

# Prediction of Diabetes using Fuzzy Ontology Approach

Mrs. C. Gomathi

Assistant Professor

Department of Computer Science  
and Engineering

Anna University, BIT campus  
Tiruchirappalli, India.

Dr. V. Rajamani

Professor & Principal

Department of ECE,

Vel Tech Multi Tech Engineering  
College,  
Chennai, India.

Ms. K. Jeya

M.E Student

Department of Software  
Engineering

Anna University, BIT campus  
Tiruchirappalli, India.

**Abstract:-** Discover the knowledge from healthcare system is used to improve the health and healthcare delivery. There are many data mining techniques available to discover the knowledge from clinical dataset. This system provides the knowledge about diabetes through fuzzy based ontology. Fuzzy logic is applied to the dataset to extract the knowledge. Fuzzification, fuzzy inference, defuzzification are the steps used in fuzzy logic. Here Trapezoidal membership function is applied to convert crisp value into fuzzy value or linguistic terms in the fuzzification process. This extracted knowledge is represented in the form ontology. Ontology provides the relationship between symptoms and how such factors are depend one another to predict the disease at early stage.

## I. INTRODUCTION

Medical decision is a highly specialized and challenging job due to various factors, especially in case of diseases that show similar symptoms, or in case of rare diseases. Medical diagnosis process for enhancing health-related decisions and actions with, organized healthcare knowledge and patient data to improve health and healthcare delivery. Artificial intelligence in machine learning together with biomedical engineering improves the available clinical dataset into healthcare knowledge to build the clinical decision support system. The proposed system provides healthcare knowledge about diabetes Mellitus disease. Diabetes Mellitus (DM) is the most common metabolic disorder and its prevalence varies worldwide. In developing countries, the prevalence of diabetes is increasing, where there are, as estimated by the World Health Organization (WHO), around 70 million people suffering from diabetes mellitus. In 2013 it was estimated that over 382 million people throughout the world had diabetes. Thus, it is essential that every country attempts to assess the magnitude of the problem and takes steps to control and prevent diabetes mellitus and provide appropriate care. Diabetes is a severe metabolic disorder marked by high blood glucose level, excessive urination, and persistent thirst, caused by lack of insulin actions. There are usually three forms of diabetes—Type1, Type 2, and gestational. Diabetes can produce terrible complications like blindness, kidney failure, and so forth, so need to know how to identify potential cases quickly. There is an urgent need for early diagnosis and treatment of

diabetes to further improve the survival rate and make usual life. Most statistical studies focus on predicting patient survival by analyzing relationships between newly developed or found biomarkers and clinic pathological data. This is mainly due to the fact that patient survival rate is not a simple issue, but related to various factors, such as genetic background of the patient, the lifestyle behavior of the patient, the age of the patient and so on. Knowledge discovery from medical records consists of two phases are data mining, ontology construction. The data mining phase is related to data mining process including data preparation, selection, and extraction of knowledge. The ontology construction phase is related to the process of building the ontology from the extracted knowledge which represents the output of the data mining. Ontologies represent knowledge about the classes of individuals, properties of individuals, and relations between individuals that are possible in a specified domain of knowledge. An ontology of a domain is beneficial in establishing a common vocabulary for the describing the domain of interest. This is important for unification and sharing of knowledge about the domain and connecting with other domains. Ontology represents in the form hierarchical level of graph. Many data mining technique are used to extract the knowledge. This technique provides diagnosis disease with high complexity. The proposed system provides an accuracy of diagnosis with less complexity and to provide good quality healthcare knowledge. Proposed method use fuzzy logic algorithm to construct the ontology graph. Fuzzy logic algorithm is used for extract the knowledge in the form of rule. Based on these rules weightage is assigned to the symptoms in ontology graph construction. Rules contain antecedent and consequent part. Consequent part represents the class that is disease contain or not and antecedent represents symptoms. Ontology constructed by using languages are Web Ontology Language (OWL), XML (extensible markup language). consequent part represent individuals in ontology graph and antecedent part represent class and subclass ,property in ontology. Knowledge accessible and sharable on the Web environment, must convert it into a semantic representation that can be embedded into the contents of Web pages.

