*Ontology based Disease Prediction and Consultant Recommendation System*

Pirinthika Amirthalingam   
Faculty of Information Technology  
University of MoratuwaColombo, Sri Lanka  
[pirinthika.15@itfac.mrt.ac.lk](mailto:pirinthika.15@itfac.mrt.ac.lk)

Mathangki Sundaram   
Faculty of Information Technology  
University of MoratuwaColombo, Sri Lanka  
[sundaram.15@itfac.mrt.ac.lk](mailto:sundaram.15@itfac.mrt.ac.lk)

Sabthavi Jatharthanathan   
Faculty of Information Technology  
University of MoratuwaColombo, Sri Lanka  
[sabthavi.15@itfac.mrt.ac.lk](mailto:sabthavi.15@itfac.mrt.ac.lk)

Sagara Sumathipala   
Faculty of Information Technology

University of MoratuwaColombo, Sri Lanka  
[sagarasns@gmail.com](mailto:sagarasns@gmail.com)

*Abstract*—Since people today are very busy, they don’t have time to take care of their health. Sometimes consulting a doctor is very time-consuming. People seek online solutions to avoid time wastage. But this is not advisable as these online predictors are considering only the general symptoms and no interdependencies of symptoms are considered are to provide that prediction. This prediction mostly lack accuracy and misleading sometimes. Normally for a differential diagnosis other attributes such as age, family history, risk factors should be taken into consideration. And it will be easy for people if they could express their symptom in natural language than in medical terms. For this purpose, we propose a system where a user can interact with the chatbot in natural language through a series of questions and could predict the disease condition the user has, and the type of consultant user should consult.

Keywords—Natural Language Processing, Ontology, Fuzzy Logic, Fuzzy Inference System, Machine Learning

# Introduction

FINDING THE RIGHT doctor isn't easy—and it shouldn't be. When you put your life in someone else's hands, you need to feel confident that it was a right choice. As common people lack medical knowledge, they face difficulties in choosing the right doctor for their symptoms. It leads them to ignore their symptoms which is sometimes fatal or else mislead to wrong consultants which results in wastage of money and time.

Ontology based disease prediction and consultant recommendation system facilitates the user to interact with the system through a chatbot. It categorizes the user according to age and gender and then asks for symptoms. Then the system asks further questions and predict the possible health condition of the user and suggest the possible consultants. Further the system helps the user to take care of the disease by giving medical advises such as food habits and physical activities for a healthy lifestyle and recommend additional laboratory tests which could be asked by the consultant.

This system mainly contains 3 modules. Those are Disease and symptom extractor, Medical Ontology, Disease prediction module. It is planned to use Natural Language Processing to extract the medical data entered by the user through the chatbot and interact through a series of questions about the symptoms. Ontology will be created using large amount of reliable web data including diseases, symptoms, age range, patient history, impact of family history on a particular disease, risk factors and consultants. Fuzzy Inference System is used for disease prediction.

The structure of the paper is as follows. Section 2 gives an overview of the existing approaches in Disease and symptom extraction, Medical Ontology and Disease prediction with their positives and critiques. Our methodology is presented in Section 3.Then the evaluation details are shown. Finally, the paper concludes with the conclusion and the further works which the next level of research will be carried on.

# Literature Review

## Disease Symptom Extraction

Disease and Symptom extraction in medical domain is generally a difficult task since the medical terminology has different words in day to day life. Existing Disease and Symptom extraction systems can detect the exact medical terminologies which can be understood only by people in the medical field. The major drawback of the existing system is the words that can be tagged are restricted only for clinical documents. The proposed system interact with the texts which is entered by ordinary people.

The study by Qikang Wei et al. (2016) developed a system for Disease Named Entity Recognition (DNER) by combining Bidirectional Recurrent Neural Networks (Bi-RNN) with Conditional Random Field (CRF). The combination of Recognized entities has been done by Support Vector Machine(SVM) algorithm [1]. 500 abstract test set has been used to test the performance of the system. The system achieved F1 performance of 78.04%. However, the system deals with scientific articles. Therefore, it has a boundary detection for the input. This led to this system into performance loss. This system does not have any algorithms to detect misspelled or partially typed word because it is only for scientific articles.

Another system proposed by Qin Xiaona et al (2015) has used Conditional Random Field (CRF) that combines the features of Maximum Entropy Markov Model (MEMM) and Hidden Markov Model (HMM) [2]. This system is suitable for symptom extraction in Traditional Chinese Medical Records. 500 TCM medicine records were used in the experiment to test the system. The system has reached 83.98% of F1 measure. However, this system only for extracting the symptoms which are in the Medical records.

The study by Fan Tong et al. (2017) presented a Deep learning approach. The system has Bi-directional Long Short-Term Memory layers to capture long term context information and fully connected layers to improve the performance [3]. The system has achieved the F-Score of 89.58%. The entire system is only for Disease Named Entity Recognition in high level biomedical texts. The system does not detect other entities like symptoms.

Another approach by Hyeju Jang et al. (2016) presented an approach by using Hidden Markov Model (HMM) to tag the entities in clinical documents [4]. This system assigns concepts like Symptom, Therapy, Performance to the phrases in the medical records. The training corpus of this system was built manually, and this is the serious limitation of this system. The training corpus is not enough to show that the system is efficient. The result of the system is not good as expected.

