Framework for a Knowledge-Based Intelligent Clinical Decision Support System to Predict Prostate Cancer

JIMOH R.G [1] , ADEKUNLE Y.A [2] , EBIESUWA SEUN [3]

[1] Computer Science Department, University of Ilorin, Ilorin, Kwara State, Nigeria

[2] Computer Science Department, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

[3] Computer Science Department, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

ABSTRACT

This paper proposes a novel framework for a Knowledge Based Intelligent Clinical Decision Support System for the prediction of prostate cancer which is one of the most deadly illnesses that has a deleterious effect on people afflicted with it and has for long remained a perennial health problem affecting a significant number of people the world over. In the framework the patient information is fed into the system and the Knowledge base stores all the information to be used by the Clinical Decision Support System and the classification/prediction algorithm chosen after a thorough evaluation of relevant classification algorithms for this work is the C4.5 Decision Tree Algorithm with its percentage of correctly classified instances given as 61.0734%; it searches the Knowledge base recursively and matches the patient information with the pertinent rules that suit each case and thereafter gives a most precise prediction as to whether the patient is susceptible to prostate cancer or not. This approach to the prediction of prostate cancer is new and has not existed in other literature related to this subject as it employs a very efficacious solution to the problem of determining if a person has the likelihood of developing this dreaded illness or is almost not susceptible to the ailment.

Keywords: Prostate Cancer, CDSS, AI, PCa, K-NN, SVM and SMO.

INTRODUCTION

In recent times, the development of intelligent decision making applications is fast gaining a lot of ground. This concept is known as Artificial Intelligence (AI). Artificial Intelligence has different sub-fields which include expert systems, machine vision, machine learning and natural language processing amongst others.

A Decision Support System is an interactive computer-based system intended to help decision makers utilize data and models in order to identify and 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 foster 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 generate case-specific recommendation(s) [2]. 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.

Cancer of the prostate (PCa) is now recognised as one of the most crucial medical problems facing the male population. In Europe, PCa is the most common solid neoplasm, with an incidence rate of 214 cases per 1000 men, outnumbering lung and colorectal cancer [3]. In addition, PCa is currently the second most common cause of cancer death in men [4]. Furthermore, since 1985, there has been a slight increase in the number of deaths from PCa in most countries, even in countries or regions where PCa is not common [5].

Prostate cancer affects elderly men more often than young men. It is therefore a bigger health concern in developed countries with their greater proportion of elderly men. Thus, about 15% of male cancers are PCa in developed countries compared to 4% of male cancers in developing countries [6]. It is worth mentioning that there are large regional differences in incidence rates of PCa. For example, in Sweden, where there is a long life expectancy and mortality from smoking-related diseases is relatively modest, PCa is the most common malignancy in males, accounting for 37% of all new cases of cancer in 2004 [7].

Prostate cancer screening is now common practice despite the lack of evidence for mortality reduction. One of the few established risk factors for prostate cancer is a family history of the disease, particularly for men with a family history of an early-onset prostate cancer [8]. In the United States, prostate-specific antigen (PSA)–based screening has been recommended by the American Urological Association and the American Cancer Society, especially for men with affected first-degree relatives [9]. Little is known, however, about the impact of PSA screening among men with a family history of prostate cancer. Selective screening of subgroups of the population with an increased risk of prostate cancer may improve program performance, that is, increase the detection rate in the high-risk population and program specificity in the target population (effectively identify men free of cancer), but has the disadvantage of low program sensitivity (only a small proportion of cancers in the target population are detected) [10].

RELATED WORKS

- **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.

