Thyroid Diagnosis based Technique on Rough Sets with Modified Similarity Relation

Elsayed Radwan [1,3]
[1]Computer Science Department P.O.Box: 715
[3]Institute of Scientific Research and Revival of Islamic Heritage, Umm Al-Qura University, KSA

Adel M.A. Assiri [2,3]
[2]Biochemistry Department,
Faculty of Medicine,Umm Al-Qura University, KSA
[3]Institute of Scientific Research and Revival of Islamic Heritage, Umm Al-Qura University, KSA

***Abstract*—Because of the patient's inconsistent data, uncertain Thyroid Disease dataset is appeared in the learning process: irrelevant, redundant, missing, and huge features. In this paper, Rough sets theory is used in data discretization for continuous attribute values, data reduction and rule induction. Also, Rough sets try to cluster the Thyroid relation attributes in the presence of missing attribute values and build the Modified Similarity Relation that is dependent on the number of missing values with respect to the number of the whole defined attributes for each rule. The discernibility matrix has been constructed to compute the minimal sets of reducts, which is used to extract the minimal sets of decision rules that describe similarity relations among rules. Thus, the rule associated strength is measured.**

Keywords— Thyroid Disease - Rough Sets - Data Discretization – Knowledge Reduction; Modified Similarity Relation MSIM

I. INTRODUCTION

Hypothyroid is one of the most common diseases that frequently misunderstood and misdiagnosis. Thyroid is a small butterfly-shaped gland, located in the neck. The thyroid produces several hormones; two of them are important: triiodothyronine (T3) and thyroxin (T4). Each of them must be produced by thyroid in normal rang; to help cells convert oxygen and calories into energy [6,7,8]. It is based almost exclusively upon measuring the amount of thyroid hormone in the blood. So there are normal ranges for thyroid hormones which have been calculated by our application in this paper: T4, T3, TSH, TBG, T4U and FTI. Doctors faced many problems during dealing with patient's data such as huge data needed from patients, misdiagnosis, missing data when applying patient history and the required human efforts.

According to the previous research, various neural network methods including Multi-Layer Perception with Back-Propagation method (MLP), Radial Basis Function (RBF) and adaptive Conic Section Function Neural Network (CSFNN) are used to help diagnosis of thyroid disease, their classification accuracies are separately 88.3%, 81.69% and 85.92% [7]. Also, five different methods including Linear Discriminant Analysis (LDA), C4.5 with default learning parameters (C4.5-1), C4.5 with parameter c equal to 5 (C4.5-2), C4.5 with parameter c equal to 95 (C4.5-3) and DIMLP with two hidden layers and default learning parameters(DIMLP) to perform classification, and the accuracies reached 81.34%, 93.26%, 92.81%, 92.94% and 94.86% respectively [8]. Moreover, an accuracy of 81% was obtained with the application of artificial immune recognition system (AIRS) [7]. Furthermore, diagnosed thyroid diseases with an expert system that called ESTDD (expert system for thyroid disease diagnosis), whose accuracy was 95.33%[14]. Finally, swarm optimization optimized support vector machines with fisher score (FS-PSO-SVM) CAD system for thyroid disease, and the average accuracy of 97.49% was achieved.

As a result, all diagnosis algorithms currently in use depend on different kinds of heuristics where the resulting recodes contain missing bases due to the presence of gaps which may be misclassifying the diseases. Also, a continuous dataset due to the measuring range are discovered. Thus, a good and effective tool to deal with vagueness, missing and uncertainty of information is needed in the presence of continuous data which should be discretized. Rough sets [11,12] deals with the classificatory of data tables and focus on structural relationships in data sets. Rough Sets theory constitutes a framework for inducing minimal decision rules, these rules in turn can be used to perform a classification task. The main goal of the rough set analysis is to search large databases for meaningful decision rules and finally acquire new knowledge. Rough sets has been successfully applied in many different fields, particularly the medical field [2]. A rough set investigates structural relationship in data rather than probability distribution and produce decision table rather than trees [5].