The systems mentioned above are proposed only for clinical documents and medical journals which contains high level medical text that are used by people or experts related to medical field. In the proposed system, the Disease and Symptom extraction is used to extract the words from the text of ordinary people who are not professionally related to the medical field.

## Medical Ontology

The Ontology provides semantics for the search and always tend to give meaningful results. As because of this ontology development attracted all the domains. Medical domain is not an exception. Medical domain covers a vast knowledge area about the symptoms, diseases, animals ,microorganisms, treatments , checkups ,precautions , jargon etc. It exists in unstructured manner all over the web. Prabath Chaminda Abeysiriwardana et al. (2012) [7] proposed a method for Gather medical data from different sources and creating a huge knowledge domain. They defined classes and relationships and created the ontology using skeleton methodology. It consists diseases and symptoms only. They created a simple website with searching capabilities to access the ontology and evaluated it with the help of the medical domain experts.

Osama Mohammed et al(2012) [14] integrated two existing medical ontologies manually using alignment algorithm. It was limited to a smaller number of diseases. But they believe this method could be repeated for any large number of diseases and symptoms. Linda Mhadhbi et al(2017) [5] proposes creation of disease symptom ontology with the help of the existing biomedical ontologies to help the physician to find all the related symptoms of a disease. Heiner Oberkampf, Turan(2015) [15] proposes a model for disease symptom relations using existing ontology mappings to propagate semantic type information for disease and symptom across ontologies. Here they increase the semantics of the ontology by clustering diseases and symptoms and giving them a common label. But these approaches never given 100% content coverage.

## Disease Prediction Module

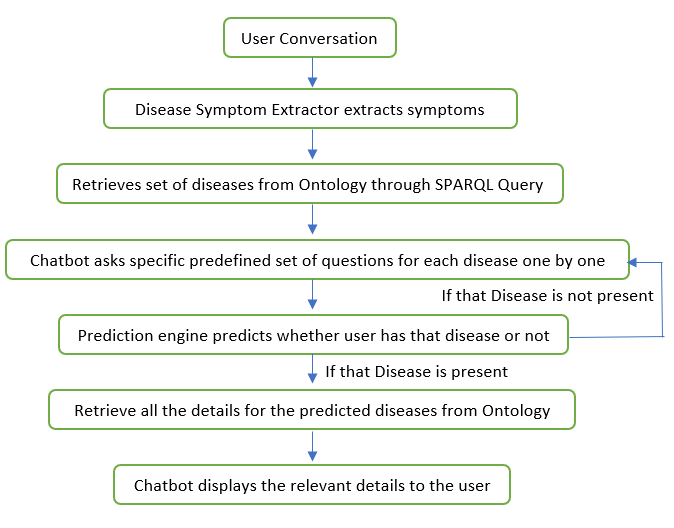
In existing systems, there are many classification methodologies used to predict diseases such as Decision tree, Neural Network, Support Vector Machine, Naïve Bayes etc. Ebenezer Obaloluwa Olaniyi et al (2015)[9] used feedforward multilayer perceptron, and support vector machine to analyze heart disease dataset which was taken from UCI machine learning repository. They later compared the performance of these two techniques through Specificity, Sensitivity, Accuracy, Positive Prediction Value (PPV), Negative Prediction Value (NPV). The accuracy obtained are 85%, 87.5% for feedforward multilayer perceptron, and support vector machine respectively.

Asma A. AlJarullah (2011) [10] proposed a decision tree technique to predict diabetes by analyzing Pima Indians Diabetes Data Set. His system mainly has two phases such as data preprocessing phase which includes attribute identification and selection, handling missing values, and numerical discretization and diabetes prediction phase which is constructed using J48 decision tree classifier. The accuracy of 78.1768% was obtained from this system.

B. Venkatalakshmi et al (2014) [11] used Decision tree and Naive Bayes to analyze 13 attribute structured clinical database from UCI Machine Learning Repository. They compared the performance of these two techniques through Specificity, Sensitivity, Accuracy, Precision and Recall. The accuracy obtained are 85.03%, 84.01% for Naïve Bayes and Decision Tree respectively. P.B. Khanale et al (2015) [12] used the fuzzy verdict mechanism to predict diabetes and achieved the accuracy of 87.2% by analyzing Pima Indians Diabetes Data Set. Azian Azamimi Abdullah et al (2011)[13] designed a Fuzzy Expert System for the diagnosis of Hypertension. They performed the analysis on the sample database which consists records of 10 people from both genders and different working backgrounds. This system used expert’s knowledge and existing literature for diagnosis and modeling linguistic variables.

# Methodology

In this system chatbot interacts with the user and collects user responses. Then, Disease Symptom Extraction module extracts the symptoms from user responses and these extracted symptoms hit the ontology through SPARQL query and retrieve set of diseases. Then, Chatbot asks specific predefined set of questions for each retrieved disease one by one and sends the responses to prediction engine. Prediction engine predicts whether the user has that disease or not. If the disease is present, then predicted disease hits the ontology through SPARQL query and retrieve all the details and chatbot displays those details to the user. If the disease is not present, then the chatbot asks specific predefined set of questions of next disease and the process continues.

