VOL. 4, NO 2, DECEMBER, 2024
ISSN 2756-4444
E-ISSN 2756-4452

SUPERVISED LEARNING TECHNIQUE FOR PREDICTING PROGNOSTIC RISK OF MALARIA AND TYPHOID FEVER

BY

Abiodun, O. Adepoju[1] & Otuniyi, M. Otukelu[2]

[2] Department of Computer Science, D.S Adegbenro ICT Polytechnic, Ogun State, Nigeria
[1] Department of Computer Engineering, D.S Adegbenro ICT Polytechnic, Ogun State, Nigeria
Email: pojubiodun@gmail.com / tuxyt2000@yahoo.com

Abstract

Malaria and typhoid are serious infectious diseases of the digestive system, caused by ingesting food or water contaminated with the bacillus salmonella. The rate at which malaria and typhoid fever kills particularly in West Africa is alarming, and thereby require solution development and knowledge-based decision support. At times, rarely available facilities could be occasioned by inadequate medical personnel among other germane issues that require urgent redress. In this study, supervised learning technique was adopted to improve the prognostic prediction of malaria and typhoid fever. Improved predictive system for malaria and typhoid prognosis was developed using Naïve Bayes model, and experimented with localized clinical dataset using Visual Studio.net and ActiveX Data Object. Naive Bayes approach was adopted as supervised learning to sequence of symptoms for every observation; by analyzing parameter values into a particular diagnostic label being the likely predictor in a training set. Experimental evaluation of the system was measured; giving seventy-two percent (72%) accuracy and eighty-one percent (81%) precision as performance metrics.

Keywords: Malaria, Typhoid, Data Mining, Supervised Learning, Prediction, Model

Introduction

Malaria is such an infectious disease that is caused by the bite of some types of mosquito which brings about fever and shivering. It could be a serious disease of the digestive system, caused by ingesting food or water contaminated with the bacillus salmonella is known as typhoid (Cheng et al., 2016). Knowledge-based diagnostic system is a software system or combination of software and hardware technologies; capable and competent of exerting a specific clinical task usually performed by human medical experts (Ayushi et al., 2017). Computational techniques can be applied to real-time biological situations for solving problems, in accordance with technological changes. Computer assisted analysis and decision modeling tools in the domain of knowledge or specific walk of life and human endeavor have continued to evolve (Lamesgin & Sirajudeen, 2016). Emergence of prognostic systems that possess those properties associated with domain experts like reasoning process, understanding, and critical thinking is now imperative for health services. These are highly specialized computer systems capable of simulating that element of a human specialist's knowledge and reasoning that can be formulated into knowledge chunks characterized by a set of facts and heuristic rules (Ayeni et al., 2020).

An intelligent agent can process or type match and achieve a quality which is likely to be real and similar to human natural intelligence or latent physical skills. The aim of true data representation is to facilitate reasoning. A quality factual representation must be declarative and procedural pattern (Cai et al., 2020). Regardless of cognitive emulation, complexity for human entities in diagnostic reasoning is a major challenge in the implementation of computer-based programs for diagnosis and prognosis support. Extended view unveils that diagnosis and prognosis still remains the major application of intelligent agents and expert systems. Therefore, expert system development and application has served the dual purpose of diagnosis and treatment (Singla et al. 2014). Public health of people is a direct index of development of a nation which affects the productivity of potential workforce and allocation of resources within a family, community and nation (Ayeni, 2020). In West Africa, rarely available health facilities are occasioned by inadequate medical personnel among other germane issues that require urgent redress. Hence, supervised learning technique for improving the prognostic prediction of malaria and typhoid fever, using Naïve Bayes model and localized clinical dataset for experimental analysis with Visual Studio and ActiveX Data Object is the focus of this study.

