Intelligent Automation And Soft Computing, 2020
Vol. 26, no. 3, 465–478
DOI: 10.32604/iasc.2020.013923

Rough Set Based Rule Approximation and Application on Uncertain Datasets

L. Ezhilarasi[1,*], A.P. Shanthi[2] and V. Uma Maheswari[1]

[1] Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai, Tamil Nadu,600 025, India
[2] Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai, Tamil Nadu, 600 025, India.

ABSTRACT
Development of new Artificial Intelligence related data analysis methodologies with revolutionary information technology has made a radical change in prediction, forecasting, and decision making for real-world data. The challenge arises when the real world dataset consisting of voluminous data is uncertain. The rough set is a mathematical formalism that has emerged significantly for uncertain datasets. It represents the knowledge of the datasets as decision rules. It does not need any metadata. The rules are used to predict or classify unseen examples. The objective of this research is to develop a rough set based classification system that predicts and classifies unseen examples by learning from the minimal subset of decision rules extracted from uncertain datasets using rule approximation. This paper proposes a novel rule approximation classifier, Weighted-Attribute Significance Rule Approximation (WASRA) that uses a subset of the decision rules generated by any rule induction algorithm, to compute the concept weights of the condition attributes. The concept weights and the significance of condition attributes are used to design a novel classifier. This classifier is implemented and initially tested on a few benchmarked datasets of the UCI repository. The classifier is subsequently tested on a real-time dataset and compared to other standard classifiers. The experimental results illustrate that the proposed WASRA performs well and shows an improvement in the prediction accuracy compared to other classifiers. This classifier can be applied to any dataset which has uncertainty.

KEYWORDS: Rough set theory, concept weight, rule approximation, attribute significance

1 INTRODUCTION

IN many real-life situations, data stored in the form of datasets plays a pivotal role in analysis, decision making, prediction, and forecasting. There are many traditional algorithms and statistical methods in the literature for the reasoning of precise data (Berzuini 1988; Michalski and Chilausky, 1980). The precise data analysis uses techniques that are based on either strong assumptions of probability distributions or knowledge about dependencies. The representation of the right information at the right time provides better knowledge. When the data stored is incomplete or imprecise, it results in uncertain datasets. Data analysis and prediction of uncertain datasets have attracted many researchers, practitioners, and philosophers.

Artificial Intelligence provides methods, algorithms, and techniques to analyze and extract useful information from imprecise data for decision making or prediction. This is handled by two prominent research areas of Artificial Intelligence, Machine Learning, and Data Mining. Machine learning introduces new algorithms to learn from the set of examples or from previous experiences. Data mining is used to extract valuable information and hidden patterns for analysis, classification, and prediction. Many theories, algorithms, and techniques have been developed to deal with imprecise data. Among them are fuzzy set, Dempster-Shafer theory of evidence and rough set theory.

Rough set theory is a mathematical framework proposed by Pawlak in 1982 to deal with uncertain, imprecise, inconsistently erratic and vague descriptions of examples (Pawlak, 1981; Pawlak, Grzymala-Busse, Slowinski, Ziarko, et al. 1995). In rough set theory, patterns in data can be characterized by means of approximations or equivalently by decision rules induced by the data. The decision rules express the hidden patterns about the learning examples. There are many algorithms found in the

CONTACT L.Ezhilarasi ezhilvijayanand@gmail.com

literature for inducing rules from examples. Some of them are ID3, PRISM, LEM2, MLEM2 and MODLEM (Quinlan, 1986; Cendrowska, 1987; Grzymala-Busse, 1997; Grzymala-Busse, 2002; Stefanowski, 1998). A comparison of ID3, PRISM, and LEM2 is presented in (Chan and Grzymala-Busse, 1991). PRISM is a rule-based classifier which uses local covering for rule generation. ID3 is a decision tree based rule learner algorithm. LERS (Learning from Examples using Rough Set) is a data mining system that computes rules using rough set concepts from uncertain datasets (Grzymala-Busse, 1992). Rule sets generated can be used for classification of new examples or in the interpretation of knowledge.

An important application of rough sets that has become a significant research trend is in the area of classification (Pawlak, 1992). This paper presents work that leads to the development of a rough set-based classification system, Weighted- Attribute Significance Rule Approximation, (WASRA) using rule approximation. The novel rule approximation approach uses only a subset of rules generated by any rule induction algorithm. This work defines and associates weights with condition attributes. These weights are useful in the approximation of the antecedent of a rule when it is not matched exactly by an example. Another concept proposed in this work is the determination of the class label based on the attribute-value pair. This is also used for the approximation and assignment of decision rules.

This paper is organized as follows. Section 2 presents the current work related to rule generation frameworks using rough set theory. The basic concepts of rough set theory and rule induction algorithms are given in section 3. The architecture of the proposed classifier is explained in section 4. Section 5 gives the dataset details, tabulated results of the experiments, and comparison with other classifiers. Section 6 presents the conclusion of the work done.

2 RELATED WORK

THIS section briefly discusses the existing works related to rule generation frameworks based on rough set theory on attribute reduction and hybridization.

2.1 Rough set frameworks based on attribute reduction

Sumalatha et al have proposed a rule generation framework based on rough set theory to find the behavioral pattern of customers (Sumalatha, Uma Sankar, Sujatha, et al., 2016). The reducts are found using the discernibility matrix and decision rules are generated for the prediction of the decision class. The Experiments have been conducted on a Portuguese banking institution that predicts the deposit nature of customers.

Nandhini et al (Nandhini and Sivanandam, 2015) have proposed a predictive associative rule-based classifier, Classification based on Predictive Association Rule (CPAR) algorithm, using gain ratio in health care data diagnosis. Reducts by rough set and T-test techniques are used for dimensionality reduction. Laplace accuracy is used in k- best rule selection method to choose the best rule. This classifier considers only binary class datasets.

Hassanien et al (Hassanien and Ali, 2004) have designed a simplification algorithm based on rough set methodology by generating classification rules for breast cancer data. The reduct set is computed and the simplification algorithm is applied on the reduced system to generate simplified rules. The significance of the rule is evaluated by applying the statistical significance test of attributes and the rules are pruned. The pruned rules are used for the classification of breast cancer dataset.

2.2 Rough set frameworks based on hybridization

Kusiak (Kusiak, 2001) has proposed a rule structuring algorithm based on the concept of data mining and rough set theory for semiconductor applications. Evolutionary computation approach is expanded to rule structuring and data engineering for the longevity of decision rules. Rough set produces a minimal set of rules from the available data. The rough set rules generated were suitable to form meta-structures that led to decision making with transparent knowledge analysis in semiconductor application.

Bazan et al (Bazan, Peters, Skowron, Son, Szczuka, et al., 2003) have proposed a rough set approach for incremental concept approximation. The relevant pattern is searched by defining conflict resolution strategies. Tuning of the parameter in a conflict resolution strategy extracts patterns for concept approximation.