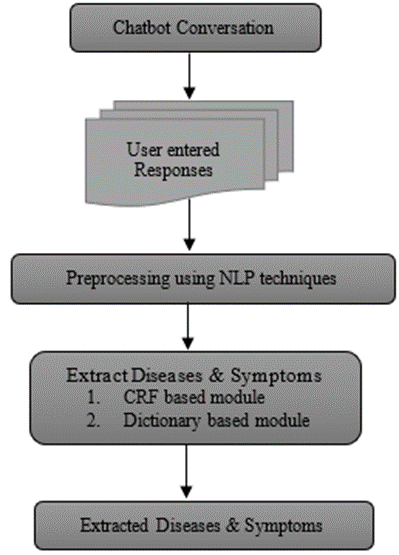


## Disease Symptom Extraction

According to the Literature review, the existing systems are dealing with high level medical text which can be found in Medical journals or Medical textbooks which can only be understood by the people who are in the Medical field. Some of the systems are dealing with only Diseases and some of the systems are only dealing with symptoms. Some systems are dealing with genes, proteins, drugs etc.

In our system we are going to deal with the ordinary texts which is entered by the patients (people who are not from the medical field). Also, our system deals with extraction of both Diseases and Symptoms from the text. All other systems are dealing with only diseases or only symptoms. Another novelty is, in our system we have bidirectional approach which contains a machine learning module which uses Conditional Random Fields and a dictionary-based module.

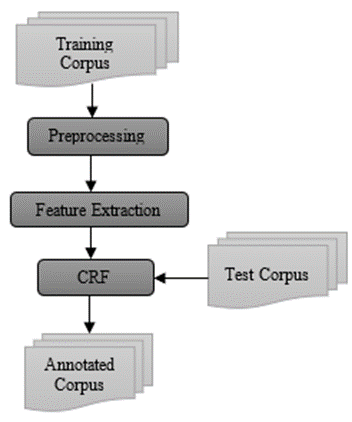
However, the system is dealing with the user entered text, therefore there is a chance of misspelled words and words that have same meanings but different in the spelling. Here we are overcoming the issues by using a dictionary-based module. There we are having two algorithms to deal with that problem. Fuzzy string searching technique and Minimum edit distance calculation is used to deal with those words. Therefore, the system can catch the diseases and symptoms even though the user makes a mistake. It can extract the exact symptom or disease from the user text. By this approach the system achieves the novelty in Disease and Symptom extraction.

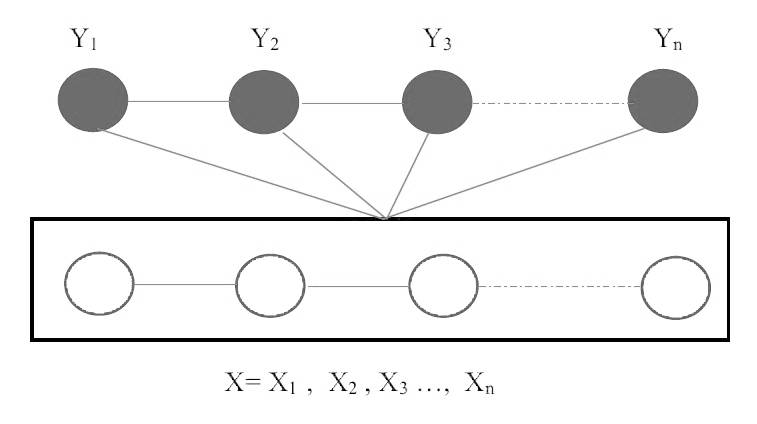


The above methodology is the approach of the Disease and Symptom extraction from the user responses. The Chatbot will build a conversation with the user and the responses from the user will be separated. The separated responses will be sent to the preprocessing module.

The Disease and Symptom Extraction consist of two modules.

### 1) Conditional Random Field based Module



Since Conditional random Field model has achieved good results in identification of Entities tasks, the feature based module uses the model. This model is proposed as an undirected graph model as shown in Figure 3. The conditional probability of the output is calculated based on the input nodes The corpus used to train and test the model is the NCBI Disease corpus.

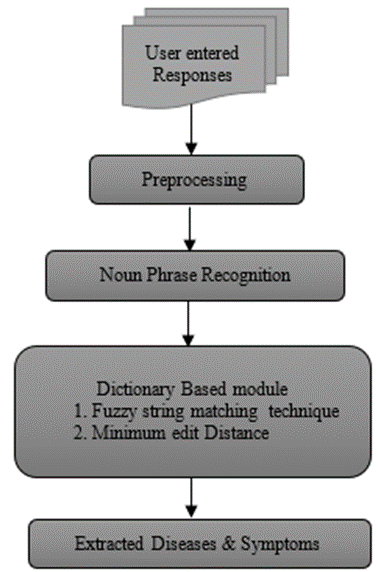
The above graph represents the first-order linear chain Conditional random Field. X indicates the sequence of observations and Y indicates the sequence of Output. When there is a set of sequences given as x = {x1,x2,……xn} the Conditional Random Field tries to model the probability p(y|x) of the output sequence of y = {y1,y2,……yn}. The conditional probability is calculated by the following equations.