- **Decision Support System for Heart Disease Based on sequential Minimal Optimization in Support Vector Machine**

Here, Vadicherla & Sonawane (2013), the proponents of this system claim that computer based Medical Decision Support System (MDSS) is useful for the physicians with its fast and accurate decision making process. They opined that predicting the existence of heart disease accurately, results in saving the lives of patients followed by proper treatment. Their objective was to present a MDSS for heart disease classification based on sequential minimal optimization (SMO) technique (which incorporated its features like high accuracy and high speed) in support vector machine (SVM). In using this method, they illustrated the UCI (University College Irvine) machine learning repository data of Cleveland heart disease database and consequently trained the SVM by using SMO technique. Hence, they also claim that given the ease of use and better scaling with the training set size, SMO is a strong candidate for becoming the standard SVM training algorithm. Training a SVM requires the solution of a very large QP (Quantum Platform) optimization problem. SMO algorithm breaks this large optimization problem into small sub-problems. Both the training and testing phases give the accuracy on each record. The results proved that the MDSS is able to carry out heart disease diagnosis accurately in a fast way and it was reported to show good ability of prediction on a large dataset.

- **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.

- **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.

- **Decision Support System for the Diagnosis of Schizophrenia Spectrum Disorders.**

In 2013, Kahn, Perkins and Lieberman developed a decision support system for the diagnosis of schizophrenia spectrum disorders.

The development of this system is described in four-stages: knowledge acquisition, knowledge organization, the development of a computer-assisted model, and the evaluation of the system's performance. The knowledge is extracted from an expert through open interviews. These interviews aimed at exploring the expert's diagnostic decision making process for the diagnosis of schizophrenia. A graph methodology was employed to identify the elements involved in the reasoning process. Knowledge was first organized and modeled by means of algorithms and then transferred to a computational model created by the covering approach. The performance assessment involved the comparison of the diagnosis of 38 clinical vignettes between an expert and the decision support system. The results showed a relatively low rate of misclassification (18-34%) and a good performance by the decision support system in the diagnosis of schizophrenia, with an accuracy of 66-82%.

METHODOLOGY

A very comprehensive dataset (Almond 2009) consisting of 90,000 instances compiled from the UCI (University of California, Irvine) data repository was used. This dataset was translated into the Attribute Relational File Format (ARFF) in which the five distinct attributes used for this work were highlighted. These attributes are Genetic Mutation, Genome Build, Chromosome, Map and Log P. Hence, the dataset consists of five major columns, each representing the respective attribute.

The dataset was then induced with Classification algorithms namely C4.5 decision trees, Support Vector Machine (SVM), K-Nearest neighbor algorithm and Bayes Classifier Algorithm. The Classification algorithms were evaluated using the Waikato Environment for Knowledge Analysis software version 3.6.7 based on the percentage of correctly classified instances with the C4.5 decision trees having 61.0734%, the Support Vector Machine (SVM) algorithm had 50.0515%, the Bayes Classifier Algorithm had 50.2045% and the K-Nearest Neighbor algorithm had 50.1235%. Sequel to the result obtained from this evaluation the C4.5 decision trees turn out as the Classification algorithm of preference for this research. Thereafter, a decision tree program was written in Java with 28 lines of code for the core program to implement the C4.5 decision tree algorithm that will provide the intelligence for this Clinical decision support system and help it make the right decisions when supplied with patient information.

THE PROPOSED FRAMEWORK

The proposed framework is composed of five basic components namely patient information, knowledge representation/processing, knowledge base warehouse, classification/prediction algorithm and the intelligent generator system.

The patient information is fed into the system by making used of a user or communication interface through which the user can effectively communicate with the decision support system. Thereafter, this information passes through the knowledge representation/processing phase in which the knowledge is identified, synthesized, formalized and aligned.

The knowledge base warehouse stores the preconditions or symptoms necessary for the existence of prostate cancer. This component has a direct connection with the C4.5 decision tree algorithm which emerged as the classification/prediction of preference after the evaluation process that was carried out. This C4.5 decision tree algorithm serves as the brain behind the smooth operation and accurate prediction of the Clinical Decision Support System and enables the system to carry out prediction in the most accurate manner. In carrying out this prediction, the decision tree algorithm recursively matches the patient information with the rules stored in the knowledge base warehouse and selects the most appropriate decision in each case.

The Intelligent Generator System is responsible for the patient feedback and eventual treatment should the patient develop prostate cancer.

