Transaction A: Civil Engineering
Vol. 16, No. 2, pp. 116–125

Making a Decision Between the Rehabilitation and Reconstruction of Asphalt Pavements Using the Rough-Set Theory

M. Arabani[1,*], A.K. Haghi[1] and B. Amani[2]

Abstract. *Every year a great amount of money is expended for the rehabilitation and reconstruction of roads and pavements in most countries. Besides, making an ideal decision on this based on the types of failures determined, takes too long. The rough-set theory is an effective tool for the analysis of information systems in a Pavement Management System (PMS) database gained by both objective and subjective methods. A rough-set based analysis acts like a knowledge engineer who sits between data and the user. This approach appears to capture information on uncertainty, imprecision and ambiguity along with precise values in a PMS database. This paper explores a new approach to the rough-set theory in a PMS database that enables pavement engineers to discover the shortest subsets of condition attributes having quality equal to the general quality of defined characteristics in the information system, to assess and describe pavement conditions, and to derive decision rules for rehabilitation and reconstruction of the pavements. To evaluate the results, the best algorithm of defined attributes in the information system is determined by making use of a stepwise linear regression method and the result is compared with rough-set ones. The results of the research indicate that the rough-set theory has a better and stronger operational capability in identifying the effective parameters for the severity evaluation of typical distresses in asphalt pavements and in decision-making for selecting the type of repair.*

Keywords: *Pavement Management System (PMS); Pavement Condition Index (PCI); Asphalt pavement; Rough-set theory; Pavement distress.*

INTRODUCTION

Pavement distresses are classified into two different categories. The first is known as the functional failure. In this case, the pavement does not carry out its intended function without either causing discomfort to passengers, or high stresses to vehicles. The second known as the structural failure, includes a collapse of pavement structure or the breakdown of one or more components of the pavement with such magnitude that the pavement becomes incapable of sustaining the loads imposed upon its surface [1]. In some cases one type of failure may be accompanied by the other type but mostly there is only one type of failure.

1. *Faculty of Engineering, The University of Guilan, Rasht, P.O. Box 3756, Iran.*
2. *Faculty of Engineering, Azad University of Anzali, Guilan, P.O. Box 1451, Iran.*
*. *Corresponding author. E-mail: m_arbani@yahoo.com*

Received 12 September 2006; received in revised form 4 March 2008; accepted 12 May 2008

The functional failure depends primarily upon the degree of surface roughness. Structural failure in a flexible pavement may be a result of fatigue, consolidation or shear, developing in the subgrade, subbase, base course or surface [2].

Two methods can be used for the evaluation of pavement distresses. The first evaluates the effect of failure on the intended function of the pavement, that is, its serviceability under daily traffic. The second is a mechanical evaluation with visual inspection determining the physical conditions of pavement and the problems causing these conditions [3].

The difference between the two methods is due to the correlation of pavement behaviour to constructional procedures. For instance, a crack in the surface of the pavement may have no effect on the serviceability of the pavement to the traffic, but this crack can let water into the base course and increase the deflection of the pavement and cause serious failure. Any change in pavement behavior can cause some type of failure and may be a result of constructional functions [4].

One of the most appropriate methods for evaluating pavement distresses is the Pavement Condition Index (PCI) procedure. This procedure has been recommended by the U. S. Corps of Engineers and is considered as a standard method in many organizations all over the world [5]. The PCI is in fact a number between 0 and 100 that shows pavement conditions from poor to excellent. The PCI number is calculated by evaluating several segments of a pavement to determine the severity of distress. The information gained in this procedure can provide a complete reorganization of the main causes of failure and their relation to traffic load, climatic conditions or other effective factors [6].

After evaluating the distresses by the PCI method, it is important to make an economical decision about the rehabilitation or repair of the pavement. In most cases, it is difficult to select which type of repair is suitable for the pavement, because the type and severity of distress may be different in each segment of the road and cannot be repaired using one procedure. On the other hand, a general PCI number cannot provide sufficient information to make the ideal decision for different parts of a pavement [7]. Thus, development of an adequate algorithm to reduce the number of variables effective on decision making with the PCI evaluation results can be very helpful.

The present paper describes a method for applying the rough-set theory to determine the severity level of distresses in flexible pavements with experimental knowledge. For this purpose, the rough-set theory was applied to diagnostic cases, by experts in this field. This method is applied to remove conditional attributes and classes of each conditional attribute insignificant in the diagnosis to extract the smallest decision algorithm(s), that is (are) still capable of making a diagnosis equal to those of experts [8]. Finally, to evaluate the results, the best algorithm of defined attributes in the information system is determined by making use of a stepwise linear regression method and the result is compared with rough-set ones.