Khoo et al (Khoo, Tor, Zhai, et al., 1999) have built a prototype system, Rough set based classification system (RClass) that integrates rough set theory and a statistical-based inductive learning algorithm. Statistical based inductive learning algorithm uses an entropy-based inductive approach for decision making and choosing of attributes. Conflict in training data is identified and resolved using rough set. A reliability index is calculated for every possible rule. The classification rules induced by RClass system are simple and logical.

Sharma et al (Sharma, Kumari, Kar, et al., 2019) have proposed a forecasting model based on hybridization of rough set and double exponential smoothing for air passengers time series dataset.

Ziarko et al (Ziarko and Shan, 1996) have presented a rough set methodology based on decision matrix and Boolean decision function. This method finds all minimal or maximally simplified rules for target classification. The simplification of Boolean decision function is used to reduce the problem of the finding of all deterministic rules. The simplification problem can be easily split into disjoint sub-problems and this facilitates its implementation in the multiprocessor system. Identification of minimal rule length for a decision class enables to choose between many possible solutions for target classification.

By using rough sets and belief function theory, Trabelsi et al (Trabelsi and Elouedi, 2008) have developed a learning approach to derive decision rules from uncertain data. The rules generated by rough sets are used in the Transferable Belief Model (TBM). TBM indicates that uncertainty exists in decision attribute values. This model consists of credal and pignistic levels. In the credal level, belief is represented by a belief function. In the pignistic level, beliefs are used to make a decision by attaching a pignistic probability function. The drawback of this method is that the belief rules generated are not optimal.

As observed in the literature, it can be seen that researchers have generally used rough sets for hybridization and attribute reduction. The proposed system provides a novel approach for rough set-based rule approximation and application on uncertain datasets. The proposed system is different in that, the system uses a minimal subset of decision rules generated by a rule induction algorithm and for rule approximation, the concept weight and significance of the condition attributes are used.

3 PRELIMINARIES

THIS section briefly discusses the basic notion of rough set theory and rule induction algorithms.

3.1 Rough set theory

In rough set philosophy (Pawlak, 1984; Pawlak, 1996), every example of the universe is associated with some amount of information and is expressed as attributes used for example descriptions. The main advantage of rough set theory is that it does not need any preliminary information about data (Pawlak, Grzymala-Busse, Slowinski, Ziarko, et al., 1995). Rough set theory can effectively resolve the problems related to the dependency between attributes, reducing all redundant attributes, evaluating the significance of attributes, identifying the hidden pattern and inducing decision rules for the given dataset. The rough set theory has been applied successfully in many real-life problems in engineering, medicine, banking, financial and other areas.

3.1.1 Information system and Decision system

Let U be a universe of examples. Each example $x \in U$ is described by a set of characteristic attributes, A. Let the set of attributes be divided into two subsets C and D called condition attributes and decision attributes, respectively, such that $A = C \cup D$ and $C \cap D \neq \emptyset$. Let $C = \{a_1, a_2, .., a_n\}$ and $D = \{d_1, d_2, .., d_m\}$, where n, m are the number of condition attributes and where m is the number of decision attributes, respectively. The set of values that can be taken by condition attribute a_i and decision attribute, d_j are V_{a_i} and V_{d_j}.

An information system is a set of examples that are described only with condition attributes. When a set of examples are described with both condition and decision attributes, they form a decision system. Each decision attribute value called class label defines a concept that contains all the examples of U having the same class label. A decision system is said to be inconsistent (uncertain) if any two examples have the same values for all condition attributes but different values for decision attributes.

3.1.2 Rough set

The rough concept is a set of examples that cannot be defined precisely by the given set of attributes. Rough set theory approximates the rough concept using the two definable sets, the greatest definable set completely contained in the concept (lower approximation), and the least definable set containing the concept (upper approximation) (Pawlak, 1992). In rough set approach, the decision rules are induced from a decision system and a minimal set of rules are generated. The rules induced by lower and upper approximation are called certain and possible rules, respectively.

3.1.3 Dependency of attributes

Finding out of the dependencies between attributes of a decision system is one of the major concerns in data analysis. The positive region is the union of the lower approximations of all the concepts. This is indicated by $POS_C(D)$. It includes only those examples which belong to the corresponding concepts without any ambiguity.

A set of attributes D depend on a set of attributes C if all values of attributes in D are uniquely determined by the values of attributes in C.Otherwise, there is a partial dependency between C and D (Pawlak, 1981; Ziarko, 1999). The degree of dependency between C and D or the consistency measure of the universe U is represented by $\gamma(C, D)$ or γ and is given by

$$\gamma(C, D) = \frac{|POS_C(D)|}{|U|} \quad (1)$$

A state of total dependency of Don C is represented by $\gamma = 1$. This specifies that its positive region is the entire universe and U is consistent.

3.1.4 Significance of attributes

The significance of an attribute reflects the degree of decrease of the positive region as removing an attribute 'a' from C. It is in the real value, in the range of closed interval [0, 1] (Pawlak, 1982). The consistency measure as given in Eq. (1) is used to calculate the significance of an attribute. The significance of an attribute 'a' is defined in Eq. (2).

$$\sigma(a) = 1 - \frac{\gamma(C-\{a\},D)}{\gamma(C,D)} \quad (2)$$

3.1.5 Decision rules

The decision rules are one way of representing knowledge in a data set (Ziarko, 2002). A decision rule has an antecedent part and a consequent part. The consequent part gives the concept predicted and the antecedent part specifies the necessary conditions, to have this concept predicted. In rule induction and classification process a measurement is required to determine the rule quality (Yao and Zhong, 1999; Lavrac, Flach, Zupan, et al., 1999). The rule quality calculated is a function that determines the strength of each rule induced. The measures of rule quality that are considered in this work are Laplace accuracy and support.

Laplace accuracy determines the quality of the concept rules by considering its rule coverage (Niblett, 1987). The formula for Laplace accuracy is defined below in Eq. (3).

$$Laplace\ accuracy = \frac{n_m(R)+1}{n(R)+m} \quad (3)$$

'm' is the number of concepts in the dataset, '$n_m(R)$' is the number of examples in the predicted concept m covered by the rule R and '$n(R)$' is the total number of examples covered by the rule R.

Support of a rule x is the number of examples matching the rule conditions and decision attribute (Pawlak, 2004). In other words, the set of attribute-value pairs occurring on the antecedent part of rule x is denoted as rule condition part and symbolized as $C(x)$. The consequent of a rule is the decision part and is denoted by $D(x)$. The support of a rule x is given below in Eq. (4).

$$Supp_x(C,D) = card\big(C(x) \cap D(x)\big) \quad (4)$$

3.2 Rule Induction algorithms

Learning from Examples using Rough Sets (LERS) is a rule induction system introduced by Jerzy Grzymala Busse and developed at the University of Kansas. LERS has developed four different options of rule induction methods and the most popular of them is LEM2 algorithm.

