Volume 5, Issue 3, March 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

The Prediction, Diagnosis and Treatment of Diabetes Mellitus Using an Intelligent Decision Support System Framework

Adekunle Y.A
Department of Computer Science, Babcock University,
Ilishan-Remo, Ogun State, Nigeria

Abstract— In recent times, diabetes mellitus has become a highly dreaded illness which has negatively impacted the health of its victims the world over. This paper proposes a framework for the prediction, diagnosis and treatment of diabetes mellitus so as to provide a long sought proactive solution to the diabetes menace. In the framework, the patient information is fed into the clinical decision support system through the user interface; the knowledge base stores the rules and data to be used by the system and the pattern classification/prediction algorithm that emerged as the best after a thorough evaluation of some relevant classification algorithms is the C5.0 decision tree algorithm which had its percentage of correctly classified instances given as 78.4534%; the C5.0 algorithm searches the knowledge base recursively and matches the patient information with the pertinent rules that suits each case and thereafter gives the correct prediction as to whether the patient in question is susceptible to diabetes mellitus or not; it also provides a measure of the likelihood of the patient developing diabetes . The problem of accurately determining if a patient is prone to diabetes mellitus or not which has been a perennial problem in the domain of medicine will be most likely effectively combated with the solution provided by this framework.

Keywords— Diabetes Mellitus, Hyperglycemia, Type 1 diabetes, Type 2 diabetes, Gestational diabetes mellitus (GDM), Clinical Decision Support System (CDSS), Artificial Intelligence (AI), Support Vector Machine (SVM), Decision Trees (DT) and Bayes Classifier.

I. INTRODUCTION

In this recent age, the development of intelligent decision making applications is becoming the order of the day. This concept is universally described as Artificial Intelligence (AI). Artificial Intelligence has different sub-fields which include machine vision, expert systems, machine learning and natural language processing, speech recognition and a host of others.

A Decision Support System is an interactive computer-based system built to aid decision makers in utilizing data and models so as to identify, solve problems and make decisions [1]. According to the Clinical Decision Support (CDS) Roadmap project, CDS is “providing clinicians, patients, or individuals with knowledge and person-specific or population information, intelligently filtered or present at appropriate times, to engender better health processes, better individual patient care, and better population health.”

A Clinical Decision Support System (CDSS) is an active knowledge system, where two or more items of patient data are used to produce case-specific recommendation(s) [2]. This implies that a CDSS is a decision support system (DSS) that utilizes 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 enormous.

Diabetes mellitus is a group of metabolic diseases characterized by elevated blood glucose levels
(hyperglycemia) resulting from defects in insulin secretion, insulin action or both. Insulin is a hormone manufactured by the beta cells of the pancreas, which is required to utilize glucose from digested food as an energy source. Chronic hyperglycemia is associated with microvascular and macrovascular complications that can lead to visual impairment, blindness, kidney disease, nerve damage, amputations, heart disease, and stroke [3]. In 1997 an estimated 4.5% of the US population had diabetes and direct and indirect health care expenses were estimated at $98 billion.

A Clinical Decision Support System (CDSS) is an active knowledge system, where two or more items of patient data are used to produce case-specific recommendation(s) [2]. This implies that a CDSS is a decision support system (DSS) that utilizes 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 enormous.

Diabetes mellitus is a group of metabolic diseases characterized by elevated blood glucose levels
(hyperglycemia) resulting from defects in insulin secretion, insulin action or both. Insulin is a hormone manufactured by the beta cells of the pancreas, which is required to utilize glucose from digested food as an energy source. Chronic hyperglycemia is associated with microvascular and macrovascular complications that can lead to visual impairment, blindness, kidney disease, nerve damage, amputations, heart disease, and stroke [3]. In 1997 an estimated 4.5% of the US population had diabetes and direct and indirect health care expenses were estimated at $98 billion.

Blacks are 1.7 times as likely to develop diabetes as whites. The prevalence of diabetes among blacks has quadrupled during the past 30 years. Among blacks age 20 and older, about 2.3 million have diabetes – 10.8 percent of that age group.

Blacks with diabetes are more likely than non-Hispanic whites to develop diabetes and to experience greater disability from diabetes-related complications such as amputations, adult blindness, kidney failure, and increased risk of heart disease and stroke; Death rates for blacks with diabetes are 27 percent higher than for whites.