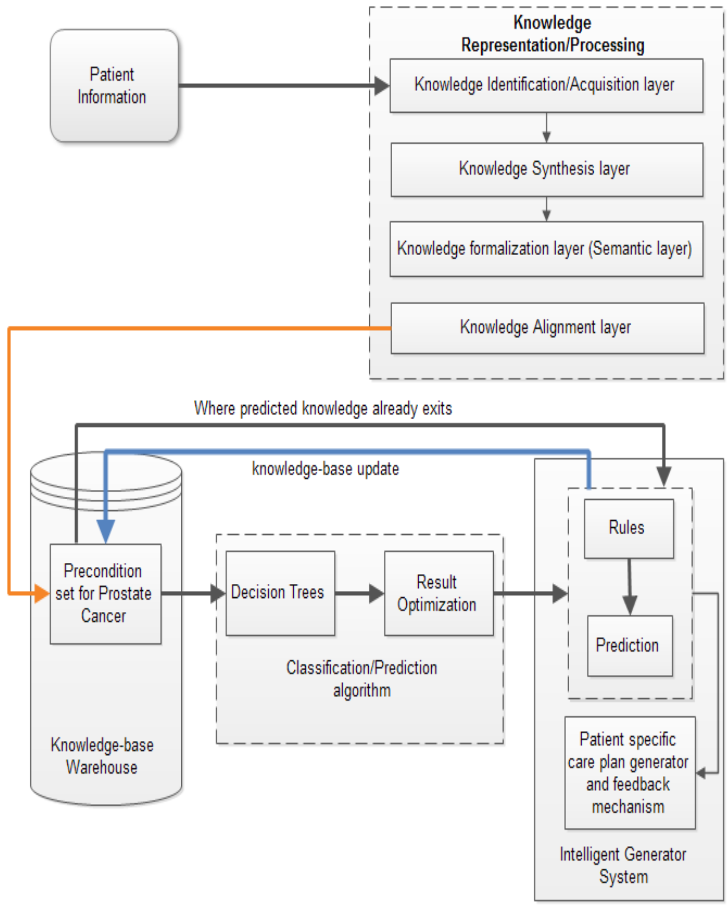

Figure 1: Proposed Prostate Cancer Modeling and Execution (PCME) Framework (Source: Adaptive, 2014)