LEM2 algorithm [Grzymala-Busse (2015)] computes a single local covering for each concept from the decision system which uses only categorical attributes. It generates a minimal set of decision rules. The strength of the rule is determined with the measures of support, consistency, and coverage associated with each rule.

4 PROPOSED WASRA CLASSIFICATION SYSTEM

IN this section, a detailed description of the proposed architecture, Weighted Attribute Significance Rule Approximation (WASRA) is presented. WASRA is a rule approximation algorithm which takes as input a rule subset and computes the concept weights of attributes which are then used for rule approximation. The approximated rules are used for classification of datasets. The overall architecture of the WASRA classification system is shown in Figure.1.

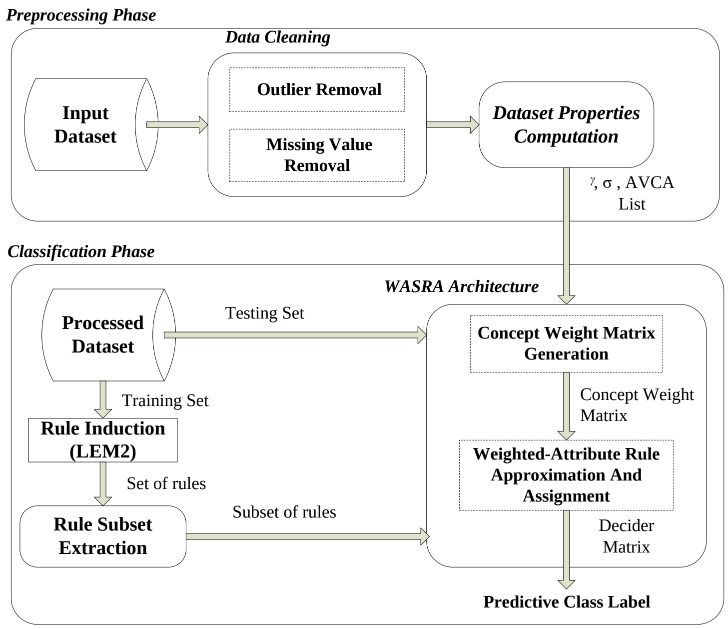

Figure 1. WASRA Classification System

The system involves two phases, namely, preprocessing phase and classification phase. The preprocessing phase includes cleaning of data and computation of dataset properties.

4.1 Pre-Processing Phase

In the preprocessing phase, a dataset is given as input to the data cleaning module that performs two tasks: outlier removal and missing value removal.

In any multi-class dataset, the number of examples belonging to each class may vary significantly. Those examples of a class label which are very minimally present compared to those of other class labels are not represented adequately. Hence, these are removed from the dataset and not considered for classification. This task is referred to as outlier removal. The examples that do not have values for all the attributes are also not considered for classification. The removal of these examples is performed by the missing value removal task.

The preprocessed dataset is given as input to the Dataset Properties Computation module. This module computes the consistency measure, the significance of attributes and builds the Attribute-Value Class Affinity list (AVCA) as given in Figure 2.

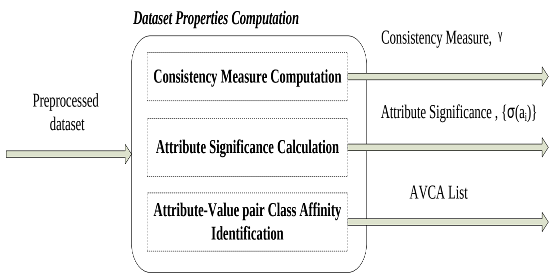

Figure 2. **Dataset properties computation module**

The consistency measure γ of the dataset is defined in Eq. (1) that is given in section 3.1.3. The significance of each attribute a_i , $\sigma(a_i)$ is computed using Eq. (2). The Attribute-Value pair Class Affinity (AVCA) is a list whose cardinality equal the number of condition attributes in the decision system. Each element of the AVCA list points to another list with the size equal to the number of values that the corresponding attribute can take. Each entry of this list has a value of the decision attribute.

The label d_j of d that has the maximum number of examples with value v_{a_i} for a_i among all the labels of d is taken as the class affinity of the pair (a_i, v_{a_i}). The representation of the AVCA list is given in Figure3.

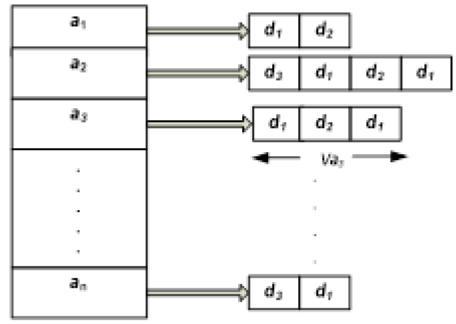

Figure 3. **AVCA List**

4.2 Classification Phase

The input to the classification phase is the preprocessed dataset, which is split into training and test dataset based on stratification selection. This is done to retain the same class distribution in each fold as in the whole dataset while performing cross-validation. The rule induction module receives the training dataset and the rules are generated. The rule set R generated is given as input for rule subset extraction. Each rule 'r' in the rule set R has two parameters, namely, Laplace accuracy 'l_r' and support 's_r'. The rules in the rule set R are arranged in decreasing order of l_r values. The rules with support threshold > 1 are extracted and the new rule set R_p is obtained.

4.2.1 Concept weight matrix generation module

This module receives a new rule set R_p as input and generates the concept weight matrix CWM of dimension $n \times m$ is given in Figure 4.

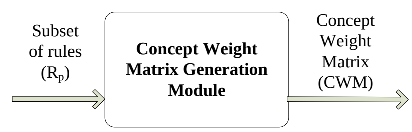

Figure 4. **Concept Weight Matrix Generation Module**

Here, 'p' is the number of rules in the rule set R_p, 'n' denotes the number of condition attributes and 'm' denotes the number of concepts or class labels. The rule set R_p is used to generate the concept weight $con\_wt_i^j$ of each attribute a_i corresponding to the j^{th} value of the decision attribute d. Let $|R_i^j|$ be the number of rules for the j^{th} value of the decision attribute, d in which the attribute, a_i is present. Let $|R^j|$ be the total number of rules for the decision value d_j. The concept weight $con\_wt_i^j$ is computed and is given in Eq. (5) as follows.

$$con\_wt_i^j = \frac{|R_i^j|}{|R^j|} \qquad (5)$$

The concept weight matrix generation $\mathrm{CWM}(i,j)$ is one of the inputs to the weighted attribute rule approximation and assignment module and is given in Figure 5.

