

Fuzzy Inspired Case based Reasoning for Hematology Malignancies Classification

Odigie . B¹ and Iminavan A.A.²

¹ Federal Polytechnic, Auchu

Received: 11 December 2014 Accepted: 5 January 2015 Published: 15 January 2015

Abstract

Conventional approaches for collecting and reporting hematological data as well as diagnosing hematologic malignancies such as leukemia, anemia, e.t.c are based on subjective professional physician personal opinions or experiences which are influenced by human error, dependent on human-to-human judgments, time consuming processes and the blood results are non-reproducible. In the light of those human limitations identified, an automatic or semi-automatic classification and corrective method is required because it reduces the load on human observers and accuracy is not affected due to fatigue. Case-Based Reasoning (CBR) as a multi-disciplinary subject that focuses on the reuse of past experiences or cases to proffer solution to new cases was adopted and combined with the power of Fuzzy logic to design a software model that will effectively mine hematology data. This study aim at helping the medical practitioners to diagnose and provide corrective treatment to both normal patients and patients with hematology disorder at the early stage which can reduce the number of deaths. This aim is achievable by developing an intelligent expert system based on fuzzy logic and case-based reasoning for classification of hematology malignancy.

Index terms— fuzzy logic, case base reasoning, hematology

1 Introduction

he most commonly ordered or frequently requested test in medicine is hematology tests otherwise known as the Full Blood Picture (FBP). Larry et al ??2012) indicated that the initial Hemogram results are Red Blood Cell (RBC) count, Hemoglobin (HB) concentration, Packed Cell Volume (PCV), Hematocrit, and total White Blood Cell count (WBC), to which platelet count, Lymphocytes and Neutrophils were added as the necessary vital dye stains and the resolution were substantially improved. Hematological or blood data are tightly regulated trails with high clinical relevance. They provide clinical indicators of health and diseases and they are affected by a number of factors even in apparently healthy patients (Kelada et al, 2012;). These factors include age sex, ethnic background, body build and social nutritional and environmental factors especially altitudes and genetic factors.

Several studies have shown that some of these hematological parameters exhibit considerably variations at different periods in life. Higher at birth than at any other period of life, the levels of these parameters decreases as human cells became hypochromic with the development of physiologic iron deficiency hematologic malignancies such as Anaemia. This explains that values outside the normal range are diagnostic for health disorders including cancer, immune diseases and other infections (Kelada et al, 2012). These parameters in turn have to be comprehended in context with vital signs and symptoms.

Conventional approaches for collecting and reporting hematological data as well as diagnosing and classifying hematologic malignancies using techniques such as gene expression signatures are based on subjective professional physicians opinions or experiences which are influenced by human error, dependent on human-to-human judgments, time consuming process and the blood results are non reproducible ??Armstrong et al ,2002; ??even

et al 2003; ??ascal et al, 2008; ??inakshi and Sourabh, 2013). So, an automatic or semi-automatic classification method is required because it reduces the load on human observer and accuracy is not affected due to fatigue. Based on this evidence, the application of soft computing techniques such as fuzzy logic and case base reasoning have continued to expand the horizons of intelligent system development in the medical domain.

Hullermeyer (2007) opted that Case based reasoning (CBR) methodology presents a foundation for new technologies for building intelligent computer aided diagnoses systems. These Technologies directly address the problems found in the traditional Artificial Intelligence (AI) techniques, e.g. the problems of knowledge acquisition, remembering, robust and maintenance. CBR solves a new problem by retrieving and adapting solutions or parts of solutions of a previously solved problem. The idea of CBR has strong appeal because it is recognized that much of human expertise in experimented database and CBR is considered to capture this idea. The main concept that characterized CBR is that expertise reports in a particular field is collected as a repository of cases, and each experience and solution or outcome is confined and archived in the case base for reuse and future reference ??Hoda, 2008).

In another improvement, Livi et al., (2015) maintained that ??adeh (1988) has pioneered the concept of information granularity (IG) which has been conceived with the aim of developing the so-called human-centered computation. Pedrycz (2013) upheld that human perceive and process complex information by organizing existing knowledge along with available experimental evidence and structuring them in a form of some meaningful, semantically sound entities, which are central to all ensuring supporting decision-making activities. Zadeh has expanded IG to include fuzzy logic. Fuzzy Logic (FL) helps computers to paint vivid pictures of the uncertain world by representing and manipulating data that are not precise but rather fuzzy. Fuzzy logic provides an inference methodology that helps appropriate human reasoning capabilities to be applied to knowledge-based system (Obi and Imianvan, 2014).

