IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 3, Issue 2, April-May, 2015
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

A Genetic Algorithm Based Approach To Analyze The Risk Factor Of Heart Disease

Ashish Vashist[1], Ms. Kirti Bhatia[2], Mrs. Asha Vashist[3]

[1]Student M.TECH CSE, S.K.I.T.M., Bahadurgarh(Haryana)
[2]H.O.D in CSE Deptt, S.K.I.T.M. College OF Engineering (Bahadurgarh)
[3]Lecturer in CSE Deptt, D.G.I.T, Barhana(Jhajjar)

Abstract---The purpose of this thesis was to examine heart disease Angina risk factors. In particular, this Thesis was organized around the central theme of adiposity, which is a prevalent Complication following SCI. Study focused on understanding the relationships between activities of daily living (ADL) and risk factors including central adiposity, lipoproteins, and triglycerides. Using genetic algorithm, while controlling for pertinent covariates such as sex, age, and leisure time physical activity (LTPA), it was found that Mobility ADL (wheeling and transferring) were negatively associated with total and LDL-cholesterol. Study also examined whether individuals who considered themselves to be overweight subsequently had less favorable subjective well-being, and were more likely to report specific secondary complications than individuals who did not consider themselves to be overweight. In summary, the findings suggest that a) participation in specific types of ADL (i.e. Mobility ADL) are associated with a lower risk and should be further explored) elevated perceived adiposity is associated with specific secondary complications and lower subjective well-being. Overall thesis findings support the overwhelming evidence of the benefits of daily physical activity and maintaining a healthy bodyweight in the SCI population.

1.INTRODUCTION

Data Mining

How does it work in medical systems? :-

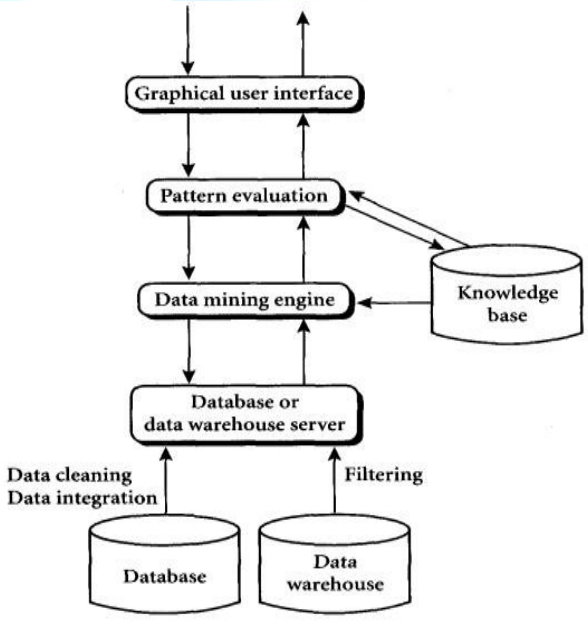
Data Mining is an analytic process designed to explore data in search of consistent patterns or systematic relationships between variables and to validate the findings. Data mining tools help us to predict behavior and future trends, allowing to make proactive, knowledge-driven decisions. it simply means to extract data from huge amount of data for the simplification of the processes to make the processes easier several types of analytical software's are available like neural networks, machine learning, artificial intelligence. Statistics and database system. Four types of relationships found in such systems are:

Classes – using the stored information to access the data for determining a need. For example if a GARMENT shopkeeper wants to increase his sales he can use the predetermined data to study the consumer pattern.

Clusters: Data grouped to a particular set of consumers. For example, data can be mined to determine the pattern of female customers visiting a store to buy a particular product.

Associations: we find link between the data when mining, to find associations among them.

Sequential pattern: we can determine trends and patterns to anticipate behavior of consumers. We can



determine the household purchases of a particular family for a month basis.

Major elements in data mining are as follows:

- We first extract information, then transform it and load the required data into the warehouse.

- Secondly we store and manage the data in multidimensional system.
- We provide access to technology professionals and business analyst's
- Data can now be analyzed using various application software.
- at the last stage we can present the data in a useful manner which may contain charts, graphs and tables.

