Adaptive Rule Learning using Ontology and Neural Network for Blood Diagnosis

Ronnarong Kaewprasert [1] and Surasak Mungsing [2]
School of Information Technology Sripatum University
kaow9kaow@gmail.com[1], surasak.mu@spu.ac.th[2]

Abstract

This paper presents a specific model for medical diagnosis developed with ontology based on neural network techniques. The model provides in the clinical laboratory domain with the possibility to design diagnostic applications. Given a set of blood chemistry test results have identified which diseases could justify the particular findings. Semantic rule approach for constructing neural networks has been applied, which was developed to facilitate their training by supervised learning, the combinatorial ontology neural network model based on adaptive semantic rule for the blood diagnosis using SWRL. In 10 fold cross validation; the model performance result was measured with average 95.66 % accuracy, 94.56 % recall and 93.59 % F-measure.

Keywords: Neural network, Ontology, Blood diagnosis, Semantic rule.

1. Introduction

In clinical laboratory, decision making is a complex task as it depends on a variety of interrelated functions. We are concentrating not only on the accuracy and precision of the laboratory result, but also on the clinical laboratory interoperability of the test result from the physicians who use decision support system. Making the right decision at the right time is the most important factor in healthcare systems, especially in medical diagnosis systems.

Clinical laboratory results are the important factor in diagnosis systems. The blood test of personalization refers to the delivery of right diagnosis and health status of every individual people. Diagnosis decision support systems do not provide reasoning and mainly focus on integration of data and knowledge in medical. In order to improve knowledge representation and reasoning facility, the ontology acts as a stepping stone to improve the personal risk interpretation systems.

In this paper, we combined the implementation of ontology and artificial neural networks with semantic rule learning techniques by SWRL in blood diagnosis. The dataset used for experiments was trained and tested which

was obtained from the Srisangwornsukhothai Hospital in 2013. The performance measurements of the model were used as accuracy, recall and F-measure.

The rest of this paper is organized as the following; Section 2 is dedicated as literature review. In this section, brief reviews of previous works in medical diagnosis are presented. In Section 3, the methodology of diagnosis model, ontology development, semantic rule and training process of extreme learning machine are explained. In Section 4, research results are provided. Finally, conclusions and future works are given in Section 5.

2. Literature review

Artificial neural networks (ANNs) is currently a very active research domain in medicine and it is believed that it will be more widely used in health care systems in the next few years. Neural networks (NN) techniques have recently been applied to many medical diagnosis problems. For example, Fernandez de Canete et al. have proven useful in the analysis of blood and urine samples of diabetic patients [1]. Koushal Kumar and Abhishek have been investigated for diagnosis of kidney stones diseases using neural network algorithms [2]. In 2011, Bekir used Back-propagation and the Navie Bayes classifier for hepatitis diagnosis [3], the study has been presented a comparison between Back-propagation and Naive Bayes Classifiers to diagnose hepatitis disease. The diagnostic system is considered a Multi-layer Perceptron (MLP) ANN using back- propagation learning algorithm to classify Thyroid disease [4]. Ganesan et al. applied neural network to diagnose cancer [5] in 2010. Elveren and YumuŞak have been used to analyze demographic data from lung cancer patients with a view to develop diagnostic algorithms that might improve triage practices in the emergency department. Tuberculosis diagnosis was realized by using multilayer neural networks (MLNN) [6]. Bahar et al. analyzed ANN for prediction of headache in this year [7]. This analysis has been done to investigate the ability of neural networks to detect and classify the complete improvement of headache in elderly patients during the follow- up period. Sumathi et al. constructed ANN to pre-diagnosis of hypertension [8]. The authors used ANN for solving the problems of hypertension diagnosing using Back-propagation learning algorithm.

In the recent decades, Artificial Neural Networks (ANNs) are considered as the best solutions to achieve this goal and involved in widespread researches to ontology mapping. In previous approach, they used the integration of ontology mapping with indexing techniques and rules based techniques, adopted the multilayer

feed-forward neural network couple with back propagation algorithm to learn the similarity of ontology patterns [9]. Artificial neural network approach to learning and adjusting the above weights, and thereby support a new ontology matching algorithm, with the purpose to avoid some of the disadvantages in both rule-based and learning-based ontology matching approaches [10]. One of the network structures that have been give a new algorithm for measuring the ontology similarity and ontology mapping using fast algorithm for learning large scale preference relations. The new algorithm has less complexity and also has high equality [11].

2.1 Blood Test Results Interpretation

A chemistry test is a blood test that measures the levels of several substances in the blood (such as sugar, cholesterol). A chemistry test tells the person about general health, helps looking for certain problems, and finds out whether organ for a specific problem is working [12,13].

2.2 Glucose

This is the chief source of energy for all living organisms. A level greater than 110 mg/dl in someone who has fasted for 12 hours suggests a diabetic tendency. If this level is elevated even in a non-fasting setting one must be concerned that there is a risk for developing diabetes. This is an incredibly powerful test and can predict diabetes ten years or more before one develops the strict definition of diabetes.