In this paper Rough sets try to classify Thyroid in the presence of missing bases and build the Modified Similarity Relations that is dependent on the number of missing bases with respect to the number of the whole defined attributes for each rule [15]. The Thyroid relation attributes are converted to suitable representation for rough set analysis by discretizing and then constructing a matrix where each row corresponding to the similarity score between Thyroid attributes and each column corresponding to a defined attribute that describe the position of bases inside the rule. The discernibility matrix is used to discern similarity relation among rules in the presence of gaps and deduction of decision rules which describe Thyroid relation attributes with a minimal set of attributes.

According to the previous discussion, the paper is organized as follows; in section II a brief introduction of important field (rough set) is discussed. Section III describes the fundamentals of our method where the an approach of the supervised learning for incomplete Thyroid dataset using rough

sets and its performance are given. Section IV examine the application and guide the user using it. Then we conclude with section V the purpose of that paper and its results.

II. PRELIMINARIES

This section briefs on the basic notions of rough sets that is used in this paper and the detailed definitions can be referred to some related papers [5, 11,12, 13].

A. Rough Set Theory

Rough set theory proposed by Pawlak [11] is an effective approach to imprecision, vagueness, and uncertainty. Rough set theory overlaps with many other theories such that fuzzy sets, evidence theory, and statistics. From a practical point of view, it is a good tool for data analysis. The main goal of the rough set analysis is to synthesize approximation of concepts from acquired data. The starting point of Rough set theory is an observation that the objects having the same description are indiscernible (similar) with respect to the available information. Determination of the similar objects with respect to the defined attributes values is very hard and sensible when some attribute values are missing. This problem must be handled very carefully. The indiscernibility relation is a fundamental concept of the rough set theory which used in the complete information systems. In order to process incomplete information systems, the indiscernibility relation needs to be extended to some equivalent relations.

The starting point of rough set theory which is based on data analysis is a data set called an information system (IS). IS is a data table, whose columns are labeled by attributes, rows are labeled by objects or cases, and the entire of the table are the attribute values. Formally, $IS = (U, AT)$, where U and AT are nonempty finite sets called "the universe" and "the set of attributes," respectively. Every attribute $a \in AT$, has a set V_a of its values called the "domain of a ". If V_a contains missing values for at least one attribute, then S is called an incomplete information system, otherwise it is complete. Any information table defines a function ρ that maps the direct product $U \times AT$ into the set of all values assigned to each attribute. The example of incomplete information system depicted in Table I where set of objects in the universe corresponding to set of instances and set of attributes corresponding to set of values inside each instance. The values of attributes are corresponding to the values of bases inside each object such as the value of instance1at attribute1is defined by ρ(instance1, attribute1) =[1,2].

TABLE I. EXAMPLE OF INCOMPLETE INFORMATION SYSTEM

	attribute1	***attribute2***	***attribute3***
instance1	[1,2]	0	*
instance2	[1,3]	1	A
instance3	[3,4]	2	A

The concept of the indiscernibility relation is an essential concept in rough set theory which is used to distinguish objects described by a set of attributes in complete information systems. Each subset A of AT defines an indiscernibility relation as follows:

$$IND(A) = \{(x,y) \in U \times U : \rho(x,a) = \rho(y,a) \ \ \forall a \in A, A \subset AT\} \quad (1)$$

Obviously, $IND(A)$ is an equivalence relation, the family of all equivalence classes of $IND(A)$, for example, a partition determined by A which is denoted by $U / IND(A)$ or U / A [11]. Obviously $IND(A)$ is an equivalence relation and:

$$IND(A) = \bigcap IND(a) \ \ where \ a \in A \quad (2)$$

A fundamental problem discussed in rough set is whether the whole knowledge extracted from data sets is always necessary to classify objects in the universe; this problem arises in many practical applications and will be referred to as knowledge reduction. The two fundamental concepts used in knowledge reduction are the core and reduct. Intuitively, a reduct of knowledge is its essential part, which suffices to define all basic classifications occurring in the considered knowledge, whereas the core is in a certain sense it's most important part. Let A set of attributes and let $a \in A$, the attribute a is dispensable in A if:

$$IND(A) = IND(A - \{a\}) \quad (3)$$

Otherwise a is indispensable attribute. The set of attributes B , where $B \subset A$ is called reduct of A if:

$$IND(B) = IND(A) \quad (4)$$

and A may have many reducts. The set of all indispensable attributes in A will be called the core of A , and will be denoted as $CORE(A)$:

$$CORE(A) = \bigcap RED(A) \quad (5)$$

Recently, many researches have proposed to represent knowledge in a form of discernibility matrix [12]. This representation has many advantages because it enables simple computation of the core and reduct of knowledge.