Fig.1 Data Mining Elements

In Medical Fields

With advancement in technology several methods are available to computerize the process of medical diagnosis. It includes programs that employs probabilistic, statistical methods, knowledge-based systems that use artificial intelligence methods. Variety of people at various premier institutes like MIT and the New England Medical Center used this approach to diagnose and treat people based on the theories of probability and utility. Their main aim was to build systems that could give optimal medical solutions. Rule-based expert systems gained more popularity with changes in the medical field. CASEY is a real world application based on Cased Based Reckoning (CBR) methodology to give a diagnosis for the heart disorders. It functions as a digital therapy advisor. MYCIN (with 450 rules) was developed to diagnose blood infections by Buchanan, Feigenbaum and Shortliff. Other scientists like Phan and Chen designated the use of logic to healthcare diagnostic systems. With the help of a set of sensors the system could monitor the heart rate, blood pressure and body temperature.

Using logic in medical diagnosis is a hopeful technique that could easily capture the necessary medical information and come up with sound diagnosis solutions catering to the needs of the consumers. The present work launch a simple and effective methodology to develop specialist systems for medical diagnosis. The methodology is extensive and can be used in diagnosing a overall of diseases. However to embellish the concept we consider in this paper a set of eight upper respiratory contamination to develop a prototype computer program that can infer proper diagnosis resolution based on patient data.

Medical Diagnosis

Medical artificial intelligence is firstly concerned with the establishment of AI programs that perform diagnosis and make therapy recommendations. Unlike medical applications based on another programming method such as purely statistical or probabilistic methods, medical AI programs based on symbolic models of disease and their relationship with patient's factors and clinical exhibitions . Medical expert systems contain medical knowledge, if we enter the particular symptoms referring to a particular disease it will automatically link the patients' data i.e.; the symptoms with the disease . logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth – truth values between "completely true" and "completely false". It was introduced by Zadeh in 1965 as a means to model the uncertainty of natural language.

In this paper we have developed an expert system that uses a collection of membership functions and rules, instead of Boolean logic to search about data i.e. the patient's symptoms and diseases associated with it. Leung, Lau and Kwong described a general structure of a system that can be used as the core part of the application. The structure can be summarized in the following four steps, carried out in order:
(1) **Composition**: all of the subsets assigned to a particular output variable are pooled together to form a single subset for a particular output variable.
(2) **Fuzzification**: the membership functions defined on the input variables are applied to their actual values, to determine the degree of truthiness for each rule premise.
(3) . **Deffuzification:** is an elective step which can be used when it is useful to convert the output set to a crisp number
(4) **Inference**: the true value for the premise of each rule is calculated and applied to the conclusion part of each rule. The result of one subset is assigned to each output variable for each rule.

Defuzzification methods are available [9], however two of the more common techniques are the CENTROID (center of area) and the MAXIMUM methods. In the CENTROID method, the crisp value of the output variable is computed by finding the variable value of the center of gravity of the membership function for the value. In the MAXIMUM method, one of the variable values at which the subset has its maximum truth value is chosen as the crisp value for the output variable.

2. LITERATURE REVIEW

1 HEART DISEASE RISK FACTORS IN INDIVIDUALS WITH SPINAL CORD INJURY

It only takes an instant to acquire a spinal cord injury (SCI), yet the devastating effects last a life time.[1] At least some degree of paralysis almost always occurs following a SCI. In addition to paralysis, individuals with SCI also often suffer from a myriad of other SCI related illnesses and comorbidities.[2] Overwhelming evidence suggests that individuals with SCI often have a higher level of adiposityi than able bodied counterparts.[3] This increased level of adiposity combined with lower levels of physical activity place individuals with SCI at an increased risk for comorbidities such as coronary heart disease (CHD) and diabetes. Possible ways to mitigate this elevated risk are greatly needed. Leisure time physical activity (LTPA) has been identified a strategy to lower adiposity and CHD risk[.4],[5] However, the role of other daily activities, particularly activities of daily living (ADL), in decreasing adiposity and CHD risk is unknown.