2.3 BUN (Blood Urea Nitrogen)

BUN is a waste product derived from protein breakdown in the liver. Increases can be caused by excessive protein intake, kidney damage, certain drugs, low fluid intake, intestinal bleeding, exercise, heart failure or decreased digestive enzyme production by the pancreas. Decreased levels are most commonly due to inadequate protein intake, mal-absorption, or liver damage.

2.4 Creatinine

Creatinine is also a protein breakdown product. Its level is a reflection of the bodies muscle mass. Low levels are commonly seen in inadequate protein intake, liver disease, kidney damage or pregnancy. Elevated levels are generally reflective of kidney damage and need to be monitored very carefully.

2.5 Uric Acid

Uric acid is the end product purine metabolism. High levels are seen in gout,

infections, high protein diets, and kidney disease. Low levels generally indicate protein and molybdenum (trace mineral) deficiency, liver damage or an overly acid kidney.

2.6 Cholesterol

Group of fats is vital to cell membranes, nerve fibers and bile salts, and a necessary precursor for the sex hormones. High levels indicate diet high in carbohydrates/sugars. Low levels indicate low fat diet, mal-absorption, anemia, liver disorders, and carbohydrate sensitivity.

2.7 Triglycerides

These are fats used as fuel by the body, and as an energy source for metabolism. Increased levels are almost always a sign of too much carbohydrate intake and hyperlipidism. Decreased levels are seen in hyperthyroidism, malnutrition and mal-absorption.

2.8 Alkaline Phosphatase (ALP)

ALP is an enzyme that is found in all body tissue, but the most important sites are bone, liver, bile ducts and the gut. A high level of ALP in your blood may indicate bone, liver or bile duct disease. Certain drugs may also cause high levels. Growing children, because of bone growth, normally have a higher level than adults do.

2.8 Transaminases (ALT & AST)

These are enzymes that ~~are~~ primarily found in the liver. Drinking too much alcohol, certain drugs, liver disease and bile duct disease can cause high levels in the blood. Hepatitis is another problem that can raise these levels. Low levels of AST and ALT may indicate deficiency of vitamin B6.

3. Research Methodology

3.1 Laboratory Diagnosis Ontology Development

We take Protégé-OWL as the ontology editor that is an extension of Protégé supporting the Web Ontology Language (OWL) [14]. It enables users to load and save OWL and RDF ontologies, edit and visualize classes, properties, and SWRL rules, define logical class characteristics as OWL expressions, execute reasoners such as description logic classifiers, edit OWL individuals for semantic web markup. SWRL (Semantic Web Rule Language) [15] based on a combination of the OWL DL and OWL Lite with the Unary/Binary Datalog RuleML sublanguages of the Rule Markup Language is developed in order to make up for the OWL language that is imperfect in the ability of expression. SWRL described the knowledge of OWL ontology by highly abstract syntax expression, which realized the combination between the horn-like

rules and OWL knowledge base. We take SWRL to express formally spatial reasoning rules.

The methodology used for this approach was a combination of text analysis and an expert approach. The first step was to extract terms document as much laboratory diagnosis based knowledge as possible from the literature and to collect and to review the related knowledge resources and categorize them systematically [16,17].

- **Build the concept of terms**: The first task is to build a concept of terms. The main goal of this task is the creation of a glossary of terms and terms identified by UMLS which have emerged along the knowledge acquisition process [18].
- **Build the property and relation**: The task consists of building relation diagrams. The relationship creates a semantic link between two concepts providing meaningful information. The relate diagram permits to represent graphically the various relations that exist between the different concepts in the same or different hierarchies.
- **Build the logical-axioms**: In this step, we will define the ontology concepts by using the logical expressions which are always true. We define for each axiom, its description in natural language, the name of the concept to which the axiom refers and the logical expression.
- **Build the instances**: The instances describe some instances (inclusive in the field processes of concept dictionary) with their attributes and values.

3.2 Linguistic Variable in Semantic Rule

The concept of a linguistic variable was first introduced by Zadeh [19] as a model of how words or labels can represent vague concepts in natural language. We are discussing here as modified in semantic rule construction. A linguistic variable is a quadruple [X, T(X), U, M] in which X is the name of the variable, T(X) is a countable term set of labels or words (i.e. the linguistic values), U is an universal set, and M is the set referential of the values T(X).

Linguistic variables are used for system input and output, and are represented by words such as FBS, BUN, Creatinine. A linguistic attribute *FBS* may take linguistic set values *very low , low, Normal, high*, very high and appear in a semantic rule as logical condition, for example *FBS=low*. An alternative way is to use a predicate function *FBS(x)*. Depending on the type of variable *x* predicate function may have different interpretation. For example, if *x* is the blood sugar level of blood test and $x \in$ [60 mg/dl,70 mg/dl] then *FBS(x)* is *low*, i.e. logical

condition *FBS(x)=low* is true. One may also introduce predicates for each FBS defined by logical functions *FBS-low(x), FBS-normal(x), FBS-high(x)*. Such logical predicate functions are linguistic variables, mapping symbolic or real values of *x* into binary 0, 1 or *false, true*.

