*Research Article*

# Rough Sets Data Analysis in Knowledge Discovery: A Case of Kuwaiti Diabetic Children Patients

**Aboul Ella Hassanien,[1, 2] Mohamed E. Abdelhafez,[3] and Hala S. Own[4]**

[1] *Information Technology Department, Faculty of Computer and Information, Cairo University, 5 Ahmed Zewal Street, Orman, Giza 12613, Egypt*

[2] *Quantitative and Information System Department, College of Business Administration, Kuwait University, P.O. Box 5468, Safat 13055, Kuwait*

[3] *Quantitative Methods and Information Systems Department, College of Business Administration, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait*

[4] *Department of Solar and Space Research, National Research Institute of Astronomy and Geophysics, Helwan, Cairo 11421, Egypt*

Correspondence should be addressed to Aboul Ella Hassanien, abo@cba.edu.kw

The main goal of this study is to investigate the relationship between psychosocial variables and diabetic children patients and to obtain a classifier function with which it was possible to classify the patients on the basis of assessed adherence level. The rough set theory is used to identify the most important attributes and to induce decision rules from 302 samples of Kuwaiti diabetic children patients aged 7–13 years old. To increase the efficiency of the classification process, rough sets with Boolean reasoning discretization algorithm is introduced to discretize the data, then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. Finally, the rough sets dependency rules are generated directly from all generated reducts. Rough confusion matrix is used to evaluate the performance of the predicted reducts and classes. A comparison between the obtained results using rough sets with decision tree, neural networks, and statistical discriminant analysis classifier algorithms has been made. Rough sets show a higher overall accuracy rates and generate more compact rules.

## 1. INTRODUCTION

Recently, diabetes has become one of the most common chronic diseases among children. Several studies have shown that diabetes has a great impact on the children life. In particular, it has been shown that diabetic children are highly exposed to emotional and behavioral problems compared with normal children [1]. The relationship of diabetes to psychological problems among adolescents has been investigated by many authors [2].

Medical databases have accumulated large quantities of information about patients and their medical conditions. Relationships and patterns within these data could provide new medical knowledge [3–6]. Analysis of medical data is often concerned with treatment of incomplete knowledge, with management of inconsistent pieces of information and with manipulation of various levels of representation of data. Existing intelligent techniques of data analysis are mainly based on quite strong assumptions (some knowledge about dependencies, probability distributions, large number of experiments), that are unable to derive conclusions from incomplete knowledge or cannot manage inconsistent pieces of information.

The classification of a set of objects into predefined homogenous groups is a problem with major practical interest in many fields, in particular, in medical sciences [5, 7, 8]. Over the past two decades, several traditional multivariate statistical classification approaches, such as the linear discriminant analysis, the quadratic discriminant analysis, and the logit analysis, have been developed to address the classification problem. More advanced and intelligent techniques have been used in medical data analysis such as neural network, Bayesian classifier, genetic algorithms, decision trees, fuzzy theory, and rough set. Fuzzy sets [9] provide a natural

framework for the process in dealing with uncertainty. It offers a problem-solving tool between the precision of classical mathematics and the inherent imprecision of the real world. Neural networks [10] provide a robust approach to approximating real-valued, discrete-valued, and vector-valued functions. The well-known backpropagation algorithm, which uses gradient descent to tune network parameters to best fit the training set with input-output pair, has been applied as a learning technique for the neural networks. Other approaches like case-based reasoning and decision trees [11, 12] are also widely used to solve data analysis problems. Each one of these techniques has its own properties and features including their ability of finding important rules and information that could be useful for the medical field domain. Each of these techniques contributes a distinct methodology for addressing problems in its domain. Rough set theory [13–15] is a fairly new intelligent technique that has been applied to the medical domain, and is used for the discovery of data dependencies, evaluates the importance of attributes, discovers the patterns of data, reduces all redundant objects and attributes, and seeks the minimum subset of attributes. Moreover, it is being used for the extraction of rules from databases. One advantage of the rough set is the creation of readable if-then rules. Such rules have a potential to reveal new patterns in the data material.

In this paper, we present an intelligent data analysis approach based on rough sets theory for generating classification rules from a set of observed 302 samples of the diabetic Kuwaiti children patients to investigate the relationship between diabetes and psychological problems for Kuwaiti children aged 7–13 years old.

This paper is organized as follows. Section 2 gives a brief introduction to the rough sets. Section 3 discusses the proposed rough set data analysis scheme in detail. The motivation and characteristics of diabetic datasets are presented in Section 4. Experimental analysis and discussion of the results are described in Section 5. Comparative analysis with statistical discriminant analysis, neural networks, and decision trees are given and discussed in Section 6. Finally, conclusions are presented in Section 7.

## 2. ROUGH SETS: FOUNDATIONS

Rough sets theory is a new intelligent mathematical tool proposed by Pawlak [13–15]. It is based on the concept of approximation spaces and models of the sets and concepts. In rough sets theory, the data is collected in a table, called a decision table. Rows of a decision table correspond to objects, and columns correspond to features. In the data set, we assume that a set of examples with a class label to indicate the class to which each example belongs are given. We call the class label a decision feature, the rest of the features are conditional. Let $\mathcal{O}, \mathcal{F}$ denote a set of sample objects and a set of functions representing object features, respectively. Assume that $B \subseteq \mathcal{F}$, $x \in \mathcal{O}$. Further, let $[x]_B$ denote

$$[x]_B = \{y : x \sim_B y\}. \tag{1}$$

Rough sets theory defines three regions based on the equivalent classes induced by the feature values: lower ap-

proximation $\underline{B}X$, upper approximation $\overline{B}X$, and boundary $\mathrm{BND}_B(X)$. A lower approximation of a set $X$ contains all equivalence classes $[x]_B$ that are subsets of $X$, and upper approximation $\overline{B}X$ contains all equivalence classes $[x]_B$ that have objects in common with $X$, while the boundary $\mathrm{BND}_B(X)$ is the set $\overline{B}X \setminus \underline{B}X$, that is, the set of all objects in $\overline{B}X$ that are not contained in $\underline{B}X$. So, we can define a rough set as any set with a nonempty boundary.

The indiscernibility relation $\sim_B$ (or by $\mathrm{Ind}_B$) is a mainstay of rough set theory. Informally, $\sim_B$ is a set of all objects that have matching descriptions. Based on the seletion of $B$, $\sim_B$ is an equivalence relation partitions a set of objects $\mathcal{O}$ into equivalence classes. The set of all classes in a partition is denoted by $\mathcal{O}/\sim_B$ (also by $\mathcal{O}/\mathrm{Ind}_B$). The set $\mathcal{O}/\mathrm{Ind}_B$ is called the quotient set. Affinities between objects of interest in the set $X \subseteq \mathcal{O}$ and classes in a partition can be discovered by identifying those classes that have objects in common with $X$. Approximation of the set $X$ begins by determining which elementary sets $[x]_B \in \mathcal{O}/\sim_B$ are subsets of $X$.