(1)

Where

(2)

### 2) Dictionary based module



This module is important to get the user entered symptoms by matching the identified Noun phrases with the dictionary. For the diseases CTD Disease dictionary is used. For the symptoms, a manually created dictionary is used. In this module only the Diseases and Symptoms in the dictionary will be identified.

As the user entered responses may contain misspelled words, there are two techniques used to overcome the issue.

#### Fuzzy Sting matching technique

This technique finds the strings that match with a pattern approximately. This finds a disease or symptom even the user misspelled the word or enter only the partial word of the disease or symptom. Exact string matching can cause errors due to misspelling of words. Therefore, this technique has been used in the dictionary-based module.

#### Minimum Edit Distance

The Edit distance of the Noun phrases will be calculated with the Dictionary words. The word which has the Edit Distance less than or equal to three, the word from the dictionary will be taken.

## Medical Ontology

In this study ontology act as a knowledgebase. It provides all the information needed by the system for disease prediction and the recommendation purposes. There are ontologies available specifically for certain diseases. But there is none for diseases, symptoms, disease prevention methods, laboratory tests and consultants combined in a single ontology.

We have chosen ontology as the knowledgebase because it is reusable, extendable, semantic and reliable. When compared to relational database ontology can be added with meaning and comprehension also it is interoperable with heterogenous information sources. The goal is to provide a huge ontology consists information about diseases, symptoms, risk factors, impact of family history and recommendations like food habits, physical activities and consultants.

Ontology creation divided into two main parts. They are,

1. Creation of base ontology
2. Integration of other existing ontologies

**Base ontology**

This is built from scratch and it acts as the skeleton of the whole knowledgebase. For now, we have created this base ontology with six main diseases such as diabetes, kidney stone, tuberculosis, hypertension, myocardial infarction and nephrolithiasis.

There are many methodologies used in ontology development such as skeleton methodology, METHONTOLOGY, top-down or bottom-up or combination of both development processes, KACTUS engineering, TOVE, Enterprise Model Approach etc. Within this METHONTOLOGY methodology shows some success in building complex ontologies [16]. So, this approach is used to build the proposed medical ontology. METHONTOLOGY starts by identifying the following activities that are involved in the development of an ontology,

### 1) **Specification**: Identifying the purpose of the ontology.Here the purpose is to provide a vast knowledge on medical domain and to interact with the prediction engine.The intended users of the system is patients and medical students. Firstly term dertermination should be done which means identifying the list of terms in the domain.(Eg: Demographics, disease, symptom, labtest,consultants)

### 2)**Knowledge acquisition**: As this is in medical domain accuracy and reliability is very important. So, this ontology is created using data from medical documents, standard medical books with the help of knowledge from medical experts and clinicians and from following websites which allow their content for secondary research purposes.

* MedlinePlus and Daily Med by the National Institute of Health’s US, National Library of Medicine.
* Mayo Clinic and Jhon Hopinks. Both are world’s leading institution of medicine and science.

### 3) **Conceptualisation**: Domain terms are identified as concepts, instances, verbs relations or properties and each are represented using an applicable informal representation. Classes, subclasses, property,subproperty and annotations are defined.

Diseases and symptoms were classified and grouped according to the medical standards (following the taxonomy of disease ontology and symptom ontology available in BioPortal).

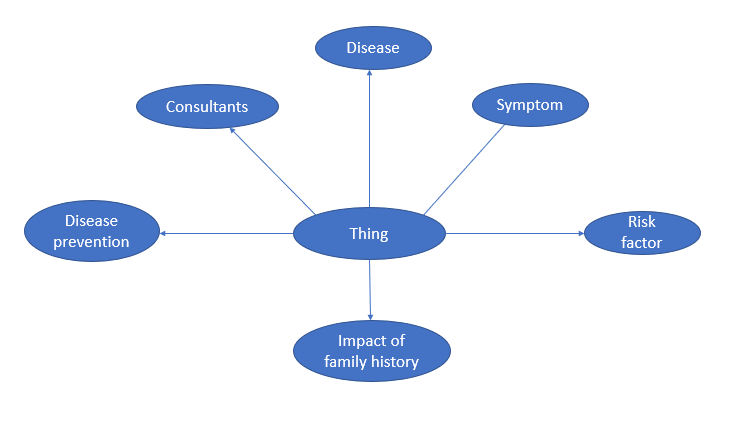


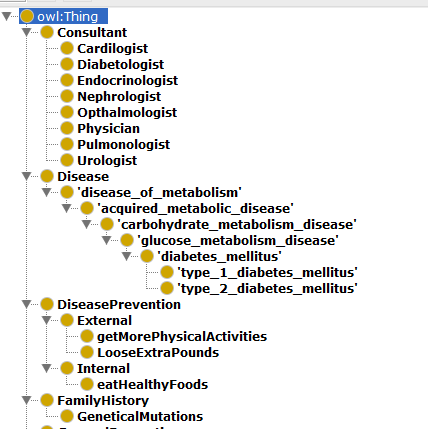
Fig 6: Main classes

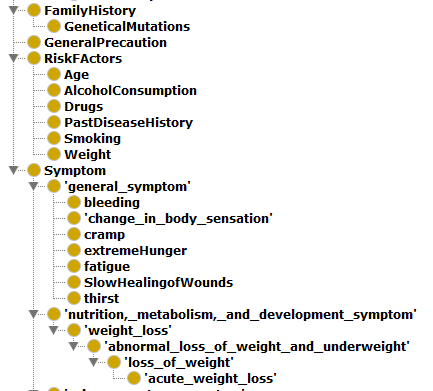
The major classes are defined as shown in the diagram and subclasses and relationships are defined according to the standardization of disease ontology and symptom ontology.