If the input $x \in X$ is given as a real number or a large number of integer or symbolic values linguistic variables are created dividing the data range *X* into distinct (for crisp logic) sets X_i and introducing variables $s_i(x)$ =F unless $x \in X_i$, when $s_i(x)$ =T. For $X \subseteq R$ sets X_i are usually intervals and linguistic variables are binary functions mapping *x* into 0 or 1. A typical linguistic variables associated with the tire pressure attribute will be *low* if $x < 70$, *normal* if $71 \leq x \leq 115$ and *high* if $x > 115$. A rule may then have conditions of the form *high(x)*, which is usually written as *x=high*, meaning that $x > 115$.

3.3 Semantic Rule Structure Model

The semantic rule structure in this study used a neural network structures consisting of an input layer and an output layer as shown in Figure 1. In this system, real valued 9 input blood test results is linguistic variable, and six outputs are index of six classes (diabetes, renal disease, gout, heart disease, liver disease and normal).

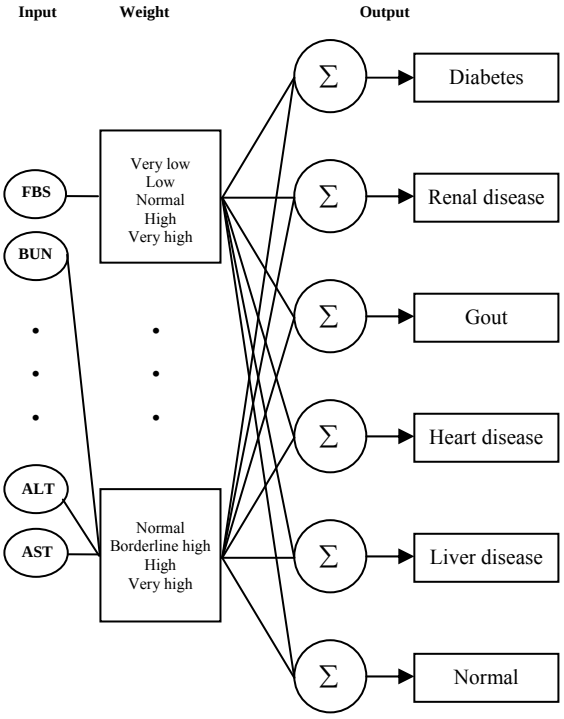

Fig. 1. Semantic rule structure with neural network

3.4 Building Semantic Rule in Blood Diagnosis

Through the investigation and analysis on blood test of personalized risk disease, we summarize the following semantic rules that constitute the basis of the blood diagnosis as shown in Figure 1. To classify the blood test result of blood diagnosis, the semantic rule set is designed in rule language shown in the following examples:

Rule 1: IF (FBS is Normal) and (BUN is Normal) and (Creatinine is Normal) and (Uric acid is Normal) and (Cholesterol is Normal) and (Triglyceride is Normal) and (ALP is Normal) and

(ALT is Normal) and (AST is Normal) THEN Diagnosis and Disease are presents.

Rule 2: IF (FBS is High) and (BUN is Normal) and (Creatinine is Normal) and (Uric acid is Normal) and (Cholesterol is Normal) and (Triglyceride is Normal) and (ALP is Normal) and (ALT is Normal) and (AST is Normal) THEN Diagnosis and Disease are presents.

⋮

Rule n: IF (FBS is Very high) and (BUN is Very high) and (Creatinine is Very high) and (Uric acid is Very high) and (Cholesterol is Very high) and (Triglyceride is Very high) and (ALP is Very high) and (ALT is Very high) and (AST is Very high) THEN Diagnosis and Disease are presents.

3.4 Dataset Training

In this study, Srisangwornsukhothai Hospital in 2013 laboratory results were used. Data of 1000 personal of blood chemistry have been used with 9,000 instances and has 9 attributes. Dataset contains 138 risk disease and 882 normal data and has 9 input attributes and 2 output of classification (diagnosis and disease). The data was further divided into 80% for training and 20 % for testing the data. All data have nine features. These features are laboratory examination: fasting blood sugar (FBS), blood uria nitrogen (BUN), creatinine, uric acid, cholesterol, triglyceride, alkaline phosphatase (ALP), alanin aminotransferase (ALT) and aspartat aminotransferase (AST).

3.5 Dataset Cleaning

The outlier detection and noise elimination is an important issue in data analysis. The exclusion of outliers improves data quality and therefore classification performance. Several researchers have proposed various approaches for data cleaning. G.H. John [20] proposed a technique that removes a misclassified training instance from training data and reconstructs the trees, the process is repeated till all such instances are removed from training data. Misclassified instances are recognized using tree classifier as a filter. The resulting classifier enhances classification performance accuracy. Broadly and Friedl [21] proposed a method for detecting mislabeled instances. The method uses a set of learning algorithms to construct classifiers that act as a filter for the training data. The technique removes outliers in regression analysis.