In this research using a CBR system will assist the physician to quickly diagnose the problem, provide pre-hospitalization aid to the patient and hence reduce the risk of death due to blood parameter in-balance, while fuzzy logic provides the conceptual framework for systematically analysis this un-précised and complex data of hematology parameter. It can recognize the types of blood parameter deficiency based on the evidence indentified from the blood data attributed, symptoms and their severity. The result and decision will be a suggestion that the human expert may accept or consider for further analysis, or completely reject it. We intend to implement this expert system using Matrix Laboratory (MATLAB) Software.

2 II.

3 Review of Related Works

Artificial intelligent prediction is based on human-like learning ability in pattern recognition and generalization better known as machine learning and many machine learning algorithms have been designed and used to analyze medical datasets and proved to be indispensable tools for intelligent data analysis ??Rosma, 2009; Garibaldi and Ifeachor, 2000; Butcher, 2004; ??endonca, 2004; ??iu et al., 2005). Obi and Imianvan (2014) presented an interactive neural fuzzy expert system for leukemia diagnosis. They combine the strength of soft computing techniques of Neural network and fuzzy system to model a medical procedures for indentifying one of the four types of blood malignancies leukemia based on available signs and symptoms. Bendi et al., (2011) performed a critical study of algorithms for Automatic classification used in medical fields and patients with liver diseases were analyzed. KNN, Back propagation and Support Vector Machine (SVM) are giving better results with all these features set combinations. Shahina et al., (2009) designed a computer-aided decision support system for analyzing and diagnosing stress-related disorders based upon finger temperature signals. The method of case-based reasoning is employed to make recommendations for stress diagnosis by retrieving and comparing with previous similar cases in terms of features extracted.

Moreover, fuzzy techniques were incorporated to better accommodate uncertainty in clinicians reasoning as well as imprecision in case indexes. Hoda(2008) proposed a fuzzy case based reasoning for poison classification. The solution presented utilizes computer science field of artificial intelligence and it was realized through the Fuzzy Logic techniques and case based reasoning. The system developed was a complete stand-alone entity that can identify the types of poisoning and its' percentage. It still utilizes the human expert's intelligence. The identification it supplies is a suggestion that the human expert may accept or consider for further analysis, or completely reject. Guessoum et al. (2012) presented the combination of CBR and RBR in order to gather their powers within the same system. RBR has proved its performance in modeling of reasoning which can be explain by humans, that is why we adopted it to conceive the reuse phase of CBR process by an expert system. The inference engine associated to knowledge base and forwarding chaining ensure adaptation task by drawing inferences starting from the diagnosis found in the retrieval phase and basing on some attributes having highest weights whose values will compose the set of facts of our expert system. They also propose some heuristics traced for estimating the similarity on missing data and symbolic descriptors in the retrieval phase. Results show also that these heuristics functions are benefic for the system because they optimize the result of the retrieval phase. Nilsson et al. ??2003) addressed the domain of psychophysiological dysfunctions, a form of stress. The system is classifying physiological measurements from sensors. The system is divided into smaller distinct parts. Measurements, like signals from an ECG, are filtered and improved. A case library of models of distortions etc.

is applied to the filters. Features are extracted from the filtered signals (measurements). An additional set of features are extracted from the first set, for trend analysis etc. The features from the first and second set, and patient specific data, are used as a case. The cases are classified with a k-nearest neighbour match. Schmidt and Gierl (2002) proposed a prognostic model to forecast waves of influenza epidemics, based on earlier observations done in previous years. TeCoMED combines CBR with Temporal Abstraction to handle the problem of the cyclic but irregular behaviour of epidemics.