II. CLASSIFICATION OF DIABETES MELLITUS

There were several classification systems established for diabetes mellitus by the WHO Expert Committee on Diabetes (1980, 1985). The current WHO classification system has been established in co-operation with the National Diabetes Data Group (USA). It is mainly based on the aetiology of diabetes mellitus (Table 1).

Table 1: Classification of diabetes mellitus

Type 1 diabetes mellitus	
	Immune mediated
	Idiopathic
Type 2 diabetes mellitus	
Other specific types of diabetes	
	Genetic defects of islet ß-cell function
	Genetic defects of insulin action
	Diseases of the exocrine pancreas
	Endocrinopathies
	Drug- or chemical- induced diabetes
	Infections
	Uncommon forms of diabetes
	Other genetic syndromes
Gestational diabetes mellitus	

The terms IDDM (insulin dependent diabetes mellitus) and NIDDM (non-insulin dependent diabetes mellitus) were used previously but have now been abandoned. Presently, the terms "Type 1" and "Type 2" diabetes are used. The more prevalent form is Type 2 diabetes.

A. Type 1 diabetes

(Insulin-dependent diabetes, juvenile diabetes)
Type 1 diabetes is characterized by cellular-mediated autoimmune destruction of islet ß-cells.
Markers:
- islet cell antibodies (ICAs)
- auto-antibodies to insulin (IAAs)
- auto-antibodies to glutamic acid decarboxylase (GAD65)
- auto-antibodies to tyrosine phosphatases IA-2 and IA-2ß

Association with HLA: DQA and DQB genes: HLA-DR/DQ alleles may be protective
Environmental factors are poorly defined. Virus infectious and nutritional factors are discussed.
Age: Onset predominantly in childhood and adolescence, but occurs at any age
Idiopathic diabetes in African or Asian people. This form of diabetes is strongly inherited, has permanent insulinopenia, is prone to ketoacidosis without antibodies to ß-cells.
Laboratory findings:
- Hyperglycaemia
- Ketonuria
- Low or undetectable serum insulin and C-peptide levels
- Auto-antibodies against components of the islet ß-cells.

B. Type 2 diabetes

(Maturity-onset diabetes, non-insulin dependent diabetes).
Type 2 diabetes is due to insulin insensitivity combined with a failure of insulin secretion to overcome this by hypersecretion, resulting in relative insulin deficiency. There is a strong genetic predisposition. Type 2 diabetes is more common in individuals with family history of the disease, in individuals with hypertension or dyslipidaemia and in certain ethnic groups.
The risk of developing Type 2 diabetes increases with:
- Family history of diabetes (in particular parents or siblings with diabetes)
- Obesity ($\geq$ 20% over ideal body weight or BMI $\geq$ 25.0 kg/m²)
- Membership of some ethnic groups
- Age $\geq$ 45 years
- Previously identified IFG or IGT
- Hypertension ($\geq$ 140/90 mmHg in adults)

- HDL cholesterol level <1.0 mmol/L (<0.38 g/L) and/or a triglyceride level ≥ 2,3 mmol/L (≥2,0 g/L)
- Reduced physical activity
- History of gestational diabetes mellitus (GDM) or delivery of babies >4,5 kg

MODY is a form of youth onset diabetes which is not insulin-dependent, with a strong dominant family history, and is associated with abnormal hepatic nuclear factor (HNF) or glucokinase genes.

Table 2: General characteristics Type 1 and Type 2 diabetes (Source: WHO, 2002)

Characteristics	Type 1 diabetes	Type 2 diabetes
Typical age of onset (years)	< 35	> 35
Genetic predisposition	low	high
Antibodies to ß-cells	yes (90 – 95%)	no
Body habitus	normal/ wasted	obese
Plasma insulin/C-peptide	low/absent	high
Main metabolic feature	insulin deficiency	metabolic syndrome with insulin insensitivity
Insulin therapy	responsive	high doses required
Insulin secretagogue drugs	unresponsive	responsive

Laboratory findings:
-hyperglycaemia
- hyperlipidaemia
- high serum insulin/C-peptide level
-defective insulin secretion
- insulin resistance

C. Gestational diabetes mellitus (GDM)

Definition: Any degree of clinical glucose intolerance with onset or first recognition during pregnancy.

GDM complicates the pregnancy: The following problems may develop with GDM: altered duration of pregnancy placental failure hypertension / pre-eclampsia high birth weight of the newborn.

Therapy: nutrition therapy insulin.