ROUGH-SET THEORY

The rough-set theory was proposed by Pawlak [9] as a new mathematical tool for reasoning about vagueness, uncertainty and imprecise information [8]. Rough-set has been successfully applied to extracting laws from decision tables [10], automated extraction of rules from clinical databases [11], learning conceptual design rules, rules for water demand prediction [12], data mining and knowledge discovery, and the pavement management database [13]. Recently, Kryszkiewicz presented a general rough-set framework for dealing with incomplete information systems [14,15].

Rough-set analysis is essentially a nonparametric statistical method that is able to handle a diverse and less immediate tangible set of factors. It provides a formal tool for transforming a data set, such as a collection of past observations or a record of experiences, into structured information, in such a way that it can classify objects having distinctive patterns of attributes. It is not always possible to distinguish objects on the basis of available information (descriptors). The imperfect information causes indiscernibility of objects through the values of the attributes describing them and prevents them from being unambiguously assigned to a given single set. In this case, the only sets that can be precisely characterized, with regard to the values of ranges of such attributes are lower and upper approximations of the set of objects [6]. We will now set out the basic principles of this method. (For more details, see [9,16-18].)

Human knowledge based on experience (e.g. concerning decision making in a specific field) is often recorded in a structure called an information system. This information system contains information about particular cases (objects, states, observations, events) and factors and attributes effective on them (features, variables, characteristics, symptoms). The set of attributes consists of two kinds. The first kind (called condition attributes) concerns the results of some tests or measurements, data from observations, anamnesis, symptoms of cases, states etc. The other kind (called decision attributes) concerns some expert's decisions, diagnoses, classified results of a treatment, etc. [19].

With reference to a certain finite set of objects, U, it is assumed possible to perceive the differences existing between them by observing some information associated with each of them. A finite set, Q, of attributes, which serves to identify and characterize these objects, is identified. Since the rough-set theory aims to classify and distinguish data on the basis of different values their attributes assume with reference to each object, each attribute, $q \in Q$, must be able to assume different values in its domain, U_q.

There must be, therefore, at least two of these values for the attribute to be a significant basis for the required characterization. If an attribute is quantitative, its domain is, in practice, partitioned into a suitable number of sub-intervals, which give a good description of the phenomenon studied, so as to avoid ending up with a distribution of values with a high number of modalities, which would not be useful for the analysis intended. The difficult choice of the bounds (called norms) used to define these sub-intervals is important to ensure a correct application of this approach and that too much information is not lost in the translation of original quantitative attribute-values into qualitative coded values.

At this point, every $x \in U$ may be introduced as a vector whose components are the distinct evaluations of x, with respect to every attribute of Q, and called a description of x, in terms of attribute-values from set Q. The table, containing the descriptions of every $x \in U$ by means of the attributes of the set Q, is known as the information table. It is also possible to obtain a description of $x \in U$, in terms of any one subset of attributes $P \subseteq Q$.

A fundamental concept of the rough-set theory is that of the binary relation of indiscernibility, denoted by I_P. Two objects x, $y \in U$, are said to be P-indiscernible by means of the set of attributes $P \subseteq Q$, if they have the same description. Thus, the binary relation, I_P, is reflexive, symmetric and transitive (equivalence relation); its classes, that is, the subsets of U containing all the objects with the same description in terms of the attributes from subset P, and only these, are called P-elementary sets. If all the attributes of Q are considered, the Q-elementary sets are called atoms. The P-elementary sets, $P \subseteq Q$, generate a partition of U, in which every object, $x \in U$, belongs to one and only one P-elementary set.

To explain the rough-set theory, it is necessary to introduce two other key concepts. Let $P \subseteq Q$ be a subset of attributes and $X \subseteq U$ a subset of objects of U. P-lower approximation of X, denoted by $P_L X$ is a subset of U with the elements as the objects belonging to the P-elementary sets contained in the set X, and only these. In other words, the elements of $P_L X$ are all the elements of U belonging to all the classes generated by the indiscernibility relation, I_P, and contained in X.

We define the P-upper approximation of X, denoted by $P_U X$, as the subset of U, the elements of which are all the objects belonging to the P-elementary sets having at least one element in common with the set X, and only these. In other words, the elements of $P_U X$ are all the elements of U belonging to all the classes generated by the indiscernibility relation, I_P, that have at least one representative belonging to X, and only these.

The difference between these sets is known as the P-boundary of X, denoted by $Bn_P(X) = P_U X - P_L X$. Therefore, $P_L X \subseteq X \subseteq P_U X$ results and, consequently, if object x belongs to $P_L X$, it is also an element of X; if x belongs to $P_U X$, it may belong to the set X; therefore, $Bn_P(X)$ constitutes the "doubtful region" (with reference to its elements, nothing can be said with certainty about its belonging to the set X). The indiscernible classes generated by I_P, therefore, constitute the basic instrument of the rough-set theory for better recognition and evaluation of data. This knowledge is intended as a family of partitions of U, generated by the indiscernibility relation I_P on U, $P \subseteq Q$.