Concept Weight Matrix Generation $CWM(n \times m)$

For each attribute $a_i \in A$
Begin
    For each value $d_j \in v_d$ where $d \in D$
    begin
        Let R^j denote the rules of R_p with decision value d_j for d
        Let $|R^j|$ denote the cardinality of R^j
        For each rule in R^j
        Begin
            If a_i is present in the rule then
                add the rule to R_i^j
        end for;
        $con_{wt}{}_i^j = \frac{|R_i^j|}{|R^j|}$
        $CWM(i,j) = con\_wt_i^j$
    end for;
end for;

Figure 5. Concept weight matrix generation

4.2.2 Weighted Attribute Rule Approximation and Assignment module

The rule subset, concept weight matrix, and dataset properties are the inputs to this module, which outputs the decider matrix as shown in Figure 6. This module performs rule approximation and finds the decision class label of each example with respect to all rules in the rule subset. The class label decided by each rule for all the examples is stored in the decider matrix.

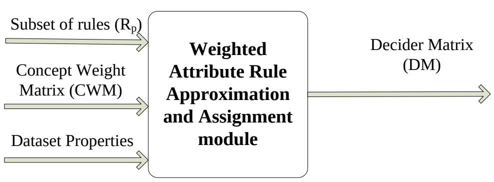

Figure 6. Weighted Attribute Rule Approximation and Assignment module

Let E be the set of examples in the test dataset. The decider matrix of dimension $t \times p$ is denoted as $DM(t,p)$ where 't' denotes the number of examples in E and 'p' denotes the number of rules in R_p. Each entry in the decider matrix is a value (class label) of the decision attribute d or NA (indicates that rule cannot be used). For example $e_t \in E$, this module finds the concept, to which the example belongs based on each rule $r \in R_p$. Let d_j denote the decision label of rule r. When rule r is applied to example e_t, only the weights of attributes corresponding to the concept d_j are considered.

The attributes of the example whose values match with those of condition attributes present in the antecedent of the rule are found. Let '$\mathrm{W_{match}}$' be the sum of the concept weights for the matched attributes. The sum of the weights '$\mathrm{W_{mismatch}}$' of the attributes of the example, whose values differ with the values of the condition attributes of the rule is also computed.

$\mathrm{W_{match}}$ and $\mathrm{W_{mismatch}}$ are used for assigning a class label to e_t using r.

1. If $\mathrm{W_{match}} = 0$ *or* $\mathrm{W_{match}} < \mathrm{W_{mismatch}}$ then r cannot be used and the entry corresponding to row e_t and column r is 'NA'. [Not available]
2. If $\mathrm{W_{match}} > \mathrm{W_{mismatch}}$ then entry corresponding to e_t and r is d_j.[Rule triggered]
3. If $\mathrm{W_{match}} = \mathrm{W_{mismatch}}$ then the following method is adopted. An attribute a in the example and not present in the antecedent is considered. The significance of the attribute is used when more than one such attribute is found in the example.

Let v_a be the value of a in e_t. Use $(a,\ v_a)$ as the indices to AVCA list to obtain the class affinity of $(a,\ v_a)$. This is used as an entry for e_t and r in the decider matrix $DM(t,p)$. These steps are performed for all rules of R_p and for each example e_t. For each example, the class label c_{pr} recommended by the majority of the rules is assigned. The average prediction accuracy is calculated. The proposed rule approximation and assignment procedure is given in Figure 7.

The above proposed rule approximation algorithm is tested and implemented on some benchmark datasets of UCI repository (Dua and Graff, 2019) and also on a real-time soil dataset. The following section provides the experimental illustration of the proposed rule approximation system.

5 EXPERIMENTAL ILLUSTRATION

5.1 *Environment specification*

ALL experiments are carried out on a computer with processor Intel (R) Core (TM) i7-2600 CPU@ 3.40 GHz, 4.0 GB RAM and 250 GB HDD. The program was developed using RoughSets Package in R programming language version 3.4.1 in Windows 7 Professional environment.

5.2 *Dataset specification*

The experiments are performed on four nominal datasets of medical domains obtained from the UCI Machine Learning Repository and on a real-time soil dataset. The UCI datasets are postoperative dataset, mammography dataset, SPECT Heart dataset and contraceptive dataset The dataset details, such as number of examples, number of condition attributes, number of examples after preprocessing, concepts, class distribution ratio and consistency measure are given in Table 1, followed by the dataset description.

Weighted Attribute Rule Approximation and Assignment

For each example $e_t \in E$
  begin
    $w_{match} = 0,\ w_{mismatch} = 0;$
    For each rule $r \in R_p$
      begin
        Consider class label d_j of this rule r
        For each a_i in r
          if the value of a_i in r equals value of a_i in e_t
            $w_{match} = w_{match} + con\_wt_i^j;$
          else
            $w_{mismatch} = w_{mismatch} + con\_wt_i^j;$
          end if;
        end for;
    // Decider matrix construction
        if $((w_{match} = 0)\ OR\ (w_{match} < w_{mismatch}))$
          $DM\ (t,r) = NA$
        else if $\left(w_{match} > w_{mismatch}\right)$
          $DM\ (t,r) = d_j$
        else
        begin
          Among all the attribute a_i not present in r.
          Let a denote the attribute with the maximum value of $\sigma(a_i)$ for all i.
          Let v_a denote the value of a in e_t.
          Use $(a,\ v_a)$ and obtain the class label d_j from the AVCA list.
        end else;
      end for;
    $c_{pr} = d_j$ which appears the maximum number of times in row t.
    In case of ties, the label suggested by the first column is used.
end for;
return c_{pr};

Figure 7. Weighted attribute rule approximation and assignment procedure

5.2.1 SPECT heart dataset

The dataset describes diagnosing of cardiac Single Proton Emission Computed Tomography (SPECT) images. It uses radioactive tracers that are injected into the blood to produce pictures of the heart. Doctors use SPECT to diagnose coronary artery disease and find out if a heart attack has occurred. For feature patterns extraction of cardiac patients, SPECT images are converted into 22 binary attributes and each of the patients is classified with two decision class labels: 0-normal and 1-abnormal. The varieties of heart diseases are grouped in a single class label abnormal. The dataset consist of 22 condition attributes $F_1, F_2, \cdots, F_{22}$ that have binary values as yes or no. The overall diagnosis of the decision attribute has binary value 0- normal and 1- abnormal.

5.2.2 Postoperative dataset

The postoperative dataset gives the characteristic symptoms of a patient after surgery. The dataset has 8 condition attributes and one decision attribute. The set of values the decision attribute can take is $V_{Adm-Decs} = \{A, I, S\}$. The attributes are explained in Table 2. In the preprocessing of the dataset, three examples that had missing values and the examples with decision values 'I' which has a very minimal (1) out of 87examples are removed. Hence, the number of class labels gets reduced to $\{A, S\}$.