Diagnosis of GDM:

Fasting plasma glucose level >7,0 mmol/L (>1,26 g/L) or casual plasma glucose >11,1 mmol/L (>2,00 g/L), confirmed on a subsequent day.

Laboratory strategy to diagnose GDM:
One step approach: OGTT (75 g glucose)
Two step approach: 1. First OGTT with 50 g glucose load; cut-off value after 1 hour plasma glucose >7,8 mmol/L (>1,40 g/L)
Second OGTT with 75 g glucose load and evaluation as the standard OGTT

Six weeks after pregnancy or later the woman should be re-examined for the presence of diabetes mellitus or IGT.

III. PREVALENCE OF DIABETES MELLITUS

The prevalence of diabetes in Western life-style countries is estimated to be between 6.0 and 7.6 %. In some developing countries the prevalence is more than 6%. The mean percentage prevalence varies between ethnic groups (American Indians, Hispanics, and others). Between 1995 and 2025 there is predicted to be a 35% increase in the world-wide prevalence of diabetes; the rising number of people with diabetes will occur mainly in populations of developing countries, leading to more than 300 million people with diabetes globally by 2025 [5]. Presently as many as 50% of people with diabetes are undiagnosed. Since therapeutic intervention can reduce complications of the disease, there is a need to detect diabetes early in its course. The risk of developing Type 2 diabetes increases with age, obesity, and lack of physical activity.

IV. SCREENING FOR DIABETES

Screening for diabetes is an analytical, organizational, and financial challenge. The organizational and financial aspects are the biggest limiting factors. Several strategies have been suggested and evaluated for community screening. If possible community screening should occur within the local health-care system so that individuals with positive findings get appropriate follow-up investigations and treatment.

Screening strategy will depend on the underlying prevalence of diabetes, structure of the local health-care system, and the economic condition of the country. The aim of screening is to identify asymptomatic individuals who are likely to have diabetes. There are two strategies that may be applied for screening

1. Detect all people with diabetes in a population.
2. Detect diabetes amongst those people who are mostly likely to have diabetes (selective, or opportunistic screening)

In a recent Danish study the authors stated that no randomized control trials are available to advise on the question of opportunistic versus systematic screening. These authors favour economic models which give preference to opportunistic screening rather than systematic screening. In other countries with a higher prevalence of diabetes, systematic screening may be more cost-effective.

A. Opportunistic screening:

Detection of people with diabetes who contact health services for other reasons, by physical and laboratory examination.

B. Selective screening:

A verbal or written questionnaire is distributed in the population. This questionnaire should identify those individuals who are at high risk of having diabetes. They should be referred to a physician for consideration of diagnosis.

Selective screening should consider individuals :

- with typical symptoms of diabetes
- with a first-degree relative with diabetes
- who are members of a high risk ethnic group
- who are overweight (BMI ≥ 25.0 kg/m²)
- who have delivered a baby >4.5 kg or had GDM
- who are hypertensive (≥ 140/90 mmHg)
- with raised serum triglyceride and cholesterol levels
- who were previously found to have IGT or IFG [6].

C. Systematic screening:

Identification of people with new diabetes will be low at follow-up examinations at regular intervals (e.g. 3 years) because the incidence of new disease is low. This will give rise to problems of specificity and motivation. For the systematic screening of diabetes the recommendation of the American Diabetes Association may be followed. In this, screening should begin at an age of 45 years and be repeated at intervals of 3 years.

The basic laboratory measures for screening are:

1. Fasting capillary blood glucose
2. Glucosuria
3. HbA1c
4. OGTT

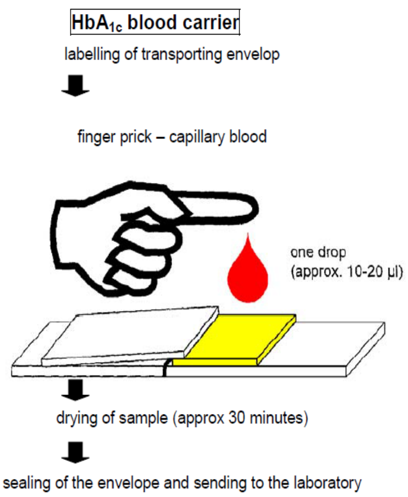
The common and best indicator for estimating diabetes prevalence and incidence is fasting blood glucose (FPG). FPG concentration of >7,0 mmol/L (>1,26 g/L) is an indication for retesting. For centralized screening the analysis of glycated haemoglobin (HbA1c) from a blood drop is recommended, though this approach is more expensive than FPG.