4 III.

5 Methodology

The traditional process for the medical diagnosis of hematology malignancies or diseases such as Anemia, Leukemia, e.t.c starts when an individual consults a physician (doctor) and presents a set of complaints (symptoms). The physician then requests further information from the patient or from others close to him who knows about the patient's medical history in severe cases. Medical laboratory results such as the full blood count which comprises of the White Blood Cell, Red blood Cell, Packet Cell Count (PCV), Hemoglobin count, Hematocrit count e.t.c is requested and compared with values shown in table 1. Other data collected include patient's previous state of health, living condition and other medical conditions. A physical examination of the patient condition is conducted and in most cases, a medical observation along with other medical test(s) is carried out on the patient prior to medical treatment. From the symptoms presented by the patient, the physician narrows down the possibilities of the illness that corresponds to the apparent symptoms and make a list of the conditions that could account for what is wrong with the patient. These are usually ranked in the order (Low, Moderate and high). The physician then conducts a physical examination of the patient, studies his or her medical records and ask further questions, as he goes in an effort to rule out as many of the potential conditions as possible. When the list has been narrowed down to a single condition, it is called differential diagnosis and this provides the basis for a hypothesis of what is ailing the patient. Until the physician is certain of the condition present; further medical test are performed or schedule such as medical imaging, scan, X-rays in part to conform or disprove the diagnosis or to update the patient medical history. Other Physicians, specialist and expert in the field may be consulted (sought) for further advices. Despite all these complexities, most patient consultations are relatively brief because many diseases are obvious or the physician's experience may enable him to recognize the condition quickly. Upon the Completion of the diagnosis by the physician, a treatment plan is proposed, which includes therapy and follow-up (further meeting and test to monitor the ailment and progress of the treatment if needed). Review of diagnosis may be conducted again if there is failure of the patient to respond to treatment that would normally work. The procedure of diagnosing a patient suffering from hematology malignancies is synonymous to the general approach to medical diagnosis. The physician may carry out a precise diagnosis, which requires a complete physical evaluation to determine whether the patient have the diagnosed disease.

Since this traditional method involve the physician to continuously repeat this processes for several patients which will over time form a case base, adopting our intelligent fuzzy inspired case based reasoning (CBR) model depicted in figure ?? will help in providing solutions to new problem based on past experiment can accurately and promptly assist medical professional in drawing accurate conclusion on type of hematology disease.

IV.

6 Hematological Data Selection

Hamilton and Bickle (2013) and Kim (2011) indicated that most frequently requested hematology test requested in medicine is the full blood picture (FBP). This contains a wealth of information about the components of blood. The typical constituent parts of the FBP are Hemoglobin concentration (Hb), Mean Cell Volume (MCV), Mean Corpuscular HEMOGLOBIN (MCH), Packed Cell Volume (PCV) Red Blood Cell (RBC) distribution width, White Blood Cell (WBC) Count incorporating a differential white cell count, Platelet Count (PC), and Reticulocyte Count (RC) ad their acceptable or normal values are shown in table ??.

Any hematology data outside these ranges are either caused by blood infections or other sources of hematology malignancies (Kim, 2011). V.

7 Global

8 Result and Discussion

To design our fuzzy inspired case base expert system for hematology classification and corrective treatment (see figure ??), we designed a software system which consists of a set of symptoms (physical observations) and hematology parameters of patient which form a new case needed for the diagnosis.

The case base model consists of the database, which consist of cases of past experiences of hematology parameters (Full Blood Count), available signs and symptoms, identified disease and possible corrective treatment (solutions). Signs and symptoms are pre-medical data collected to assist the physicians carry an in-depth medical test on the patient. But the true picture of the patient heath condition is determined by collecting hematology parameter values. The values of the parameters are often vague (fuzzy) and imprecise hence the adoption of

