**Communications in Mathematics and Applications**
Vol. 14, No. 3, pp. 1255–1262, 2023
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
http://www.rgnpublications.com
DOI: 10.26713/cma.v14i3.2412

***Special Issue***
**Recent Trends in Mathematics and Applications**
**Proceedings of the International Conference of Gwalior Academy of Mathematical Sciences 2022**
***Editors*: Vinod P. Saxena and Leena Sharma**

Research Article

# Rough Set Based Decision Support for Feature Extraction of Rice Data

Hemlata Saxena[1], Leena Sharma[2] and Minakshi Panchal*[1,2]

[1] *Department of Mathematics, Career Point University, Kota, Rajasthan, India*
[2] *Pimpri Chinchwad College of Engineering, Nigdi Pradhikarn, Pune, Maharashtra, India*
***Corresponding author:** minakshi.panchal@pccoepune.org

**Received:** February 16, 2023 **Accepted:** June 11, 2023

**Abstract.** The use of technology in the agriculture sector makes it more productive. Different technologies are used for quality control, classification, and prediction of grains in agriculture. It is challenging for decision-making with big agricultural data to concentrate more when many features are given in the data. It is difficult for users to scrutinize the market's excellent quality rice. Determining rice quality by the visual judgment of human inspectors is neither practical nor reliable. Therefore, a proven methodology is essential for the rice quality classifying system, which will overcome the manual quality classification process. In this study, the concept of *Rough Set Theory* is applied to find a set of minimal attributes and generate a set of decision rules for predicting rice type.

**Keywords.** Rough sets, Decision rules, Feature extraction, Decision support

**Mathematics Subject Classification (2020).** Primary 94A16, Secondary 68T09



## 1. Introduction

Rice, the world's most important food crop, is the primary food of nearly 50% of people on the globe. Nutrition, economy, employment, culture, and history are all these areas where it plays a significant role. It is one of the most consumed cereal grains. However, the seed quality impacts crop yield and productivity a lot. Thus, quality seed selection is crucial in guaranteeing the seed quality for a healthy and robust crop. The primary purpose of the proposed approach is to offer a solution for quality control and analysis, which reduces the required effort to classify the rice seeds.

The present study used data that includes different features of rice grains like major axis, minor axis, eccentricity, length, and breadth, just to name a few. Dimensionally, big data contains valuable hidden information which is to be discovered. In this paper, too, such voluminous data is considered. This research uses five different varieties of rice: Basmati, Jasmine, Arborio, Ipsala, and Karacadag, often grown in Turkey. For classification, 12 morphological and four shape features of rice grain are considered.

## 2. Related Work

An effective and efficient method to classify the agricultural data is proposed using A rough set-based ensemble learning algorithm. It is proven to be the best method for accuracy in the experimental comparison of different algorithms conducted on four agricultural datasets (Shi *et al*. [12]). In order to expedite the rough set and increase its accuracy in compared to other accelerating techniques, the theorem for the stability of the local redundancy of attributes and the model of relative importance are provided (Xia *et al*. [15]). The rough set theory-based attribute selection technique may efficiently minimize the time complexity, its associated domains have significantly advanced in recent decades (Chen *et al*. [1]). It has been demonstrated that pattern dimensionality can be greatly reduced using rough set approaches. They have demonstrated their viability as data mining methods in front of neural network classifiers (Swiniarski and Skowron [14]). Rough sets can be combined with other methods when they alone cannot produce better results. Some of the hybridizations of rough sets with benchmark techniques like neural networks, fuzzy sets, and evolutionary processes are presented with state-of-the-art therein (Saxena *et al*. [11]). An algorithm with the modification in (LEM2) is proposed, leading to the formation of essential rules from agricultural data, which is useful to the farmers of the state to build their farming plans, allowing them to crop production (Sabu and Raju [10]).

The concept of rough set theory is applied to data of Portuguese banking institution to reduce the dimensionality of data and to predict the deposit nature of the customer, and it is observed that it give 905 accuracy (Sumalatha *et al*. [13]).

To describe the property of the rule set in general, evaluation metrics of the rules set can be used. The single-rule evaluation metric measures of an individual rule (Chen *et al*. [2]). Three-way decisions, i.e., positive, boundary, and negative rules, are offered as a different approach to interpreting rules in rough set theory. Statistical analysis techniques can benefit from rough set theory (Yao [17]). Rough set theory is applied in the medical field for feature reduction, like only two parameters are sufficient in different therapy-related groups of PD patients (Przybyszewski *et al*. [9]). An irregular set model of weighted neighborhood probabilistic rough sets is proposed, and it is proven more competitive than other existing attribute reduction algorithms (Xie *et al*. [16]).

