A Web-Based Clinical Decision Support System for the Management of Diabetes Neuropathy using Naïve Bayes Algorithm

Shade KUYORO, Folasade OSISANWO, Monday EZE, and Atisi FRANKLYN

Department of Computer Science, Babcock University, Illishan-Remo, Nigeria.
afolashadeng@gmail.com, osisanwof@gmail.com, franklynatisi@yahoo.com,eze_monday@yahoo.com

Abstract— Diabetes Neuropathy is a chronic health problem with devastating, yet preventable consequences. Due to this shortage of specialists, there is a need for a Clinical Decision Support System that will diagnose and manage diabetes neuropathy. This work therefore aimed at designing a web-based Clinical Decision Support System for the management of early diabetes neuropathy. Four pattern classification algorithms (K-Nearest Neighborhood, Decision Tree, Decision Stump and Rule Induction) were adopted in this work and were evaluated to determine the most suitable algorithm for the clinical decision support system. Datasets were gathered from reliable sources; two teaching hospitals in Nigeria, these were used for the evaluation Benchmarks such as performance, accuracy level, precision, confusion matrices and the models building's speed were used in comparing the generated models. The study showed that Naïve Bayes outperformed all other classifiers with accuracy being 60.50%. K-Nearest Neighborhood, Decision Tree, Decision Stump and Rule induction perform well with the lowest accuracy for x- cross validation being 36.50%. Decision Tree falls behind in accuracy, while k-nearest neighbour and Decision Stump maintain accuracy at equilibrium 41.00%. Therefore, Naïve Bayes is adopted as optimal algorithm in the domain of this study. The rules generated from the optimal algorithm (Naïve Bayes) forms the back-end engine of the Clinical Decision Support System. The web-based clinical decision support system was then designed. The automatic diagnosis of diabetes neuropathy is an important real-world medical problem. Detection of diabetes neuropathy in its early stages is a key for controlling and managing patients early before the disabling effect present. This system can be used to assist medical programs especially in geographically remote areas where expert human diagnosis not possible with an advantage of minimal expenses and faster results. For further studies, researchers can improve on the proposed clinical decision support system by employing more than one efficient algorithm to develop a hybrid system.

Index Terms—Accuracy, Algorithm, Classification, Diabetes, Naïve Bayes, Neuropathy, Precision

———————————————————— ◆ ————————————————————

1 INTRODUCTION

THE world is fast evolving and in order to cope with the insatiable demand of the human race for the kind of living that can be described as top-notch in which people have all they need at their beck and call, there is the need to develop intelligent decision making applications that will drive systems or devices to carry out tasks that require human intelligence. This concept is known as Artificial Intelligence (AI). In science and technology, the desire for improvement is a constant subject which triggers advancements. Technology has changed civilization in many different ways. Humans have always been on a path of progression through the help of technology, the twentieth and twenty-first centuries have seen a number of advancements that revolutionized the way people work, live and play.

A Clinical Decision Support System (CDSS) is an active knowledge system, where two or more items of patient data are used to generate case-specific recommendation(s) (Chen et al, 2002).[1] This implies that a CDSS is a Decision Support System (DSS) that uses knowledge management to achieve clinical advice for patient care based on some number of items of patient data. This helps to ease the job of healthcare practitioners, especially in areas where the number of patients is overwhelming. Clinical decision support system (CDSS) provides clinicians, staff, patients or other individuals with knowledge and person-specific information, intelligently fil-

tered or presented at appropriate times, to enhance health and health care. A CDSS can also be seen as an application that analyses data to help healthcare providers make clinical decisions (Rouse, 2014).[2]

Computer based methods are increasingly used to improve the quality of medical services. Though, the fields in which computers are being used have very high complexity and uncertainty; the uses of intelligent systems such as fuzzy logic, artificial neural network and genetic algorithm have been developed (Jimoh et al, 2014).[3] Recent advances in the field of Artificial Intelligence have led to the emergence of expert systems and computational tools; designed to capture and make available the knowledge of experts in a field. Hence, this work focuses on the design and implementation of a web-based clinical decision support system for the management of early diabetes neuropathy.