A P-rough-set is the family of all subsets of U, which includes the same lower and upper P-approximations. The intention is, thus, to approximate a set, X, $X \subseteq U$, by means of a pair of sets associated with it, called lower approximation, $P_L X$, and upper approximation, $P_U X$, of X, which can be then considered as a particular case of an interval set. Only if $P_U X = P_L X$ does X prove to be equal to the union of a certain number of P-elementary sets and is called P-definable. Clearly, in this case (and only in this case), it is possible to affirm with certainty whether x, $x \in U$, belongs to X, $X \subseteq U$, using the set of attributes P. Moreover, the accuracy of the approximation of X, equal to:

$$\frac{\mathrm{card}(P_L X)}{\mathrm{card}(P_U X)}, \tag{1}$$

will be at the maximum value (i.e., equal to 1). In general, therefore, the aim of the rough-set analysis is to establish whether x is an element of X, based on the lower and upper approximations of X, rather than directly by means of a specific characteristic function.

Let $Y = (Y_1, Y_2, \cdots, Y_n)$ be a certain classification of U. Regarding the classification of Y P-lower approximation and P-upper approximation respectively are the sets in Y that having as their elements the P-lower and P-upper approximations, that is $P_L Y = (P_L Y_1, P_L Y_2, \cdots, P_L Y_n)$ and $P_U Y = (P_U Y_1, P_U Y_2, \cdots, P_U Y_n)$. The quality of the approximation of the partition Y considering the set of attributes P, denoted by $\gamma_P(Y)$, can be gained by the ratio of the total number of P-correctly classified objects (i.e., belonging to the P-lower approximations of Y_i, $i = 1, 2, \cdots, n$), to the total number of objects considered. This ratio is called the quality of the classification and will have its maximum value (equal to one) if, and only if, each class Y_i of Y proves P-definable.

Another fundamental concept in this theory is that of attribute reduction (i.e., given a classification Y of the objects of U, the goal is searching for a minimal possible set of independent attributes (R) that has the same quality of classification as the original set of attributes P). The minimal subset $R \subseteq P \subseteq Q$ that $\gamma_R(Y) = \gamma_P(Y)$ is called Y-reduct of P and denoted by $\mathrm{RED}_Y(P)$. (Note that a single table of information may have more than one reduct.) The intersection of all the Y-reducts is known as Y-core of P, that is, $\mathrm{CORE}_Y(P) = \bigcap \mathrm{RED}_Y(P)$. Naturally the core contains all the attributes of P, which are the most important attributes in the information table (i.e., the most relevant for a correct classification of the objects of U).

In another words, in order to analyze the information table, it is sufficient to use any one of the reduced attributes $R \subseteq Q$. So, the classification Y of the objects of U may be characterized only without eliminating any information and there is no need to any other defined attributes of $Q - R$. On the other hand, each of the attributes not belonging to the core may be neglected without deteriorating the quality of the classification considered, but if any one attribute belonging to the core were eliminated from the information table, it will not be possible to obtain the highest quality of approximation with the remaining attributes.

Consequently, as mentioned above, rough-set theory is essentially a classification method devised for non-stochastic information and by making use of it, the following results are obtainable:

- Evaluation of the relevance of particular condition attributes;
- Construction of a minimal subset of variables ensuring the same quality of description as the whole set (i.e., reducts of the set of attributes);
- Intersection of those reducts giving a core of attributes that cannot be eliminated without disturbing the quality of description of the set of attributes;
- Identification and elimination of irrelevant attributes [20].

THE APPLICATION OF ROUGH-SET THEORY IN SEVERITY EVALUATION OF TYPICAL PAVEMENT DISTRESSES

The rough-set theory has the attention of researchers and theoreticians worldwide and has been successfully applied to different fields. It is to be noted that the rough-set theory is not basically a MCDM method and its concept is different. This method uses the definition of mathematical sets and subsets in analyzing procedures.

The first step towards analyzing the information by rough-set method and recognize the severity level for pavement distresses is the construction of an information system.

Information System

The information system is, in fact, a finite data table; columns of which are labelled by attributes and rows by objects and the entry in column q and row x has the value $\rho(x, q)$. Each row in the table represents the information about an object in U. Thus, the application of the rough-set analysis in the severity evaluation of typical pavement distresses may proceed in two successive steps:

- Construction of an information survey;
- Classification of information contained in the survey [20].

Information Survey

In this research, the information survey consists of studies based on 35 different cases of information obtained from parts of an asphalt pavement and extracted by experts, in accordance with the diagnostic method prepared by the American Army Corps of Engineers in 1984. Then, the collected information is formed into an unclassified table, such as the information system. Each row in this table represents information about the characteristics of one case of pavement studied. Each column represents one of the defined characteristics and the last one specifies its severity to be reconstructed, rehabilitated or to need minor repair.