Screening strategies from a laboratory technical perspective. Decentralized screening

In decentralized screening fasting blood glucose is the appropriate analyte, followed by retesting FPG and/or by urine glucose. The comparability of glucose analyses must be verified by internal and external quality control. HbA1c may also be used in decentralized screening although the results may vary when different chromatographic methods are used. The OGTT is not recommended as the first step of screening but rather as a confirmation test.

Centralized screening

This is dependent on easy specimen collection, specimen stability and specimen transport. These conditions are met by capillary blood collection, preservation of the specimen as dry blood on a filter paper and HbA1c analysis by an immunological procedure at a central laboratory. Chromatographic methods are less suitable for HbA1c measurement in dried blood samples since some HbA1c may be partially degraded during transportation whist still having preserved its antigenicity.



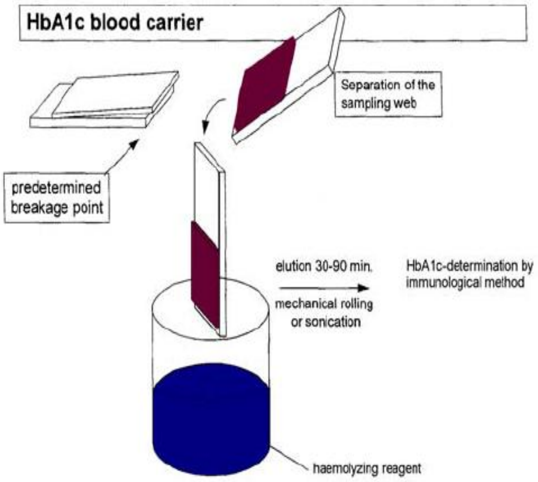


Fig. 1 Specimen collection device for centralized analysis of HbA1c (Barron, 2006)

V. RELATED WORKS

A. Emergent Frameworks for Decision Support Systems

Ioana Andreea Stanescu & Florin Gheorghe Filip (2011). The authors of this work opined that knowledge is generated and accessed from heterogeneous spaces. Furthermore, recent advances in Information Technologies have led to enhanced tools and techniques for improving the efficiency of knowledge-based decision support systems. Frameworks were presented for developing the optimal blend of technologies required in order to better the knowledge acquisition and reuse in large scale decision-making environments.

A case-study in the field of clinical decision support systems based on emerging technologies was presented. Such intelligent information systems were developed to be able to run sophisticated models at the back end, but remain friendly enough at the front-end to be used comfortably by any user regardless of the domain of operation.

The study presented frameworks of decision support systems (DSS) under the impact of emergent technologies, such as bi-dimensional (2D) barcodes and mobile infrastructures. It is a case study of a computer-based clinical diagnostic tool called MEDIS developed in Advanced Technology Systems, Targoviste, Romania, Romanian Academy, Bucharest, Romania in 2011. MEDIS is a pilot development of a clinical diagnostic decision support system (CDDSS) designed to collect and reuse knowledge from heterogeneous sources, accessible in desktop and mobile environments. The research analyzed the problems encountered in the development, implementation and maintenance of the clinical decision support systems and discussed innovative alternatives and solutions.

The study has a strength of presenting frameworks for the structure of Decision Support Systems in general and Clinical DSS in particular. It was tested using the MEDIS platform.

The test case employed for the framework was based on a local product which was just a prototype and a pilot test bed. It is not yet available universally especially in the developing country environment. The frameworks have never been tested on cervical cancer which is very predominant in the developing countries and also the subject of this research work.

B. Development of an Expert System as a Diagnostic Support of Cervical Cancer (CC) in A Typical Glandular Cells, Based on Fuzzy Logics and Image Interpretation

Karem, R. (2013): This research described an expert system (ES) which is able to provide a diagnosis to CC through the integration of fuzzy logics and image interpretation techniques. The ES comprises three segments viz:

- risk diagnosis which consists of the interpretation of a patient's clinical background and the risk of contracting CC from specialist perspectives
- cytology image detection which consists of image interpretation
- determining of CC precursor injuries which consists of retrieving the information from the prior phases and integrating the ES by means of a fuzzy logic

The result of the study revealed and gave 100% effectiveness and correctness as 21 cases already diagnosed were subjected to the ES which also diagnosed positive. A major weakness identified with this study among others is non-suitability and non-workability of the ES in sub Saharan nations and also the implementation cost which is enormous.