Related Literatures

Cai et al. (2020) developed intelligent clustering system, which is frequently used to segment database into separate groups of like-points. Its algorithm uses visualization approach by quantization vector that allows sample instance to maintain low-dimensional grid in an organized manner, as k-means technique handles clustering and classification perspective through unsupervised learning method. A global approach to optimization was leveraged upon for partitioning the symptoms of breast cancer; having trained the breast cancer data to build a predictive knowledge base for prognostic systems with accurate classification (Mashhour, 2019). The reported method produced an intelligent solution that determines the segments' focal point; space and period needed for storage and execution yardstick depends on the dimension of problem domain. Rajamani & Rathika (2015) enumerated the essential parameters for regulating fuzzy inference system from graphical user interface; editors for handling input/output elements, membership functions; the rule and to monitor behaviour of specific rules for improving system performance. Fuzzy inference mechanism is specifically applicable to knowledge-based and control system, with components formation of fuzzification, knowledge base, inference rules, defuzzification, and fuzzy logic in intelligent reasoning (Amjad, 2014).

Multi-layer inference system was incorporated in adaptive neuro-fuzzy mechanism for classifying data for liver diseases; though its prediction reliability for illness of the liver was not evaluated (Cheng et al., 2016). Machine learning algorithm had been widely used in the past for disease classification in comparison manner; its projection of hybridized method is expected to produce perfect prediction with effectiveness (Rahmon et al., 2018). Neuro-fuzzy system takes a form of integrated model in computational intelligence, which brings the benefit of neural network and fuzzy logic together through estimation technique to approximate the input parameters (Prem-Pal & Ranjit, 2013). Clustering and classification go together and consequently used interchangeably because clustering is needed for pre-processing, so as to group dataset into distinct classes (Ayeni et al., 2020). Synthetic intelligence in medical practice is importantly associated with the development of man-made entity or technological artifact that matches disease traits with observed symptoms by classification model so as to provide prognosis and therapeutic basis (Ayushi *et al.*, 2017). Support Vector Machine (SVM) is a supervised learning algorithm for classifying and creating the extent of relationship in predictive situation. Meanwhile, Naive Bayes classifier can produce proper classification with a desired and improved performance when use on medical dataset for liver disease (Ayeni, 2020).

Methodology

Supervised learning technique of data mining and predictive model namely: Naïve Bayes classifier was employed in the analysis and determination of prognostic risk of malaria and typhoid fever for clinical decision support. Datasets were obtained from localized repository of tertiary health care in Nigeria, through selected teaching hospitals in South West, Nigeria. These datasets splits into two: sixty percent (60%) and forty percent (40%) for training and testing respectively. The characteristics of the datasets are presented in table 1, and experimental samples in table 2.

Table 1

Datasets Characteristics

Features	**Dataset**	
	Before Pre--Processing	**After Pre-processing**
Dataset Characteristics	Multivariate	Multivariate

Attribute Characteristics	Text, Nominal, Boolean	Text, Nominal, Boolean
Number of Instances	400	350
Number of Attributes	Four (4)	Four (4)
Associated Tasks	Prediction	Prediction
Missing Values (?)	Yes (50)	Nil

Table 2

Symptom 1	**Symptom 2**	**Risk**	**Prognosis**
High	Low	High	True
High	Mild	Low	True
High	Low	Mild	True
High	Mild	High	True
Mild	Mild	Mild	True
Mild	High	Mild	False
Mild	Low	High	False
Low	Mild	Low	False

Pre-processed extracts of **Observation** Data **for Training Set**

The sequence of symptoms as observation data as shown in table 2, are analyzed by parameter values into a particular prognostic category through likely predictor in training set. The classifier is trained to classify a sequence of nominal values into an appropriate ailment by likely predictor and yield a corresponding output. Accordingly, observed incidence for the predictor with the highest nominal value and likelihood range must be selected or chosen for a probable outcome. The classification algorithm usually refers to the training instances, in order to compute probabilities of every class based on the possibility distribution in the training dataset as described in table 1. In the first instance, when computing the possibility of the diagnosis class TRUE, the classifier will count the number of data cells in which the SYMPTOM 1 scale equals to HIGH when the observation sequence in data cells is classified as TRUE. Only four (4) data cells exist with this yardstick, meanwhile, there are five (5) data cells in which the SYMPTOM 1 equals to or greater than MILD and the data cell is classified as TRUE. Therefore, relative possibility of SYMPTOM 1 equals to HIGH given TRUE equals to 4/5; while the relative possibility of SYMPTOM 1 equals to MILD given TRUE equals to 1/5.