Here we provide a brief explanation of the basic framework of rough set theory, along with some of the key definitions. A review of this basic material can be found in sources such as [13–15].

### 2.1. Information system and approximation

*Definition 1* (information system). Information system is a tuple $(U, A)$, where $U$ consists of objects and $A$ consists of features. Every $a \in A$ corresponds to the function $a : U \rightarrow V_a$, where $V_a$ is value set of $a$. In applications, we often distinguish between conditional features $C$ and decision features $D$, where $C \cap D = \varnothing$. In such cases, we define decision systems $(U, C, D)$.

*Definition 2* (indiscernibility relation). Every subset of features $B \subseteq A$ induces indiscernibility relation:

$$\mathrm{Ind}_B = \{(x, y) \in U \times U : \forall_{a \in B} \, a(x) = a(y)\}. \tag{2}$$

For every $x \in U$, there is an equivalence class $[x]_B$ in the partition of $U$ defined by $\mathrm{Ind}_B$.

Due to imprecision which exists in real world data, there are sometimes conflicting classification of objects contained in a decision table. Here conflicting classification occurs whenever two objects have matching descriptions, but are deemed to belong to different decision classes. In that case, a decision table contains an inconsistency.

*Definition 3* (lower and upper approximation). In the rough sets theory, the approximations of sets are introduced to deal with inconsistency. A rough set approximates traditional sets using a pair of sets named the lower and upper approximations of the set. Given a set $B \subseteq A$, the lower and upper approximations of a set $Y \subseteq U$ are defined by, respectively,

$$\underline{B}Y = \bigcup_{x:[x]_B \subseteq X} [x]_B,$$
$$\overline{B}Y = \bigcup_{x:[x]_B \cap X \neq \varnothing} [x]_B. \tag{3}$$

*Definition 4* (lower approximation and positive region). The positive region $\text{POS}_C(D)$ is defined by

$$\text{POS}_C(D) = \bigcup_{X:X \in U/\text{Ind}_D} \underline{C}X; \qquad (4)$$

$\text{POS}_C(D)$ is called the positive region of the partition $U/\text{Ind}_D$ with respect to $C \subseteq A$, that is, the set of all objects in $U$ that can be uniquely classified by elementary sets in the partition $U/\text{Ind}_D$ by means of $C$ [16].

*Definition 5* (upper approximation and negative region). The negative region $\text{NEG}_C(D)$ is defined by

$$\text{NEG}_C(D) = U - \bigcup_{X:X \in U/\text{Ind}_D} \overline{C}X, \qquad (5)$$

that is, the set of all objects that can be definitely ruled out as members of $X$.

*Definition 6* (boundary region). The boundary region is the difference between upper and lower approximations of a set $X$ that consists of equivalence classes having one or more elements in common with $X$. It is given by the following formula:

$$\text{BND}_B(X) = \underline{B}X - \overline{B}X. \qquad (6)$$

### 2.2. Reduct and core

An interesting question is whether there are features in the information system (feature-value table) which are more important to the knowledge represented in the equivalence class structure than other features. Often we wonder whether there is a subset of features which by itself can fully characterize the knowledge in the database. Such a feature set is called a reduct. Calculation of reducts of an information system is a key problem in RS theory [14, 15, 17]. We need to get reducts of an information system in order to extract rule-like knowledge from an information system.

*Definition 7* (reduct). Given a classification task related to the mapping $C \rightarrow D$, a reduct is a subset $R \subseteq C$ such that

$$\gamma(C, D) = \gamma(R, D) \qquad (7)$$

and none of proper subsets of $R$ satisfies analogous equality.

*Definition 8* (reduct set). Given a classification task mapping a set of variables $C$ to a set of labeling $D$, a reduct set is defined with respect to the power set $P(C)$ as the set $R \subseteq P(C)$ such that $\text{Red} = \{A \in P(C) : \gamma(A, D) = \gamma(C, D)\}$. That is, the reduct set is the set of all possible reducts of the equivalence relation denoted by $C$ and $D$.

*Definition 9* (minimal reduct). A minimal reduct $R_{\text{minimal}}$ is the reduct such that $\|R\| \leq \|A\|$, for all $A \in R$. That is, the minimal reduct is the reduct of least cardinality for the equivalence relation denoted by $C$ and $D$.

*Definition 10* (core). Attribute $c \in C$ is a core feature with respect to $D$, if and only if it belongs to all the reducts. We denote the set of all core features by $\text{Core}(C)$. If we denote by $R(C)$ the set of all reducts, we can put

$$\text{Core}(C) = \bigcap_{R \in R(C)} R. \qquad (8)$$

The computation of the reducts and the core of the condition features from a decision table is a way of selecting relevant features. It is a global method in the sense that the resultant reduct represents the minimal set of features which are necessary to maintain the same classification power given by the original and complete set of features. A straighter manner for selecting relevant features is to assign a measure of relevance to each feature and choose the features with higher values. Based on the generated reduct system we will generate list of rules that will be used for building the classifier model of the new object with each object in the reduced decision table (i.e., reduct system) and classify the object to the corresponding decision class. The calculation of all the reducts is fairly complex (see [12, 18–20]).

### 2.3. Significance of the attribute

Significance of features enables us to evaluate features by assigning a real number from the closed interval [0,1], expressing how important a feature in an information table is. Significance of a feature $a$ in a decision table DT can be evaluated by measuring the effect of removing of the feature $a$ in $C$ from feature set $C$ on a positive region defined by the table DT. As shown in Definition 2, the number $\gamma(C, D)$ expresses the degree of dependency between feature $C$ and $D$ or accuracy of approximation of $U/D$ by $C$; the formal definition of the significant is given as follows.

*Definition 11* (significance). For any feature $a \in C$, we define its significance $\zeta$ with respect to $D$ as follows:

$$\zeta(a, C, D) = \frac{|\text{POS}_{C \setminus \{a\}}(D)|}{|\text{POS}_C(D)|}. \qquad (9)$$

Definitions 7–11 are used to express the importance of particular features in building the classification model. For a comprehensive study we refer to [21]. One of importance measures is to use frequency of occurrence of features in reducts. Then, one can also consider various modifications of Definition 7, for example approximate reducts, which preserve information about decisions only to some degree [12]. Further, positive region in Definition 4 can be modified by allowing for the approximate satisfaction of inclusion $[x]_C \subseteq [x]_D$, as proposed, for example, in VPRS model [22]. Finally, in Definition 2, the meaning of $\text{IND}(B)$ and $[x]_B$ can be changed by replacing equivalence relation with similarity relation, especially useful when considering numeric features. For further reading, we refer to, for example, [14, 17].