2 EXPERT SYSTEM FOR SUPPORTING DIAGNOSIS OF HEART DISEASES

Case-Based Reasoning (CBR) is a general artificial intelligence prototype for reasoning from experience. CBR methodology has been investigated in improving human decision-making and has received much attention in developing knowledge-
based systems in medicine [16]. A special issue that includes papers on CBR theory and applications was published [8, 9]. Unlike the traditional rule-based approach in which expert knowledge must be represented in "if-then" rules, a case based approach allows knowledge to be grouped and stored as cases. The development of this approach has surged as a key tool for developing a new generation of expert systems [5]. Following to the CBR approach, when a new problem is introduced to the system, theproblem is indexed, and subsequently, the indexes are used to retrieve past cases from case memory. CBR has already been applied in a number of different applications in medicine. CBR is appropriate in medicine for some important reasons; cognitive adequateness, explicit experience, duality of objective and subjective knowledge, automatic acquisition of subjective knowledge, and system integration [3]. Some real CBR-systems are: CASEY that gives a diagnosis for the heart disorders [6], GS.52 which is a diagnostic support system for dysmorphic syndromes, NIMON is a renal function monitoring system, COSYL that gives a consultation for a liver transplanted patient [2] and ICONS that presents a suitable calculated antibiotics therapy advise for intensive care patients [15].

3 Expert knowledge and data mining in a medical diagnosis domain.

It describes a medical diagnosis system in the field of physiotherapy and, more specifically, muscle function assessment based on isokinetic machine data, using an expert system and data mining techniques[5]. An isokinetic machine can be described as apparatus on which patients perform strength exercises. This machine has the peculiarity of limiting the range of movement and the intensity of effort at a constant speed (which explains the term isokinetic). Data concerning the strength exerted by the patient throughout the exercise are recorded and stored in the machine so that physicians can visually analyze the results using specialized computer software.

4 Research on a Cancer Information System

NCI (The National Cancer Institute) is responsible for managing an immense collection of cancer-related information. Part of that information management responsibility involves finding innovative ways to share information in as timely, efficient, and intuitive manner as possible. NCI has therefore instituted a series of small information-sharing initiatives which are publicly available on-line through various links to their World Wide Web (WWW) pages. NCI also shares its digitized collections in a variety of formats (including CD-ROM) as test beds for data mining investigations.

Some of NCI's on-line initiatives involving cancer information include:

– **CancerNet** (http://www.nci.nih.gov) – provides information about cancer, including state-of-the-art information on cancer screening, prevention, treatment and supportive care, and summaries of clinical trials.

– **CancerNet for Health Professionals** – includes access to PDQ and related information on: treatments; screening, prevention and genetics; supportive care and advocacy issues; clinical trials; a directory of genetic counselors; CancerLit topic searches; cancer statistics; and the *Journal of the National Cancer Institute*. (http://wwwicic.nci.nih.gov/health.htm);

NCI's International Cancer Information Center (ICIC) clearly considers that it is essential for the cancer information that it manages to be easily accessible to all levels of medical information users from the very naive to the extremely expert. "Other novel channels of information distribution will be explored to bring cancer information to those who require it, whether health professionals, patients, or policy makers. Appropriate choice cannot be made unless the full range of options is available to these decision makers"

3. EXPERIMENTAL COMPUTATION

The present work introduces a simple and effective procedure to develop expert systems for medical diagnosis. The methodology used is general and can be used in diagnosing a wide range of diseases. However to illustrate the concept, we have just considered Heart Diseases to develop a prototype computer program that can deduce proper diagnosis decisions(**risk factor**) based on patient's data.

1. Existing Work

Medical diagnostic related expert system is one of the favorite's research areas now-a-days. Lot of work is being done in this field now to find new solutions. Each work has its own level of significance. There are no. of disease with lot of symptoms and the diagnosis of these problems with help of one system is difficult to design. Each disease and its symptoms, diagnosis is itself a research area.