Let $K = (U, A)$ be a knowledge representation system with $U = \{x_1, x_2, \cdots x_n\}$ by a discernibility matrix of K denoted by $M(k)$, which means $n \times n$ matrix defined by:

$$(c_{ij}) = \{a \in A : \rho(x_i, a) \neq \rho(x_j, a)\} \ for \ i, j = 1,2....,n \quad (6)$$

Thus entry c_{ij} is the set of all attributes which discern objects x_i and x_j .

The core can be defined now as the set of all single element entries of the discernibility matrix, i.e.

$$CORE(A) = \{a \in A : c_{ij} = (a) \ for \ some \ i,j\}. \quad (7)$$

It can be easily seen that $B \subset A$ is the reduct of A if B is the minimal subset of A such that $B \bigcap c \neq \phi$ for any nonempty

entry $c\,(c \neq \phi)$ in $M(k)$.In other words reduct is the minimal subset of attributes that discerns all objects discernible by the whole set of attributes. Let $C, D \subset A$ be two subsets of attributes, called condition and decision attributes respectively. *KR*- system with distinguished condition and decision attributes will be called a decision table and will be denoted $T = (U, A, C, D)$ .Every $x \in U$ associate a function $d_x : A \to V_a$, such that $d_x(a) = a(x) = \rho(x, a)$, for every $a \in C \bigcup D$; the function dx will be called a decision rule, and x will be referred to as a label of the decision rule d_x [5].

III. Supervising Learning Based on Rough Sets with Modified Relation

Transforming non categorical attributes in decision table into categorical ones is done by using Rough Sets Boolean Reasoning RSBR discretization Algorithm [1,3,9,10]. Using reduct algorithm to generate reducts and induce decision rules that associated from the discretized decision table with two factors strength and certainty factors. Moreover, the medical dataset suffers from missing attribute values that cause more complex problem. Thus , the similarity matrix is used to solve these problems, wherever the clustering process groups elementary sets, making the problem less complex than the original one.

A. Discretizing Continuous Features with Missing Values

Discretization means that a notion of "distance" between attribute values is not needed in contrast to many other machine learning techniques. Non-categorical attributes should be discretized as a preprocessing step. The discretization step thus determines how coarsely we want to view the world. In other words, Discretization is the process for transforming continuous feature into qualitative features [10]. For numerical attributes, this amounts to search for cut-off points that define intervals. In the medical domain, there are often values that are "natural" to use as cut-off points and that can be used to manually discretize variables. Such cut-off points may not be found in the literature, thus, the existed algorithms can be used to suggest them [12]. A given number k could be considered as an upper bound for the number of cut point. In practice, k is set to be much less than the number of instances, assuming no repetition of continuous values for a feature [3]. The number of decision rules is affected by the number of values of the attributes. If many attributes have many vales, the number of decision rules increases. Therefore the number of cut points has to be evaluated carefully in the discretization process.

Although there are effective methods of discretization of real-valued attributes like entropy, frequency binning, naïve, semi naïve, different results by using different discretization methods are obtained. The results of discretization affect directly the quality of the discovered rules. In this work rough sets theory can be applied to compute a dependency measure considering the partitioning generated by the cut points and the decisional feature in order to obtain a better set of cut points [1,9,10].

Unlike some of discretization methods that totally ignore the effect of the discretized attribute values on the performance of the induction algorithm, Rough Set Boolean Reasoning, RSBR, combines discretization of real-valued attributes and classification. The basic concepts of the discretization based on the RSBR can be summarized as follows:

a) Sort the continuous values of the features to be discretized

b) Discretization of a decision table, where $V_c = [\upsilon_c, \omega_c)$is an interval of real values taken by attribute c; is a searching process for a partition $\mathcal{P}_c$ of V_c for any $c \in \mathcal{C}$ satisfying some optimization criteria (depend only of the computation of the dependency measure) while preserving some discernibility constraints. Any partition of $V_c$is defined by a sequence of the so-called cuts $v_1 < v_2 < \cdots . < v_k$ from V_c;

c) Proposing an algorithm to do so using the Scott's formula to obtain a family of partitions$\{\mathcal{P}_c\}_{c \in \mathcal{C}}$ which can be identified with a set of cuts.