The classification processes in figure 1 computes all the relational probabilities for visual analysis by conversion. Raw data being acquired about observations and given symptoms from clinical records cannot be used directly by the Naive Bayes classifier. Defining the labelled ranges in accordance with data distribution for observation parameters, and within short intervals will improve the accuracy of the classification process and intelligent reasoning. A system prototype was created to experiment with the classification model and to measure its experimental performance, by approximating the numeric data of chosen parameters and converting them on a nominal scale for a likely predictor. Figure 1 shows the architecture of the improved prognostic system for risk determination.

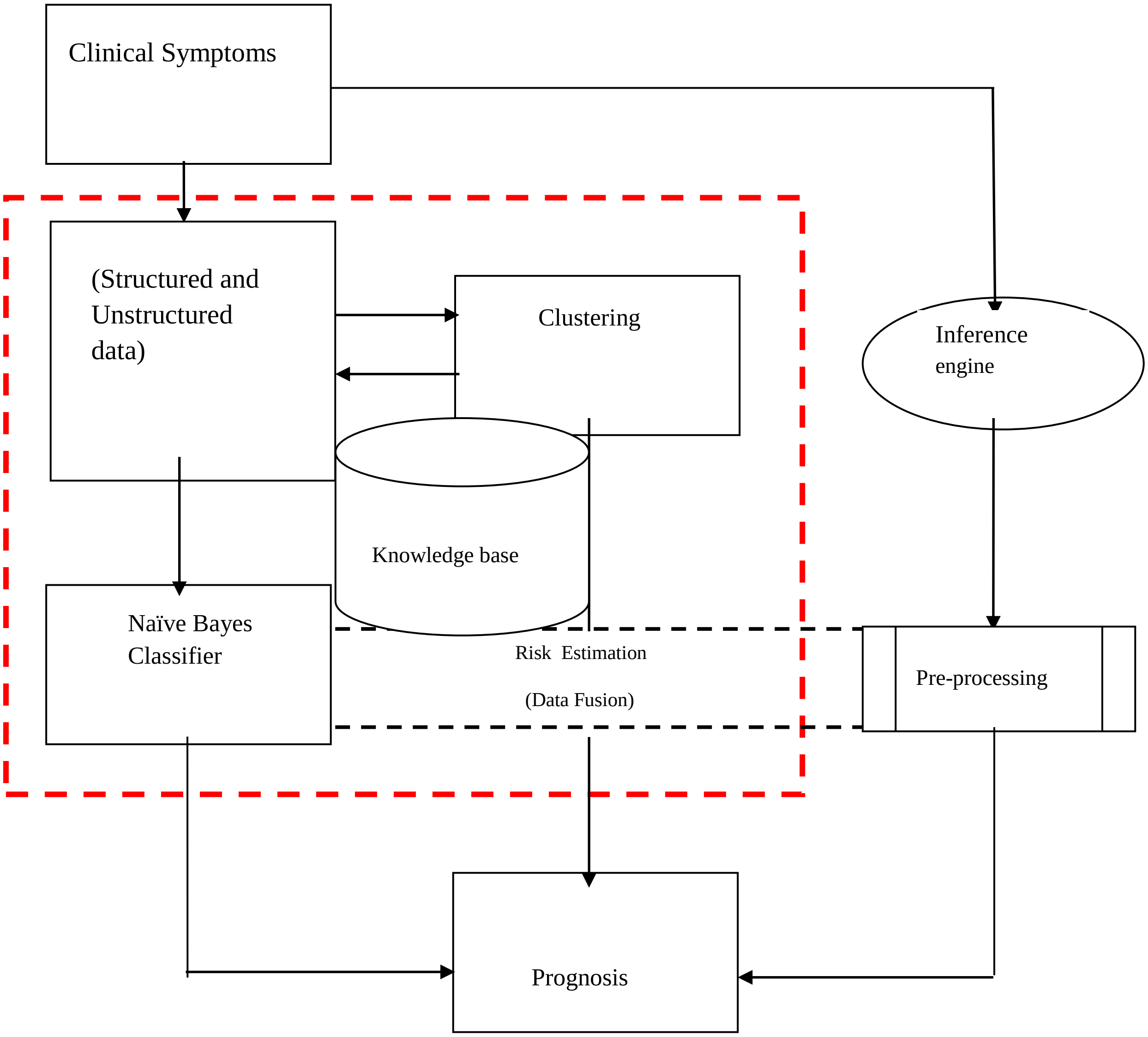
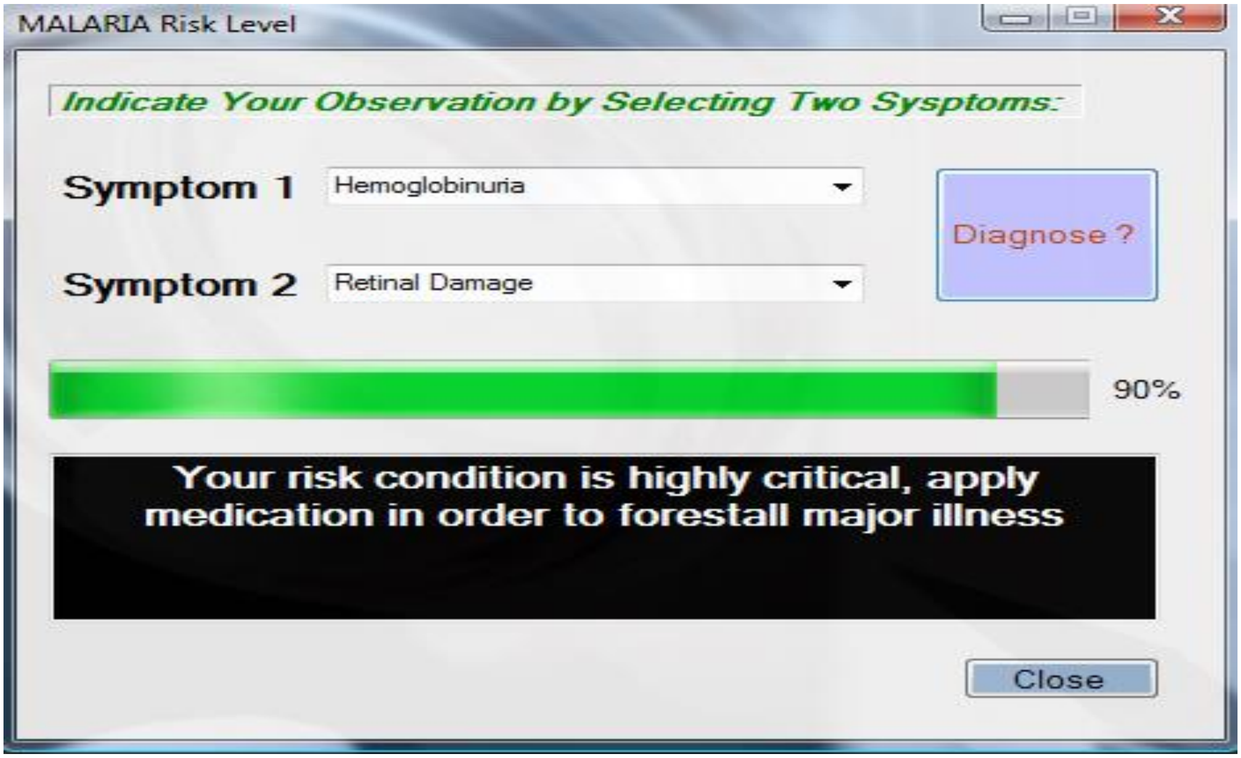