Diabetes Neuropathy is a chronic health problem with devastating, yet preventable consequences. Over 135 million people worldwide suffer from diabetes, with 25% developing podiatric problems related to the disease, such as diabetic neuropathy. This neuropathy often causes severe pain and can be incapacitating. Globally, rates of diabetes neuropathy were 15.1 million in 2000 (Skaff et al, 2003).[4] According to Nau et al (2007)[5], 7.8% of 23.6 million people of USA population

were recorded for having type 2 diabetes that resulted to diabetes neuropathy. [5](Nau et al, 2007). The World Health Organization (WHO) estimates the number of diabetes patients will reach 300 million by 2025. Four to five percent of health budgets are spent on diabetes-related illnesses, such as the management of diabetic neuropathy and its consequences.

According to the World Health Organization, there is one diabetes specialist available for 10,000 Nigerians (WHO, 2012).[6] Methods are needed to quantitatively evaluate the integrity of both small and large-caliber sensory nerve fibers in order to detect and manage this condition early in its progression. Since diabetic neuropathy is an irreversible condition, early detection is a key factor for controlling and managing patients early before the disabling effects present. Due to this shortage of specialists, there is a need for a Clinical Decision Support System that will diagnose and manage diabetes neuropathy. Existing Clinical Decision Support Systems make use of patients symptoms, medical history, physical exam and the blood sugar level to diagnose patients and come up with a valid result as regards whether such person(s) have diabetes neuropathy or not.

However, very little research have been carried out using the genetic predisposition, loss of nerve function and plasma insulin level of the individual in question to predict his or her susceptibility to diabetes, thus making this aspect presently pose itself as a gray area. This research consequently proposes a web-based adaptive clinical decision support system to diagnose and manage diabetes neuropathy based on the genetic predisposition, loss of nerve function and plasma insulin level of the person(s) in question. The main objective is to develop a Web-based Clinical Decision Support System to diagnose and manage Diabetes Neuropathy using Naïve Bayes Algorithm. The specific objectives are to:

1. design a web-based clinical decision support system to diagnose and manage diabetes neuropathy;

2. classify diabetes neuropathy using Naïve Bayes Theorem and

3. carry out performance evaluation between Naïve Bayes theorem and four well known classification algorithms (K-nearest neighbor (KNN), Decision Tree (DT), Decision Stump (DS), and Rule Induction (RI) based on appropriate dataset using the Rapidminer 6.2 software.

2 REVIEW OF RELATED WORKS

2.1 A Novel Analysis of Diabetes Mellitus by Using Expert System Based on Brain Derived Neurotropic Factor (BDNF) Levels

In 2013, Devarapalli, Apparao, Kumar & Sridhar [7] proposed a novel concept of designing and building intelligent expert systems for the detection and diagnosis of Diabetes Mellitus. The expert system classification is based on critical diabetic parameters like Brain-Derived Neurotropic Factor (BDNF) levels, and Fasting Blood Glucose (FBG). The proposed rule-based expert system constructs large-scale knowledgebase based on the behavior of the BDNF related diabetic data. The system will give an expert decision taking into consideration all the valid ranges of diabetic parameters. The proposed expert system can work effectively even for large sets of patient data. The gap in this work is that the maximum size of patient datasets which can be processed by the expert system at once was not specified.

2.2 Web Based Intelligent Decision Support System for Type 2 Diabetes Patients

In this study, Sahar (2013) [8] described a web based intelligent decision support system (IDSS) for type 2 diabetes patients, where more emphasis is put on patient empowerment and self-management in treatment. The IDSS is based on a well-documented decision support system, designed mainly for type 2 diabetes patients. The scope of IDSS is broad from being mainly used by type 2 diabetes patients as a tool for their empowerment, self-management, communication, and education. The result of the study revealed and gave 100% effectiveness and correctness as 21 cases already diagnosed were subjected to the Expert System which also diagnosed positive. A major weakness identified with this study among others is non-suitability and non-workability of the Expert System in sub Saharan nations and also the implementation cost which is enormous.