## II. RELATED WORK

Regression-based data mining technique is used for predictive analysis of diabetic treatment. Datasets of Non Communicable Diseases (NCD) risk factors in Saudi Arabia was studied and analyzed to identify effectiveness of different treatment types for different age groups (old, young). Oracle Data miner software was used for predicting modes of treating diabetes and support vector machine algorithm was used for experimental analysis [4]. This system describes drug treatment for patients in the young age group can be delayed to avoid side effects. In contrast, patients in the old age group should be prescribed drug treatment immediately, along with other treatments, because there are no other alternatives available. Decision tree induction from the data to extract set of rules in disjunctive normal form and formulate crisp model [5]. Transform crisp set of rules into a fuzzy model and optimize the parameters of the fuzzy model. This system provides automated diagnosis CAD (Coronary Artery Disease) based on easily and noninvasively acquired features, and is able to provide interpretation for the decisions made. Semantics driven semiautomatic method (Intellego) is used for identify the absence of causal relationships between symptoms and disorders in background knowledge and suggests plausible relationships that can rectify missing relationships. In this approach build the initial knowledge base and semantically annotate the EMR documents with concepts from the knowledge base [2]. Symptoms and disorder are to be identified for each document and then generate an intellego: coverage for each document that is aggregates the symptoms for set of disorder. It suggests plausible candidate relationships between disorders and symptoms synthesized from the EMR documents that can rectify the inconsistencies. Validate the suggested relationships by consulting a domain expert. Update the knowledge base based on expert feedback. Vast storage of information was used for diagnosis disease It focuses on computing the probability of occurrence of a particular ailment from the medical data by mining it using a unique algorithm which increases accuracy of such diagnosis by combining the key point of neural networks, Large Memory Storage, and Retrieval, *k*-NN, and differential diagnosis all integrated into one single algorithm [1]. This algorithm can be used in solving a few common problems that are encountered in automated diagnosis these days, which include diagnosis of multiple diseases showing similar symptoms, diagnosis of a person suffering from multiple diseases, receiving faster and more accurate second opinion, and faster identification of trends present in the medical records.

## III. ARCHITECTURE DESIGN

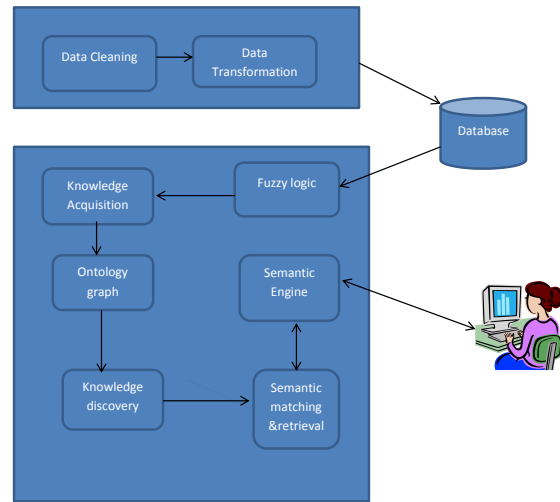


Fig.1 Architecture Diagram

### A. Preprocessing

Pima Indian Diabetes dataset is collected from UCI machine learning repository website which contains missing values, inconsistent and redundant data. These inconsistent data is removed by using data cleaning and transformation method. In data cleaning fill up the missing values, remove inconsistent data, smoothing noisy data and select the relevant attribute. In transformation method transform the inconsistent data values into consistent data values for processing by using normalization that is high values are converted into binary value between 0 to 1. In cleaning method, fill up the missing values by Ignore the tuple, Fill in the missing value manually, a global constant to fill in the missing value, attribute mean to fill in the missing value, attribute mean for all samples belonging to the same class as the given tuple, the most probable value to fill in the missing value. Here missing values fill up by zero.