Classification of Information

The rough-set approach can effectively handle quantitative data, but this data must first be converted into qualitative or categorical data by means of an adequate codification. This is done by means of a set of thresholds called norms, which discretize the measurement scales, by which the quantitative data are expressed. This is applied to both categorical and ratio information. The observations or samples are classified into various categories for each attribute separately. From the researcher's viewpoint, the introduction of the thresholds could mean a methodological advantage, because the discretization of the measurement scale for quantitative attributes should represent the researcher's perception of the analyzed phenomenon that can be represented and analyzed in a form understandable to the researcher. However, this step is one of the most problematic issues in the application of a rough-set analysis.

First, the use of thresholds implies some loss of information. Second, thresholds are chosen subjectively. For example, thresholds are often those that produce some satisfactory approximation of the considered categories. This is the case in this survey for most attribute variables. In general, some sensitivity analysis on the classification used is meaningful, as a balance needs to be found between homogeneity and class size. This classification exercise leads then to a decision table, in which all objects are subdivided into distinct categories for each relevant attribute [20]. The categories used are presented in Table 1 and the resulting coded information is listed in Table 2.

Outputs of Rough-Set Analysis

By applying the information system of pavement distresses, four main sets of indicators and outputs can be calculated.

A. The Reducts

The reducts are all combinations of explanatory or independent attributes that can completely determine

Table 1. Conditional attributes of asphalt pavement failure.

Conditional Attributes	Classification of Individual Situations	Severity Level
(*a*) *Upheaval & shoving*	1. Upheaval & shoving cause high severity discomfort for riding 2. Upheaval & shoving cause medium severity discomfort for riding 3. Upheaval & shoving cause low severity discomfort for riding 4. None	H M L N
(*b*) *Block cracking*	1. High severity 2. Medium severity 3. Low severity 4. None	H M L N
(*c*) *Alligator or fatigue cracking*	1. Progressed cracked pieces, severely spalled at the edges 2. Development of cracking in the pattern of pieces, lightly surface spalled 3. Longitudinal disconnected hairline cracks running parallel to each other 4. None	H M L N
(*d*) *Longitudinal and transverse cracking*	1. High severity 2. Medium severity 3. Low severity 4. None (note to the type of cracking and width of cracking)	H M L N
(*e*) *Lane/shoulder drop off or heave*	1. The difference in elevation between traffic lane & the shoulder 2. The difference is between 51 mm to 102 mm 3. The difference is between 25 mm to 51 mm 4. The difference is less than 25 mm	H M L N
(*f*) *Slippage cracking (crescent or half-moon-shaped-cracks)*	1. The average width of cracking is more than 38 mm 2. The average width of cracking is betwen 10 mm to 38 mm 3. The average width of cracking is less than 20 mm 4. None	H M L N
(*g*) *Swell*	1. Swell causes excessive bounce of vehicle which creates substantial discomfort or safety risk 2. Swell causes significant bounce of vehicle which creates some discomfort 3. Swell causes some bounce of vehicle which creates no discomfort 4. None	H M L N
(*h*) *Potholes*	1. High severity 2. Medium severity 3. Low severity 4. None (note to the depth of potholes)	H M L N
(*I*) *Rutting*	1. Mean rut depth is more than 25 mm 2. Mean rut depth is between 13 mm to 25 mm 3. Mean rut depth is between 6 mm to 13 mm 4. None	H M L N
(*J*) *Corrugation*	1. Corrugations cause excessive vibration of the vehicle which creates substantial discomfort or safety risk 2. Corrugations cause significant vibration of the vehicle which creates some discomfort 3. Corrugations cause some vibration of the vehicle which creates no discomfort 4. None	H M L N
(*k*) *Raveling and weathering*	1. High severity 2. Medium severity 3. Low severity 4. None	H M L N
(*L*) *Depression*	1. Mean depth of depression is more than 51 mm 2. Mean depth of depression is between 25 mm to 51 mm 3. Mean depth of depression is between 13 mm to 25 mm 4. Mean depth of depression is less than 13 mm	H M L N

(or explain) the variation in the dependent attribute, without any need for other explanatory variables [20], so the reducts resulted from the rough-set analysis have the same values of accuracy and quality of description as the whole attributes. In other words, these algorithms can explain and evaluate the variations in the pavement conditions and their severity level of distresses, without any need for other independent parameters.

The reducts are given in Table 3. There appear to be, on the basis of the chosen set of characteristics and classification of these characteristics, six competitive

Table 2. Observation data for diagnosis of pavement distresses severity levels.