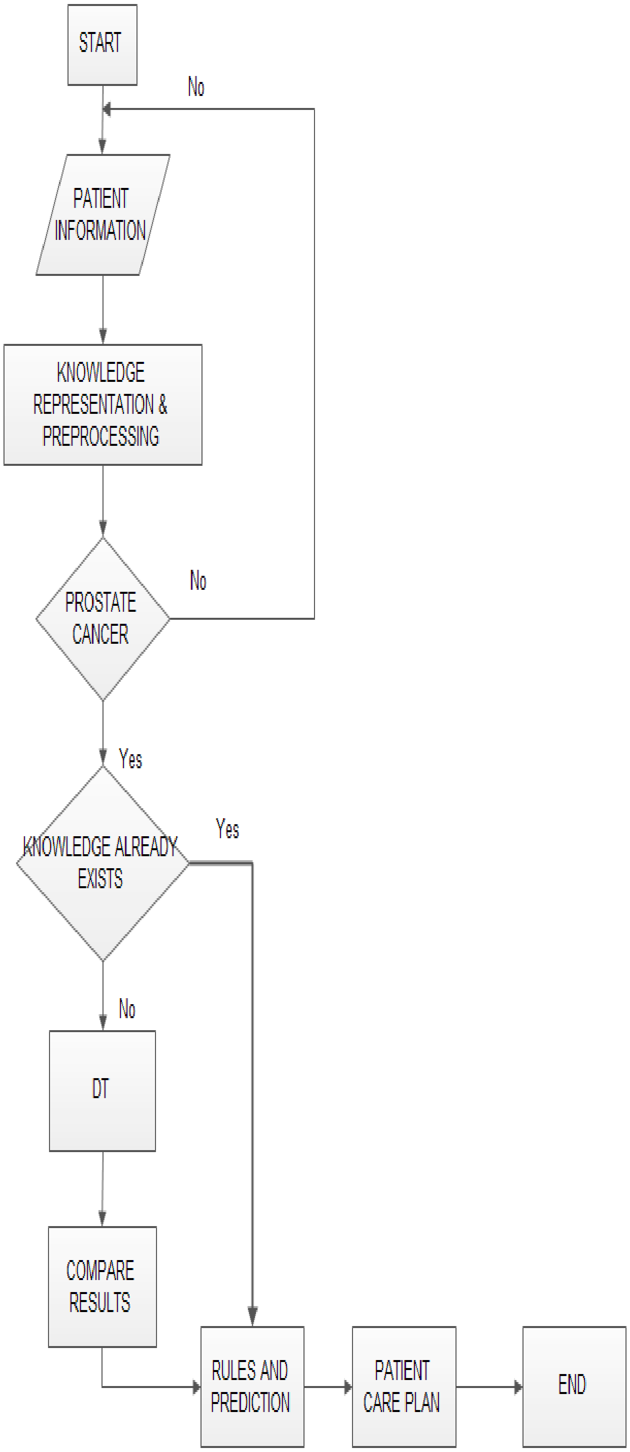

Figure 2: Flowchart for Adaptive Prostate Cancer Modeling and Execution (PCME) Framework

CONCLUSION AND RECOMMENDATION

This research work finds relevance in all parts of the world where men live with the health challenge called prostate cancer, thus it is a very germane work that will help improve the quality of health men affected or who have the tendency to be affected by this menace can have. 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 practitioner that are burdened by the overwhelming and seemingly intimidating number of patients they need to attend to on a regular basis. 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 of this Knowledge Based Intelligent Clinical Decision Support System for the prediction of prostate cancer 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 tremendous results gotten from the use of these clinical decision support systems.

For further work another tenacious researcher can go another step further in this work by introducing other highly efficacious algorithms that can be used alongside the C4.5 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 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.2002 Elsevier Science B.V.

[3] Chi, K.N, Bjartell, A, Dearnaley, D, et al. (2009). Castration-resistant prostate cancer: from new pathophysiology to new treatment targets. Eur Urol. 56(4): pp. 594-605.

[4] Attard, G, Cooper, C.S, De Bono J.S. (2009). Steroid hormone receptors in prostate cancer: a hard habit to break? Cancer Cell 16(6):458-62.

[5] Schröder F.H. (2008). Progress in understanding androgen-independent prostate cancer (AIPC): a review of potential endocrine-mediated mechanisms. Eur Urol 53(6):1129-37.

[6] Haldar, S, Basu, A, Croce, C.M. (1997). Bcl-2 is the guardian of microtubule integrity. Cancer Res 57(2):229-33.

[7] Stapleton, A.M, Timme T.L, Gousse, A.E, et al. (1997). Primary human prostate cancer cells harboring p53 mutations are clonally expanded in metastases. Clin Cancer Res 3(8):1389-97.

[8] Pienta, J.E, Radford, M.J. & Krumholz, H.M (2003). Integrating Clinical Practice Guidelines into Routine of Everyday Practice. Crit Pathway Cardio, Vol. 4, Issue 1, pages 161-167.

[9] Smith, J. L., Dombroski, B. A., Nath, S. K., Lasseter, V. K., Wolyniec, P. S., Nestadt, G., & Pulver, A. E. (2000). Prostate cancer susceptibility loci on chromosomes 13q32 and 8p21. Nature genetics, Vol. 20, Issue1, pages 70-73.

[10] Hakama, N.C (2001). Negative symptoms in prostate cancer: definition and reliability. Archives of Prostate Cancer. Vol. 39, Issue 7, pages 784–788.

[11] 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.

[12] Vadicherla, D. &Sonawane, S. (2013). Decision support system for heart disease based on sequential minimal optimization in support vector machine. International Journal of Engineering Sciences & Emerging Technologies, Feb. 2013. ISSN: 2231-6604 Volume 4, Issue 2, pp: 19-26.

[13] 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). Volume 2, Issue 1, January 2013. ISSN: 2278-1323.

[14] 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. ISSN: 1545-5955, Issue: 99.

[15] Kahn, R., Perkins, D., Lieberman, J.(2012). Predictors of treatment response in patients with first-episode prostate cancer disorder. The British Journal of Gynecology, Vol. 185, Issue 1, pages 18-24.