Unfortunately, the Thyroid dataset contains missing values, gaps. These gaps should be violated before measuring the rules' dependencies. As the result, similarity measures are used to find similar pairs of objects. Besides the discretization approach [3] , the missing attribute value has simultaneously been solved during learning algorithm, as illustrated in Fig. 1.

Fig. 1. the discretization computing algorithm

Input: Knowledge Representation System with n instance and f feature , $K = (U, A, d)$

Output: minimal sets of decision rules d_x

For each continuous feature υ_k ($k = 1 \ldots p$)

A. For j= 2 to m_k, ($m_k$is nclass.scott(υ_k))

a. Calculate the partition considering j equal interval then define a set of Boolean variables BV(U) .

b. Create a new decision table $T^{\mathcal{P}}$by using the set of Boolean variables defined in step a, where $T^{\mathcal{P}}$ is called P-discretization of T.

c. Treat the missing attribute values in the decision attribute system $T^{\mathcal{P}} = (U, A^{\mathcal{P}}, d)$

d. Compute the dependency measure $\gamma_j(\upsilon_k) = \frac{\text{card}(\text{Pos}(\upsilon_k, d))}{\text{card}(U)}$

e. If(number of partitions $p_k = \arg\max_j \gamma(\upsilon_k)$ OR $\gamma_j(\upsilon_k) = 1$)

Stopping Criteria ($T^{\mathcal{P}}$construct a new data matrix with discrete values)

B. *Endfor* j

C. Divide the range of υ_k considering p_k interval

Endfor k.

The algorithm depends only on the computation of the dependency measure that needs to compute the complexity of missing attribute value algorithm. Thus, in the worst case the order of the algorithm is $O(n^2 p) * O(T^{\mathcal{P}} = (U, A^{\mathcal{P}}, d))$, where n is the number of instance and p is the number of attributes. To treat the missing values appeared in the Thyroid dataset, the second part of our approach consists of rough set analysis for discovering the gaps (missing attribute values).

The rough set approach used here is modified to deal with Incomplete Information System, where *IIS = (U, AT)*, where

U and AT are nonempty finite sets called "the universe" and "the set of attributes," respectively. Every attribute $a \in AT$, has a set of values called V_a and this set contains missing values for at least one attribute. In order to process Incomplete Information Systems (IIS), the indiscernibility relation has been extended to some equivalent relations, for example, tolerance relation, similarity relation, valued tolerance relation, and so forth. Similarity relation $SIM(A)$ denotes a binary relation between objects that are possibly indiscernible in terms of values of attributes and in the case of missing values the modified relation is:

$$SIM(A) = \{(x,y) \in U \times U,\ a \in A, \rho(x,a) = \rho(y,a) \ or\ \rho(x,a) = * \ or\ \rho(y,a) = *\} \quad (8)$$

$$SA(x) = \{y \in U : (x,y) \in SIM(A), A \in AT\} \quad (9)$$

SA(x) is the maximal set of objects which are possibly indiscernible by A with x.

The modified similarity relation ($MSIM$) can be defined as follows[15]:

1) $(x,x) \in MSIM(A)$ where A $\subset$ AT, for allx $\in$ U;
2) (x, y) $\in MSIM(A)$ where A $\subset$ AT, N = |A| ≥ 2 if and only if
   a. $\rho_x^a = \rho_y^a, \forall a \in A$ where ρ_x^a, ρ_y^a are defined values,
   b. EP(x, y) $\geq \frac{N}{2}$ if N is even
   c. EP(x, y) $\geq \frac{N+1}{2}$ if N is odd (10)

Where EP(x, y) = $\left|(\rho_x^a, \rho_y^a)\right|$ for all $a \in A$, $A \subseteq AT$ is the number ofequal pairs for the attribute "a" for all $a \in A$ for the objects "x," "y," respectively, where ρ_x^a, $\rho_y^a$are defined values. (10)