### 4)**Integration**: No integration in this phase.

### 5)**Implementation**: This ontology is implemented using Protege 5 and OWL 2.

#### Ontology capture: Key concepts and relationships are identified in the domain of interest. Identified classes are arranged in a hierachy by specilization and generalization. Relationships between instances, data properties and object properties are given. Annotations are given using the standard medical thsaures.





### 6)**Evaluation**: Validating the ontology using medical domain experts using SPARQL queries. Consistency of the ontology is checked with HermiT and FaCT++ resoners.

### 7)**Documentation**: Document ontologies according to the type and purpose.

**Integrating existing ontologies**

Standard ontologies for several diseases are available in the BioPortal. To get a huge ontology with all the needed content these standard ontologies for the selected diseases are integrated with the base ontology. This is done using overlapping concepts that has semantic equivalence.

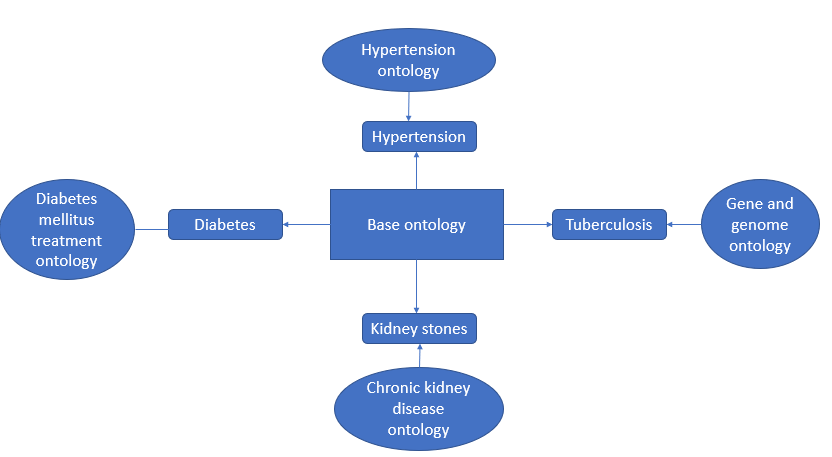


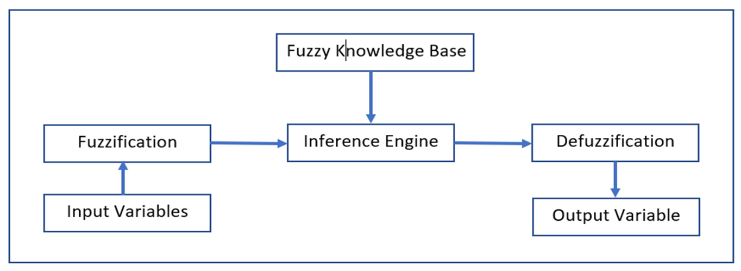
Fig 10: Ontology integration

Ontology mapping is used to align the base ontology with each standard ontology. For this alignment there should be equivalent entity in the base ontology and the corresponding target ontology. This is done manually but this can be automated.

This ontology has achieved hundred percentage content coverage for both disease prediction and recommendation.

## Disease Prediction Engine

In this study, Fuzzy Inference System (FIS) is being developed for disease prediction. Instead of other similar approaches, Fuzzy Inference System is the most suitable one in Medical diagnosis where uncertainty and vagueness plays the major role. This system relies on expert’s knowledge for determining fuzzy sets and selecting most appropriate fuzzy rules. In this study, as a prototype we have developed a system to predict diabetes.



This module uses Mamdani type and consists 7 main steps. Those are data collection, determine fuzzy sets, build membership functions, Fuzzy rule base development, Fuzzification, Fuzzy Inference and defuzzification.

### 1) Data Collection: This is the first step of this proposed module. A survay (<https://forms.gle/uTR7mSe5LhmAKUvq9>) which was prepared under the supervision of domain expert. was conducted among 65 patients which includes both males and females at age range between 18 to 80 years. This module includes 6 input variables which consists Age, BMI, Polyuria (Increased urination in day time), Nocturia (Increased urination in night time), Polydipsia (Increased thirst), Polyphagia (Increased hunger) and 1 output variable which shows whether the patient has diabetes or not.

### 2) Determine fuzzy sets for all variable*:* Fuzzy sets have been determined for each variable with the aid of domain expert. Table 1 shows Fuzzy sets, Range, Fuzzy triangular values for each variable.