### 2.4. Decision rules

In the context of supervised learning, an important task is the discovery of classification rules from the data provided in the decision tables. The decision rules not only capture patterns hidden in the data as they can also be used to classify new unseen objects. Rules represent dependencies in the dataset, and represent extracted knowledge which can be used when classifying new objects not in the original information system. When the reducts were found, the job of creating definite rules for the value of the decision feature of the information system was practically done. To transform a reduct into a rule, one only has to bind the condition feature values of the object class from which the reduct originated to the corresponding features of the reduct. Then, to complete the rule, a decision part comprising the resulting part of the rule is added. This is done in the same way as for the condition features. To classify objects, which has never been seen before, rules generated from a training set will be used. These rules represent the actual classifier. This classifier is used to predict to which classes new objects are attached. The nearest matching rule is determined as the one whose condition part differs from the feature vector of re-image by the minimum number of features. When there is more than one matching rule, we use a voting mechanism to choose the decision value. Every matched rule contributes votes to its decision value, which are equal to the times number of objects matched by the rule. The votes are added and the decision with the largest number of votes is chosen as the correct class. Quality measures associated with decision rules can be used to eliminate some of the decision rules.

### 3. ROUGH SETS DATA ANALYSIS TECHNIQUES

In this section, we will discuss in details the proposed rough sets scheme to analyze the diabetic children patient's databases. The scheme used in this study consists of two main stages: preprocessing and processing. Preprocessing stage includes tasks such as data cleaning, completeness, correctness, attribute creation, attribute selection and discretization. Processing includes the generation of preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, and classification processes. These stages leading towards the final goal of generating rules from information or decision system of the diabetic database. Figure 1 shows the overall steps in the proposed rough sets data analysis scheme.

### 3.1. Preprocessing stage

In order to successfully analyze data with rough sets, a decision table must be created. This is done with data preparation. The data preparation task includes data conversion, data cleansing, data completion checks, conditional attribute creation, decision attribute generation, discretization of attributes, and data splitting into analysis and validation subsets. Data conversion must be performed on the initial data into a form in which specific rough set tools can be applied. Data splitting created two subsets of size 252 objects for the
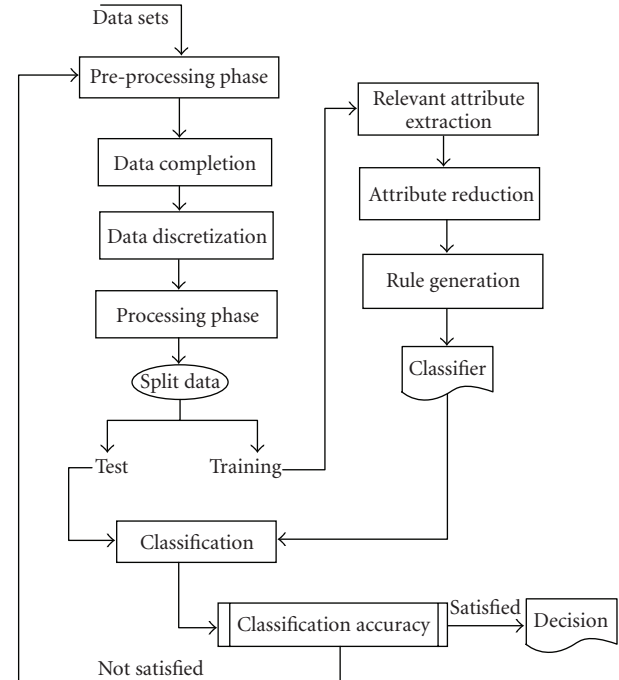


Figure 1: Overall rough sets data analysis scheme.

data analysis set and 50 objects for the validation set using a random seed. More details will be discussed later in the data characteristics and its description section.

### Data completion and discretization of continuous-valued attributes

Data completion is often the case in which real world data will contain missing values. Since rough set classification involves mining rules from the data, objects with missing values in the data set may have an undesirable effect on the rules that are constructed. The aim of this procedure is to remove all objects that have one or more missing values. Incomplete information systems exist broadly in practical data analysis, and approaches to complete the incomplete information system through various completion methods in the preprocessing stage are normal in data mining and knowledge discovery. These methods may result in distortion of original data and knowledge, and can even render the original data mining system unminable. To overcome these shortcomings inherent in traditional methods, we used the decomposition approach for incomplete information system proposed in [23].

When dealing with attributes in concept classification, it is obvious that they may have varying importance in the problem being considered. Their importance can be preassumed using auxiliary knowledge about the problem and expressed by properly chosen weights. However, in the case of using the rough set approach to concept classification, it avoids any additional information aside from what is included in the information table itself. Basically, the rough set approach tries to determine from the data available in the information table whether all the attributes are of the same

strength and, if not, how they differ in respect of the classifier power. Therefore, some strategies for discretization of real value attributes have to be used when we need to apply learning strategies for data classification with real value attributes (e.g., equal width and equal frequency intervals). It has been shown that the quality of learning algorithm is dependent on this strategy, which has been used for real data discritization [24]. Discretization which uses data transformation procedure that involves finding, cuts in the data sets which divide the data into intervals. Values lying within an interval are then mapped to the same value. Doing this process will lead to reduce the size of the attributes value set and ensures that the rules that are mined are not too specific. In this paper, we adopt the rough sets with boolean reasoning (RSBR) algorithm proposed by Zhong et al. [23] for the discretization of continuous-valued attributes. The main advantage of RSBR is that it combines discretization of real-valued attributes and classification. The main steps of the RSBR discretization algorithm are provided below.

### 3.2. Processing stage

As we mentioned before, processing stage includes generating preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, and classification processes. These stages lead towards the final goal of generating rules from information or decision system of the diabetic database.

#### Relevant attribute extraction and reduction

One of the important aspects in the analysis of decision tables extracted from data is the elimination of redundant attributes and identification of the most important attributes. Redundant attributes are any attributes that could be eliminated without affecting the degree of dependency between remaining attributes and the decision. The degree of dependency is a measure used to convey the ability to discern objects from each other. The minimum subset of attributes preserving the dependency degree is termed reduct. The computation of the core and reducts from a decision table is a way of selecting relevant attributes [20, 25]. It is a global method in the sense that the resultant reducts represent the minimal sets of features which are necessary to maintain the same classificatory power given by the original and complete set of attributes. A straighter manner for selecting relevant attribute is to assign a measure of relevance to each attribute and choose the attributes with higher values.