Segments	**Conditional Attributes**												**Severity**
	a	***b***	***c***	***d***	***e***	***f***	***g***	***h***	***i***	***j***	***k***	***l***	
S_1	4	2	2	1	1	3	1	1	2	1	3	2	A*
S_2	4	3	2	2	1	3	1	1	2	3	1	2	A
S_3	4	4	2	1	1	2	1	1	2	3	2	1	A
S_4	4	4	2	1	2	4	3	3	2	3	3	2	C
S_5	3	4	2	2	2	4	2	1	1	1	1	2	B**
S_6	4	4	2	2	1	4	3	1	2	1	3	2	B
S_7	3	4	2	1	1	4	2	1	2	3	1	2	C***
S_8	4	3	2	1	1	2	1	2	1	3	3	2	A
S_9	4	3	2	2	3	4	3	3	2	3	3	2	C
S_{10}	4	3	2	1	1	3	3	3	2	3	2	2	B
S_{11}	3	3	3	2	1	4	3	3	2	3	1	2	B
S_{12}	4	3	2	2	1	4	3	1	2	3	2	2	C
S_{13}	4	3	2	2	1	3	3	1	1	3	2	2	B
S_{14}	3	3	4	1	1	4	3	3	2	2	1	2	B
S_{15}	4	4	2	2	2	3	3	3	2	3	3	2	C
S_{16}	4	4	4	2	1	4	3	3	1	3	2	2	C
S_{17}	4	4	4	2	2	3	3	3	1	2	1	2	B
S_{18}	4	3	2	2	1	4	3	1	2	1	2	2	B
S_{19}	4	4	2	2	2	4	1	1	2	1	1	1	A
S_{20}	4	3	4	2	1	4	3	3	2	3	2	2	C
S_{21}	3	4	4	1	1	3	3	3	2	3	1	2	B
S_{22}	3	4	4	1	1	3	2	3	2	3	1	2	B
S_{23}	4	4	2	2	2	4	2	3	2	3	2	2	C
S_{24}	4	4	2	2	2	4	2	3	2	3	2	2	C
S_{25}	3	3	2	2	1	4	3	3	2	3	1	2	B
S_{26}	4	4	2	1	1	2	1	1	2	3	2	1	A
S_{27}	2	3	3	2	1	4	2	3	2	3	2	1	B
S_{28}	4	3	3	2	2	4	3	3	1	3	1	2	C
S_{29}	4	2	2	2	1	4	2	2	2	2	2	2	B
S_{30}	4	2	2	1	1	2	2	1	2	3	2	2	A
S_{31}	2	3	4	2	1	2	3	1	2	3	3	2	B
S_{32}	4	3	4	2	2	4	3	3	2	2	2	1	C
S_{33}	3	2	3	1	2	2	2	3	1	3	1	2	B
S_{34}	3	4	3	2	3	4	3	2	2	3	2	2	C
S_{35}	4	2	4	1	3	2	1	1	1	2	1	1	A

*: A: High level severity (should be reconstructed)

**: B: Medium level severity (should be overlayed)

***: C: Low level severity (minor repair is needed)

theories for explaining the variance in the severity level of pavement distresses.

These algorithms are the shortest sets of attributes, resulted from the rough-set analysis. In Table 3, the first decision-making algorithm determines the variations in the severity level by combining and evaluating the parameters of upheaval & shoving, block cracking, slippage cracking, swell, and corrugation. Also the second reduct suggests combining and evaluating the parameters of upheaval & shoving, block cracking, slippage cracking, corrugation and raveling and weathering to determine the severity level of pavement distresses. Generally, all the reducts can be used in the severity evaluation of pavement distresses and they have the same values of quality and accuracy of approximation.

B. Core

The core is the set of variables (parameters or attributes) that are in all reducts, as discussed above,

Table 3. Reducts and core.

Reducts	
Algorithm no. 1	{*Upheaval & shoving, block cracking, slippage cracking, swell, corrugation*}
Algorithm no. 2	{*Upheaval & shoving, block cracking, slippage cracking, corrugation, raveling & weathering*}
Algorithm no. 3	{*Upheaval & shoving, lane/shoulder drop off or heave, slippage cracking, swell, corrugation*}
Algorithm no. 4	{*Upheaval & shoving, slippage cracking, swell, corrugation, raveling & weathering*}
Algorithm no. 5	{*Upheaval & shoving, slippage cracking, potholes, corrugation, raveling & weathering*}
Algorithm no. 6	{*Lane/shoulder drop off or heave, slippage cracking, swell, corrugation, raveling & weathering*}
Core	{*Slippage cracking, corrugation*}

or that are part of all theories [20]. This means that these variables strongly influence the severity level of pavement distresses. In this study, the core consists of the slippage cracking and corrugation.

C. Accuracy of Classes

The accuracy for each value class of the decisional variable is calculated by dividing the lower to the upper approximation of each class [20]. The results are shown in Table 4 and for all classes of pavement distresses the accuracy is equal to 1. Also, the accuracy and quality of classification are equal to 1. This value is the maximum and it means that on the basis of the chosen characteristics in this study, samples are fully discernible regarding the three classes of pavement distresses.