1. FUZZY SETS OF VARIABLES

| ***Variables*** | ***Range*** | ***Fuzzy sets*** | ***Fuzzy Triangular Values*** |
| --- | --- | --- | --- |
| Age | 18-34  33-55  54-80 | Young  Mid Age  Old | [18, 26, 34]  [33, 44, 55]  [54, 67, 80] |
| BMI (kg/m2) | 0-18.5  18.4-24.9  24.8-29.9 29.8-50.0 | Underweight  Normal  Overweight  Obesity | [0, 9, 18.5]  [18.4, 21 , 24.9]  [24.8,27 , 29.9]  [29.8, 39, 50] |
| Polyuria (Frequency of Urination per day) | 0-4  3-9  8-15 | Low  High  Very High | [0, 2, 4]  [3, 6, 9]  [8, 11, 15] |
| Nocturia (Frequency of Urination per Night) | 0-2  1-4  3-8 | Low  High  Very High | [0, 1, 2]  [1, 2, 4]  [3, 5, 8] |
| Polydipsia (Frequency of thirst per day) | 0-4  3-7  6-14 | Low  High  Very High | [0, 2, 4]  [3, 5, 7]  [6, 10, 14] |
| Polyphagia (Frequency of hunger per day) | 1-3  2-5  4-10 | Low  High  Very High | [1, 2, 3]  [2, 4, 5]  [4, 7, 10] |
| Output (Blood Sugar Level - mmol/L) | 4-5.8  5.5-30 | Non-Diabetes  Diabetes | [4, 7, 9]  [7.8 , 17, 30] |

### 3) Build membership Functions: Membership function have been determined for each variable. Here we have used triangular membership function as it is more accurate than the other membership functions such as trapazoid, sigmoid. It was expressed using three points as (a, m, b).

0 if x<a

x-a/m-a if a<=x<=m

μA(x) **=**  b-x/b-m if m<x<=b

0 if x>b (3)

The triangular membership function equation for fuzzy set Age (young) = [18,26,34] is shown below.

0 if x<18

x-18/26-18 if 18<=x<=26

μA(x) = 34-x/b-26 if 26<x<=34

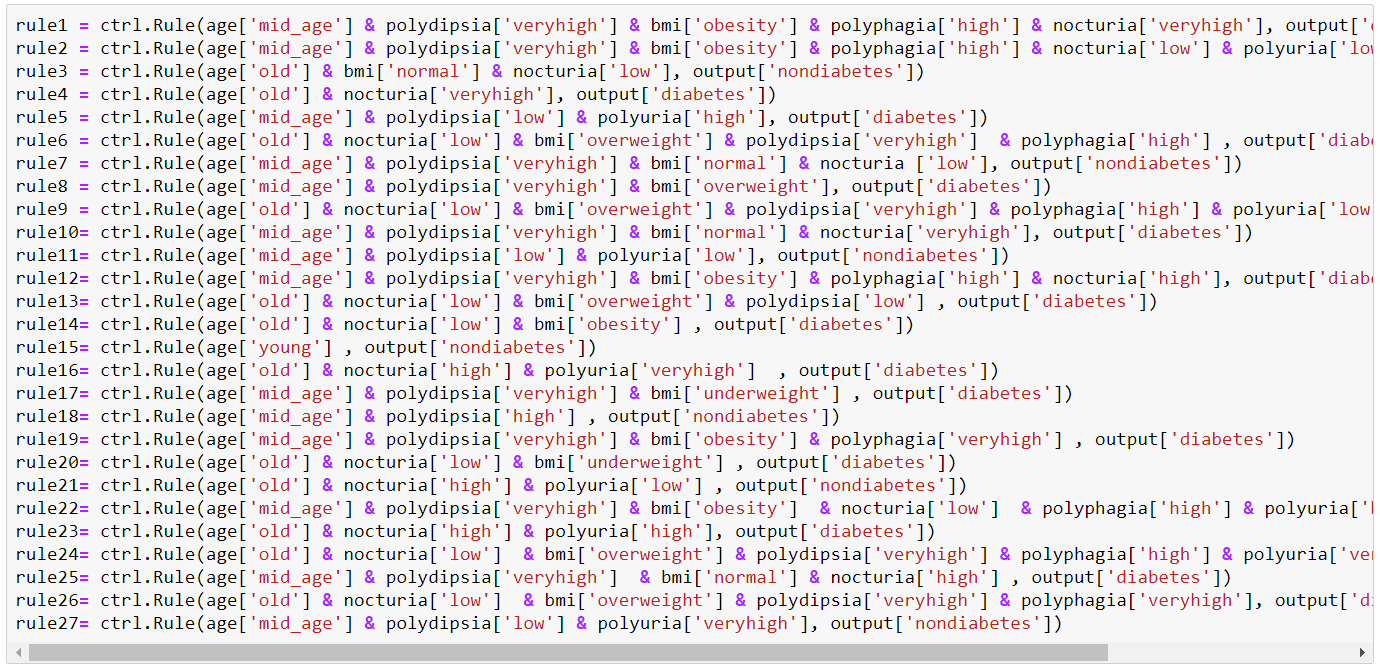
0 if x>34 (4)

### 4) Fuzzy Rule Base Development: Fuzzy rules have been created to map fuzzy inputs with the fuzzy output. Here Fuzzy decision Tree was generated from the training data (50 patients) using a standard decision tree learning algorithm called FID3 and Fuzzy rules were extracted from the generated decision tree[21]. Then the most accurate rules are selected by the domain expert. This module contains 27 fuzzy rules which are in the form of R1: If a is X then b is Y, where X and Y are linguistic values of defined fuzzy set. Some of the extracted rules are given below.