C. Data Mining in Clinical Decision Support Systems for Diagnosis and Treatment of Heart Disease

According to Amin, Agarwal & Beg (2013) medical errors are both costly and harmful. Medical errors cause thousands of deaths worldwide each year. Hence, a clinical decision support system (CDSS) would offer opportunities to reduce medical errors as well as to improve patient safety. They affirm that one of the most important applications of such systems is in diagnosis and treatment of heart diseases (HD). This is because statistics have shown that heart disease is one of the leading causes of deaths all over the world (CDC Report). Data mining techniques have been very effective in designing clinical support systems because of its ability to discover hidden patterns and relationships in medical data. Here, the proponents also undertook a comparative analysis of the performance and working of six CDSS systems which use different data mining techniques for heart disease diagnosis. They conclude by asserting based on their findings that there is no system to identify treatment options for Heart disease patients. They further claimed that in spite of having a large amount of medical data, it lacked in the quality and the completeness of data thereby creating the need for highly sophisticated data mining techniques to build up an efficient decision support system. They claim that even after doing this, the overall reliability and generalization capability might still be questionable. Hence, the need to build systems which will be accurate, reliable as well as reduce cost of treatment and increase patient care. More so, the building of systems which are understandable and which could enhance human decisions are very germane.

D. An Intelligent Decision Support System for the Operating Theater
In 2013, Sperandio, Gomes, Borges, Brito and Almada-Lobo asserted that decision processes inherent in operating theatre organization are often subjected to experimentation, which sometimes lead to far from optimal results. They further affirm that the waiting lists for surgery had always been a societal problem, with governments seeking redress with different management and operational stimulus plans partly due to the fact that the current hospital information systems available in Portuguese public hospitals, lack a decision support system component that could help achieve better planning solutions. As such they developed an intelligent decision support system that allows the centralization and standardization of planning processes which improves the efficiency of the operating theater and tackles the fragile situation of waiting lists for surgery. The intelligence of the system is derived from data mining and optimization techniques, which enhance surgery duration predictions and operating rooms surgery schedules.

E. HIROFILOS: A Medical Expert System for Prostate Diseases (Constantinos Koutsojannis, Maria Tsimara & Eman Nabil, 2008)
In this study a fuzzy expert system for diagnosing, and learning purpose of the prostate diseases was described. HIROFILOS is a fuzzy expert system for diagnosis and treatment of prostate diseases according to symptoms that are realized in one patient and usually recorded through his clinical examination as well as specific test results. The user-friendly proposed intelligent system is accommodated on a hospital web page for use as a decision support system for resident doctors, as an educational tool for medical students, as well as, an introductory advisory tool for interested patients. It is based on knowledge representation provided from urology experts in combination with rich bibliographic search and study ratified with statistical results from clinical practice. Preliminary experimental results on a real patient hospital database provide an acceptable performance that can be improved using more than one computational intelligence approaches in the future.

VI. CLINICAL DECISION SUPPORT SYSTEM (CDSS)
The clinical decision support system is another example of a knowledge based system. A clinical decision support system is an active knowledge system where two or more items of patient data are used to generate case specific recommendations [2].

A. Target Area of care
CDSSs assist doctors in assessing various clinical issues from accurate diagnosis of a particular disease to the treatment of the disease. The general target areas of care for CDSS are:
Preventive care which has to do with screening and disease management
Diagnosis which is done based on the patients' signs and symptoms
Follow-up management which has to do with frequent checkups
Hospital Provider Efficiency [11].

B. System Design
The system design for CDSS will usually include the following subsystems:
Communication which handles notification and alerts
Knowledge discovery which deals with rules and regulations
Knowledge repository which contains problem solving knowledge [12].

C. Factors leading to successful CDSS implementation
The following under listed factors lead to the successful implementation of CDSS:
Simple, user friendly interface
Automated decision support
Timely result
Workflow integration
Continuous Knowledge-base and update support [13].

VII. PATTERN CLASSIFICATION METHODS
Pattern classification refers to the theory and algorithms of assigning abstract objects into distinct categories, where these categories are typically known in advance. For this research, the pattern classification methods considered are Decision Trees (DTs), K-Nearest Neighbor (KNN), Naïve bayes Classifier and Support Vector Machine (SVM).