Figure 1: Architecture of the Improved Prognostic System

Results and Discussion

The system was tested against some proven cases of Typhoid and Malaria for experimental analysis, which obtained seventy-two percent (72%) accuracy and eighty-one percent (81%) precision when the symptoms from clinical data are given as input. Figures 2,3,4 and table 3 shows the outcome of valid and invalid data in usability testing as dependent input variables; as well as classifier output in relevance to presence of fever and risk level in this study.

Figure 2: Main Module of Improved Clinical Prognostic System



Indicate Your Observation by Selecting Two Sysptoms:

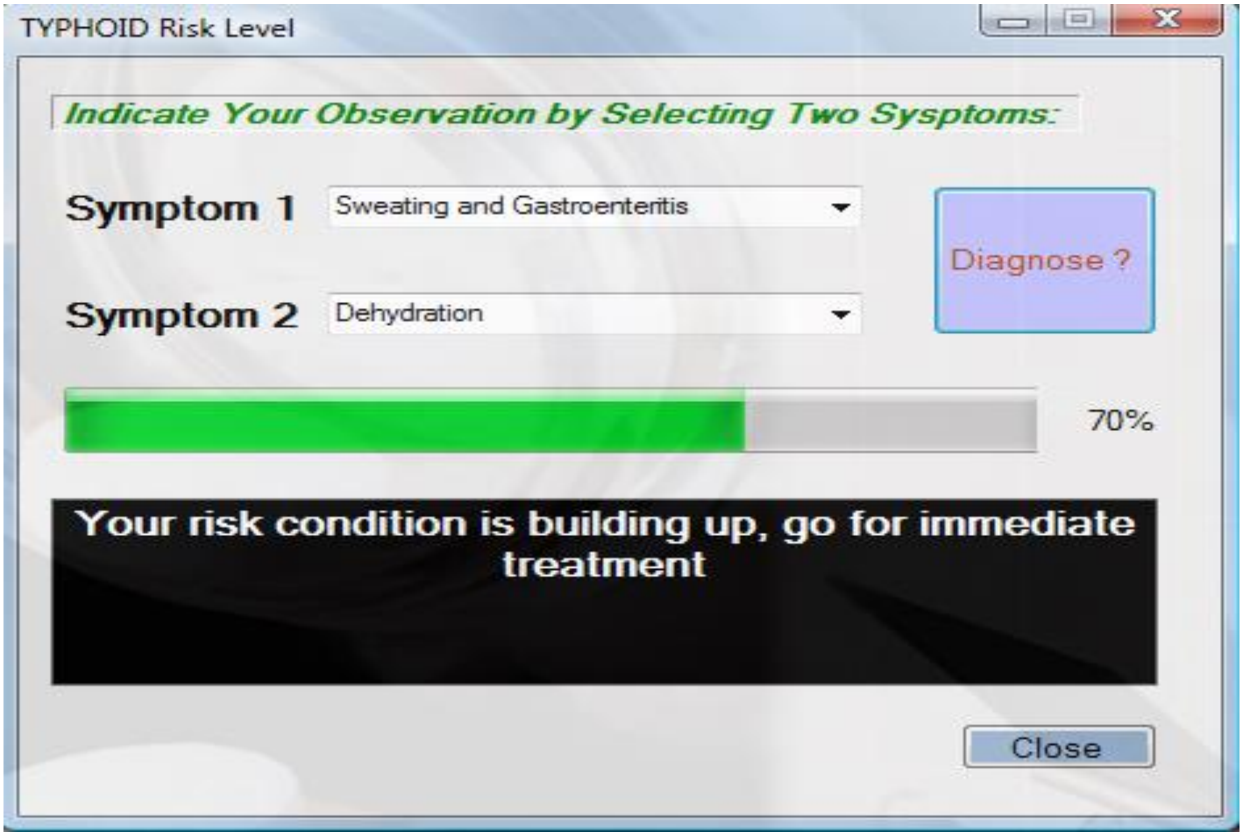
Symptom 1: Hemoglobinuria

Symptom 2: Retinal Damage

90%

Your risk condition is highly critical, apply medication in order to forestall major illness

Figure 3: Predicting Malaria Risk during Prognosis



Indicate Your Observation by Selecting Two Sysptoms:

Symptom 1: Sweating and Gastroenteritis

Symptom 2: Dehydration

70%

Your risk condition is building up, go for immediate treatment

Figure 4: Predicting Typhoid Risk during Prognosis

Thus, posterior probability is computed by classification algorithm which also resides on middleware module, and the possibility estimation of malaria fever or otherwise is returned to output terminal, unlike Rahmon et al. (2018) which does not distinguish between malaria and typhoid bacteria for operational efficacy of detection system. Consequently, dynamic scenarios were established to ensure confidence in the prediction framework to discover relationship likelihood between rare symptoms of Pretyohid and Haemoglobirua with highest risk level; thus emanating as improvement to the work of Cai et al. (2020) that has static knowledge base for observable symptoms.