D. Rules

The rules are exact or approximate relationships between explanatory variables and dependent variables. They may be considered like "if... then..." statements. A rule may be exact (deterministic), or approximate (non-deterministic). An exact rule guarantees that the values of decision attributes correspond to the same values of condition attributes (same conditions, same decisions); an approximate rule, on the other hand, states that more than one value of decision attributes corresponds to the same values of condition attributes (same conditions, different decisions). Therefore, only in the case of exact rules, using the information contained in the decision table, it is always possible to state with certainty whether an object belongs to a certain class of the decision variable or not. An exact rule, therefore, offers a sufficient condition of belonging to a decision class while an approximate rule (only) admits the possibility of this [20]. Table 5 shows the rules generated from the present research data set.

Table 4. Accuracy and quality of the classification of the severity level.

Severity Level of Pavement Distresses	Accuracy
1 (high or A)	1
2 (medium or B)	1
3 (low or C)	1
Accuracy of approximation	1
Quality of approximation	1

ANALYZING THE INFORMATION AND DETECTING PARAMETERS, USING LINEAR REGRESSION METHOD

The linear regression method is a mathematical tool reflecting the linear correlation between independent and dependent parameters. Through the said method, the changes existed in the dependent parameters can be evaluated. In this method, mathematical correlative equations are used. The selection of the best form of the said equations and the determination of their parameters require a great deal of study, concerning the aforementioned issues. The current and common form of the correlation model used in the linear regression model is as follows:

$$Y = a_o + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n. \quad (2)$$

Here, Y is the dependent variable, $X_{1 \cdots n}$ are independent variables and $a_{0 \cdots n}$ are coefficients of the equation, which have already been determined.

Since the purpose of this project is the recognition of the most important and effective attributes, concerning the severity evaluation of pavement distresses, the dependent parameters ought to be the indicator of the selective level of pavement distresses severity. Independent parameters must be the indicator of pavement characteristics.

However, there is a great problem in the construction and formation of regression equations being linked with the utilization of the parameters representing the changes, the best in the severity level of pavement distresses, because using all the parameters and factors to form the regression equation is practically difficult and even at times impossible. Consequently, it is very hard and even infeasible to gather a lot of input information, required for the said issue. Thus, the

Table 5. Rules generated by the rough-set analysis.

Deterministic Rules	
Rule 1	$(g = 1)$ => (severity = 1)
Rule 2	$(d = 1)\&(f = 2)\&(h = 1)$ => (severity = 1)
Rule 3	$(a = 3)\&(h = 3)$ => (severity = 2)
Rule 4	$(J = 1)\&(k = 2)$ => (severity = 2)
Rule 5	$(e = 1)\&(g = 3)\&(k = 3)$ => (severity = 2)
Rule 6	$(f = 3)\&(i = 1)$ => (severity = 2)
Rule 7	$(d = 2)\&(e = 1)\&(g = 2)$ => (severity = 2)
Rule 8	$(i = 1)\&(J = 1)$ => (severity = 2)
Rule 9	$(b = 3)\&(d = 1)\&(f = 3)$ => (severity = 2)
Rule 10	$(a = 4)\&(f = 4)\&(J = 3)$ => (severity = 3)
Rule 11	$(b = 4)\&(f = 4)\&(J = 3)$ => (severity = 3)
Rule 12	$(a = 4)\&(d = 2)\&(h = 3)\&(i = 2)$ => (severity = 3)

stepwise method can be used to determine the shortest and the most suitable and possible combinations of attributes or to detect the most important defined attributes in the information system which are considered to define the observed changes concerning the values of the dependent parameter.

In this method, different parameters are used to gain the best linear correlation in the model in question to obtain the highest value of R^2 by dependent parameters. In this process, at first, the value of the correlation coefficient between each independent parameter and dependent variable is evaluated. This is accomplished to determine which independent parameter enjoys the highest degree of correlation coefficient with the dependent parameter. In the next step, the said process is continued by adding each independent parameter to the primary one, under the framework of a linear regression equation with two independent variables. So, at each step, the value of R^2 is assessed. This trend continues until the best secondary parameter of the independent attributes is gained (and they do not have a high value of correlation with each other). This process continues until, with the addition of another independent parameter to the model, some changes happen in the value of R^2, which are trivial and also negligible. Consequently, the existed parameters in the linear regression equation obtained from this method, are treated as the most significant defined parameters in the information system. According to the linear regression method, these parameters can analyze the observed changes at the levels of dependent parameters in the best way and can be used in the severity evaluation of pavement distresses.