2.3 Web Based Medical Diabetes Diagnosis System Using ANN-ARM For Diabetes Mellitus

In this study, Sridar & Shanthi (2013) [9] fused Artificial Neural Network (ANN) with Association Rule Mining (ARM) for better accuracy. The study was focused on designing and developing a Web based Artificial Neural Network with Association Rule Mining for the diagnosis of diabetes. The data gathered comprised of the following attributes: glucose level, Body Mass Index (BMI), Systolic Blood Pressure, Diastolic Blood Pressure, cholesterol and fasting blood sugar. The web based system combined ANN and ARM to give better accuracy in diagnosing Diabetes Mellitus using the above mentioned input parameters. The gap in this work is that the claim made about the accuracy of the ANN and ARM was not substantiated with test cases.
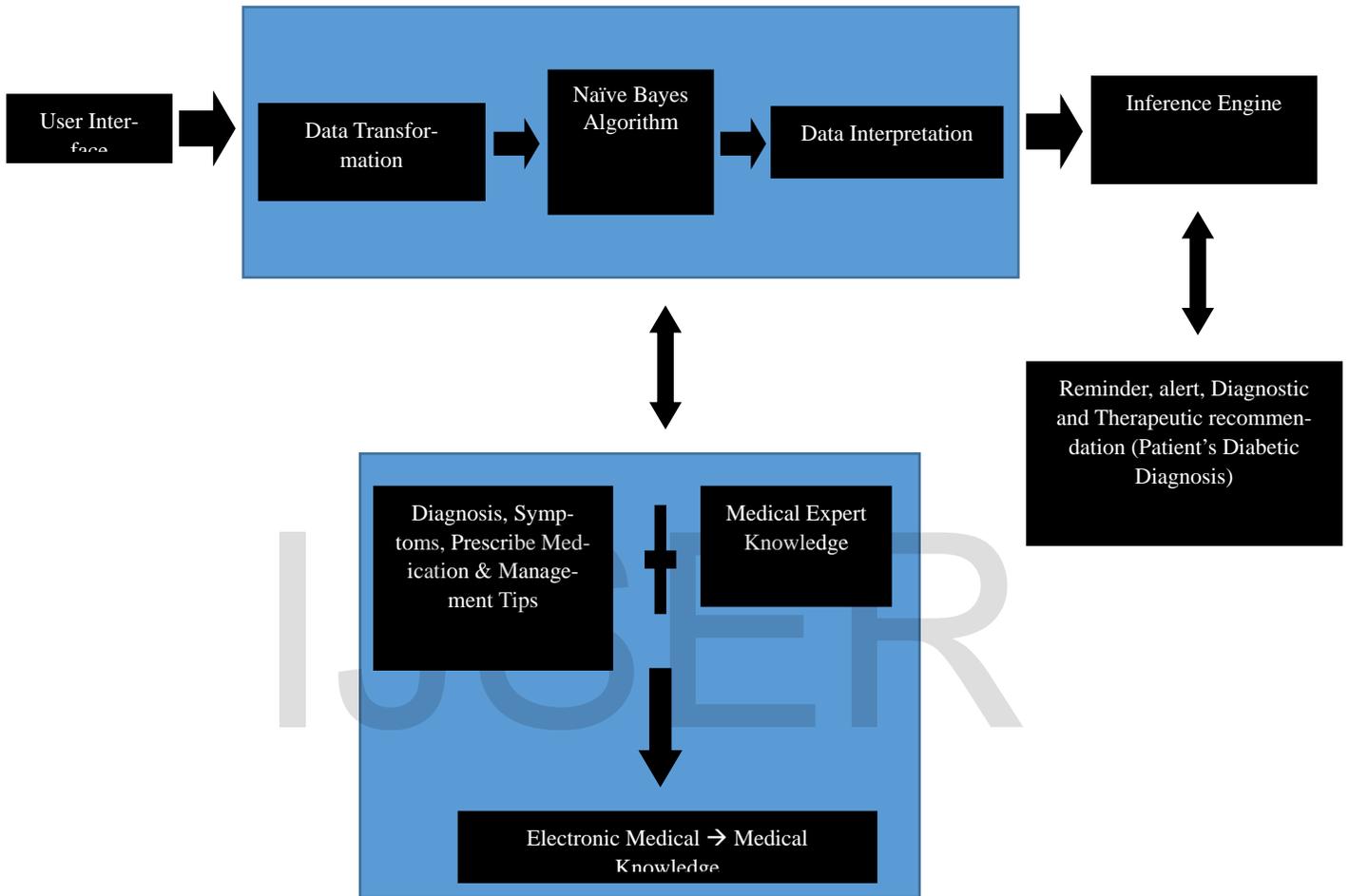
Figure 1: Architecture of the Design CDSS for Diabetes Neuropathy

3   METHODOLOGY

The methodology employed in achieving the objectives of the research is discussed in this section.

3.1 Architecture of the CDSS

Figure 1 depicts the architecture of the Clinical Decision Support System (CDSS) for Diabetes Neuropthy.