In decision tables, there often exist conditional attributes that do not provide (almost) any additional information about the objects. So, we should remove those attributes since it reduces complexity and cost of decision process [17, 20, 25, 26]. A decision table may have more than one reduct. Any of them can be used to replace the original table. Finding all the reducts from a decision table is NP-complete. Fortunately, in applications, it is usually not necessary to find all of them—one or few of them are sufficient. A natural question is which reducts are the best. The selection depends on the optimality criterion associated with the attributes. If it is possible to assign a cost function to attributes, then the selection can be naturally based on the combined minimum cost criteria. In the absence of an attribute cost function, the only source of information to select the reduct is the contents of the table. In this paper, we adopt the criteria that the best reducts are the ones with the minimal number of attributes and—if there are more such reducts—with the least number of combinations of values of its attributes (cf. [25, 27]). We introduce a reduct algorithm based on the degree of dependencies and the discrimination factors. The main steps of the reduct generation algorithm are provided below (refer to Algorithm 2).

#### Rule generation and classification

The generated reducts are used to generate decision rules. The decision rule, at its left side, is a combination of values of attributes such that the set of (almost) all objects matching this combination have the decision value given at the rule's rough side. The rule derived from reducts can be used to classify the data. The set of rules is referred to as a classifier and can be used to classify new and unseen data. The main steps of the Rule Generation and classification algorithm are provided below (refer to Algorithm 3).

When rules are generated, the number of objects that generate the same rule is typically recorded. The quality of rules that are generated based on attributes included in the reduct is connected with its quality. We would be specially interested in generating rules which cover possibly largest parts of the universe. Covering the universe space with more general rules implies smaller size of a rule set. We could therefore use this idea in measuring the quality of a reduct. If a rule is generated more frequently across different rule sets, we say this rule is more important than other rule. The rule importance measure [28] $R_I$ is used as an evaluation to study the quality of the generated rule. It is defined by

$$R_I = \frac{\tau_r}{\rho_r},\qquad(12)$$

where $\tau_r$ is the number of times a rule appears in all reduct and $\rho_r$ is the number of reduct sets. The quality of rules is related to the corresponding reduct(s). The generating rules which cover largest parts of the universe $U$ which means more general rules implies smaller size of a rule set.

## 4. CASE STUDY: DIABETES PATIENTS

### 4.1. Motivation

The present study is a cross-section study conducted at Kuwait University in the period September to November 2005. The sample comprised 302 children aged 7–13 years. Trained interviewers administered questionnaires to parents and caretakers.

Input: Information system table ($S$) with real valued attributes $A_{ij}$ and $n$ is the
      number of intervals for each attribute.
Output: Information table ($ST$) with discretized real valued attribute
1: **for** $A_{ij} \in S$ **do**
2:    Define a set of boolean variables as follows:

$$B = \left\{ \sum_{i=1}^{n} C_{ai}, \sum_{i=1}^{n} C_{bi} \sum_{i=1}^{n} C_{ci}, \ldots, \sum_{i=1}^{n} C_{ni} \right\} \qquad (10)$$

3: **end for**
    Where $\sum_{i=1}^{n} C_{ai}$ correspond to a set of intervals defined on the variables of
    attributes $a$
4: Create a new information table $S_{\text{new}}$ by using the set of intervals $C_{ai}$
5: Find the minimal subset of $C_{ai}$ that discerns all the objects in the decision class
    $D$ using the following formula:

$$\Upsilon^{u} = \wedge \{ \Phi(i,j) : d(x_i \neq d(x_j)) \} \qquad (11)$$

Where $\Phi(i,j)$ is the number of minimal cuts that must be used to discern two
different instances $x_i$ and $x_j$ in the information table.

ALGORITHM 1: RSBR discretization algorithm.

## 4.2. Data characteristics and its description

The data for this study were collected by the Statistical Consultation unit in the College of Business Administration at Kuwait University, Kuwait. Participants (parents or guardians) of diabetic children were interviewed to complete the questionnaire. The interviews have been conducted at a governmental hospital. The questionnaire consists of two parts; the first part has socio-demographic and clinical characteristics of the subjects and the second part is the *strengths and difficulties questionnaire (SDQ)* [29].

The socio-demographic and clinical characteristics include respondent nationality, respondent gender, child gender, respondent age, child age, respondent education, child education, family income, duration of diabetes (in years), does any of the parents have diabetes, number of brothers who have diabetes, number of times the child entered the hospital because of diabetes, the Hemoglobin level, and type of diabetes. The strengths and difficulties questionnaire (SDQ) is widely used as a useful tool for screening emotional and behavioral problems in children aging 4–16 years [1, 24, 30–37]. The SDQ has been translated into more than 40 languages, being available in the internet at www.sdqinfo.com. The SDQ has 25 items, some are positive and some are negative. The 25 items in the SDQ comprise 5 subscales of 5 items each. The five subscales are emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behavior.

Emotional symptoms scale. Often complains of headaches, stomach-ashesetc, worried, often unhappy, nervous, many fears, easily scared.

Hyperactivity scale. Restless, overactive, cannot stay still for long, constantly fidgeting or squirming, easily distracted, concentration wanders, thinks things out before acting, sees tasks through to the end, good attention span.

TABLE 1: SDQ classified scores.

| Parenet completed | Normal | Borderline | Abnormal |
|---|---|---|---|
| Total diffficulties score | 0–13 | 14–16 | 17–40 |
| Emotional symptoms score | 0–3 | 4 | 5–10 |
| Conduct problems score | 0–2 | 3 | 4–10 |
| Hyperactivity score | 0–5 | 6 | 7–10 |
| Peer problems score | 0–2 | 3 | 4–10 |
| Prosocial behavior score | 6–10 | 5 | 0–4 |

Conduct problems scale. Often has temper tantrums or hot tempers, generally obedient, often fights with other children or bullies them, often lies and cheats, steals from home, school, or elsewhere.

Peer problems scale. Rather solitary, tends to play alone, has at least one good friend, generally liked by other children, picked on or bullied by other children, gets on better with adults than with other children.

Prosocial scale. Considerate of other people's feeling, shares readily with other children, helpful if someone is hurt, upset of feeling ill, kind to younger children, often volunteers to help others.