Data cleaning has many in common although they are different disciplines. Some pattern recognition algorisms used in data mining are also applied in data cleaning. The differences are that the data cleaning has more specific and concrete jobs, which are to detect

and remove data with error, noisy, missing values, outlier, redundancy and inconsistence in order to improve the database quality before dataset used training. So we say data cleaning is pre-process of dataset or data analysis. In some situations, an alternative of data cleaning is date filtering, which retrieves or deletes our intended data or data pattern from the original database, and forms a new desired dataset. In this paper, we summarized above classifications and defined data problems needing elimination as shown in Figure 2.

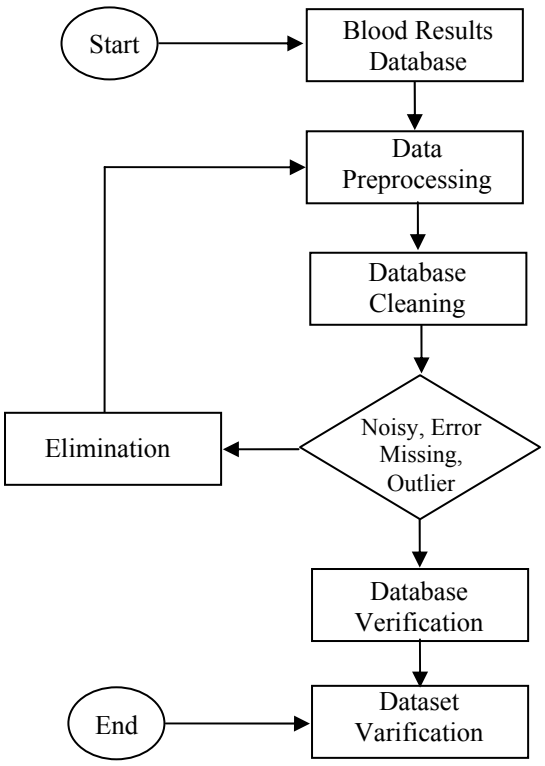

Fig. 2. Dataset preprocessing

3.6 Modified Back Propagation Algorithm in Semantic Rule Learning

The back propagation algorithm is a technique used in developing multilayer neural networks in a supervised manner. The back propagation algorithm, also known as the error back propagation algorithm, is based on the error-correction learning rule [22]. The modified algorithm has two passes through the different process of the semantic rule learning: a forward pass and a backward pass. In the forward pass, an activity pattern is applied to the input linguistic variable of the semantic rule, and its effect propagates through the rule. Finally, a set of outputs is produced as the actual response of the semantic rule layer. During the forward pass the linguistic weights of the semantic rule are all fixed. During the backward pass, the linguistic weights are all adjusted in accordance with an accuracy-correction rule. The actual response of the semantic rule layer is subtracted from a desired response to produce an accuracy value. This accuracy value is then propagated backward through the semantic rule layer. The linguistic weights are adjusted to make the actual response of the semantic rule layer move closer to the desired response in a statistical sense. The linguistic weight adjustment is made according to the generalized semantic rule to

maximize the accuracy. An example of learning in semantic rule layer is shown in Figure 3.

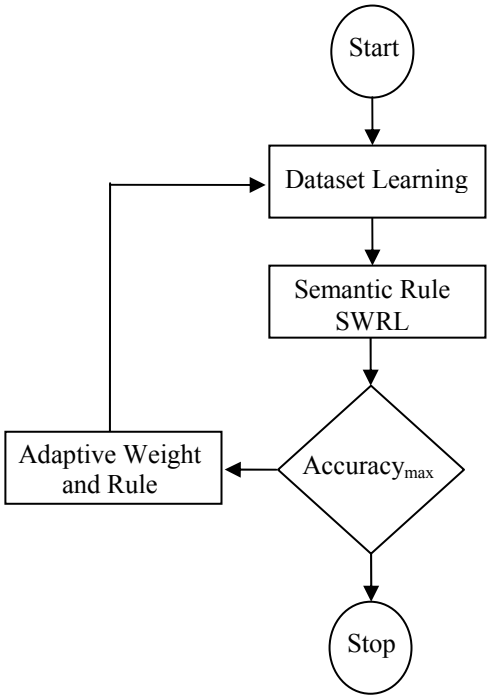

Fig. 3. Semantic rule learning

3.7 Performance Measurement

Performance of model can be evaluated by using some very well-known statistical measures: classification accuracy, recall, and F-measure. These measures are defined by true positive (TP), true negative (TN), false positive (FP) and false negative (FN) in form of confusion matrix. The accuracy, recall and F-measure values in the results that test results are calculated as follows [23].

$$accuracy = \frac{TP + TN}{(TN + TP + FP + FN)} \quad (1)$$

$$recall = \frac{TN}{(TN + FP)} \quad (2)$$

$$F - measure = \frac{TP}{(TP + FN)} \quad (3)$$

3.8 Cross validation

In order to calculate the performance of the each model, the whole laboratory results data are divided into training and test sets. Specifically, we used the following two step 10-fold cross-validation procedure to estimate the prediction accuracy, recall and F-measure. Step one: We randomly divided the dataset (1,000 records) into 10 disjoint subsets (folds), with each fold containing approximately the same number of records (100 records). The sampling was stratified by the class labels to ensure that the subset class proportions are roughly the same as those in the whole dataset. Step two: For each subset, a dataset classification model was constructed using the nine of the ten folds and tested on the tenth one to obtain a cross-validation estimate of its prediction results [24].