There are several kinds of reduct considering for decision tables. In this paper, Let $\mathcal{K} = (U, A', d)$ be a decision system. The generalized decision in $\mathcal{K}$ is the function $\partial_A: U \to \mathcal{P}(V_d)$ which is defined by

$$\partial_A'(x) = \{i : \exists\, y \in U, y\ MSIM(A')x \text{ and } d(y) = i\}, \quad (11)$$

A decision system $\mathcal{K}$ is called consistent, if $|\partial_{A'}(x)| = 1$ for any $x \in U$ otherwise $\mathcal{K}$ is inconsistent. Any set consisting of all objects with the same generalized decision value is called the generalized decision class.

The decision relative reduct may be found from the modified discernibility matrix $M^d(\mathcal{K}) = (c_{ij}^d)$, the elements in the discernibility matrix can be defined as follows:

$$c_{ij}^d = \begin{cases} c_{ij} - \{d\} & |\partial_{A'}(x_i)| = 1\ or |\partial_{A'}(x_j)| = 1\ and\ |\partial_{A'}(x_i)| \neq |\partial_{A'}(x_j)| \\ \emptyset & otherwise \end{cases} \quad (12)$$

Where c_{ij} is computed by equation (6).

The minimal reducts for the Incomplete Information System IIS are as follows: a set $B \subset A'$ is reduct of IIS if and only if:

$$\partial_B = \partial_{A'}, \forall C \subset B, MSIM(C) \neq MSIM(A') \quad (13)$$

Construction of the decision relative discernibilty function Δ from the discernibilty matrix $M^d(\mathcal{K})$ shows that the prime implicants, DNF, of the Boolean function representation of discernibility matrix. Δ is a discernibility function for IIS if and only if:

$$\Delta = \wedge \{\vee c_{ij},\ (x_i, x_j) \in U \times U, c_{ij} \neq \phi\} \quad (14)$$

These analyses try to cluster similar Thyroid attributes due to the presence of missing bases (gaps) inside it. The following algorithm is used to analysis the similarity of rules using rough sets:

Rough Sets Algorithm: Compute minimal sets of decision rules

Input: The Decision Representation System $K = (U, A', d), A' = A^P$ (n instance and p attributes)

Output: minimal sets of decision rules d_x

1. Compute $S_{A'}(x)$ for each object x in U
2. Compute $MSIM(A')$ for the set A' of attributes
3. $reduct_{\min} \leftarrow A$
4. $N \leftarrow |reduct_{\min}|$
5. For $i = 0$ to $N-1$ do
   Remove the i^{th} attribute a_i from the set $reduct_{\min}$
   If $MSIM(A) \neq MSIM(reduct_{\min})$
   $reduct_{\min} \leftarrow reduct_{\min} \cup a_i$
   endif
   endfor
6. Construct discernibility matrix $(c_{ij})_A$
7. Compute discernibility function Δ
8. Describe sets of d_x specified by Δ

Fig. 2. the computation of gaps using MSIM

The complexity of the algorithm of missing attribute value computation should be of order $O(n^2)$. Then the whole algorithm should be of order $O(n^4 p)$.

B. Rule Generation

Decision algorithm is a finite set of "if..then" decision rules. With every decision rule three coefficients are associated: the strength, the certainty and the coverage factors of the rule. The coefficients can be computed from the data or can be a subjective assessment. It is shown that these coefficients satisfy Bayes' formula. Bayesian inference methodology consists in updating prior probabilities by means of data to posterior probabilities, which express updated knowledge when data become available. The strength, certainty

and coverage factors can be interpreted either as probabilities (objective), or as a degree of truth. Moreover, they can be also interpreted as a deterministic flow distribution in flow graphs associated with decision algorithms. This leads to a new look on Bayes' theorem and its applications in reasoning from data, without referring to its probabilistic character.