Rule 1: If (Age is mid\_age) and (Polydipsia is very\_high) and (BMI is obesity) and (Nocturia is very\_high) and (Polyphagia is high) then ( Output is Diabetes )

Rule 2: If (Age is mid\_age) and (Polydipsia is very\_high) and (BMI is obesity) and (Nocturia is low) and (Polyphagia is high) and (Polyuria is low) then ( Output is Diabetes )

Rule 3 : If (Age is old) and (BMI is normal) and (Nocturia is low) then ( Output is NonDiabetes )



### 5) Fuzzification: It is the process to map the crisp input to fuzzy values from 0 to 1 using built membership functions. These inputs are mapped into fuzzy values by drawing a line up from the inputs to the input membership functions above and marking the intersection point.

### 6) Fuzzy Inference: Then these IF-THEN rules are evaluated to produce fuzzy output from the fuzzy input by using inference engine. It has 3 operations.

#### Applied AND operator in the antecedent of each rule to build rule strength.

#### Then the rule strength is combined with output membership function by using min method to find the consequence of each rule.

#### Finally, the consequences are combined to get the output distribution using max method.

Figure 24 illustrates these operations.

### 7) Defuzzification: It is the last step of this module which transforms output distribution to the final crisp output. There are many defuzzification techniques such as Center of Gravity, Mean-Max Membership, Weighted Average Method. Here, we have used Center Of Gravity method. This method finds the point z\* where a vertical line would slice the output distribution into two equal masses.

Mathematically defuzzified value z\* defined as :

z\* = (5)

where is the membership in class A at value .

# Evaluation

Each module is evaluated as follows.

## Disease Symptom Extraction

The Disease and Symptom extraction module’s accuracy is measured by using Precision (P), Recall (R) and F-measure (F-score). The formulas for the evaluation are given below.

P = TP / (TP + FN) (6)

R = TP / (TP + FP) (7)

F – Score = (2 x P x R) / (P + R) (8)

Where TP indicates True Positive, FP indicates False Positive and FN indicates False Negative. F – measure is the average of combined Precision and recall.

Table II shows the performance measure of the Disease and Symptom extraction module.

1. PERFORMANCE MEASURE OF DISEASE SYMPTOM EXTRACTION

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision(%)** | **Recall(%)** | **F-measure(%)** |
| Bi-RNN, CRF | 76.57 | 79.57 | 78.04 |
| MEMM, HMM | 84.68 | 83.37 | 83.98 |
| Bi-directional LSTM | 89.16 | 90.0 | 89.58 |
| HMM | The Paper didn’t provide | | |
| Our Approach | 89.31 | 89.0 | 89.62 |

## Medical Ontology

Here the medical ontology is evaluated in two ways.

1. Technical evaluation- Using ontoclean methodology [17] correctness of taxonomy, defined classes, equivalent classes and relationships were evaluated.
2. Content evaluation- This was done by hitting the ontology with different SPARQL queries and the results was checked by medical experts. This showed 100% accuracy.

## Disease Prediction Module

The performance of this module is evaluated with the help of sensitivity, specificity, accuracy. In order to find these measurements first we want to compute True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). These values are calculated by comparing the prediction result of this system with expert diagnosis result.

Figure 26 shows the Confusion Matrix of training data which consists 15 patients.

A. Sensitivity: The ratio of actual positives which are predicted correctly.

Sensitivity = TP/ (TP + FN) = 0.875 (9)

B. Specificity: The ratio of actual negatives which are predicted correctly.

Specificity = TN/ (TN + FP) = 1 (10)

C. Accuracy: The sum of True Positive and True Negative divided with the total number of patients.

Accuracy = (TN + TP)/ (TN + TP + FN + FP) = 0.93 (11)

By analyzing the dataset of 15 patients, this module achieved the 93.33%. of Accuracy.

Table 3 shows the comparison of our proposed module with the existing systems with similar fuzzy approach.

1. COMPARISON OF PROPOSED MODULE ACCRACY WITH THE EARLIER SYSTEMS

|  |  |  |  |
| --- | --- | --- | --- |
| ***Authors*** | ***Year*** | ***Method*** | ***Accuracy*** |
| V. Jain et al [12] | 2015 | Fuzzy | 87.2% |
| Kemal Polat [22] | 2016 | FCMAW | 91.41% |
| Hanslal Prajapati et al [23] | 2017 | Fuzzy | 91.6% |
| Mohammed Benamina et al [24] | 2018 | Fuzzy & CBR | 91.67% |
| I.K. Mujawar et al [25] | 2019 | Fuzzy | 84% |
| **Proposed Method** | **2020** | **Fuzzy** | **93.33%** |

# Conclusion

Disease prediction systems are used to predict the disease of a person using the symptoms. In this proposed system, the prediction of the disease will be done by interacting with the patient through a Chatbot. In this system, the prediction of the disease will be done by gathering patient’s data through the chatbot and these data along with the information available in the ontology is used to predict the disease. Our research focuses on the disease prediction in a frame-based manner. This increase the accuracy of disease prediction.

# Further Work

In this paper, we showed an experiment, how the medical data of the patient can be used to predict whether the patient has diabetes or not. Further, our research will be conducted forward for other diseases. In addition to that, the system will recommend a consultant according to the most probable disease of the patient.