Table 1. Dataset details

S. No	Dataset	Number of examples	Number of examples after preprocessing	Condition attributes	Concepts	Class Distribution ratio	Consistency measure
1.	Postoperative	90	86	8	2	7:3	0.84
2.	Mammography	961	830	5	2	5:5	0.53
3.	SPECT Heart	267	267	22	2	2:8	0.82
4.	Contraceptive	1473	1473	9	3	3:1:6	0.74
5.	Soil	350	350	11	3	3:3:4	0.64

Table 2. Description of postoperative dataset

S. No	Attribute name	Attribute description	Domain values
1.	L-Core (Lc)	patient's internal temperature in ^{0}C	high (> 37), mid (>= 36 and <= 37), low (< 36)
2.	L-Surf (Ls)	patient's surface temperature in ^{0}C	high (> 36.5), mid (>= 36.5 and <= 35), low (< 35)
3.	L-O2 (Lo)	oxygen saturation in %	excellent (>= 98), good (>= 90 and < 98), fair (>= 80 and < 90), poor (< 80)
4.	L-Bp (Lb)	last measurement of blood pressure	high (> 130/90), mid (<= 130/90 and >= 90/70), low (< 90/70)
5.	Surf-Stbl (Ss)	stability of patient's surface temperature	stable, mod-stable, unstable
6.	Core-Stbl (Cs)	stability of patient's core temperature	stable, mod-stable, unstable
7.	Bp-Stbl (Bs)	stability of patient's blood pressure	stable, mod-stable, unstable
8.	Comfort (Cf)	patient's perceived comfort at discharge	measured as an integer between 0 and 20
9.	Adm-Decs (D)	discharge decision	I (patient sent to Intensive Care Unit), S (patient prepared to go home), A (patient sent to general hospital floor)

5.2.3 Mammography dataset

This dataset is used to predict the severity of mammographic mass lesion from BI-RADS attributes and the patient's age. This system helps physicians in decision making of whether to perform a biopsy on suspicious lesion seen in a mammogram or perform a short-term follow-up examination. The data was collected at the Institute of Radiology of the University Erlangen-Nuremberg between 2003 and 2006. The dataset has five condition attributes and one decision attribute. The set of values the decision attribute can take is $V_{Severity} = \{1, 2\}$. The dataset description is given in Table 3.

5.2.4 Contraceptive method choice dataset

This dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey, which consists of 1473 examples of married women. The problem is to predict the current contraceptive method choice of a woman based on her socio-economic and demographic characteristics. The dataset consists of nine condition attributes and one decision attribute. The set of values the decision attribute, Contraceptive method used can take is $V_{\text{contraceptive method used}} = \{1, 2, 3\}$. Table 4 gives the dataset description of the contraceptive method choice.

5.2.5 Real-time soil dataset

The soil dataset has 350 soil samples taken from villages in Thanjavur district of Tamil Nadu. The dataset correlates the pH value with the other nutrients to assess the fertility of the soil. The eleven condition attributes are discretized with the range of the nutrient index specified by the domain expert as given in Table5.

Table 3. Description of Mammography dataset

S. No	Attribute name	Attribute description	Domain values
1.	BI-RADS (BR)	BI-RADS assessment	1 to 5
2.	Age (A)	patient's age in years	Integer values
3.	Shape (Sh)	Shape of mass	round=1, oval=2, lobular=3, irregular=4
4.	Margin (Mr)	mass margin	circumscribed=1, microlobulated=2, obscured=3, ill-defined=4, spiculated=5
5.	Density (De)	Density of mass	high=1, iso=2 low=3, fat-containing=4
6.	Severity (D)	Level of severity	benign=1 or malignant=2

The range of fertility rating of pH specified by domain expert is given in Table 6. The set of values the decision attribute pH can take is $V_{pH} = \{\text{acidic}, \text{neutral}, \text{alkaline}\}$ that is labeled as 1, 2 and 3.The dataset consists of 119 samples for acidic, 98 samples for neutral and 133 samples for alkaline.

Table 4. Contraceptive method of choice dataset description

S. No	Attribute name	Attribute description	Domain values
1.	Wife's age (Wa)	Wife age in years	Integer values
2.	Wife's education (We)	Wife education	1=low, 2, 3, 4=high
3.	Husband's education (He)	Husband education	1=low, 2, 3, 4=high
4.	Number of children ever born (Nc)	Number of children born	Integer values
5.	Wife's religion (Wr)	Religion of wife	0=Non-Islam, 1=Islam
6.	Wife's working (Ww)	Working status of wife	0=Yes, 1=No
7.	Husband's occupation (Ho)	Occupation of husband	1, 2, 3, 4
8.	Standard-of-living index (St)	Life style	1=low, 2, 3, 4=high
9.	Media exposure (Me)	Exposure to media	0=Good, 1=Not good
10.	Contraceptive method used (D)	Contraceptive method used	1=No-use, 2=Longterm, 3=Short-term

Table 5. Nutrient Index Value of soil dataset by experts

Nutrients	Low (mg/kg)	Medium (mg/kg)	High (mg/kg)
Nitrogen as Nitrate (N)	< 14.0	14.0 – 20.0	>20.0
Phosphorus (Alkaline soil) (P1)	< 15.0	15.0 – 22.0	>22.0
Phosphorus (Acidic soil) (P2)	< 100	100	100+
Potassium (K)	< 150	150	>150
Calcium (Ca)	< 2000	2000	>2000
Magnesium (Mg)	< 500	500	>500
Sulphur (S)	< 20.0	20.0 – 30.0	> 30.0
Zinc (Zn)	< 2.5	2.5 – 3.5	> 3.5
Manganese (Mn)	< 10.0	10.0 – 20.0	>20.0
Copper (Cu)	< 2.5	2.5 – 3.5	> 3.5
Iron (Fe)	< 9.0	9.0 – 20.0	> 20.0

Table 6. Fertility rating of pH value for soil dataset

Parameter	Acidic	Neutral	Alkaline
pH	< 6.5	6.5-7.5	> 7.5

5.3 Experimental Results

This section illustrates the performance of the proposed classification system with the postoperative dataset. The computation of the dataset properties is carried out, which is followed by the classification and comparison with other standard classifiers.

5.3.1 Computation of dataset properties

The computation of dataset properties includes consistency measure computation, the calculation of significance of attributes and identification of the attribute-value pair class affinity of the dataset. The consistency measure for postoperative dataset calculated is 0.84. The attribute significance value $\sigma(a_i)$ for each attribute a_i is computed. The AVCA list for the postoperative dataset is identified. The AVCA list specifies the class affinity identified by each attribute-value pair of the dataset and is shown in Figure 8.

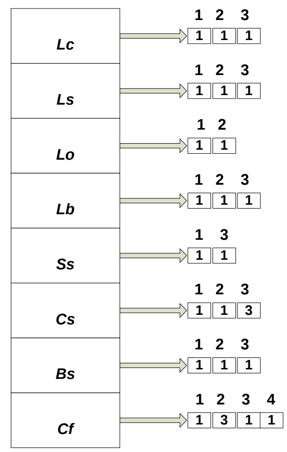

Figure 8. AVCA List for the postoperative dataset

The attributes are arranged in the order of most significant to least significant and tabulated in Table 7.