In this investigation, the analysis of the information has been accomplished via a stepwise method and it is concluded that a *swell* parameter is the most significant attribute, enjoying the highest degree of correlation coefficient with a dependent parameter. By the same taken, the parameters regarding *slippage cracking* and *corrugation* are treated as the second and third most important parameters in the information system respectively. The said parameters form a four-variable regression equation, which has the highest value of R^2 between the other four-variable equations. Table 6 indicates the gained linear regression equations. At the end of this step-by-step method, the result indicates that the *lane/shoulder drop off* or *heave*, after the three above-mentioned parameters, is the most significant factor among independent parameters, for estimating the severity of pavement distresses. As it is, the set of these four parameters resulted from the stepwise regression analysis is considered as the best algorithm for the severity evaluation of pavement distresses.

COMPARING THE RESULTS OF ROUGH-SET THEORY WITH THE STEPWISE REGRESSION METHOD

In order to investigate and evaluate the results of a rough-set analysis, the linear regression equations have been formed. Therefore, as the first criterion, the R^2 is obtained from the multiple linear regression equations by various decision-making algorithms. These are accomplished by statistical software, the results of which are reflected in Table 7.

As mentioned earlier, the algorithms presented in this investigation are decision-making algorithms resulted from the rough-set analysis and stepwise regression method. As indicated in Table 7, the degree of difference within the mean values of R^2 obtained from the multiple linear regression of decision-making algorithms of a rough-set analysis and the stepwise method is about 5.16%.

Table 6. Stepwise regression equations.

Step	Parameters	Equations	R^2
1	Swell	severity $= 0.548 + 0.668\ g$	50%
2	Slippage cracking	severity $= -0.300 + 0.486\ g + 0.378\ f$	62.5%
3	Corrugation	severity $= -1.40 + 0.373\ g + 0.505\ f + 0.365\ j$	73.5%
4	Lane/shoulder drop off or heave	severity $= -1.79 + 0.377\ g + 0.477\ f + 0.376\ j + 0.301\ e$	80.2%

Table 7. Coefficient of determination, quality and accuracy of approximation of different algorithms.

Algorithm	R^2	Accuracy of Approx.	Quality of Approx.
$\{a, b, f, g, j\}^*$	75.5%	1	1
$\{a, b, f, j, k\}^*$	67.1%	1	1
$\{a, e, f, g, j\}^*$	80.6%	1	1
$\{a, f, g, j, k\}^*$	75.8%	1	1
$\{a, f, h, j, k\}^*$	69%	1	1
$\{e, f, g, j, k\}^*$	82.2%	1	1
$\{e, f, g, j\}^{**}$	80.2%	0.66	0.8

* The shortest decision-making algorithms resulted from rough-set analysis

** The algorithm resulted from stepwise regression analysis

a = *Upheaval & shoving,*

b = *Block cracking,*

c = *Alligator or fatigue cracking*

d = *Longitudinal and transverse cracking,*

e = *Lane/shoulder drop off or heave,*

f = *slippage cracking,*

g = *swell,* h = *potholes,* i = *rutting,*

j = *corrugation,* k = *Raveling and weathering,* l = *depression*

The R^2 value resulted from the linear regression of all parameters (each 15 parameters) is 85.3%.

In the second part, to evaluate the results of the stepwise method, the values of the accuracy and quality of approximation have been studied. These values represent how precisely the independent defined parameters, regarding the algorithms, can predict the dependent parameter. In other words, they represent the accuracy and quality of the algorithms studied in the analysis of pavement distresses and their failure characteristics. In essence, although the stepwise algorithm studied confirms the accuracy and quality of approximation resulted from the rough-set analysis, the differences between these values with the algorithms concerning the rough-set analysis are 34% and 20%, respectively.

Therefore, the three important indicators (Coefficient of Determination, Accuracy and Quality of Approximation) studied in this research are true for the reducts resulted from the rough-set analysis. First of all, the multiple linear regression equations resulted from these decision-making algorithms confirm the desirable values of R^2, to estimate the severity of pavement distresses and to explain the reasons for the pavement distress. Second, the above-mentioned algorithms confirm the highest values of accuracy and quality of approximation in the processing of information and the evaluation of pavement distresses severity. Consequently, reducts resulted from the presented rough-set analysis can better explain the differences in pavement distresses. So, these algorithms can be considered as the most confident in the information process.

A comparison is accomplished with the studies of Attoh-Okine [13]. He defined 7 parameters in order to express and describe the existing conditions of the pavement, but in the present investigation, 12 parameters are verified. Also, in Attoh-Okine's information table, 21 samples were examined, while in this study there are 35 and this indicates that this survey is more precise. In addition, in Attoh-Okine's studies, there is no comparison to old methods, but algorithms gained in this study by the rough-set theory are also analyzed by a linear regression method.