Each of the 25 items is marked as not true, somewhat true, or certainly true. Except for five positive items, written in italic, the scores for each item are 0 for not true, 1 for somewhat true and 2 for certainly true. The five positive items, written in italic, are scored in the opposite direction, 2 for not true, 1 for somewhat true, and 0 for certainly true. The sum of the scores of each subscale ranges from 0 to 10. The sum of the scores of the first four subscales of the SDQ gives a total difficulties score, ranging from 0 to 40. High scores for each of the first four subscales indicate difficulties, whereas high scores on the prosocial subscale indicate

```
Input: information table (ST) with discretized real valued attribute.
Output: reduct sets R_final = {r_1 ∪ r_2 ∪ ··· ∪ r_n}
1: for each condition attributes c ∈ C do
2:     Compute the correlation factor between c and the decisions attributes D
3:     if the correlation factor > 0 then
4:         Set c as relevant attributes.
5:     end if
6: end for
7: Divide the set of relevant attribute into a different variable sets.
8: for each variable sets do
9:     Compute the dependency degree and compute the classification quality
10:    Let the set with high classification accuracy and high dependency as an initial
       reduct set.
11: end for
12: for each attribute in the reduct set do
13:    Calculate the degree of dependencies between the decisions attribute and that
       attribute.
14:    Merge the attributes produced in previous step with the rest of conditional
       attributes
15:    Calculate the discrimination factors for each combination to find the highest
       discrimination factors
16:    Add the highest discrimination factors combination to the final reduct set.
17: end for
18: repeat
       statements 12
19: until all attributes in initial reduct set is processed
```

ALGORITHM 2: Reduct generation algorithm.

```
Input: reduct sets R_final = {r_1 ∪ r_2 ∪ ··· ∪ r_n}
Output: Set of rules
1: for each reduct r do
2:     for each correspondence object x do
3:         Contract the decision rule (c_1 = v_1 ∧ c_2 = v_2 ∧ ··· ∧ c_n = v_n) ⟶ d = u
4:         Scan the reduct r over an object x
5:         Construct (c_i, 1 ≤ i ≤ n)
6:         for every c ∈ C do
7:             Assign the value v to the correspondence attribute a
8:         end for
9:         Construct a decision attribute d
10:        Assign the value u to the correspondence decision attribute d
11:    end for
12: end for
```

ALGORITHM 3: Rule generation and classification.

strengths. The author of the SDQ classified scores for each of the subscales and for the total difficulties as normal, borderline, and abnormal (clinical). These classified scores are shown in Table 1.

Table 2 shows the percentages of boys and girls and total percentages of children whose scores are in the normal, borderline, and abnormal classes.

The cutoffs were chosen so that roughly 80% of children in the community are categorized as normal, 10% are borderline and 10% are abnormal. Goodman (1997) in [29]

pointed out that the "borderline" cutoffs can be used with high-risk samples where false positives are not a major concern and the "abnormal" cutoffs can be used for studies of low-risk samples where it is more important to reduce the rate of false positives. In the present study, abnormal and borderline cases were considered positive for mental health problems. According to the total difficulties scores, the results showed that 69.1% of the children have overall mental health problems (55% abnormal and 14.1% borderline). The highest percentage (82.4%) was for the emotional problems

TABLE 2: Percentage of children in normal, borderline, and abnormal groups.

| Scale | Normal | Borderline | Abnormal |
|---|---|---|---|
| Total difficulties score | | | |
| (boys, girls, total) | (24.1, 36.6, 30.9) | (17.5, 11.2, 14.1) | (58.4, 52.2, 55) |
| Emotional symptoms score | | | |
| (boys, girls, total) | (20.9, 14.9, 17.7) | (9.4, 11.8, 10.7) | (69.8, 73.3, 71.7) |
| Conduct problems score | | | |
| (boys, girls, total) | (23.2, 27.3, 25.4) | (12.3, 12.4, 12.4) | (64.5, 60.2, 62.2) |
| Hyperactivity score | | | |
| (boys, girls, total) | (47.9, 64.6, 56.8) | (25, 14.3, 19.3) | (27.1, 21.1, 23.9) |
| Peer problems score | | | |
| (Boys, girls, total) | (69.8, 70.2, 70) | (12.2, 9.3, 10.7) | (18, 20.5, 19.3) |
| Prosocial behavior score | | | |
| (boys, girls, total) | (92.8, 90.7, 91.7) | (4.3, 4.3, 4.3) | (2.9, 5, 4) |

whereas the lowest percentage (30%) was for peer relationship problems. In the present study, abnormal and borderline cases were considered positive for mental health problems.

### 4.3. The internal consistency

The internal consistency of the SDQ for this sample, using Cronbach's alpha coefficient [38], was 0.59. For the SDQ subscales, Cronbach's alpha coefficient for the total difficulty scores was 0.72, whereas for the five subscales were prosocial subscale 0.51, hyperactivity subscale 0.49, emotional subscale 0.53, conduct subscale 0.61, and the peer subscale 0.27. The internal consistency of the SDQ has also been investigated in many countries (e.g., UK [30], Germany [34, 39], Holland [40], Sweden [36], Bangladeshi [2], Finnish [41]). The results across different countries supported the internal consistency of the SDQ. The SDQ has been used as a useful tool to identify the clinical cases. The range of correct identification was 81–91% [35]. The study comprised 302 respondents, (75.8%) citizen, and (24.2%) noncitizen, (32.1%) males and (67.9%) females. The age of the respondents ranges from 17 to 75 year-old and the respondent education ranges from elementary school (7%), intermediate (21%), high school (20.3%), diploma (19.7%), university level (31.7%) to master level (0.3%). Family income ranges from less than 500 KD (15.6%), 500–700 KD (17.3), 701–900 KD (14.6%), 901–1100 KD (9%) to more than 1101 KD (43.5%). The total sample of children was 302: (53.5%) girls and (46.5%) boys; the child age ranges from 7 to 13 years old with mean 10.2 years and SD (2.1); the child education ranges from elementary (49%), intermediate (50%), and high school (0.7%); the duration of diabetes (in years) ranges from 0 to 11 years with a mean of 3.52 years and a standard deviation of 2.343 years. Of the total sample of children (19.9%) with diabetic parents and (80.1%) their parents were not diabetic. The number of times the child entered the hospital because of high or low diabetes ranges from none (12.3%), (53.8%), 2-3 times (16.6%), 4-5 times (7.6%), to 6 or more times (9.6%). The hemoglobin readings range from 3 to 32 with a mean of 12.97 and a standard deviation of 5.88. Except two children (0.7%) having type 2 diabetes, all the children (99.3%) had type I diabetes.

## 5. EXPERIMENTAL ANALYSIS

The data set studies in this paper consists of 302 children patients with diabetes. Knowledge representation in rough set is done via information system which is a tabular form object and attributes value relation (refer to Table 1). The first analysis studies the statistical distribution of the attributes. For many data mining tasks, it would be useful to learn the more general characteristics about the given data set, like central tendency and data dispersion. Typical measure of central tendency is the mean or median. Very often in large data sets, there exist samples that are not consistent with the general behavior of the data model; such data are called outliers. Outlier detection is important since it may affect the classifier accuracy. The simplest approach of outlier detection is to use statistical measures. In our experiments we use the mean and median to detect the outliers in our data set. Tables 3 and 4 represent the statistical analysis and essential distribution of attributes, respectively.

By applying the introduced reduct generation algorithm (refer to Algorithm 2), we compute the degree of dependencies and the discrimination factors of the attributes.

