informatics J. H. V. a. M. Bemmel, M. A. , Ed. (Bohn Stafleu Van Loghum, Houten ).
- Njafa J.P. Tchapet, Nana Engo S.G. and Woafo P.(2013), "Quantum Associative Memory for the diagnosis of some tropical diseases", Cornell University, www.arxiv.org, retrieved 30/01/2016.
- 17. Oguntimilehin A, Adetunmbi A.O. and Abiola O.B.(2015), A Review of Predictive Models on Diagnosis and Treatment of Malaria Fever, International Journal of Computer Science and Mobile Computing,4(5).
- 18. Paul Han .KJ, Klein W.M.P, Arora N.K.(2011) Varieties of uncertainty in health care: a conceptual taxonomy. *Med Decis Making*.
- Peter Szolovits (2011) Uncertainty and Decisions in Medical Informatics, MedDecis Making.; 31(6): 828–838.
- 20. Reggia, J (1981), Annals of biomedical Engineering 9,605.
- Singhai M, Rawat V, Singh P, and Goyal R. (2016)
   "Fatal case of concomitant Hepatitis E and Salmonella paratyphi A infection in a sub-Himalayan patient". Ann Trop Med Public Health
- Himalayan patient". Ann Trop Med Public Health
  22. Sim I. et al.(2001), Journal of American Medical Informatics Association, 527
- 23. Smarandache F.(1998). A Unifying Field in Logics. Neutrosophy: Neutrosophic Probability, Set and Logic. Rehoboth: American Research Press.
- Smarandache F.(1999), Linguistic Paradoxists and Tautologies, Libertas Mathematica, University of Texas at Arlington, Vol. XIX,
- Smarandache F., "A Unifying Field in Logics(2002): Neutrosophic Logic, in Multiple-Valued Logic," An International Journal, 8(3), 385-438.
- 26. Smithson M. (1989): Ignorance and Uncertainty: Emerging Paradigms. New York: Springer Verlag
- 27. Uzoka F.M.E, Akinnuwesi B.A, Amoo T, Aladi F, Fashoto S., Olaniyan P, and Osuji J(2016): A framework for early differential diagnosis of tropical confusable diseases using fuzzy cognitive map engine.

Received: May 4, 2017. Accepted: May 22, 2017.