A. Decision Trees

A decision tree consists of a root node, branch nodes and leaf nodes. The tree begins with a root node, then further splits into branch nodes and each node represents a choice among various alternatives. The tree then terminates with leaf nodes which are un-split nodes that represent a decision [14]. The classification of decision trees are carried out in two phases:
Tree Building or top down: This is computationally intensive and requires the tree to be recursively partitioned until all the data items belong to the same class.
Tree pruning or bottom top: It is conducted to improve the prediction and classification of the algorithm and minimize the effects of over-fitting which may lead to misclassification of errors [15].
Some notable decision tree algorithms include Classification and Regression Trees (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0.
The advantages of decision trees include:

- They are easy to interpret and comprehend
- They can handle both metric and non-metric data as well as missing values which are frequently encountered in clinical studies.
- Little data preparation is required since data does not need to be normalized.
- They can handle data in a short time frame.
- They can be developed using common statistical techniques.

The disadvantages associated with decision trees include:

- They can over fit the data and create complex trees that may not generalize well.
- A small change in the size of a dataset could result in a completely different tree

B. K-Nearest Neighbor

K-Nearest Neighbor (k-NN) is instance based learning for classifying objects based on closest training examples in the feature space. It is a type of lazy learning where the function is only approximated locally and all computations are deferred until classification. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors. If k=1, then the object is simply assigned to the class of its nearest neighbor. The k-NN algorithm uses all labeled training instances as a model of the target function. During the classification phase, k-NN uses a similarity-based search strategy to determine a locally optimal hypothesis function. Test instances are compared to the stored instances and are assigned the same class label as the k most similar stored instances.

C. Bayes Classifier

A Bayesian network is a model that encodes probabilistic relationships among variables of interest. This technique is generally used for intrusion detection in combination with statistical schemes, a procedure that yields several advantages, including the capability of encoding interdependencies between variables and of predicting events, as well as the ability to incorporate both prior knowledge and data. However, a serious disadvantage of using Bayesian networks is that their results are similar to those derived from threshold-based systems, while considerably higher computational effort is required.

D. Support Vector Machine

Support Vector Machines have been proposed as a novel technique for intrusion detection. An SVM maps input (real-valued) feature vectors into a higher-dimensional feature space through some nonlinear mapping. SVMs are developed on the principle of structural risk minimization. Structural risk minimization seeks to find a hypothesis (h) for which one can find lowest probability of error whereas the traditional learning techniques for pattern recognition are based on the minimization of the empirical risk, which attempt to optimize the performance of the learning set. Computing the hyper plane to separate the data points i.e. training an SVM leads to a quadratic optimization problem. The implementation of SVM intrusion detection system has two phases which are training and testing. SVMs can learn a larger set of patterns and be able to scale better, because the classification complexity does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamically whenever there is a new pattern during classification.

VIII. METHODOLOGY

A thoroughly compiled dataset (Lunar, 2012) consisting of 220,000 instances obtained from the UCI (University of California, Irvine) data repository was used. This dataset was imported into the IBM Statistical Product and Service Solutions (SPSS) version 21 software and saved in the CSV format, thereafter the dataset was available on Microsoft Excel from which it was imported into the Rapid Miner version 5.3 data mining software.
The dataset was then induced with Classification algorithms namely C5.0 decision trees, Support Vector Machine (SVM), K-Nearest neighbor algorithm and Bayes Classifier Algorithm. The Classification algorithms were evaluated using the Rapid Miner software version 5.3 based on accuracy and the results showed the C5.0 decision trees having an accuracy of 78.4534%, the Support Vector Machine (SVM) algorithm had an accuracy of 61.2673%, the Bayes Classifier Algorithm had 60.2045% and the K-Nearest Neighbor algorithm had 59.1265%. Sequel to the result obtained from this evaluation, the C5.0 decision trees turn out as the Classification algorithm with the highest accuracy for this research.

Thereafter, a decision tree program was written in Java with 57 lines of code for the core program to implement the C5.0 decision tree algorithm that will provide the requisite intelligence for this Clinical decision support system and help it make the right decisions promptly when supplied with patient information. The C5.0 decision tree algorithm was thus embedded in the classification/prediction algorithm section of the clinical decision support system.