4. Research Results

4.1 Laboratory Diagnosis Ontology Representation

The laboratory diagnosis ontology presents the formalized description of concepts for the laboratory diagnosis domain. It includes basic analysis result concepts, properties that characterize are related to diagnosis disease, all relevant diagnostic

examinations and laboratory results. The ontology also includes other pre-analysis, analysis and post-analysis process related concepts when they are connected with clinical laboratory. The information presented in the ontology has been obtained by domain expert interpretation of guidelines for trueness analysis results. In its current form the laboratory diagnosis ontology presents the detailed taxonomic overview of the clinical laboratory domain with contains 270 classes describing clinical laboratory related to diagnosis disease concepts. These concepts are interconnected with super-class and sub-class relations into a hierarchical tree-like structure. At the basic level, there are eight relevant super-classes: *laboratory Interpretation, Laboratory Method, Laboratory Process, Laboratory Result, Laboratory Test, Laboratory Quality, Personal Status* and *Specimen Type*. Figure 5 presents the Protégé OWL tool displaying some of the classes from the laboratory diagnosis ontology. Individuals or instances typically present an exhaustive list of concrete concepts relevant to the class. The realized ontology includes more than 200 individuals. When possible, concepts are specified with their CUI number (Concept Unique Identifier according to UMLS) and with a list of synonyms. Finally, the ontology contains properties that introduce relations among concepts. For example, instance *Diabetes* from the class *Disease* is indicated by the individual *High Level* from the class of *FBS Test*. Or that *Hyperlipidemia* from the class *Diagnosis* may be risk to disease like *cardiovascular disease or Hypertension*. The names of these properties are *hasDiagnosis* and *hasDisease* respectively. The laboratory diagnosis ontology includes definitions of more than 50 properties.

In the design of the laboratory diagnosis ontology we have started from the terms defined in the “Guidelines for laboratory diagnosis. In order to connect the ontology concepts with the terms defined in UMLS, we have introduced the property that gives the appropriate UMLS reference for every concept in the ontology.

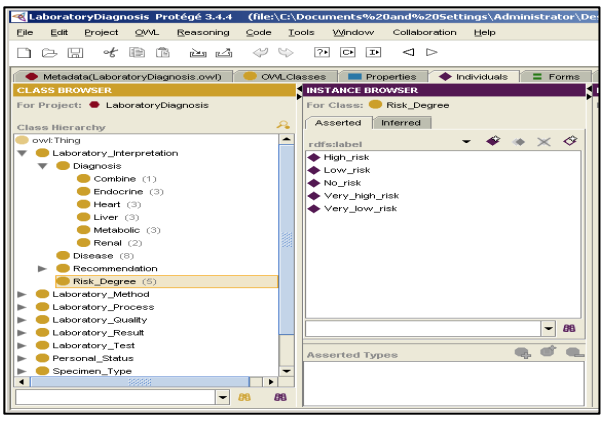

Fig. 4. Super-classes and sub-classes concepts of clinical laboratory ontology

4.2 Semantic Rule Using SWRL

In our implementation, we expressed semantic rules, such as the one described in Figure 3 within SWRL. We showed that there was a simple rule to encode the semantic rules into a crisp rule language supporting arithmetic built-in functions and, thus, in SWRL, making them directly available in current reasoners and in the Protégé OWL editor. In fact, we followed the below mentioned method to correctly deal with our semantic rule base.

Blood chemistry diagnostic (Diabetes, Renal disease, Gout, Cardiovascular disease, Liver disease)

Rule-1: Chemistry_Test (?l) ∧ hasFBS (?l, ?s) ∧ hasBUN (?l, ?b) ∧ hasCreatinine (?l, ?c) ∧ hasUric_acid (?l, ?u) ∧ hasCholesterol (?l, ?ch) ∧ hasTriglyceride (?l, ?tg) ∧ hasALP (?l, ?p) ∧ hasALT (?l, ?al) ∧ hasAST (?l, ?as) ∧ swrlb:greaterThan(?s, 69) ∧ swrlb:lessThan(?s, 111) ∧ swrlb:greaterThan(?b, 6) ∧ swrlb:lessThan(?b, 23) ∧ swrlb:greaterThan(?c, 0.6) ∧ swrlb:lessThan(?c, 1.3) ∧ swrlb:greaterThan(?u, 2.9) ∧ swrlb:lessThan(?u, 8.1) ∧ swrlb:lessThan(?ch, 200) ∧ swrlb:lessThan(?tg, 200) ∧ swrlb:greaterThan(?p, 34) ∧ swrlb:lessThan(?p, 101) ∧ swrlb:lessThan(?al, 37) ∧ swrlb:lessThan(?as, 36) ∧ hasDisease(?l, ?ds) ∧ hasDiagnosis(?l, ?d) → sqwrl:select(?d, ?ds)
Results: diagnosis = normal and disease = healthy