Rough set theory depend on philosophy of classifications so information system should be expressed by dividing non-empty finite set of attributes A into two subsets condition attribute C and decision attribute D (this process is called supervised learning) and information system in this case called decision table

$$DT = \{U, C \cup D, V\}$$

Where C is a set of condition attributes and $D \notin C$ is decision attribute

A decision rule is an expression in the form, read " if C then D ", where C and D are logical formulas called condition and decision of the rule, respectively [5,11]. Let |C|denote the set of all objects from the universe U, having the property C. If $C \rightarrow D$ is a decision rule then $supp(C, D)$ =card (|C^D|) will be called the support of the decision rule and

$$\sigma(C, D) = \frac{supp(C,D)}{card(U)} \quad (15)$$

will be referred to as the strength of the decision rule.

With every decision rule $C \rightarrow D$, the certainty factor is interpreted as the frequency of objects having the property D in the set of objects having the property C

$$Cer(C, D) = \frac{supp(C,D)}{card(C)} \quad (16)$$

IV. THYROID EXPERIMENT USING ROUGH SETS

An experimental database of thyroid records obtained from the Garvan Institute of Medical Research [4]. Two files are used for this diagnosing application. The firs file is the names file that describes the attributes; about 30 attributes information are applying for each patient: age, sex, on thyroxine, query on thyroxine, on antithyroid medication, sick, pregnant, thyroid surgery, I131 treatment, query hypothyroid, query hyperthyroid, lithium, goitre, tumor, hypopituitary, psych, TSH measured, TSH, T3 measured, T3, TT4 measured, TT4, T4U measured, T4U, FTI measured, FTI, TBG measured, TBG , referral source. The second file, the application's data file provides information on the cases for each patient. The entry for each case consists of one line that give the values for all explicitly defined attributes. If an attribute value is not known, it is replaced by "*"

Our problem is the diagnosis of hypothyroidism. The idea is to measure blood levels of T4 and TSH [6,14]. It is based almost exclusively upon measuring the amount of thyroid hormone in the blood. So there are normal ranges for thyroid hormones which have been calculated by our application in this paper. We have four types of thyroid diagnosing: Hyperthyroid, Primary hypothyroid, Compensated hypothyroid, Secondary hypothyroid. And negative. Hypothyroidism is treated by replacing the missing hormone, a hormone that is essential to the body's key functions. Diagnosis of thyroid disease is a process that depends on: Clinical evaluation, blood tests, and imaging tests. In our application, we rely on the blood test which includes the following:

- T4: Thyroxine
- T3: Triiodothyronine
- TSH: Thyroid Stimulating Hormone Test
- TT4: Total T4/ Total Thyroxine
- TBG: Thyroglobulin/Thyroid Binding Globulin.
- T4U: Thyroxine utilization rates
- FTI: Thyroid Function Tests.

The aim of the experiments is to provide some preliminary evidence on how effective the new method of feature selection is and compare the experiments after applying it. In order to evaluate the feature subset selection using Modified Similarity relation of Rough Sets, then run experiments on 2514 record of datasets.

Rough Sets classify the training Thyroid data and mostly instances were classified correctly and errors are decreased. Statistics are summarizing that accuracy is 97.49%. About 2451 instances are correctly classified. The discovered rules are described in table II.

TABLE II. THE DESCRIPTIONS OF THE RULES GENERATED FROM ROUGH SETS LEARNING WITH MSIM METHODS

RULES	DESCRIOTION
Rule 1	IF TSH <= 6 THEN Medication-diagnosis-of-Hypothyroid = Negative (supp=2246./cer=1.0)
Rule 2	IF TSH > 6 AND FTI <= 64 AND TSH-measured = Normal AND T4U-measured = Normal AND thyroid-surgery = FALSE THEN Medication-diagnosis-of-Hypothyroid = Primary-hypothyroid (supp=59.0/cer=1.0)
Rule 3	IF TSH > 6 AND FTI <= 64 AND on-thyroxine = FALSE AND TSH-measured = Normal AND thyroid-surgery = FALSE AND TT4 <= 150 AND TT4-measured = Normal AND TSH <= 47 THEN Medication-diagnosis-of-Hypothyroid = Compensated-hypothyroid (supp=126.0)
Rule 4	IF TSH > 6 AND FTI > 64 AND on-thyroxine = TRUE AND referral-source = other AND THEN Medication-diagnosis-of-Hypothyroid = Negative (supp= 31.02)
Rule 5	IF TSH > 6 AND FTI > 64 AND on-thyroxine = FALSE AND TSH-measured = Abnormal THEN Medication-diagnosis-of-Hypothyroid = Negative (supp=25.19)
Rule 6	IF TSH > 6 AND FTI > 64 AND on-thyroxine = FALSE AND TSH-measured = Normal AND thyroid-surgery = FALSE AND TT4 <=150 AND TT4-

measure= Normal AND TSH <= 39THEN Medication-diagnosis-of-Hypothyroid = Negative (15.0)