fuzzy logic in the model as means of analyzing these data. The fuzzy set of hematology parameters (attributes) is represented by 'Ha' which is defined as $Ha = \{ Ha_1, Ha_2, Ha_3, \dots, Ha_n \}$ where Ha_n represent the n th parameter or attribute of Ha and n is the total number of parameter in Ha (here $n=6$). For each of the parameter, the set of constraint have been defined which makes it ambiguous to scale properly. You will recall that for each parameter a set of acceptable range for each parameter was given in table ???. Therefore, we have modeled a set of linguistic values that ensures the proper evaluation of the constraints using the linkert scale. The scale denoted as 'HL' is given thus: $HL = \{Low, Normal, High\}$ Low if $Ha_i < \text{parameter range}$ Normal if Ha_i fall within parameter range High if $Ha_i > \text{parameter range}$ Using the above context, assuming Ha_1 for new born child is RBC with range from 5.00-6.50 ($10^6 / \text{cmm}$), Ha_1 Left and Ha_1 Right equals 5.00 $10^6 / \text{cmm}$ and 6.50 $10^6 / \text{cmm}$ respectively. A trapezium fuzzy membership function can be use to classify the hematology data into low, normal or high. If the MF for the Ha_i is at critical lab value (CLV) range (i.e $MF > 0$ as Ha_i increasing from 0 to Ha_1 Left), we say the parameter is abnormal (low). If Ha_i is within normal range ($MF=1$), we say the parameter is in normal condition. When Ha_i drift away from Ha_1 Right to the right (?), MF decreases from 1 to zero), we also experience an abnormal case (high). Both low and high are abnormal cases which indicate a particular hematology disease or malignancy (see table 2 for groupings). Table 2 present the symbolic representation of the various parameter in a search space of S (-1,0,1) which represent Low, Normal and High respectively and it can be represented using a triangular fuzzy Membership function (MF) on the specified range for each parameter (see equation 1). A total membership function value (equation 2) is computed for each case (patient). 0 0 0 1 0 0 Thrombocytosis RO3 0 0 0 -1 0 0 Thrombocytopenia RO4 1 0 0 0 1 0 Polycythemia RO5 1 0 0 0 0 1 Polycythemia RO6 -1 0 0 0 0 -1 Thalassemia RO7 0 0 -1 0 0 0 Microcytic Anemia RO8 0 0 1 0 0 0 Macrocytic Anemia RO9 0 0 0 0 -1 Cardiac Failure RO10 0 0 0 0 0 1 Spontaneous Bleeding RO11 0 0 0 0 -1 0 Heart Failure RO12 0 0 0 0 0 0 Normal/Healthy Condition $\mu_{trap}(x) = \frac{1}{n} \sum_{i=1}^n \mu_{tran}(x_i)$ $n=6$ $i=1$

?????.eq. 2

where $x = i$ th Hematology parameter value, $?$ =zero to Ha_1 Left, $?$ = Ha_1 Left to Ha_1 Right, $?$ = Ha_1 Right to ?

The CBR retrieval process takes place by the application of TMFV fitness function and k-nearestneighbor matching. The goal of retrieval in the CBR system is to retrieve not only exact matches (equivalent cases) but partial matches (similar cases) as well. During the similarity assessment, an explicit context is used; therefore, the retrieval algorithm is based on incremental context transformations. Figure ??? shows a flowchart for the case retrieval and fuzzy inference engine for hematological classification. The advantage of using fuzzy logic is that it allows one to represent the concepts that could be considered to be in more than one category (or from another point of view, it allows representation of overlapping categories).

The fuzzy inference system consists of four modules as shown in figure ???. The Fuzzification module transform the hematology parameter data of each cases (inputs features), which are crisp numbers, into fuzzy sets. This is done by applying a fuzzification function.

Here we propose to use the trapezoidal or triangular MF (see figure 4). The Knowledge base store the IF-THEN rules provided by the experts. Table ???2 shows the logic representation which will be combined with fuzzy union (OR) and Interception (AND) operators. Each rule number will be translated into fuzzy rule. Here, the reuse phase is implemented. The Inference engine simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules so as to identify solution that best describe a rule. Here, the solution is a particular hematology disease. Here, the revise phase of the case base approach is performed. The Defuzzification module transforms the fuzzy set obtained by the inference engine into a crisp value. This crisp value is stored in the case base data bank which is retained into the system case base thereby completing the cycle of the proposed fuzzy inspired case base expert system.

Further, we create the necessary pre and post processing. As inputs are received by the system, the rule based is evaluated (See table 2). The antecedent, which is the (IF X AND Y) construct test the input and produces a conclusion-solution. The consequent (THEN Z) are satisfied which identify or classify the hematology malignancy. The conclusion is combined to form logical sums. Defuzzification coverts the rules base fuzzy output into non-fuzzy numerical values. It reflects the interpretation of the logic of the different linguistic variable.