Rule-2: Chemistry_Test (?l) ∧ hasFBS (?l, ?s) ∧ hasBUN (?l, ?b) ∧ hasCreatinine (?l, ?c) ∧ hasUric_acid (?l, ?u) ∧ hasCholesterol (?l, ?ch) ∧ hasTriglyceride (?l, ?tg) ∧ hasALP (?l, ?p) ∧ hasALT (?l, ?al) ∧ hasAST (?l, ?as) ∧ swrlb:greaterThan(?s, 110) ∧ swrlb:lessThan(?s, 141) ∧ swrlb:greaterThan(?b, 6) ∧ swrlb:lessThan(?b, 23) ∧ swrlb:greaterThan(?c, 0.6) ∧ swrlb:lessThan(?c, 1.3) ∧ swrlb:greaterThan(?u, 2.9) ∧ swrlb:lessThan(?u, 8.1) ∧ swrlb:greaterThan(?ch, 250) ∧ swrlb:lessThan(?ch, 200) ∧ swrlb:greaterThan(?tg, 200) ∧ swrlb:lessThan(?tg, 200) ∧ swrlb:greaterThan(?p, 34) ∧ swrlb:lessThan(?p, 101) ∧ swrlb:lessThan(?al, 37) ∧ swrlb:lessThan(?as, 36) ∧ hasDisease(?l, ?ds) ∧ hasDiagnosis(?l, ?d) → sqwrl:select(?d, ?ds)
Results: diagnosis = hyperglycemia and disease = risk of diabetes

⋮

Rule-n: Chemistry_Test (?l) ∧ hasFBS (?l, ?s) ∧ hasBUN (?l, ?b) ∧ hasCreatinine (?l, ?c) ∧ hasUric_acid (?l, ?u) ∧ hasCholesterol (?l, ?ch) ∧ hasTriglyceride (?l, ?tg) ∧ hasALP (?l, ?p) ∧ hasALT (?l, ?al) ∧ hasAST (?l, ?as) ∧ swrlb:greaterThan(?s, 200) ∧ swrlb:greaterThan(?b, 40) ∧ swrlb:greaterThan(?c, 3.0) ∧ swrlb:greaterThan(?u, 10) ∧ swrlb:greaterThan(?ch, 400) ∧ swrlb:greaterThan(?tg, 400) swrlb:greaterThan(?p, 500) ∧ swrlb:greaterThan(?al, 500) ∧ swrlb:greaterThan(?as, 500) ∧ hasDisease(?l, ?ds) ∧ hasDiagnosis(?l, ?d) → sqwrl:select(?d, ?ds) Results: diagnosis = hyperglycemia, abnormal renal function, hyperuricemia, hyperlipidemia, abnormal liver function and disease = risk of diabetes, renal disease, gout, cardiovascular disease, liver disease

To classify the blood test result of ontology structure, the rule set is designed in semantic rule language shown in the following examples: Using SWRL to describe rules of blood test diagnosis, we obtained the following fuzzy rule expressions below and showed in SWRL rule tab as shown in Figure 5.

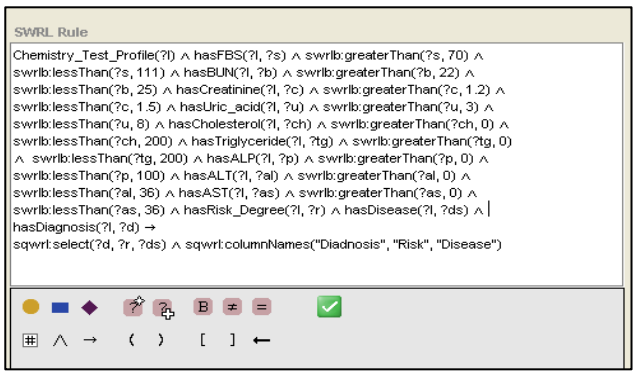

Fig. 5. Expressions of semantic rules in SWRLTab

4.3 SWRLQueryTab Displaying Blood Diagnosis Results

The SWRLQueryTab provides a convenient way to visualize the results of these SWRL queries. It has a control sub-tab that can be used to control the execution of SWRL rules containing query built-ins [26]. A decision making query can be selected from the rule table in the SWRL Editor and executed to display blood diagnosis results of the query. Users can navigate to that sub-tab to review the results displayed in tabular form as shown in Figure 6.