CONCLUSIONS

The stepwise regression method is used for the evaluation of rough-set analysis outputs in determination of the most important parameters affecting severity of pavement distresses. Meanwhile, the correlation coefficient for each multiple linear regression equation concerning the reducts of the rough-set analysis, is evaluated. In this process, it is observed that the mean difference within the values of R^2 in algorithms concerning the trough-set analysis and stepwise regression method is negligible. Therefore, it can be concluded that rough-set output reducts, in addition to their high values of accuracy and quality of approximation in processing the information and analyzing the failure characteristics, satisfy the desirable values of the correlation coefficient with the severity of pavement distresses. But, the set of parameters observed from the stepwise regression analysis satisfies the lower values of accuracy and quality of approximation. Consequently, it can be shown that the reducts resulted from the rough-set analysis are the most confident decision-making algorithms in the information processing and severity evaluation of typical pavement distresses. Thus, by using these algorithms in the severity evaluation of pavement distresses, the large amount of information necessary for previous decision-making criteria can be diminished. Also, the speed and effectiveness of information processing and the accuracy of decision-making on specified subjects are increased considerably.

The rough-set theory satisfys an excellent capability to detect the shortest possible decision-making algorithms regarding the total collection of defined parameters in the information system. Consequently, this theory can be used as a powerful tool and can simplify different sorts of decision-making problems. However, it is noticed that the exact and subtle detection of decision-making algorithms requires comprehensive and precise studies accomplished by a crew of well-versed and experienced experts.

REFERENCES

1. Smith, R.E., Darter, M.I. and Herrin, S.M. "Highway pavement distress identification manual", *Federal Highway Administration* (1979).
2. Yoder, E.J. and Witzak, M.W., *Principle of Pavement Design*, John Wiley and Sons (1975).
3. Huang, Y.H., *Pavement Analysis and Design*, Prentic-Hall, Inc. (1993).
4. American Association of State Highway Official, AASHTO, *Guide for Design of Pavement Structures*, AASHTO Committee on Design (1993).
5. Shahin, M.Y. and Walther, J.A., *Refinements of PCI Calculation Procedure*, U.S. Army Construction Engineering Research Laboratory, Champing IL (1994).
6. American Public Works Association, *Pavement Condition Index (PCI)*, Field Manual (July 1984).
7. FWA, T.F. "Remaining-life consideration in pavement overlay design", *Journal of Transportation Engineering*, **117**(6), pp. 585-601 (1991).
8. Attoh-Okine, N.O. "Combining use of rough-set and artifical neural networks in doweled-pavement-performance modeling", *Journal of Transportation Engineering, ASCE*, May/June, pp. 270-275 (2002).
9. Pawlak, Z., *Rough-Sets: Theoretical Aspects of Reasoning about Data*, Boston: Kluwer Academic (1991).
10. Skowron, A. "Extracting laws from decision tables", *J. Compute. Intelligence*, **11**(2), pp. 371-388 (1995).
11. Tumotot, S. "Automated extraction of medical expert systems rules from clinical databases based on rough-set theory", *Information Science*, **112**, pp. 67-84 (1998).
12. Arabani, M. and Lashteh Neshaei, M.A. "Application of rough-set theory as a new approach to simplify dams location", *Scientia Iranica*, **13**(2), pp. 152-158 (2006).
13. Attoh-Okine, N.O. "Rough-set application to data mining principles in pavement management database", *Journal of Computing in Civil Engineering*, **11**(4), pp. 231-237 (1997).
14. Kryszkiewicz, M. "Properties of incomplete information systems in the framework of rough-sets", *Rough-Set and Knowledge Discovery*, L. Polliowski and A. Skowron, Eds., pp. 422-450 (1998).
15. Kryszkiewicz, M. "Rules in incomplete information systems", *Information Science*, **113**, pp. 271-292 (1999).
16. Van den Bergh, J.C., Button, K., Nijkam, P. and Pepping, G., *Meta-Analysis of Environmental Policies*, Dordrecht, The Netherlands, Kluwer (1997).
17. Slowinski, R. and Stefanowski, J. "Handing various types of uncertainty in the rough-set approach", *Rough-Sets, Fuzzy Sets and Knowledge Discovery*, W.P. Ziarko, Ed., Berlin, Germany, Springer (1994).
18. Greco, S., Matarazzo, B. and Slowinski, R. "Rough-set approach to multi-attribute choice and ranking problems", ICS Research Report 38/95, Warsaw University of Technology, Warsawm (1995).
19. Predki, B. and Wilk, S. "Rough-set based data exploration using ROSE system", in *Foundations of Intelligent Systems*, Z.W. Ras and A. Skowron, Eds., Lecture Notes in Artificial Intelligence, Berlin, Springer-Verlag, 1609, pp. 172-180 (1999).
20. Nijkamp, P. and Pepping, G. "Meta-analysis for explaining the variance in public transport demand elasticities in Europe", *Journal of Transportation and Statistics*, **1**(1), pp. 1-14 (January 1998).