Table 3

Experimental Results of Supervised Learning Model for Prognosis

Risk (%)	Gender	Symptom 1	Symptom 2	Ailment Type
65	Female	Headache	Burning sensate	Typhoid
62	Male	Shivering	Sweating	Malaria
62	Male	Vomiting	Dehydration	Typhoid
58	Male	Athragra	Yellowish	Typhoid
72	Male	Convulsion	Retinal hitch	Typhoid
90	Male	Haemoglobirua	Pretyohid	Malaria

26	Female	Jaundice	Anemia	Malaria
29	Female	Anemia	Jaundice	Malaria
17	Male	Hemolytic	Haemoglobirua	Malaria
55	Male	Retinal hitch	Convulsion	Typhoid
57	Male	Yellowish	Athragra	Typhoid
72	Male	Dehydration	Vomiting	Typhoid
64	Male	Sweating	Shivering	Malaria
74	Female	Burning sensate	Headache	Malaria

Conclusion

Performance evaluation was carried out with the use of localized samples which were provided as required input and experimental data. Having acquired clinical datasets from local source as explained in the preceding section. Experimental analysis involves valid and invalid data in order to ascertain the accuracy of prognostic model in responding to observable symptoms and reasoning precision. An improved prognostic model which was provided in this study was trained and validated with data discretization. The results from experimental performance showed that classifier output is in close range.

References

Ayeni, G.A (2020). A Clinical Prognostic Framework for Classifying Severe Liver Disorders (SLDs) and Lungs' Vulnerability to Virus. *AIJR Preprints*, 212, Version 1, Accessible at www.preprints.aijr.org.

Ayeni, G.A, Ajagbe, T.T & Yekinni, W.A (2020). Electrocardiogram (ECG) Signal and Interference Filtering for Clinical Diagnostic Support. *International Journal of Research and Scientific Innovation (IJRSI)*. *7*(10), 53-56.

Ayushi, K., Sharma, A., Sharma, V. & Gupta, R. (2017). Disease Prognosis Based on Clinical Parameters using Machine Learning Approach. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 5(6), 13-20.

Amjad, M.M (2014). Naïve Bayes Classifier for Disease Detection using Smart Sensors. M.Sc Thesis: University of Agder, Norway.

Cheng, K.C, Lin, W.Y, Liu, C.S, & Lai, S.W (2016). Association of Different Types of Fever Diseases with Demographic and Clinical Factors. *Journal of BioMedicine*, 6(3), 16-22. DOI 10.7603/s40681-016-0016-2

Cai, Q, Huang, D, Ou P, Yu, H, Zhu, Z, Xia, Z, Su, Y, Ma, Z, Zhang, Y, Li, Z. & Chen J. (2020). COVID-19 in a designated infectious diseases hospital outside Hubei Province, China. *Journal of Pharmacology*, *7*(6), 1–11.

Lamesgin A. & Sirajudeen, .M. (2016). Implementation of an Expert System for Lung Disease Diagnosis. Retrieved from https://www.researchgate.net/303699292 on 2/12/2021.

Mashhour, B.A (2019). Assessment of Liver Function using Hybrid Neuro-Fuzzy Model of Blood Albumin. *International Journal of Healthcare Information System and informatics.* 5(4), 49-59.

Prem-Pal, S. & Ranjit, S. (2013). "Design of a Clinical Decision Support System using cased-based Reasoning". Proceedings of International Multidisciplinary Conference on Research, Developments and Practices in STEAMS. 4, 27-40.

Rahmon, I., Omotosho, O., & Kasali, F. (2018). Diagnosis of Hepatitis using Adaptive Neuro-Fuzzy Inference System (ANFIS). *International Journal of Computer Applications*, 108(38), 45-55.

Rajamani, R. & Rathika, M. (2015). Analysis of Cancer using Adaptive Neuro Fuzzy Inference System. *International Journal of Innovative Research in Computer and Communication Engineering*. 3(7), 7-13.

Singla, J, Dinesh, G. & Abhinav, B. (2014). Medical Expert Systems for Diagnosis of Diseases. *International Journal of Computer Applications*, *3*(7), 75-87