A web based Clinical Decision Support System (CDSS) was developed and has a user interface that was used to feed patient data into the system; the system also has a knowledge base that houses the following modules:

1. Diagnostic Module: The diagnostic module that contains symptoms, diagnostic results and prescribed medication / management techniques. The diagnostic results and prescribed medication / management techniques will be displayed through the user interface.

2. Knowledge Base: The medical expert knowledge module that houses the knowledge of the medical expert as regards type 2 diabetes neuropathy.

3. The electronic medical records that contains the details of all persons that the system has been used to diagnose as regards their type 2 diabetes neuropathy status. As such, the medical record must contain sufficient information to identify the patient to whom it relates, as well as information relevant to the patient's treatment during current and future episodes of care.

4. Classification Algorithm Module: Pattern classification algorithms refer to the theory and algorithms of assigning abstract objects into distinct categories with each category typically known in advance. The CDSS has a Classification algorithm module within which patient data is transformed before the classification is done by the algorithm. The result of classification given by the classification algorithm gives the status of the person in question as regards diabetes neuropathy.

5. Inference Engine: The inference engine serves as the processing component within the system that coordinates the activities within the system in such a way as to be able to search thoroughly through the knowledge base to match the respective patient data with the right diagnosis for such patients. It thereafter generates reminders, alerts and therapeutic recommendations for each patient the system is used to diagnose.

6. Database Design: This is the design of the SQL database which consist of the specifications of the various relations, fields that the table contains and their corresponding data type as well as the field length specifications. In the MySQL environment, the term database referred to a collection of tables (relations) and other database objects such as indexes. A table consists of rows and columns, these rows and columns the data for the table.
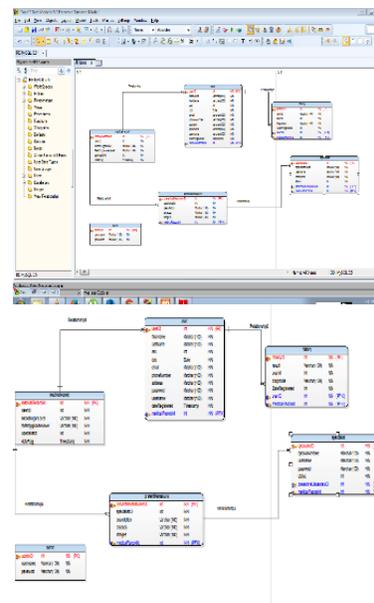

Fig. 2. Database of the Proposed CDSS for Diabetes Neuropathy

3.2 System Design

1. Use Case Diagram for the System: A use case diagram can be defined as a representation of a user's interaction with the system and depicting the specifications of a use case. A use case diagram can depict the different kind of users of a particular system and will often be convoyed by other types of diagrams as well. Fig. 3 depicts the use case diagram for the web-based Clinical Decision Support System for the diagnosis and management of diabetes neuropathy showing all the stakeholders of the system.
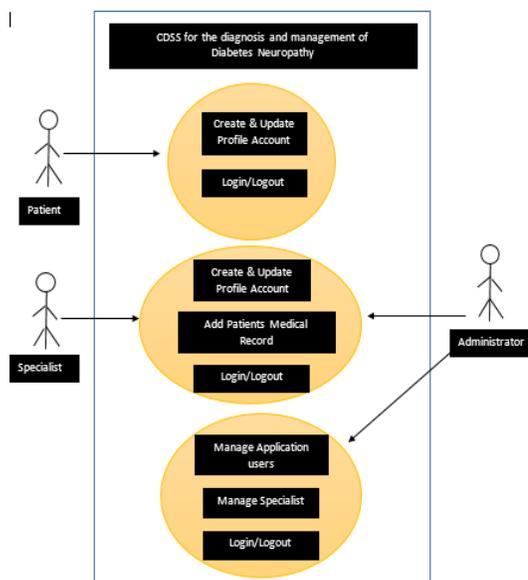

Fig. 3. Use Case Diagram for the CDSS for Diabetes Neuropathy

2. Activity Diagram: The activity diagram basically shows

the activities, associations, conditions and constraints in the system before it is developed as shown in Fig. 4.
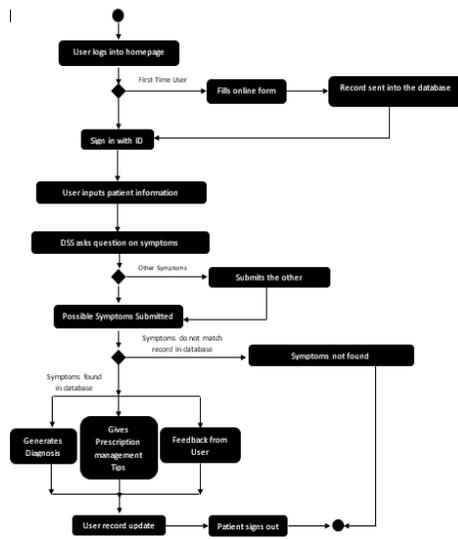

Fig. 4. Activity Diagram of CDSS for Diabetes Neuropathy.