9 Conclusion

In this study, the amazing gains of combining case based reasoning and fuzzy logic was used to design a software model that would assist health care provider to accurately classify hematology data base on both medical laboratory test and physical observation of patients. This system which uses a set of fuzzified data set incorporated case base provide more precise solution based on expert report of pervious experiences or experiment. The system designed is an interactive expert system that tells the physicians' patient's current condition as regards to evidence provided from the hematology data analysed. Hematology malignancies such as Anaemia, Leukemia. Thrombocytosis, Thrombocytopenia, Polycythemia, Polycythemia and Thalassemia can be identified and properly classified. It should however be noted that the system was able to give prescription of drugs or advice to patients as a corrective treatment. A system of this nature should be introduced in health care delivery centers and hospitals to help ease the work of physicians.

VII.

Future Work

Base on the remarkable strength of fusing Fuzzy logic and CBR technique, we shall develop a robust intelligent expert system that will intuitively provide corrective treatment of various hematology malignancies based on the causes, signs and symptoms and the hematology parameter collected.

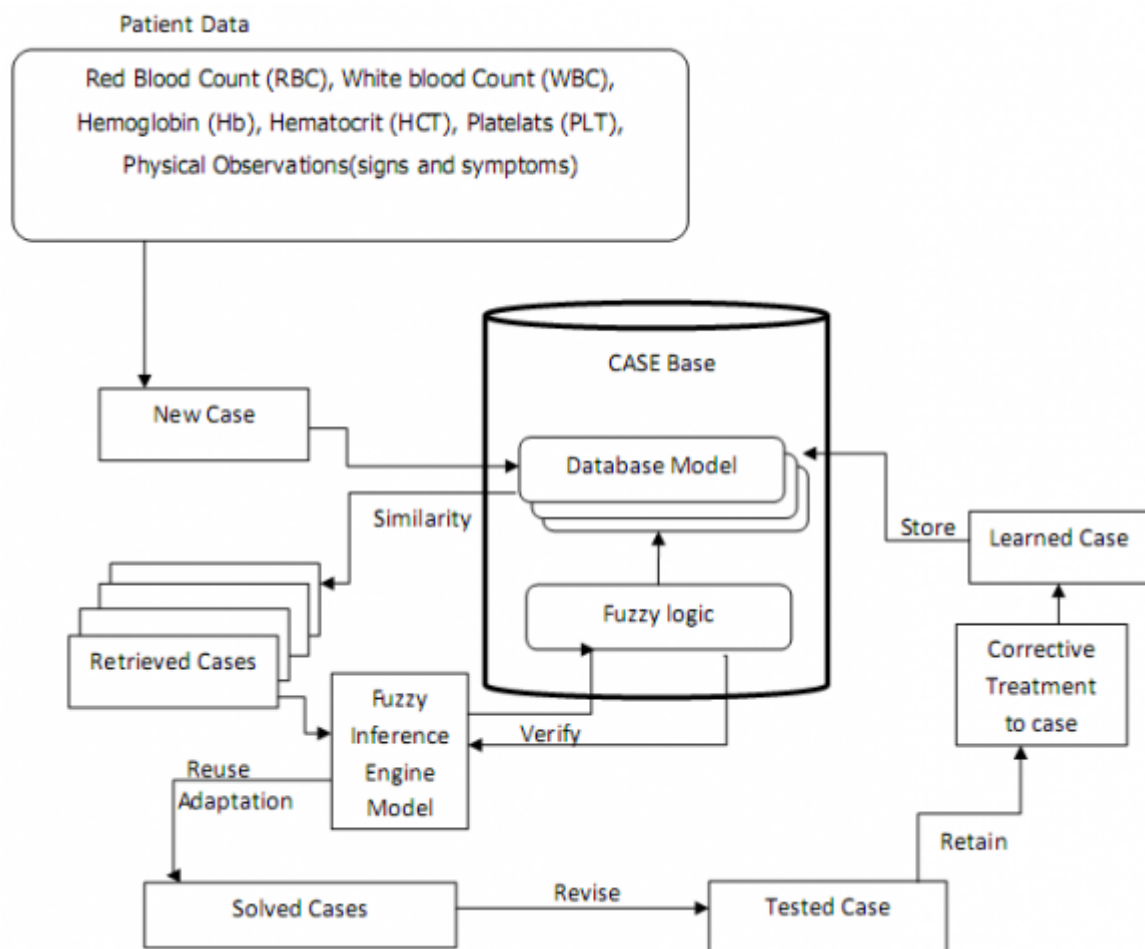


Figure 1: DFigure 1 : 1 :