Table 7. Significance of attributes values in postoperative dataset

a_i	Ls	Bs	Cf	Ss	Lb	Lo, Cs	Lc
$\sigma(a_i)$	0.25	0.21	0.18	0.17	0.10	0.07	0.04

These values of attribute significance of condition attributes are utilized in the WASRA algorithm. The number of rules generated by LEM2 algorithm and the subset of rules used by WASRA algorithm for each run of the postoperative dataset is given in Table 8.

Table 8. Number of rules generated by LEM2 and subset rules chosen in WASRA for the postoperative dataset

Algorithms	Run 1	Run 2	Run 3	Run 4	Run 5	Average (rounded)
LEM2	25	23	24	21	23	23
WASRA	19	19	18	15	16	17

73.91% of the LEM2 rules are used by the WASRA algorithm for classification. Further, the implementation of the classification system is performed.

5.3.2 Implementation of the WASRA classification system

In this section, the implementation of the proposed classification system is presented. Initially, this is carried only on the examples misclassified by LEM2. Subsequently, it is implemented on the entire test dataset of all the datasets. The number of rules generated by LEM2 algorithm used by all the datasets and subset of rules extracted for the implementation is given in Table 9.

Table 9. **Number of rules generated by LEM2 and subset of rules extracted for WASRA system for all datasets**

Runs	Postoperative		Mammography		SPECT Heart		Contraceptive		Soil	
	LEM2	WASRA	LEM2	WASRA	LEM2	WASRA	LEM2	WASRA	LEM2	WASRA
1.	25	19	102	71	32	31	483	267	60	41
2.	23	19	97	67	30	27	469	258	71	46
3.	24	18	102	68	32	27	481	264	64	44
4.	21	15	102	69	32	25	474	262	66	47
5.	23	16	101	70	31	25	485	273	56	36
Average (rounded)	23	17	101	69	31	27	478	265	63	43

The percentage of the number of rules used by WASRA is 73.91%, 68.31%, 87.10%, 55.43%, and 68.25%, respectively for each dataset. The comparison of the number of rules considered in LEM2 and WASRA algorithms for all datasets is shown in Figure 9.

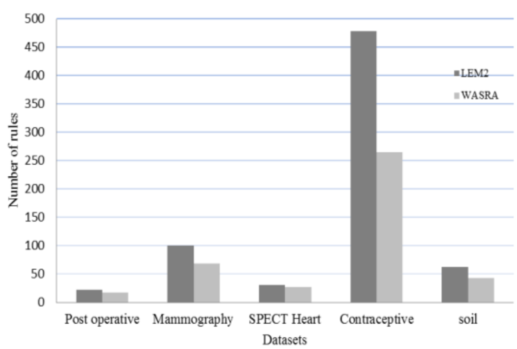

Figure 9. **The number of rules considered in LEM2 and WASRA algorithms for all datasets.**

As an illustration, the intermediate results obtained by the algorithm are given and explained for the postoperative dataset.

5.3.2.1 Implementation on misclassified examples

The implementation is carried out on the misclassified examples of all the datasets. For illustration postoperative dataset is considered.

For each concept and for each run, the concept weight matrix of the eight condition attributes is computed from the rules extracted from the LEM2 algorithm. The concept weight matrix for run 4 of the condition attributes of the postoperative dataset is shown in Table 10.

Table 10. **Concept weight matrix generated for run 4 of the postoperative dataset**

Attribute name	Concept 1	Concept 3
Lc	0.31	0.38
Ls	0.54	0.62
Lo	0.23	0.25
Lb	0.38	0.25
Ss	0.38	0.50
Cs	0.00	0.00
Bs	0.54	0.62
Cf	0.46	0.50

For each run, a decider matrix is formed with misclassified examples as rows and rule subset as columns. Table 11 shows a sample decider matrix for run 4 with misclassified examples having the following indices; 10, 76, 79, 15, 22, 2, 9 and the subset of the rules chosen for run 4 has the following rule numbers; 1, 8, 2, 4, 5, 10, 3, 9, 7, 6, 11, 12, 15, 16, 19. The class label chosen based on each rule for each misclassified example is stored in the decider matrix. The class label used for the example is the one that is found by the majority of the rules. The rules in the decider matrix are arranged in the decreasing order of Laplace accuracy that is used for breaking the ties.

The number of examples that are misclassified by LEM2, which are correctly classified by WASRA, for every run of the postoperative dataset is shown in Table 12.

The misclassified examples by LEM2 that are correctly classified by WASRA algorithms for all the datasets are given in Table 13.

Table 11. **Decider matrix for run 4 of postoperative dataset**

E \ R	1	8	2	4	5	10	3	9	7	6	11	12	15	16	19
10	1	1	1	1	1	1	1	1	1	NA	1	NA	NA	1	NA
76	1	1	1	1	1	1	1	1	1	1	NA	NA	3	NA	3
79	1	1	1	1	1	1	1	1	1	1	NA	NA	3	1	3
15	1	1	1	1	1	1	NA	1	1	NA	NA	NA	3	3	3
22	1	1	1	1	1	1	NA	1	NA	1	NA	NA	1	1	NA
2	1	1	1	1	1	1	1	1	1	1	1	NA	3	1	3
9	1	1	1	NA	1	1	1	NA	1	1	1	NA	3	1	3

Table 12. **Misclassified examples by LEM2 and correctly classified by WASRA for the postoperative dataset**

Algorithms	Run 1	Run 2	Run 3	Run 4	Run 5
LEM2	4	7	8	7	6
WASRA	2	2	2	0	1

Table 13. **Misclassified examples by LEM2 that are correctly classified by WASRA for every run of all datasets**

Runs	Mammography		SPECT Heart		Contraceptive		Soil	
	A	B	A	B	A	B	A	B
1.	45	23	11	5	160	54	30	16
2.	53	30	13	6	173	62	28	10
3.	50	31	13	1	161	41	24	15
4.	54	25	13	4	159	61	23	11
5.	47	28	10	1	156	44	27	10

A: Misclassified by LEM2; B: Correctly classified by WASRA

The percentage of misclassified examples of LEM2 that are correctly classified by WASRA algorithm is tabulated in Table 14.

From Table 14 it is seen that WASRA behaves better even for the highly inconsistent dataset, which is revealed by the mammography dataset. The results indicate that the proposed algorithm performs better than LEM2 algorithm. This enabled to proceed further in implementing the proposed algorithm on the entire test dataset.

Table 14. **Percentage of misclassified examples of LEM2 that are correctly classified by WASRA algorithm for all datasets**

S. No	Dataset names	Consistency measure (%)	Percentage of misclassified examples of LEM2 that are correctly classified by WASRA algorithm (%)
1.	Postoperative	83.7	21.87
2.	Mammography	53.3	55.02
3.	SPECT Heart	81.6	25
4.	Contraceptive	73.99	32.38
5.	Soil	64.3	46.96

5.3.2.2 Implementation on the entire test dataset

WASRA algorithm is experimented on the test dataset of all datasets. The average prediction accuracy of all dataset with stratified selection on the test dataset for five-fold validation with LEM2 and WASRA is tabulated in Table 15.