IX. CONCLUSION AND RECOMMENDATION

This research work finds relevance in all regions of the world where people live with the health challenge known as diabetes mellitus, thus it is very germane as it provides a sort of panacea to the eventual development of the illness for people who are susceptible to the condition, hence they can be aware of their predisposition ahead of time and can be able to take the necessary precautionary measures to forestall their development of the illness, thus saving them from the trauma they would otherwise have inevitably suffered.

The research is a milestone in the sub-field of health informatics as it provides a readily available Clinical Decision Support System to serve as a reliable assistant to the medical practitioners that are more often than not burdened by the enormous and seemingly intimidating number of patients they need to attend to routinely. This has culminated in a lot of fatal errors on the part of the medical practitioners which has led to the loss of innocent lives hence, the introduction, consequent adoption and deployment of this Knowledge Based Intelligent Clinical Decision Support System for the prediction, diagnosis and treatment of diabetes becomes expedient especially in the third world countries, the vast majority of who lag behind in terms of technological innovations and advancement and as a result are alien to the terrific results gotten from the use of these clinical decision support systems.

For further work another enthusiastic researcher can go another step further with this work by introducing other highly efficacious algorithms that can be used alongside the C5.0 decision tree algorithm used in this work, so as to have a hybrid system that will take decisions faster and generate more accurate decisions than those that will be given by the proposed system.

REFERENCES

[1] Power, D.J. (1999). Decision Support Systems Glossary.http://DSSResources.COM/glossary

[2] Chen, J.Q & Lee, S.M. (2002). An exploratory cognitive DSS for strategy decision making. *Elsevier Science B.V.*

[3] Harris R.E., Metzger B.E., & Coustan D.R. (2007). Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. *Diabetes. Care*, 6(8), 1345-1349.

[4] Loghmani , U, Tallarigo, L, Giampietro, O, Penno, G, Miccoli, R, Gregori, G, & Navalesi, R. (2012). Relation of glucose tolerance to complications of pregnancy in nondiabetic women. *National English Journal of Medicine*, 3(4), 12-17.

[5] Bennett, P.H, Knowler, W.C, Pettitt, D.J, Carraher, M.J, &Vasquez, B (Longitudinal studies of the development of diabetes in the Pima Indians. In *Advances in Diabetes Epidemiology*, Eschwege E, ed. Amsterdam, The Netherlands, Elsevier, 65-74.

[6] Berger,Y, Crane, R, & Farine, L. (2002). Diabetes in pregnancy in Zuni Indian women. *Diabetes Care*, 7 (9), 1231–1235.

[7] Karem, R (2010). Development of an Expert System as a Diagnostic Support of Cervical Cancer (CC) in A Typical Glandular Cells, Based on Fuzzy Logics and Image Interpretation. *English Journal of Medicine*, 2 (9), 23-34.

[8] Amin, S.U, Agarwal, K & Beg R. (2013). Data mining in clinical decision support systems for diagnosis, prediction and treatment of heart disease. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2 (1).

[9] Sperandio F, Gomes C, Borges J, Brito A.C & Almada-Lobo B. (2013). An intelligent Decision Support System for the Operating Theatre: A Case Study. Automation Science and Engineering, IEEE Transactions on Robotics & Control Systems. 32(19), 132-167.

[10] Constantinos, K, Maria, T & Eman N (2008). HIROFILOS: A Medical Expert System for Prostate Diseases. *Proc. Of The 7th WSEAS Int. Conf. On Computational Intelligence*, Man-Machine Systems and Cybernetics.

[11] Berner, E.S. (2009). Clinical Decision Support Systems: State of Art. Rockville: *AHRQ Publication*, 12(5), 90-134.

[12] Frize, M. (2005). Conceptual Framework of Knowledge Management for Ethical Decision making Support in Neonatal Intensive Care. *IEEE Transactions of Information Technology in Biomedicine*, 9(7), 205-215.

[13] Peleg, M., & Tu, S. (2011). Decision Support, Knowledge Representation and Management in Medicine. Stanford Centre for Biomedical Informatics Research. Last accessed: December 17, 2011 from http://bmir.stanford.edu/file_asset/index.php/1009/SMI-2006-1088.pdf

[14] Peng, W., Chen J. & Zhou H. (2006). An Implementation of ID3-Decision Tree Learning Algorithm. University of New South Wales. Last accessed: December17, 2011 from http://web.arch.usyd.edu.au/~wpeng/DecisionTree2.pdf.

[15] Anyanwu, M.N., & Shiva, A.G., (2009). Comparative Analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 3(3), 230-240