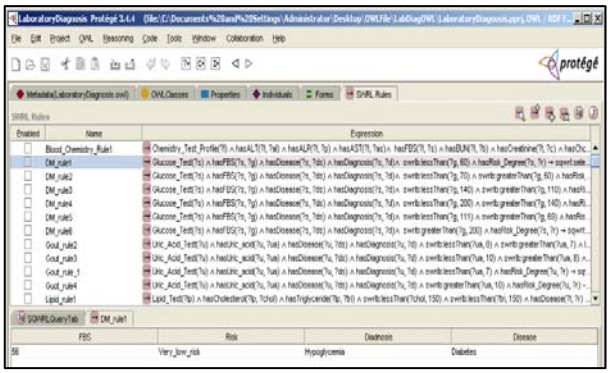

Fig. 6. Graphically displays blood diagnosis results of SWRL rules.

4.4 Adaptive Rule Learning

To learn, or adjust weights on connecting arrows between concept from input-output training samples, neural networks model, the learning algorithm are used. The SWRL rule learning algorithm consists of adjusting the antecedent and a consequent linguistic value. In the SWRL structure, the antecedent and consequents play the role of weights, is shown in Figure 7..

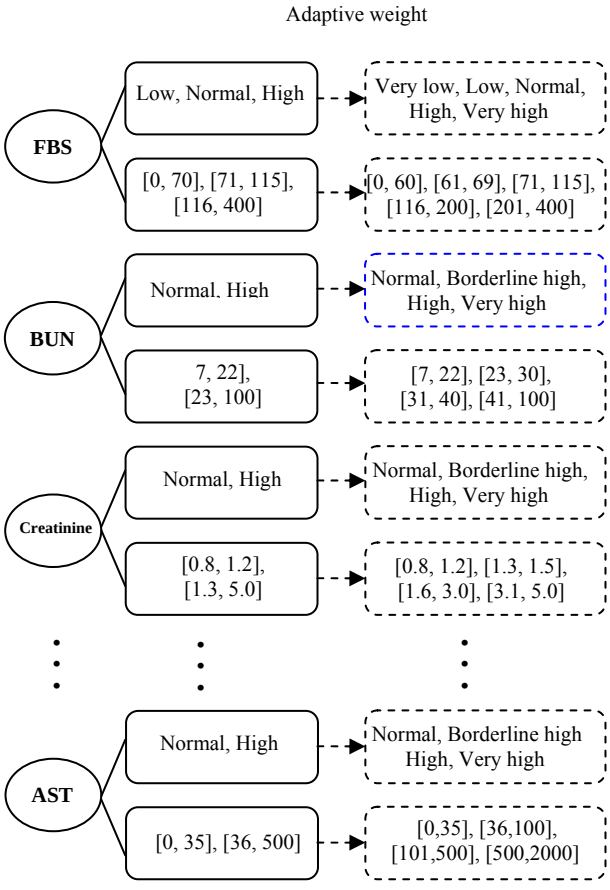

Fig. 7. Adaptive rule by learning algorithm

4.5 Adaptive Semantic Rule Evaluation

The experimental results and performance evaluation of the proposed technique as discussed, we used a considered 1,000 datasets. The evaluation metrics are calculated and the experimental results obtained for testing dataset are also tabulated. These evaluation metrics gives an overall performance categorization of the proposed system. From the results, it is clear that our proposed system performs very well. In 10-fold cross-validation, we have achieved the average results 95.66 % accuracy, 94.56 % recall and 93.59 % F-measure.

Table 1. The average results of 10-fold cross-validation.

Fold for Testing	Average		
	Accuracy	Recall	F-measure
1	94.56	94.20	93.04
2	95.60	93.52	92.46
3	95.32	94.14	93.46
4	96.12	95.06	94.56
5	95.46	94.68	93.48
6	95.68	94.42	93.42
7	94.76	94.12	93.06
8	95.84	94.26	93.22
9	96.42	95.44	94.68
10	96.86	95.72	94.54
Average	95.66	94.56	93.59

5. Conclusion

In this paper, we have presented an adaptive rule learning based on ontology and neural network for blood diagnosis. Personal risk disease detection in its early stage is the key of its cure. We have shown ontology and neural

networks are used in actual blood diagnosis with adaptive rule learning. This work described that the prediction of personal risk disease is from blood test results and gives the best result on the dataset learning. This ANN based model has been used to develop a rule base system in ontology which people would be able to self-diagnose and also it helps the doctors to plan for better medication and provide the people with early diagnosis of common risk disease. Prognosis of early diagnosis of disease with ontology combined with neural network models has the best performance in large data sets learning. The performance of the diagnosis is better comparing to physicians diagnosis as they may get higher predictive accuracy, recall, F-measure can be achieved.

Our future work will focus on healthcare applications decision support system. The diagnosis ontology can be used in combination with knowledge search engines and expert systems. As an extension, the basic idea of combining fuzzy rule is to design the architecture that using a fuzzy system to represent knowledge in imprecise data, which used in intelligent semantic decision support system.