The results show WASRA has obtained 4.92%, 13.13%, 1.11%, 2.79% and 5.71% improvement than LEM2 for the postoperative dataset, mammography dataset, SPECT heart, Contraceptive dataset, and soil dataset, respectively. The comparison of the average prediction accuracy of LEM2 and WASRA algorithms for all datasets is presented in Figure 10.

Table 15. **Prediction accuracy for all datasets with LEM2 and WASRA algorithms**

Runs	Postoperative (%)		Mammography (%)		SPECT Heart (%)		Contraceptive (%)		Soil (%)	
	LEM2	WASRA	LEM2	WASRA	LEM2	WASRA	LEM2	WASRA	LEM2	WASRA
1.	75	87.50	72.72	80	79.62	79.62	45.57	51.70	56.52	65.21
2.	61.11	55.56	68.26	82.03	75.92	81.48	41.15	45.91	60.56	69.01
3.	52.94	64.70	69.69	84.84	75.47	77.35	45.42	42.71	65.71	75.71
4.	61.11	61.11	67.66	81.43	75.47	77.35	46.10	42.71	67.14	68.57
5.	64.70	70.58	71.68	87.34	81.13	77.35	47.11	51.52	61.42	61.42
Average	62.97	67.89	70	83.13	77.52	78.63	45.07	47.45	62.27	67.98

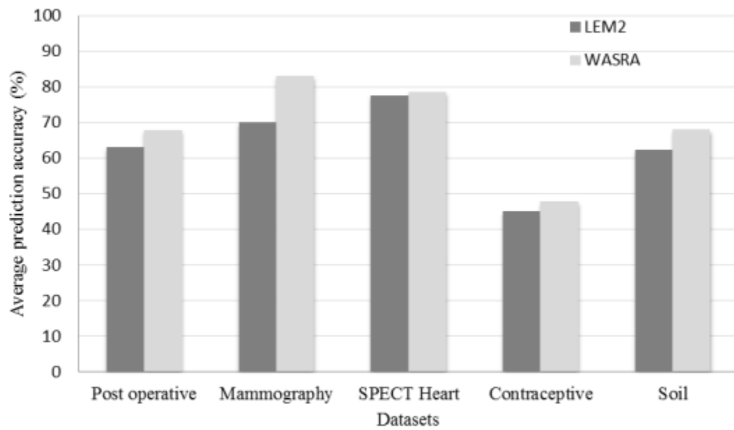

Figure 10. **Prediction accuracy for all datasets with LEM2 and WASRA algorithms**

From Figure 10, it is seen that the postoperative dataset with the positive region of 83.7% shows 4.92% improvement in the average prediction accuracy of the proposed algorithm that uses only 73.91% of the rules of LEM2. The mammography dataset with the positive region of 53.3% shows a 13.13% improvement in the average prediction accuracy of the proposed algorithm that uses only 68.31% of the rules of LEM2. The SPECT heart dataset with the positive region of 81.6% uses 68.25% of the rules of LEM2 in the WASRA algorithm shows 1.11% improvement in the average prediction accuracy.

The proposed algorithm uses 55.43% of the LEM2 rules of the contraceptive dataset with the positive region of 73.9% and shows 2.79% improvement in the average prediction accuracy. The WASRA algorithm average prediction accuracy of soil dataset with positive region 64.3% shows 5.71% improvement for 68.25% of rule subset of LEM2.

It is seen from Figure 10 that the mammography dataset displays higher improvement in the prediction accuracy with a positive region 53.3% and its class distribution ratio 5: 5, whereas, SPECT heart dataset displays more or less the same prediction accuracy as that of LEM2 with the class distribution ratio of 2: 8 and positive region 81.6%. Thus, it is seen that if the class distribution is good in a dataset, even if its positive region is not very high, WASRA shows higher improvement in the prediction accuracy, compared to LEM2.

5.3.3 Comparative analysis with standard classifiers

The proposed classifier is compared to other standard classifiers. Classifiers considered for comparison are PRISM and ID3 since these classifiers use a minimal set of rules for classification. Tab. 16 tabulates the comparison of classifiers LEM2, PRISM and ID3 with the proposed classifier, WASRA.

Table 16. **Comparison of prediction accuracy of WASRA with LEM2, PRISM, and ID3 for all datasets**

Dataset\ Classifier	LEM2 (%)	PRISM (%)	ID3 (%)	WASRA (%)
Postoperative	62.97	58.14	53.48	67.89
Mammography	70.00	66.86	74.45	83.13
SPECT Heart	77.52	73.41	75.28	78.65
Contraceptive	45.07	40.39	41.68	47.86
Soil	62.27	61.42	64.57	67.98

A graphical representation of the results in Table 16 is presented in Figure 11.

From the above experimental results and comparative analysis, it is seen that the proposed classifier WASRA is well suited for the prediction of uncertain datasets. For the mammography dataset with a positive region 53.3%, WASRA shows an improvement of 13.13%, 16.27%, and 8.68% over LEM2, PRISM, and ID3, respectively.

6 CONCLUSION

A new classification algorithm, WASRA is developed for uncertain datasets using rough set concepts. By considering and approximating a subset of rules generated by a rule induction algorithm, WASRA is able to exhibit improvement in prediction accuracy. The rule approximation algorithm uses weights of attributes, which are assigned based on their presences in the rules associated with each concept. These weights are used for approximating the rules which are used for classification. The classifier thus designed is tested on benchmark datasets of the UCI repository and its performance is comparable to those of the standard classifiers of the literature. The proposed algorithm is also tested on a real dataset. The results show WASRA has obtained 4.92%, 13.13%, 1.11%, 2.79% and 5.71% improvement than LEM2 for the postoperative dataset, mammography dataset, SPECT heart, contraceptive dataset, and soil dataset, respectively.

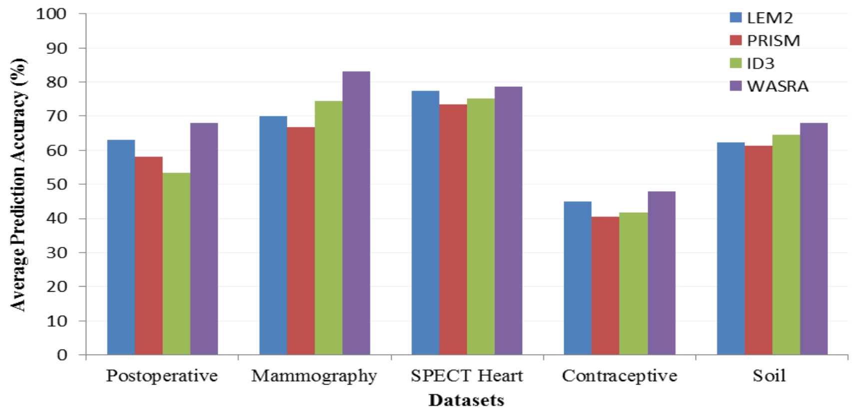

Figure 11. Comparison of WASRA with other classifiers for all datasets

7 ACKNOWLEDGEMENT

WE are grateful to the National Agro Foundation (NAF), Chennai for providing the soil samples data for this work. We are thankful to Dr. M. R. Ramasubramaniyan and Ms. P. Kalpana for the expert suggestions about the soil data and interpretation.