Tables 5 and 6 show the discrimination factor for one and five attributes.

From Table 5, we observe that the conduct attribute has the highest discrimination factor, so we choose it as the first attribute in the next combination to generate sets of two attributes. The first one will be the conduct attribute and the second one will be the rest of the conditional attributes. Then we compute the discrimination factor for all sets and choose the highest discrimination factor for two attributes. We repeat the same procedure with three attributes, four attributes, and so forth, until we reach the minimal number of reducts that contains a combination of attributes which has

TABLE 3: Statistical results of the attributes.

| Attribute | Mean | Std. Dv. | Median | Correlation with decision class |
|---|---|---|---|---|
| Respondent age | 38.83 | 7.1 | 39.0 | $-0.1103$ |
| Child age | 10.223 | 2.05 | 10.0 | $-0.0336$ |
| Respondent education | 3.494 | 1032 | 4.0 | $-0.194$ |
| Child education age | 1.522 | 0.51 | 2.0 | $-0.02723$ |
| Income class | 3.487 | 1.55 | 4.0 | $-0.141$ |
| Duration of diabetes | 3.567 | 2.555 | 3.0 | 0.015351 |
| Number of diabetes brothers | 0.19377 | 0.4828 | 0 | 0.0003 |
| Number of hospital entrance | 2.463 | 1.087 | 2.0 | 0.2307 |
| Hemoglobin | 13.01 | 5.932 | 11.07 | 0.096 |
| Emotional | 5.9100 | 20561 | 6.0 | 0.59117 |
| Conduct | 4.460 | 2.447 | 4.0 | 0.6776 |
| Hyperactivity | 4.858 | 2051 | 5.0 | 0.610 |
| Peer | 1.913 | 1.710 | 2.0 | 0.452 |
| Prosocial | 8.065 | 1.892 | 8.0 | $-0.295$ |

TABLE 4: Attribute distribution within the classes.

| Attribute | Categories | Count | Precentages |
|---|---|---|---|
| Nationality | (Class1, Class2) | (220, 69) | (76.12%, 23.87%) |
| Respondent gender | (Class1, Class2) | (96, 193) | (33.21%, 66.78%) |
| Child gender | (Class1, Class2) | (133, 156) | (46.02%, 53.97%) |
| Child education | (Class1, Class2, Class3) | (140, 147, 2) | (48.44%, 50.86%, 0.69%) |
| Diabetic parent | (Class1, Class2) | (58, 231) | (20.069%, 79.93%) |
| Decision | (Class1, Class2, Class3) | (89, 39, 161) | (30.79%, 13.49%, 55.70%) |

TABLE 5: Discrimination factor for one attributes.

| Attribute | Discrimination factor |
|---|---|
| Nationality | 0 |
| Child fender | 0 |
| Respondent age | 0.027 |
| Child age | 0.0034 |
| Respondent education | 0 |
| Child education age | 0 |
| Income class | 0 |
| Diabetes parent | 0 |
| Duration of diabetes | 0.0207 |
| Number of diabetes brothers | 0 |
| Number of hospital entrance | 0 |
| Hemoglobin | 0.086 |
| Emotional | 0.079 |
| **Conduct** | **0.121** |
| Hyperactivity | 0.034 |
| Peer | 0.027 |
| Prosocial | 0.034 |

TABLE 6: Discrimination factor for five attributes.

| Attributes | Discrimination factor |
|---|---|
| Conduct, emotional, hyperactivity, peer, duration of diabetes | 0.557 |
| Conduct, emotional, hyperactivity, peer, child age | 0.557 |
| Conduct, emotional, hyperactivity, peer, prosocial | 0.557 |

A natural use of a set of rules is to measure how well the ensemble of rules is able to classify new and unseen objects. To measure the performance of the rules is to assess how well the rules do in classifying new cases. So we apply the rules produced from the training set data to the test set data. Our measuring criteria are sensitivity, specificity, and accuracy. The sensitivity of a classifier gives a measure of how good it is in detecting that an event defined through an object has occurred, while the specificity gives us a measure of how good it is in picking up nonevent defined through the object. These evaluation measures can be calculated from confusion matrix as shown in Table 8.

Table 9 shows the number of generated rules before and after pruning process. We can observe that the number of generated rules for all algorithms is large. It makes

the same discrimination factor (see Table 6). Table 7 shows the final generated reduct sets which are used to generate the list of rules for the classification.

TABLE 7: Generated reduct sets.

| No. | Reduct |
|---|---|
| 1 | Conduct, emotional, hyperactivity, peer, duration of diabetes |
| 2 | Peer, hyperactivity, conduct, emotional, hemoglobin, child education, respondent education, child gender |
| 3 | Peer, hyperactivity, conduct, emotional, hemoglobin, child education, child gender, nationality |
| 4 | Peer, hyperactivity, conduct, emotional, hemoglobin, respondent education, child gender, nationality |
| 5 | Hyperactivity, conduct, emotional, hemoglobin, child education, respondent education, child gender, nationality |
| 6 | Peer, hyperactivity, conduct, emotional, hemoglobin, child education, respondent education, nationality |

TABLE 8: Predict performance—confusion matrix.

| Actual | Predict Class1 | Predict Class2 | Predict Class3 | Accuracy |
|---|---|---|---|---|
| Class1 | 19 | 2 | 1 | 0.904 |
| Class2 | 2 | 2 | 1 | 0.25 |
| Class3 | 0 | 0 | 27 | 0.965 |

TABLE 9: Number of generated rules before and after pruning.

| Method | Generated rule numbers | |
|---|---|---|
|  | Before pruning | After pruning |
| Decision tree | 198 | 121 |
| Neural networks | 210 | 140 |
| Rough sets | 140 | 62 |

classification unacceptably slow. Therefore, it is necessary to prune the rules during their generation.

## 6. COMPARATIVE ANALYSIS WITH STATISTICAL DISCRIMINANT ANALYSIS, NEURAL NETWORK, AND DECISION TREES

### Statistical discriminant analysis and empirical results

Discriminant analysis is aimed at finding weighted linear functions of the predictor variables. These discriminant linear functions are used to classify objects into distinct groups according to their observed characteristics. This is usually done by calculating the scores of the linear functions. In addition, it is of interest to determine the predictor variables that contribute significantly to the linear discriminating functions. The analysis was conducted using a stepwise selection procedure. Since we have three groups (normal, borderline, and abnormal), two discriminant functions were extracted. These two functions (shown down in Figure 2) were used to classify the diabetic children into one of the three groups.