References

[1] De Canetea, J.F., Gonzalez-Pereza, S., and Ramos-Diaz, J.C. (2012), "Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes", *Computer methods and programs in biomedicine,* Vol. 106, 55–66.

[2] Kumar, K., Abhishek, B. (2012), "Artificial Neural Networks for diagnosis of kidney stones disease", *International Journal Information Technology and Computer Science*, Vol. 7, 20-25.

[3] Karlik, B. (2012), "Hepatitis Disease Diagnosis Using Backpropagation and the Naive Bayes Classiers", IBU Journal of Science and Technology, Vol 1, Issue 1.

[4] Ozyılmaz, L., Yıldırım, T. (2002), "Diagnosis of thyroid disease using artificial neural network methods", in: Proceedings of ICONIP'02 9th international conference on neural information processing (Singapore: Orchid Country Club, 2002), 2033–2036.

[5] Ganesan, N. (2010), "Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data," International Journal of Computer Applications, Vol. 01, 76-85.

[6] Elveren, E., Yumuşak, N., (2011), "Tuberculosis Disease Diagnosis Using Artificial Neural Network Trained with Genetic Algorithm", *Journal of Medical Systems,* Vol.35(3), 329-332.

[7] Taşdelen, B.S., Helvaci, H. and Kaleağasi, and Özge, A., (2009), "Artificial neural network analysis for prediction of headache prognosis in elderly patients," *Turkish Journal of Medical Sciences*, Vol. 39, No. 1, 5–12.

[8] Sumathi, B., Santhakumaran, A., (2011), "Pre-Diagnosis of Hypertension Using Artificial Neural Network", *Global Journal of Computer Science and Technology*, Vol. 11, Issue 2.

[9] Nattawuttisit, S., *Sasiporn* U. (2013), "*Ontology Mapping Using Neural Networks* with *Indexing Technique*", Vol. 2, No. 1, 2277–4378.

[10] Huang, J., Dang, J., Vidal, J.M. and Huhns, M.N. (2007), "Ontology Matching Using an Artificial Neural Network to Learn Weights", *Proc. IJCAI Workshop on Semantic Web for Collaborative Knowledge Acquisition* (*SWeCKa-07*), Hyderabad, India, Jan., 80-85,

[11] Huang, X., Xu, T., Gao, W. and Jia, Z. (2011), "Ontology Similarity Measure and Ontology Mapping Via Fast Ranking Method", *International Journal of Applied Physics and Mathematics*, Vol. 1, no. 1, 54-59.

[12] Fraser, C. G. (1986), *Interpretation of clinical chemistry laboratory data*. Oxford.

[13] Corbett, J. V., Banks, A. D. and others, (1996), *Laboratory tests and diagnostic procedures with nursing diagnoses*. Appleton & Lange, 1996.

[14] Knublauch, H, Fergerson, RW, Noy, NF, Musen MA. (2004), "The Protégé OWL Plugin: An open development environment for semantic web applications", Proc Third ISWC (ISWC 2004); Hiroshima, Japan. 229–243.

[15] Boley, P.,Tabet S., Grosof B., Dean, K., (2004), SWRL: A Semantic Web Rule Language Combining OWL and RuleML.http://www.w3.org/Submission/SWRL/.

[16] Öhgren, A. (2004), "Ontology development and evolution: Selected approaches for small-scale application contexts", Jönköping University, *School of Engineering.*

[17] Fernández-López, M., Gómez-Pérez, A., Juristo, N. (1997), "METHONTOLOGY: From Ontological Art Towards Ontological Engineering", Spring Symposium on

Ontological Engineering of AAAI. Stanford University, California, 33–40.

[18] Achour, S.L., Dojat, M., Rieux, C., Bierling, P., Lepage, E. (2001), “A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development”, *Journal of the American Medical Informatics Association*, Vol. *8*, 351-360.

[19] Zadeh, L.A. (1975), "The concept of a linguistic variable and its application to approximate reasoning-I," *Information Sciences*, Vol. 8, no. 3, 199-249.

[20] John, G.H. (1995), “Robust decision trees: removing outliers from databases”, In Proceedings of the First ICKDDM, 174–179.

[21] Brodley, C.E. and Friedl, M.A. (1999), Identifying mislabelled training data, Journal of Artificial Intelligence Research, 131-167.

[22] TaŞdelen, B., Helvaci, S., Kaleağasi, H. and Özge, A. (2009), “Artificial neural network analysis for prediction of headache prognosis in elderly patients”, Turkish Journal of Medical Sciences, Vol. 39, No. 1, 5–12.

[23] Thangaraj, P., Velmurugan, L. (2012), “Improving Mobility Prediction Using Data Mining Techniques”, *International Journal of Computer Science and Telecommunications*, Vol 3, Issue 7, July.

[24] Najah, A. A., El-Shafie, A., Karim, O. A. and Jaafar, O. (2012), Water quality prediction model utilizing integrated wavelet-ANFIS model with cross-validation. *Neural Computing and Applications*, *21*(5), 833–841.