8 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

9 REFERENCES

J. Bazan, J. F.Peters, A. Skowron, N. H. Son, and M. Szczuka. (2003). Rough Set Approach to Pattern Extraction from Classifiers, *Electronic Notes in Theoretical Computer Science*, 82(4), 20–29.

C. Berzuini. (1988).Combining symbolic learning techniques and statistical regression analysis, in *Proceeding of the 7th National conference on Artificial Intelligence AAAI-88*, 612 - 617.

J. Cendrowska. (1987). PRISM: An algorithm for inducing modular rules, *International Journal of Man-Machine Studies*, 27(4), 349-370.

C. C. Chan and J. W. Grzymala-Busse. (1991). On the attribute redundancy and the learning programs ID3, PRISM, and LEM2.*Technical report*, Department of Computer Science, University of Kansas.

D. Dua and C. Graff. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

J. W. Grzymala-Busse. (1992). LERS - A system for learning from examples based on rough sets, in R.Slowinski (ed.) *Intelligent Decision Support*. Handbook of Application and Advances of the Rough Set Theory, Kluwer Academy Publishers, Dordrecht, 3-18.

J. W. Grzymala-Busse. (1993). Selected algorithms of machine learning from examples, *Fundamenta Informaticae*, 18, 193-207.

J. W. Grzymala-Busse. (1997). A new version of the rule induction system LERS, *Fundamenta Informaticae*, 31, 27–39.

J. W. Grzymala-Busse. (2002). MLEM2 A new algorithm for rule induction from imperfect data, in *Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 243–250.

J. W. Grzymala-Busse. (2015). A Comparison of Rule Induction Using Feature Selection and the LEM2 Algorithm, in Stańczyk U and L. Jain (ed.), Feature Selection for Data and Pattern Recognition, *Studies in Computational Intelligence*, 584, Springer, Berlin, Heidelberg, 163-176.

A. E. Hassanien and J. M. H. Ali. (2004). Rough Set Approach for generation of Classification Rules of Breast Cancer Data, *Informatica*,15(1), 23-38.

L. P. Khoo, S. B.Tor, and L. Y. Zhai. (1999). A Rough-Set-Based Approach for Classification and Rule Induction, *International Journal of Advanced Manufacturing Technology*, 15, 438-444.

A. Kusiak. (2001). Rough set theory - A data mining tool for semiconductor manufacturing, *IEEE Transactions on Electronics Packaging Manufacturing*, 24, 44-50.

N. Lavrac, P. Flach, and B. Zupan. (1999). Rule Evaluation Measures: A Unifying View, in ILP-99 *Proceedings of the 9th International Workshop on Inductive Logic Programming*, 174-185.

R.S. Michalski and R.L. Chilausky. (1980). Learning by being told and learning from examples: an experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis, *Policy Analysis and Information Systems Journal*, 4, 125–160.

M. Nandhini. and S. N. Sivanandam. (2015). An improved predictive association rule based classifier using gain ratio and T-test for health care data diagnosis, *Sadhana*, 40, 1683–1699.

T. Niblett. (1987). Constructing decision trees in noisy domains, in *Proceedings of the 2nd European Conference on Working Session on Learning*, 67-78.

Z. Pawlak. (1981). Information systems theoretical foundations, *Information Systems*, 6(3), 205–218.

Z. Pawlak. (1982). Rough Sets, *International Journal of Computer and Information Sciences*, 11, 341-356.

Z. Pawlak. (1984). Rough classification, *International Journal of Man-Machine Studies*, 20, 469-483.

Z. Pawlak. (1992).Rough Sets: Theoretical Aspects of Reasoning about Data, (Kluwer Academic Publishers, Norwell, MA.

Z. Pawlak. (1996). Rough Sets Present State and Further Prospects. *Intelligent Automation & Soft Computing*, 2(2), 95-102.

Z. Pawlak, J. W. Grzymala-Busse, R.Slowinski, and W. Ziarko. (1995). Rough sets, *Communications of the ACM*, 38, 88-95.

J. R. Quinlan, (1986). Induction of Decision Trees, *Machine Learning*, 1, 81–106.

J. Stefanowski. (1998).The rough set based rule induction technique for classification problems, in: *Proceedings of the 6th European conference on intelligent techniques and soft computing (EUFIT'98)*, 109–113.

L. Sumalatha, P.Uma Sankar and B.Sujatha. (2016), Rough set based decision rule generation to find behavioural patterns of customers, *Sadhana*, 41(9), 985–991.

H. K. Sharma, K. Kumari. and S. Kar. (2019). Short-Term Forecasting of Air Passengers Based on the Hybrid Rough Set and Double Exponential Smoothing Model. *Intelligent Automation & Soft Computing*, 25(1), 1-14.

S. Trabelsi and Z.Elouedi. (2008). Learning decision rules from uncertain data using rough sets, in *Conference: Computational Intelligence in Decision and Control - 8th International FLINS Conference*.

Y. Y. Yao and N. Zhong. (1999). An analysis of quantitative measures associated with rules, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 479-488.

W. Ziarko. (1999). Discovery through Rough Set Theory, Knowledge Discovery: viewing wisdom from all perspectives, *Communication of the ACM*, 42(11).

W. Ziarko. (2002). Rough Set Approaches for Discovery of Rules and Attribute Dependencies, in W. Klosgen, J. Zytkow (eds) Handbook of Data Mining and Knowledge Discovery, Oxford University Press, 328-338.

W. Ziarko and N. Shan. (1996). A method for computing all maximally general rules in attribute-value systems, *Computational Intelligence*, 12 (2), 223- 234.

10 NOTES ON CONTRIBUTORS

L. Ezhilarasi received her M.Tech degree in Computer Science and Engineering from VIT University in 2011. She is currently pursuing her Ph.D. at Anna University. Her research interests include Rough Sets, Knowledge Representation, Data Mining, Soft Computing, and Computational Intelligence. She is a member of International Rough Set Society (IRSS) and Indian Society for Rough Set (ISRS).

A. P. Shanthi is currently a Professor in the Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University. She has more than 30 years of teaching experience. Her research interests include Computer Architecture, Evolvable Hardware, Embedded systems, Machine Learning, and Soft Computing Techniques.

V. Uma Maheswari is a former Professor in the Department of Information Science and Technology, Anna University. She has more than 30 years of teaching experience. Her areas of interests are Rough Sets, Computational Intelligence, Compiler Design, Data Mining, Operating Systems, and Soft Computing.