In transformation data are transformed into consistent form for apply mining techniques. The proposed system applies transformation by min-max normalization.

#### Min-max normalization

$$V' = \frac{V - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A$$

$\min_A$  and  $\max_A$  - minimum and maximum values of an attribute

$\text{new\_max}_A$  - mapping range of maximum value

$\text{new\_min}_A$  - mapping range of minimum value

$V$  - Value of attribute A

$V'$  - New value of attribute A between  $new\_max_A$  and  $new\_min_A$

*B. Knowledge Acquisition*

After preprocessing the dataset fuzzy logic algorithm is used for mine the knowledge from dataset. Fuzzy logic has two functions derive the membership functions for input and output variables and represent them with linguistic variables. This process is equivalent to converting or mapping classical set to fuzzy set to varying degrees and generate fuzzy rule. Membership functions can have multiple different types, such as the triangular waveform, trapezoidal waveform, Gaussian waveform, bell-shaped waveform, sigmoidal waveform and S-curve waveform. Here trapezoidal waveform is used for derive membership function. Fuzzy rule is generated from these membership function.it consists of if and then part of extracted knowledge.

*C. Ontology Construction*

Extracted knowledge is represented by using ontology graph. Knowledge represented is in the form of hierarchical level of graph. Each rule can be represented as an individual (instance) of the class. Rule has antecedent part that is if part is mapped into classes, subclasses, properties in ontology graph. Consequent part mapped into class in ontology. Protege4.2 software tool is used for construct the ontology. OWL (Web Ontology Language) is the language used for construct the ontology. It represents the knowledge about diabetes is used to make decision or diagnosis easier. It is also provides detail about hidden relationship between symptoms.

*D. Query retrieval*

Knowledge accessible and sharable on the Web environment, must convert it into a semantic representation that can be embedded into the contents of Web pages. User wants to access the knowledge through query retrieval and process in semantic engine. Query retrieval consists of semantic matching and semantic indexing. Semantic matching is done by s-match algorithm. In this algorithm semantic matching is done between two trees or graphs. CL matrix represents the relations holding between concepts of labels and CN matrix represents the strongest relations holding between concepts of nodes. These two matrices provide the output for the matching algorithm

*E. Graphical user interface*

User interface is the interface used for query retrieval and query response between user and administrator (semantic engine).Graphical user interface get the input from the user and send it to the semantic engine. Semantic engine search the relevant result for the user query from the database. It provides user-friendly look and feel to the user. Java is used for create the graphical user interface. Jena API is used for provide the

connectivity between java and ontology in protégé 4.2.User enter the detail in the form of name, age, insulin, height, weight, pregnancy detail, gender, glucose, blood pressure, family history, diet, exercises, stress level. Based on these detail search the ontology and provide diagnosis of diabetes disease like type1, type2, gestational.

IV. RESULT

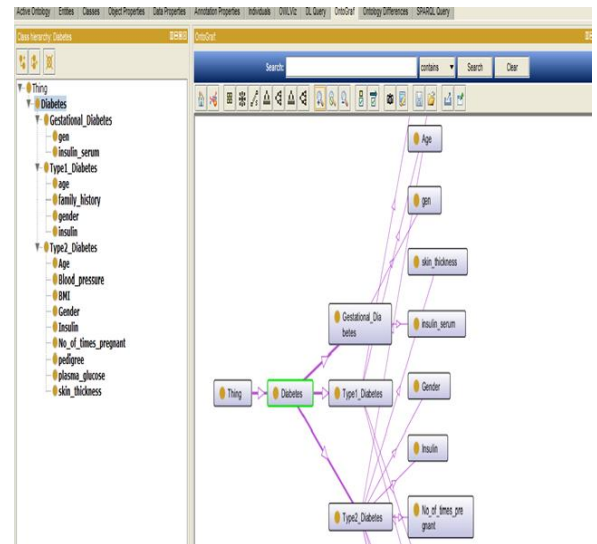


Fig.2 ontology graph for diabetes

V. CONCLUSION

The proposed system provides the accuracy of diagnosis of diabetes disease with low complexity and it also represent the knowledge is in the form of ontology. It allows simply access knowledge by the user. This is the tool to predict the diabetes and to identify people at risk. This detail will be useful for medical field in future.