3. Dataset Acquisition: The dataset used in this study was collected from University of Port Harcourt Teaching Hospital and Babcock University Teaching Hospital Repositories. The dataset contains 206 records of patients, each having 9 attributes that are used to predict the diabetes neuropathy. The acquired dataset is designed in MS Excel format and is being used for our classification exercise. The acquired data is complete and has missing values. The dataset collected consists of nine attributes or columns such as 'age', 'pregnancy' , 'patient glucose concentration', 'blood pressure' , 'insulin', 'diabetes pedigree function' , 'smoke', 'body mass index' ,'diabetes'. (Table 1) shows the nine attributes or columns of the acquired dataset in MS Excel Format.

TABLE 1: DATASET IN MS EXCEL FORMAT

'Age' indicates the age group of patients with three categorical values 'young', 'adult' and 'old'. 'Young' indicates the patients with age range 10-25, 'adult' indicates the patients with age

range 26-49 and 'old' indicates the patients with age range above 50. 'Smoke' deals with two parameters such as 'yes' and 'no' based on their habits and lifestyle. 'Insulin' deals with two parameters 'needed' and 'not-needed' which is for patients who must take insulin in order to manage their blood glucose level. 'Level' deals with three parameters 'high', 'normal' and 'low' based on the blood sugar level of the patient. 'Diabetes' deals with two parameters such as 'yes' and 'no' whether or not the person has diabetes. 'Pregnancy' deals with three parameters 'high', 'normal' and 'low' based on the number of pregnancies. 'Body mass index' deals with discrete floating point values. 'Diabetes pedigree function deals with continuous values. To maintain accuracy and to avoid errors, considerable care was taken to ensure that the datasets had correct values. To deal with the condition of zero probability values of some of the parameters, the Laplace Correction was used.
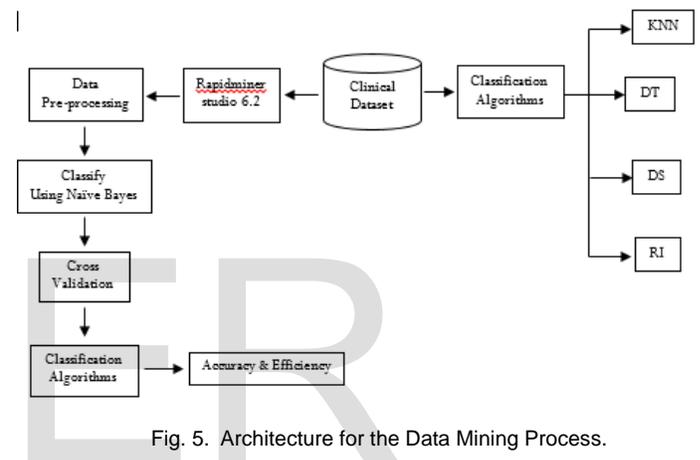
Fig. 5. Architecture for the Data Mining Process.

4 RESULT AND PERFORMANCE EVALUATION OF THE CLASSIFICATION MODELS

4.1 Performance of Classification Models

1. Performance of Naïve Bayes Model: The performance of the Naïve Bayes Classifier model using nominal cross validation shows that 60.50% of the instances in the dataset were correctly classified while 42.85% were incorrectly classified as shown in Fig. 6. The Kappa statistics reveal that the classification of this model is 0.064% reliable with a Kappa statistics mikro of 0.060% and 19.03% for the correctly classified mikro.

Variable	Nominal Cross Validation (%)
Correctly classified Instances	60.50
Incorrect classified instances	42.85
Correctly classified Instances mikro	19.03
Kappa Statistics	0.064
Kappa Statistics mikro	0.060

Fig. 6. Classification Process for Naïve Bayes Algorithm

2. Performance of K-Nearest Neighbourhood (KNN):

The performance of K-Nearest Neighbor model using nominal cross validation shows that 41.00% of the instances in the dataset were correctly classified while 19.81% were incorrectly classified asshown in Fig. 7. The Kappa statistics reveal that the classification of this model is 0.013% reliable with a Kappa statistics mikro of 0.010 and 13.00 for the correctly classified mikro.