Rule 7 IF TSH > 6 AND FTI > 64 AND on-thyroxine = FALSE AND TSH-measured = Normal AND thyroid-surgery = FALSE AND TT4 <=150 AND TT4-measure= Abnormal THEN Medication-diagnosis-of-Hypothyroid = Primary-hypothyroid (supp=6.0/cer=1.0)

Rule 8 IF TSH > 6 AND FTI > 64 AND on-thyroxine = FALSE AND TSH-measured = Normal AND thyroid-surgery = TRUE THEN Medication-diagnosis-of-Hypothyroid = Negative (supp=3.0/cer= 1.0)

Rule 9 IF TSH > 6 AND FTI <= 64 AND TSH-measured = Normal AND T4U-measured = Abnormal THEN Medication-diagnosis-of-Hypothyroid = Compensated-hypothyroid (supp=2.0)

In contrast with other previous methods that have already been used in treating Thyroid dataset [8], a comparative study between the machine learning algorithm C4.5 and rough sets with the modified similarity relation is established. The tree size created by WEKA, as depicted in figure (3) where the horizontal bar taken over 500 data element in each measure, shows that rough sets with MSIM has less tree size than that of C4.5. This will cause the reduction of the time needed for learning.

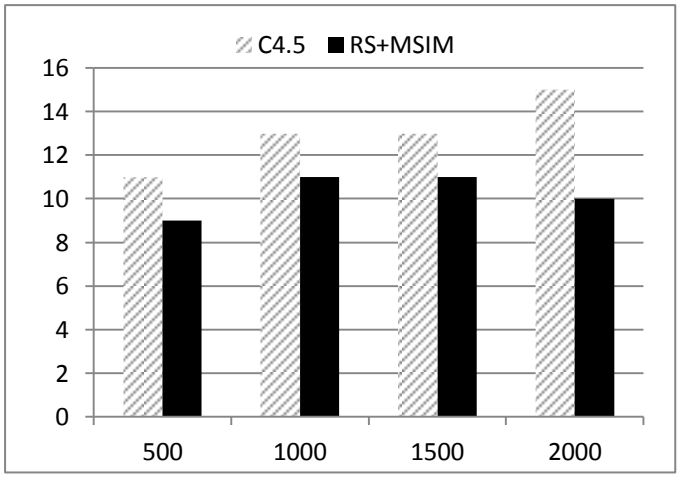

Fig. 3. a comparative study for measuring the tree size for C4.5 and rough sets with MSIM

Moreover, depending on a Thyroid dataset that are measured by [6], This study included 414 patients with thyroid diseases attending 3 main hospitals in Makkah (Al-Noor , Hera and King Abdul Aziz hospitals) over a period of one year (2007-2008).The accuracy classification is measured , about 371 patients are correctly classified.

V. Conclusions and Feature Work

Hypothyroid is one of the most common diseases. It is affects almost every aspect of health. The thyroid produces several hormones, each of them must be produced by Thyroidin normal rang; to help cells convert oxygen and calories into energy. Since Thyroid datasets are uncertain data, missing attribute values, and continuous features, Rough Sets treat these problems in the Thyroid dataset. Moreover, The MSIM, modified similarity analysis relation, is used to classify rules contain missing attribute value, gaps, with respect to the number of the whole defined attributes for each rule. Also, constructing of discernibility matrix, deduction of the production rules, and reducts in the presence of the missing attribute value are used to extract the minimal set of productions rules that describe similarity relation among rules. Hence, feature selection reduces the dimensionality of the data, the size of the hypothesis space and allows classification algorithm to operate faster and more effectively. These objectives make difference in building diagnosing algorithms more than any other machine learning algorithm. We have presented a reliable learning method and analytical study for diagnosing hypothyroid disease that can be used by doctors in other medical diagnosing algorithms. Indeed statistical results show that this evolutionary classification algorithm is the best in reducing size of tree, time, attributes and increasing accuracy. Although rough sets with modified similarity relation achieved good results that that of the machine learning algorithms, it still suffer from unsatisfied accuracy measure. Hence a hybrid model of rough sets and the machine learning should be introduced. This method uses the class information entropy of candidate patients to select the bin boundaries. Moreover, the missing attribute values are treated based on computing the information gain by dropping an attribute, then a similarity relation is measured.