Many researchers by this time have done a lot of work with diseases that can be identified by the human normal behaviors. Such diseases include lung problems, cancer etc. There is a no. of approaches being used with such problems like neural network, fuzzy approach etc.

We are offering a project for the heart disease and we are providing its risk estimation using genetic algorithm.

2 Problem Definition

The present work presents a simple and operational methodology for medical diagnosis. Diagnosis of upper respiratory infections is considered here as a vehicle to demonstrate the concept, however the developed methodology is much more suitable for application to a wider range of diseases. We define a set of features *F* relevant to the set of considered diseases *D taken by us for sampling*. The input case to be diagnosed is termed by assigning a value to each feature of the set *F*. Each disease of the set *D* is indicated by its profile in the form of a table obtained by consulting an expert physician. The inference is applied to obtain a decision set for each measured disease, and crunchy decision values are obtained to state the certainty of presence for each disease.

Risk Factor =
(no_of_factor_selected/total_factor)*100

3. Risk factors: Various risk factors causing heart diseases are as follows:-

In the United States, the main cause of death of men and women is heart problmes .Risk factors affecting it includes:

- Age
- Family history
- High blood pressure
- High cholesterol
- Smoking and drinking habits
- Lack of balanced diet
- Obesity
- Inactivity (sedentary lifestyle)
- Other health related problems
- Lack of proper exercising

4. Reference

1. Bauman WA, Spungen AM. Carbohydrate and lipid metabolism in chronic spinal cord injury. J Spinal Cord Med 2001; 24: 266–277.
2. Chen Y, Henson S, Jackson AB, Richards JS. Obesity intervention in persons with spinal cord injury. Spinal Cord 2006; 44: 82-91.
3. Jacobs PL, Nash MS. Exercise recommendations for individuals with spinal cord injury. Sports Med 2004; 34: 727-751.
4. G. Masuda, N. Sakamoto, and R. Yamamoto, "A framework for dynamic evidence based medicine using data mining," In *Proc. 15th IEEE Symposium on Computer-Based Medical Systems*, IEEE press, 2002, pp. 117-122.
5. I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, pp. 89-109, 2001.
6. F. S. Khan, R. M. Anwer, O. Torgersson, and G. Falkman, "Data mining in oral medicine using decision trees," *International Journal of Biological and Medical Sciences*, vol. 4, pp. 156-161, 2009.
7. Y. P. Yun, "Application and research of data mining based on C4.5 Algorithm," *Master thesis*, Haerbin University of Science and Technology, 2008.
8. University of California at Irvine. Reeve-Irvine Research Center. Anatomy 101: Spinal Cord and Central Nervous System. http://www.reeve.uci.edu (accessed 1 October 2008)
9. Mendoza J, Foundas AL. Clinical Neuroanatomy: A Neurobehavioral Approach. Springer; New York. 2008.
10.Maynard FM, Jr, Bracken MB, Creasey G, Ditunno JF, Jr, Donovan WH, Ducker TB Garber SL, Marino RJ, Stover SL, Tator CH, Waters RL, Wilberger JE, Young W.. International Standards for Neurological and Functional Classification of Spinal Cord Injury. American Spinal Injury Association. Spinal Cord 1997; 35: 266-274.

11. Pickett GE, Campos-Benitez M, Keller JL, Duggal N. Epidemiology of traumatic spinal cord injury in Canada. Spine 2006; 31: 799-805.
12. Sekhon LH, Fehlings MG. Epidemiology, demographics, and pathophysiology of acute spinal cord injury. Spine 2001; 26: S2–S12, 2001.
13 Canadian Paraplegia Association. www.canparaplegic.org (accessed 17 June 2008).
14 Canadian Paraplegic Association. Workforce participation survey of Canadians with spinal cord injury. Final Report, 1996. http://canparaplegic.org/vm/newvisual/attachments/676/documents/wforce.pdf
15 Anson CA, Sheperd C. Incidence of Secondary Complications in Spinal Cord Injury. Int J Rehab Res 1996; 19: 55-66.