When determining whether the two discriminant functions are significant in separating patients in the three groups
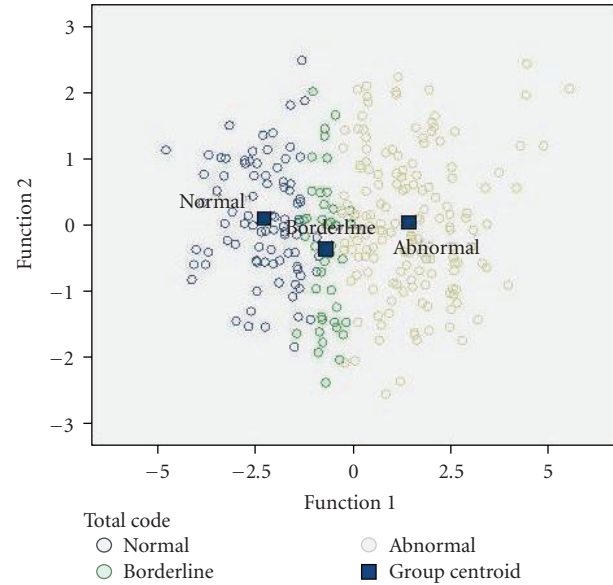


FIGURE 2: Canonical discriminant analysis.

we found that the first function explains 99.2% of the total variance and the chi-square test of Wilks' lambda is significant ($P = 0$). In contrast, the second function explains only 0.8% of the total variance and the chi-square test of Wilks' lambda is not significant ($P = .171$). To know which variables have the greater impact we examine the standardized canonical discriminant functions. Recall that the second function was not significant. For the first function, the emotional factor has the greatest impact (.644) followed by the conduct factor (.639), hyperactivity (.505), peer (.410), and child gender ($-.174$). Since the three centroids are significantly different the first function will do a good job discriminating between the three groups. This result has been illustrated in Figure 2. Table 10 shows the classification results based on the the statistical discriminant analysis and cross validation.

Figure 3 shows the overall classification accuracy of three approaches compared with the rough set approach. It shows that the rough sets approach is much better than neural networks, ID3 decision tree and statistics discriminant analysis. Moreover, for the neural networks and the decision tree classifiers, more robust features are required to improve their performance.

## 7. CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented an intelligent data analysis approach based on rough sets theory for generating classification rules from a set of observed 302 samples of diabetic Kuwaiti children patients. The main objective is to investigate the relationship of diabetes with psychological problems for Kuwaiti children aged 7–13 years old. A decomposition approach based on rough set theory to extract a complete subsets from the incomplete information system hierarchically has been used. To increase the efficiency of the classification process, rough sets with boolean reasoning (RSBR) discretization algorithm is used to discretize the data. Then, the

TABLE 10: Discriminant analysis classification results (93.3% of original grouped cases correctly classified; 91.9% of cross-validated grouped cases correctly classified).

| | | | Predicted group membership | | | |
|---|---|---|---|---|---|---|
| | | Totalcode | Normal | Borderline | Abnormal | Total |
| Original | Count | normal | 88 | 4 | 0 | 92 |
| | | borderline | 11 | 28 | 3 | 42 |
| | | abnormal | 0 | 2 | 162 | 164 |
| | % | normal | 95.7 | 4.3 | 0 | 100.0 |
| | | borderline | 26.2 | 66.7 | 7.1 | 100.0 |
| | | abnormal | 0 | 1.2 | 98.8 | 100.0 |
| Cross-validated[a] | Count | normal | 87 | 5 | 0 | 92 |
| | | borderline | 12 | 26 | 4 | 42 |
| | | abnormal | 0 | 3 | 161 | 164 |
| | % | normal | 94.6 | 5.4 | 0 | 100.0 |
| | | borderline | 28.6 | 61.9 | 9.5 | 100.0 |
| | | abnormal | 0 | 1.8 | 98.2 | 100.0 |

[a]Cross-validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
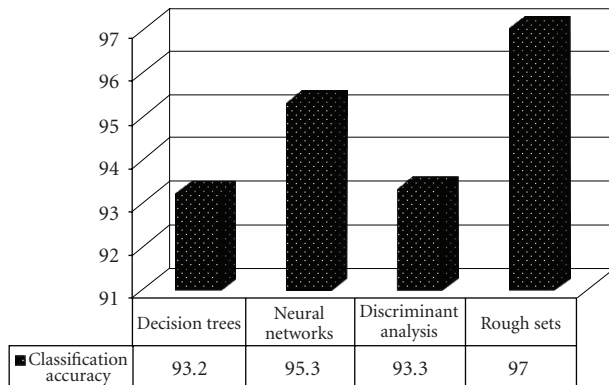


FIGURE 3: Compartive analysis in terms of the classification accuracy.

rough set reduction technique is applied to find all reducts of the data which contain the minimal subset of attributes that are associated with a class label for the classification.

The results proved that the rough set approach has higher classification accuracy with less number of generated rules compared to three different approaches, neural networks, ID3 decision tree, and statistical discriminant analysis.

In conclusion, this study shows that the theory of rough sets seems to be a useful tool for inductive learning and a valuable aid for building expert systems. Further work needs to be done to minimize the experiment duration in order to include experts in the experiments. A combination of kinds of computational intelligence techniques has become one of the most important ways of research of intelligent information processing. Neural network shows us its strong ability to solve complex problems such as the problem discussed here. From the perspective of the specific rough sets approaches

that need to be applied, an extension work of using rough sets with other intelligent systems like neural networks, genetic algorithms, fuzzy approaches, and so forth, will be considered of our future work.

## ACKNOWLEDGMENTS

## REFERENCES

[1] M. Grey, R. Whittemore, and W. Tamborlane, "Depression in type 1 diabetes in children natural history and correlates," *Journal of Psychosomatic Research*, vol. 53, no. 4, pp. 907–911, 2002.

[2] M. M. Garrison, W. J. Katon, and L. P. Richardson, "The impact of psychiatric comorbidities on readmissions for diabetes in youth," *Diabetes Care*, vol. 28, no. 9, pp. 2150–2154, 2005.

[3] N. Zhong and A. Skowron, "Rough sets based knowledge discovery process," *International Journal of Applied Mathematics and Computer Science*, vol. 11, no. 3, pp. 603–619, 2001.

[4] S. Carlin, J. Komorowski, and A. Ohrn, "Rough set analysis of medical datasets: a case of patients with suspected acute appendicitis," in *proceedings of Workshop on Intelligent Data Analysis in Medicine and Pharmacology (ECAI '98)*, pp. 18–28, Brighton, UK, 1998.

[5] K. Cios, W. Pedrycz, and R. Swiniarski, *Data Mining Methods for Knowledge Discovery*, Kluwer Academic, Boston, Mass, USA, 1998.

[6] N. Lavrajc, E. Keravnou, and B. Zupan, *Intelligent Data Analysis in Medicine and Pharmacology*, Kluwer Academic, Boston, Mass, USA, 1997.