Variable	Nominal Cross Validation (%)
Correctly classified instances	41.00
Incorrect classified instances	19.81
Correctly classified instances mikro	13.00
Kappa Statistics	0.013
Kappa Statistics mikro	0.010
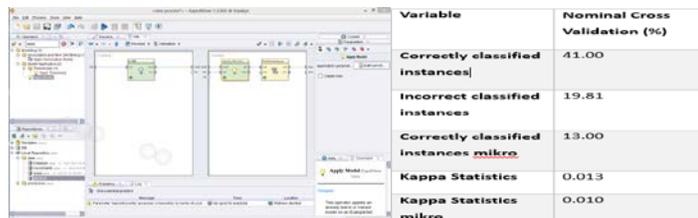
Fig. 7. Classification Process for K-nearest Neigbor Algorithm

3. Performance of Decision Stump:

The performance of Decision Stump model using nominal cross validation shows that 41.00% of the instances in the dataset were correctly classified while 39.12% were incorrectly classified as shown in Fig. 8. The Kappa statistics reveal that the classification of this model is 0.066% reliable with a Kappa statistics mikro of 0.006 and 13.00 for the correctly classified mikro.

Variable	Nominal Cross Validation (%)
Correctly classified instances	41.00
Incorrect classified instances	39.12
Correctly classified instances mikro	13.00
Kappa Statistics	0.066
Kappa Statistics mikro	0.006
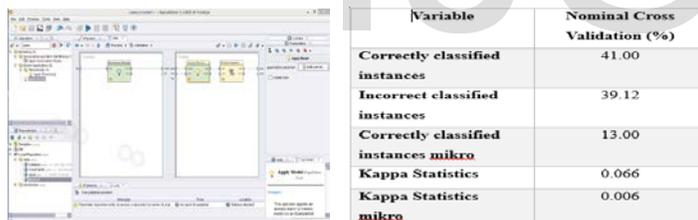
Fig. 8. Classification Process for Decision Stump Algorithm

4. Performance of Decision Tree:

The performance of Decision Tree model using nominal cross validation shows that 36.50% of the instances in the dataset were correctly classified while 22.50% were incorrectly classified as shown in Fig. 9. The Kappa statistics reveal that the classification of this model is 0.122% reliable with a Kappa statistics mikro of 0.008 and 11.41 for the correctly classified mikro.

Variable	Nominal Cross Validation (%)
Correctly classified instances	36.50
Incorrect classified instances	22.50
Correctly classified instances mikro	11.41
Kappa Statistics	0.122
Kappa Statistics mikro	0.008
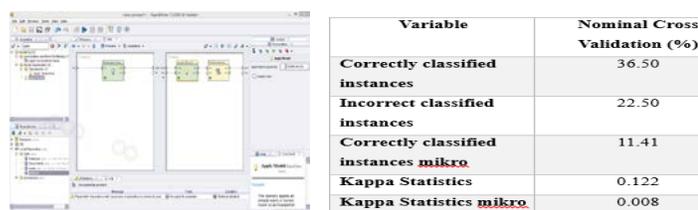
Fig. 9. Classification Process for Decision Tree Algorithm

5. Performance of Rule Induction:

The performance of Rule Induction model using nominal cross validation shows that 52.50% of the instances in the dataset were correctly classified while 38.34% were incorrectly classified. The Kappa statistics reveal that the classification of this model is 0.067% reliable with a Kappa statistics mikro of 0.006 and 28.04 for the correctly classified mikro.