##### References

1. Qikang Wei, Tao Chen, Ruifeng Xu, Yulan He, Lin Gui, “Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks,” Oxford , 2016.
2. L. Honglan, Q. Xiaona, and F. Bin, “The Symptoms and Pathogenesis Entity Recognition of TCM Medical Records Based on CRF,” 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015.
3. F. Tong, Z. Luo, and D. Zhao, “A deep network based integrated model for disease named entity recognition,” 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017.
4. Hyeju Jang ; Sa Kwang Song ; Sung Hyon Myaeng, “Semantic Tagging for Medical Knowledge Tracking,” Semantic Tagging for Medical Knowledge Tracking. IEEE, New York, NY,USA, 2016.
5. Qikang Wei, Tao Chen, Ruifeng Xu, Yulan He, Lin Gui, “Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks,” Oxford , 2016.
6. L. Honglan, Q. Xiaona, and F. Bin, “The Symptoms and Pathogenesis Entity Recognition of TCM Medical Records Based on CRF,” 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015.
7. F. Tong, Z. Luo, and D. Zhao, “A deep network based integrated model for disease named entity recognition,” 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017.
8. Hyeju Jang ; Sa Kwang Song ; Sung Hyon Myaeng, “Semantic Tagging for Medical Knowledge Tracking,” Semantic Tagging for Medical Knowledge Tracking. IEEE, New York, NY,USA, 2016.
9. Ebenezer Obaloluwa Olaniyi, Oyebade Kayode Oyedotun, Khashman Adnan, “Heart Diseases Diagnosis Using Neural Networks Arbitration", International Journal of Intelligent Systems and Applications (IJISA), vol.7, no.12, pp.75-82, 2015.
10. A. A. Al Jarullah, "Decision tree discovery for the diagnosis of type II diabetes," 2011 International Conference on Innovations in Information Technology, Abu Dhabi, 2011, pp. 303-307.
11. B. Venkatalakshmi, M.V Shivsankar, “Heart Disease Diagnosis Using Predictive Data mining”, International Journal of Innovative Research in Science, Engineering and Technology ISSN 2319-8753 Vol.3, Special Issue 3,pp. 1873-1877 ©2014 ICIET.
12. V. Jain and S. Raheja, “Improving the Prediction Rate of Diabetes using Fuzzy Expert System,” I.J. Information Technology and Computer Science, vol. 7, no. 10, pp. 84-91, 2015. DOI: 10.5815/ijitcs.2015.10.10.
13. A. A. Abdullah, Z. Zakaria and N. F. Mohamad, "Design and Development of Fuzzy Expert System for Diagnosis of Hypertension," 2011 Second International Conference on Intelligent Systems, Modelling and Simulation, Kuala Lumpur, 2011, pp. 113-117.
14. Mohammed, Osama et al. “Building a diseases symptoms ontology for medical diagnosis: An integrative approach.” The First International Conference on Future Generation Communication Technologies (2012): 104-108.
15. Heiner Oberkampf, Turan Gojayev, Sonja Zillner, Dietlind Zühlke, Sören Auer, and Matthias Hammon. From symptoms to diseases--creating the missing link. In European Semantic Web Conference, pages 652--667. Springer, 2015.
16. Wang Meiqin, Wu Qingbin," An overview of Ontology-based construction of medical knowledge base. Journal of Medical Informatics,"

2017, Vol. 38, issue 3, pp. 73-76, 2017.

1. Guarino N., Welty C.A. (2004) An Overview of OntoClean. In: Staab S., Studer R. (eds) Handbook on Ontologies. International Handbooks on Information Systems. Springer, Berlin, Heidelberg
2. Bioportal.bioontology.org. (2020). Welcome to the NCBO BioPortal | NCBO BioPortal. [online] Available at: https://bioportal.bioontology.org/ [Accessed 13 Feb. 2020].
3. Medlineplus.gov. (2020). MedlinePlus - Health Information from the National Library of Medicine. [online] Available at: https://medlineplus.gov/ [Accessed 13 Feb. 2020].
4. Mayoclinic.org. (2020). Mayo Clinic - Mayo Clinic. [online] Available at: https://www.mayoclinic.org/ [Accessed 13 Feb. 2020].
5. L. O. Hall and P. Lande, “Generation of fuzzy rules from decision trees,” J. Advanced Comput. Intell., vol. 2, no. 4, pp. 128–133, 1998.
6. Kemal Polat, “Intelligent Recognition of Diabetes Disease via FCM Based Attribute Weighting,” International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 10, no. 4, 2016, pp. 783-787
7. Prajapati, Hanslal, Anurag Jain and Sanjay Kumar Pal. “An Enhance Expert System for Diagnosis of Diabetes using Fuzzy Rules over PIMA Dataset.” , International Journal of Advance Engineering and Research Development (IJAERD),2017.
8. M. Benamina, B. Atmani, S. Benbelkacem, "Diabetes Diagnosis by Case-Based Reasoning and Fuzzy Logic", International Journal of Interactive Multimedia and Artificial Intelligence, 2018.
9. I.K. Mujawar, B.T. Jadhav, “Web-based Fuzzy Expert System for Diabetes Diagnosis,” International Journal of Computer Sciences and Engineering, Vol.7, Issue.2, pp.995-1000, 2019

.