Acknowledgment

We would like to thank the simulated research environment provided by the WEKA to machine learning group, University of Waikato, Hamilton, New Zealand. Also, a great appreciation to Gravan Institute of Medical Research for providing the medical database. Also, we should express our deep appreciation for Mrs. Nesma Elsayed for her help in preparing results.

References

[1] Ankit Gupta, Kishan G. Mehrotra,Chilukuri Mohan,"A Clustering-Based Discretization for Supervised Learning, "Statistics and Probability Letters, vol.80, pp.816-824, May 2010.

[2] Ching-Hsue Cheng & Jr-Shian Chen," Diagnosing Cardiovascular Disease Using an Enhanced Rough Sets Approach", Applied Artificial Intelligence: An International Journal,vol. 23, no. 6,pp. 487-499, 2009.

[3] Frida R. Coaquira Nina, On Applications Of Rough Sets Theory To Knowledge Discovery, Ph.D. Thesis, UNIVERSITY OF PUERTO RICO, 2007.

[4] ftp://ftp.ics.uci.edu/pub/machine-learning-databases.Accessed 9.8.2013

[5] Georg Peters, Richard Weber, Rene Nowatzke, "Dynamic Rough Clustering and its Applications", Applied Soft Computing, vol. 12, pp. 3193-3207, 2012.

[6] Hawazen A. Lamfon, Thyroid Disorders In Makkah, Saudi Arabia, Ozean Journal of Applied Sciences, vol. 1, no. 1,pp. 55-58, 2008.

[7] Li-Na Li & Ji-Hong Ouyang & Hui-Ling Chen & Da-You Liu," A Computer Aided Diagnosis System for Thyroid Disease Using Extreme Learning Machine", J Med Syst, vol. 36 no. 5, pp. 3327-3337, 2012.

[8] Nesma Ibrahim, Taher Hamza, Elsayd Radwan, "An Evolutionary Machine Learning Algorithm for Classifying Thyroid Diseases Diagnoses, "Egyptian Computer Science Journal,vol.35, no. 1, pp.73-86,Jan 2011.

[9] Nguyen H. Son and Skowron," A.: Boolean reasoning for feature extraction problems, In: Foundations of Intelligent Systems (Z.W. Ras, A. Skowron, Eds.). | Berlin: Springer, pp.117-126, 1997.

[10] Nguyen H. Son and Skowron A.: Quantization of real value attributes. | Proc. Int. Workshop Rough Sets and Soft Computing at 2nd Joint Conf. Information Sciences (JCIS'95), Durham, NC, pp.34-37, (1995).

[11] Pawlak, Z., " Rough Sets- Theoretical Aspects of Reasoning about Data", Kluwer Academic, Dordrecht, 1991.

[12] Polkowski, L., Skowron, A.(Eds), Rough Sets in Knowledge Discovery, Physica- Verlag, Heidelberg, vols: 1 and 2, 1998

[13] Roman, W. Swiniarski, Andrzej Skowron," Rough Set Methods in Feature Selection and Recognition", Pattern Recognition Letters, vol. 24, pp. 833-849, 2003.

[14] Keles, A., "ESTDD: Expert System For Thyroid Diseases Diagnosis", Expert Syst. Appl. vol.34 no. 1,pp. 242–246, 2008.

[15] Sara El-Sayed El-Metwally, Elsayed Radwan, Taher Hamza, " Multiple DNA Sequence Alignment using a Hybrid Model of GA and Rough Sets", Egyptian Computer Science Journal, vol. 34 no. 3, May 2010.