Variable	Nominal Cross Validation (%)
Correctly classified instances	52.50
Incorrect classified instances	38.34
Correctly classified instances mikro	28.04
Kappa Statistics	0.067
Kappa Statistics mikro	0.006
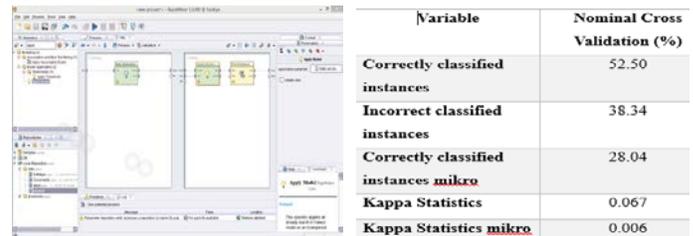
Fig. 10. Classification Process for Rule Induction Algorithm

4.2 Comparison Summary based on Classification Performance

The outcome of both x- cross validation method is similar for all the classifiers. Naïve Bayes outperformed all other classifiers with accuracy being 60.50%. k-nearest neighbor, Decision Tree, Decision Stump and Rule induction perform well with the lowest accuracy for x- cross validation being 36.50%. Decision Tree falls behind in accuracy, while k-nearest neighbour and Decision Stump maintain accuracy at equilibrium 41.00%.

TABLE 2: COMPARISON OF CLASSIFICATION MODELS BASED ON ACCURACY

	Naïve Bayes (NB)	K-NearestNeigbor (lazy.Ibk)	Decision Tree (DT)	Decision Stump (DS)	Rule Induction (RI)
Nominal X-Validation	60.50	41.00	36.50	41.00	52.50

5 CONCLUSION AND RECOMMENDATIONS

Diabetes Neuropathy has remained a very debilitating condition and is characterized by elevated blood glucose levels (hyperglycemia) resulting from defects in insulin secretion, insulin action or both. In its chronic stages, hyperglycemia is associated with micro vascular and macro vascular complications that can lead to visual impairment, blindness, kidney disease, nerve damage, amputations, heart disease, and stroke. Owing to the fact that the clinical features of diabetes in the early stages are usually not obvious even when present, a substantial number of people that have suffered from type 2 diabetes mellitus who did not get diagnosed in good time eventually develop diabetes neuropathy. It is therefore important to involve the General Practitioners in the management of diabetes

neuropathy, because they are rightly positioned to be the first set of individuals sought by patients for medical advice on health related issues. This study therefore focused on the design of a web based clinical decision support system that could aid in the management of diabetes neuropathy in order to help the people afflicted with this ailment have a healthy and normal life and thus transcend the limits that would have been imposed on their existence.

This system is recommended for clinics, hospitals, health centers and experienced health care professionals such as doctors and nurses. This project can be enhanced to cover some other operations of the hospital such as: Pharmacy: Drugs names and doses, Theatre Operations, Laboratory. This project can also be developed on mobile platforms applications.

6 REFERENCES

[1] J. Q. Chen and S.M. Lee (2002). An exploratory cognitive DSS for strategy decision making Elsevier Science B.V., 8(6), 13-32.

[2] M. Rouse, M. Clinical decision support system (CDSS) definition. Retrieved from techtarget: http://searchhealthit.techtarget.com/definition/clinical-decision-support-system-CDSS.

[3] R.G, Jimoh (2014). Simulation of Medical Diagnosis System for Malaria Using Fuzzy Logic. Cisdiar, 20-25. Retrieved from http://cisdijournal.net/uploads/V5N2P6_-__CISDIAR_JOURNAL.pdf.

[4] M.M. Skaff, J.T. Mullan, L. Fisher L, C.A. Chesla (2003). A contextual model of control beliefs, behaviour, and health: Latino and European Americans with Type 2 diabetes. Psychology Health 2003; 18(3): 295-312. Retrieved from:
http://deepblue.lib.umich.edu/bitstream/handle/2027.42/75845/hagan_3.pdf?sequence=1.

[5] D. P. Nau, J.E. Aikens, A.M. Pacholski. The Effects of Gender and Depression on Oral Medication Adherence in Persons with Type 2 Diabetes Mellitus. Gender Medicine 2007; 4(3):205-213.

[6] World Health Organization, 2012

[7] D. Devarapalli, A. Apparao, A. Kumar, G. R. Sridhar (2013), A Novel Analysis of Diabetes Mellitus by Using Expert System Based on Brain Derived Neurotrophic Factor (BDNF) Levels, Helix Vol. 1:251 - 256

[8] F. Sahar (2013) Web Based Intelligent Decision Support System for Type 2 Diabetes Patients, IACSIT International Journal of Engineering and Technology, 5(5), Forthcoming, https://ssrn.com/abstract=2321520

[9] K. Sridar, Dr. D. Shanthi (2013) Web-based Medical Diagnosis System using ANN - ARM for the Diabetes Mellitus, International Journal of Computers and Distributed Systems, 3(3), pp15-20 www.ijcdsonline.com