¹© 2015 Global Journals Inc. (US)

²Fuzzy Inspired Case based Reasoning for Hematology Malignancies Classification

³© 2015 Global Journals Inc. (US) 1



23

Figure 2: Figure 2 :DFigure 3 :

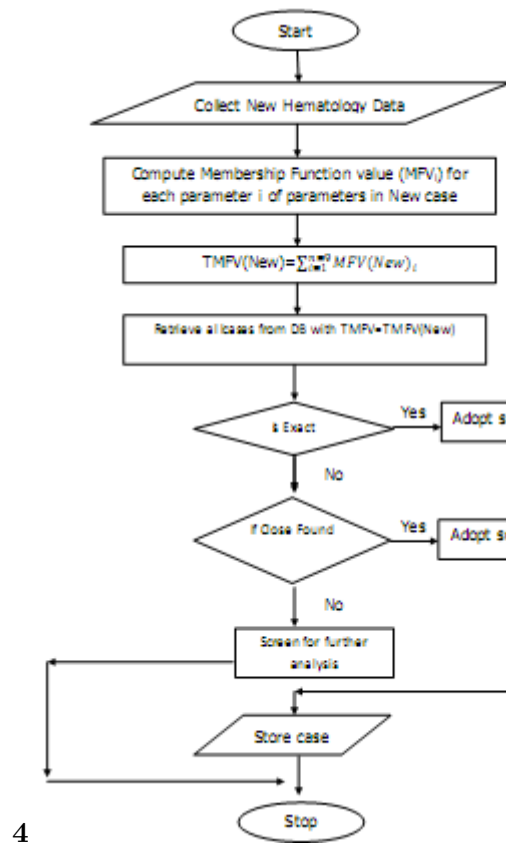


Figure 3: Figure 4 :

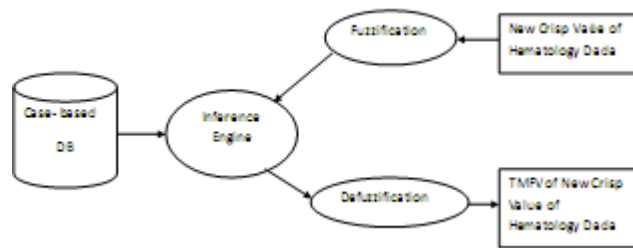


Figure 4:

2

Rule No	RBC	WBC	MCV	PLT	HB	HCT	Hematology Classification
RO0	0	1	0	0	0	0	Leukocytosis
RO1	0	-1	0	0	0	0	Leukopenia, Sepsis, Marrow hypoplasia
RO2							

Figure 5: Table 2 :

.1 International Journal of

Hybrid Information Technology Vol. 5

[Armstrong et al.] , S A Armstrong , J E Staunton , L B Silverman , R Pieters , M L Deboer , M D Minden , S Sallan .

[National Genet] , *National Genet* 30 p. .

[Bendi et al. ()] ‘A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis’. V R Bendi , M P Surendra , N Venkateswarlu . *International Journal of Database Management Systems (IJDMS)* 2011. 3 (2) .

[Guessoum et al. ()] *Combining Case and Rule Based Reasoning for the Diagnosis and Therapy of Chronic Obstructive Pulmonary Disease*, S Guessoum , M L Tayeb , D Hayet , M K Tarek . 2012.

[Butcher ()] ‘Computers in clinical assessment: historical, developments, present status and future challenges’. J Butcher . *Journal of Clinical Psychology* 2004. 60 p. .

[Lander et al. ()] *MLL transactions specify a distinct gene expression profile that distinguishes a unique leukemia*, E S Lander , T R Golub , S J Korsmeyer . 2002.

[Garibaldi and Ifeachor ()] ‘The development of a Fuzzy Expert System for the Analysis of Umbilical Cord Blood’. J M Garibaldi , E C Ifeachor . *Fuzzy Systems in Medicine* 2000. p. .