[7] S. Tsumoto, "Discovery of rules about compilations a rough set approach in medical knowledge discovery," in *Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC '99)*, pp. 29–37, Springer, Berlin, Germany, 1999.

[8] S. Tsumoto, "Automated discovery of positive and negative knowledge in clinical databases: a rule-induction method based on rough-set models that more closely represents medical experts' reasoning," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 4, pp. 56–62, 2000.

[9] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[10] S. K. Pal, L. Polkowski, and A. Skowron, *Rough-Neuro Computing: Techniques for Computing with Words*, Springer, Berline, Germany, 2004.

[11] A. E. Hassanien, "Classification and feature selection of breast cancer data based on decision tree algorithm," *International Journal of Studies in Informatics and Control*, vol. 12, no. 1, pp. 33–39, 2003.

[12] D. Ślęzak, "Various approaches to reasoning with frequency-based decision reducts: a survey," in *Rough Sets in Soft Computing and Knowledge Discovery: New Developments*, L. Polkowski, S. Tsumoto, and T. Y. Lin, Eds., Physica Verlag, Heidelberg, Germany, 2000.

[13] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Science*, vol. 11, pp. 341–356, 1982.

[14] Z. Pawlak, *Rough Sets Theoretical Aspect of Reasoning about Data*, Kluwer Academic, Boston, Mass, USA, 1991.

[15] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko, "Rough sets," *Communications of the ACM*, vol. 38, no. 11, pp. 89–95, 1995.

[16] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information Sciences*, vol. 177, no. 1, pp. 3–27, 2007.

[17] L. Polkowski, *Rough Sets: Mathematical Foundations*, Physica Verlag, Heidelberg, Germany, 2003.

[18] G. Wang, "Rough reduction in algebra view and information view," *International Journal of Intelligent Systems*, vol. 18, no. 6, pp. 679–688, 2003.

[19] A. E. Hassanien, "Rough set approach for attribute reduction and rule generation: a case of patients with suspected breast cancer," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 11, pp. 954–962, 2004.

[20] J. A. Starzyk, N. Dale, and K. Sturtz, "A mathematical foundation for improved reduct generation in information systems," *Knowledge and Information Systems Journal*, pp. 131–147, 2000.

[21] R. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognition Letters*, vol. 24, no. 6, pp. 833–849, 2003.

[22] W. Ziarko, "Variable precision rough sets model," *Journal of Computer and Systems. Sciences*, vol. 46, no. 1, p. 3959, 1993.

[23] Q. Zhang, "An approach to rough set decomposition of incomplete information systems," in *Proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA '07)*, pp. 2455–2460, Harbin, China, May 2007.

[24] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretization of continuous features," in *Proceedings of the 12th International Conference on Machine Learning*, pp. 294–301, Tahoe city, Calif, USA, 1995.

[25] J. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak, and J. Wróblewski, "Rough set algorithms in classification problem," in *Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems*, L. Polkowski, S. Tsumoto, and T. Y. Lin, Eds., pp. 49–88, Physica Verlag, Heidelberg, Germany, 2000.

[26] G. Gediga and I. Duentsch, "Statistical techniques for rough set data analysis," in *Rough Sets: New Developments*, pp. 545–565, Physica Verlag, Heidelberg, Germany, 2000.

[27] S. Ning, W. Ziarko, J. Hamilton, and N. Cercone, "Using rough sets as tools for knowledge discovery," in *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining (KDD '95)*, U. M. Fayyad and R. Uthurusamy, Eds., pp. 263–268, AAAI, Montreal, Que, Canada, July 1995.

[28] J. Li and N. Cercone, "A rough set based model to rank the importance of association rules," in *Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC '99)*, Lecture Notes in Computer Science, pp. 109–118, Springer, Berlin, Germany, 2005.

[29] R. Goodman, "The strengths and difficulties questionnaire: a research note," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 38, no. 5, pp. 581–586, 1997.

[30] R. Goodman, "Psychometric properties of the strengths and difficulties questionnaire," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 40, pp. 1337–1345, 2001.

[31] R. Goodman, T. Ford, H. Simmons, R. Gatward, and H. Meltzer, "Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample," *British Journal of Psychiatry*, vol. 177, pp. 534–539, 2000.

[32] D. J. Cornforth, H. F. Jelinek, M. C. Teich, and S. B. Lowen, "Wrapper subset evaluation facilitates the automated detection of diabetes from heart rate variability measures," in *Proceedings of the International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA '04)*, M. Mohammadian, Ed., pp. 446–455, University of Canberra, Goldcoast, Australia, July 2004.

[33] D. J. Hawes and M. R. Dadds, "Australian data and psychometric properties of the strengths and difficulties questionnaire," *Australian and New Zealand Journal of Psychiatry*, vol. 38, no. 8, pp. 644–651, 2004.

[34] H. Klasen, W. Woerner, D. Wolke, et al., "Comparing the German versions of the strengths and difficulties questionnaire (SDQ-Deu) and the child behavior checklist," *European Child & Adolescent Psychiatry*, vol. 9, no. 4, pp. 271–276, 2000.

[35] R. Goodman, D. Renfrew, and M. Mullick, "Predicting type of psychiatric disorder from strengths and difficulties questionnaire (SDQ) scores in child mental health clinics in London and Dhaka," *European Child and Adolescent Psychiatry*, vol. 9, no. 2, pp. 129–134, 2000.

[36] H. Smedje, J.-E. Broman, J. Hetta, and A.-L. Von Knorring, "Psychometric properties of a Swedish version of the strengths and difficulties questionnaire," *European Child & Adolescent Psychiatry*, vol. 8, no. 2, pp. 63–70, 1999.

[37] W. Woerner, B. Fleitlich-Bilyk, R. Martinussen, et al., "The strengths and difficulties questionnaire overseas: evaluations and applications of the SDQ beyond Europe," *European Child and Adolescent Psychiatry*, vol. 13, supplement 2, pp. 47–54, 2004.

[38] L. J. Cronbach, "Coeeifcent alpha and the internal structure of tests," *Psychometrica*, vol. 16, pp. 297–334, 1951.

[39] P. Muris, C. Meesters, and F. Van den Berg, "The strengths and difficulties questionnaire (SDQ) further evidence for its

reliability and validity in a community sample of dutch children and adolescents," *European Child & Adolescent Psychiatry*, vol. 12, no. 1, pp. 1–8, 2003.

[40] M. S. I. Mullick and R. Goodman, "Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study," *Social Psychiatry and Psychiatric Epidemiology*, vol. 36, no. 2, pp. 94–99, 2001.

[41] M. Koskelainen, A. Sourander, and A. Kaljonen, "The strengths and difficulties questionnaire among Finnish school-aged children and adolescents," *European Child & Adolescent Psychiatry*, vol. 9, no. 4, pp. 277–284, 2000.