REFERENCES

- [1] Rahul Isola, Rebeck Carvalho and Amiya Kumar Tripathy, " Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR and k-NN" IEEE Transactions on Information Technology in Biomedicine,2012. Vol.16 No:6 Page no.1287-1295 .
- [2] Sujan Perera, Cory Henson, Krishnaprasad Thirunarayan, Amit Sheth and Suhas Nair, " Semantics Driven Approach for Knowledge Acquisition From EMRs" IEEE Journal Of Biomedical And Health Informatics,2014 Vol.18 No:2 Page no.515 - 524 .
- [3] Mila Kwiatkowska, M. Stella Atkins, Najib T. Ayas, and C. Frank Ryan, " Knowledge-Based Data Analysis: First Step Toward the Creation of Clinical Prediction Rules Using a New Typicality Measure" IEEE Transactions On Information Technology In Biomedicine, 2007 Vol. 11, No. 6,Page no.651-660.
- [4] Abdullah A. Aljumah, Mohammed Gulam Ahamad and Mohammad Khubeb Siddiqui, " Application of data mining: Diabetes health care in young and old patients" Elsevier, 2013, 25 Page no. 127-136.
- [5] Markos G. Tsipouras, Themis P. Exarchos, Dimitrios I. Fotiadis, Anna P. Kotsia, Konstantinos V. Vakalis, Katerina K. Naka, and Lampros K. Michalis, "Automated Diagnosis of Coronary Artery Disease Based

- on Data Mining and Fuzzy Modeling” IEEE Transactions On Information Technology In Biomedicine, 2008, Vol.12, No. 4,Page no.447-458.
- [6] Yogachandran Rahulamathavan, Suresh Veluru , Raphael C.-W. Phan , Jonathon A. Chambers and Muttukrishnan Rajarajan “Privacy-Preserving Clinical Decision Support System Using Gaussian Kernel-Based Classification” ,IEEE Journal Of Biomedical And Health Informatics, 2014,Vol. 18, No. 1,Page no.56-66.
- [7] Saifur Rahaman,” Diabetes Diagnosis Decision Support System based on Symptoms, Signs and Risk Factor using Special Computational Algorithm by Rule Base” IEEE International Conference on Computer and Information Technology ,2012 ,ISBN : 978-1-4673-4833-1, Page no.65-71.
- [8] Widhy Hayuhardhika Nugraha Putra, Sugiyanto, Riyanarto Sarno and Mohamad Sidiq, “Weighted Ontology and Weighted Tree Similarity Algorithm for Diagnosing Diabetes Mellitus”IEEE International Conference on Computer, Control, Informatics and Its Applications,2013,page no.267-272.
- [9]Mostafa Fathi Ganji, Mohammad Saniee Abadeh ,” Using fuzzy Ant Colony Optimization for Diagnosis of Diabetes Disease” IEEE Iranian Conference on Electrical Engg,2010, ISBN:978-1-4244-6760-0,Page no.501-505.
- [10] Huan-chung li and Wei-min ko2,” Automated Food Ontology Construction Mechanism For Diabetes Diet Care” IEEE International Conference on Machine Learning and Cybernetics,2007,Page no.2953-2958.
- [11] About diabetes mellitus:types visit <http://www.m.webmd.com/diabetes/types-of-diabetes>.
- [12] <http://www.wikipedia.com/diabetes>.
- [13]UCI Machine Learning Repository. <http://www.ics.uci.edu/mllearn/MLRepository.html> for dataset.
- [14]<http://www.protege.stanford.edu/protege>.
- [15] Timothy J. Ross University of New Mexico, USA “fuzzy logic with engineering applications”[ ebook].