## 3. Mathematical Background

Z. Pawlak [6] introduced the concept of *Rough sets*. It is to formalize a crisp set. Rough set theory has become a major database mining or knowledge discovery methodology. It is a very effective technique to handle uncertainty, and vagueness of uncertainty as one of the options for fuzzy theory. Structural linkages can be found in imprecise and noisy data using a rough set technique. The following are the fundamental ideas of the rough sets:

**Information system.** An information table is also a decision table; it contains two types of attributes: conditional and decisional attributes. The decision system can be represented as

$$I = (U, A \cup \{d\}).$$

Here $d$ denotes the decision attribute (we can think about multiple decision attributes rather than just one). $U$ is a non-empty, finite set of objects (the universe), and $A$ is a non-empty, finite set of characteristics such that it is a decision attribute (instead of one, we can consider more decision attributes).

**Indiscernibility.** The decision table may contain many objects having the same attributes as features. By choosing just one representative object for each collection of objects with the same attributes, the rough set theory provides a way to decrease the size of tables. These things are referred to as tuples or indiscernible objects.

**Approximations.** Two sets — a lower approximation of $C$, and an upper approximation of $C$ — approximate a rough set definition for a given class, $C$.

All the tuples that, given the attributes, cannot be classified as not belonging to $C$ make up the upper approximation of $C$.

$$R^*X = \cup\{Y \in U/R : Y \cap X \neq \emptyset\}.$$

**Lower approximation.** All the data tuples that, given the attributes, can be used to determine with certainty that they belong toC without any doubt make up the lower approximation of $C$.

$$R_*X = \cup\{Y \in U/R : Y \subseteq X\}.$$

If a set's boundary area is not empty, it is said to be rough, otherwise, it is said to be crisp.

**Accurateness of the approximate**:
Two numerical measurements are suggested by Pawlak [7] to describe the imprecision of rough set approximations.

$$\alpha(X) = \frac{|R_*(X)|}{|R^*(X)|},$$

where $X \neq \emptyset$ and $|\cdot|$ denotes the cardinality of a set. For the empty set $\emptyset$, we define $\alpha(\emptyset) = 1$. It follows that. Based on the accuracy measure, the roughness measure is defined by,

$$\rho(X) = 1 - \alpha(X).$$

## 4. Methodology

*Rough sets* are applied to a dataset of rice grains to extract features, and decision rules for the classification of rice grains is discussed. The dataset is considered from the Kaggle website https://www.kaggle.com/. Data contains 12 morphological features and four shape features of rice grain. Table 1 provides details about the data attribute.

**Table 1.** Details of the attributes

| S. No. | Attribute | Details |
|---|---|---|
| 1 | Area | Number of pixels within the borders of the area of rice seed |
| 2 | Perimeter | Perimeter border length of rice seed |
| 3 | Major axis length | The maximum length of rice seed |
| 4 | Minor axis length | Maximum length perpendicular to the major axis |
| 5 | Eccentricity | The eccentricity of the circle, which has the same moments as the region |
| 6 | Equivalent diameter | The diameter of a circle of rice seed |
| 7 | Solidity | The ratio of pixels in the convex stem to pixels in the rice seed |
| 8 | Convex area | Number of pixels in the minimum convex polygon rice seed area |
| 9 | Extent | Bounding is the ratio of pixels in the box to pixels in the rice seed |
| 10 | Aspect ratio | Ratio of major axis length by the minor axis length |
| 11 | Roundness | Considered from area and the perimeter |
| 12 | Compactness | Ratio of the equivalent diameter by the length of the major axis |
| 13 | Shape factor | 4 shape features are calculated using area, major axis, and minor axis lengths from morpho-logical features |

The decisional table contains two types of attributes: conditional and decisional attributes. Here we will consider 16 features of the rice grain as conditional attributes and rice type as decisional attribute.

## 5. Reduct and Core

The reduction of data is *rough sets*' primary characteristic. From a set of original data, rough sets enable the extraction of a subset of portions that is sufficient to describe the knowledge in the database. A reduct is the name for this condensed set of features. Assigning significance value to each feature and choosing the characteristics with higher values is the approach for choosing significant features. We will develop a set of rules that will be applied to extract knowledge from the decisional table based on the developed reduct system. Reductions from a decisional table are computed using R software and a discernibility matrix. The set of all singleton elements in the discernibility matrix represents the core. The outcomes are as:

**Reduct-1:** A feature subset consisting of 11 attributes:
{Area, Perimeter, Major-Axis, Eqdiasq, Solidity, Extent, Aspect-Ratio, Roundness, Shapefactor_1, Shapefactor_3, Shapefactor_4}

**Reduct-2:** A feature subset consisting of 11 attributes:
{Perimeter, Major-Axis, Eqdiasq, Solidity, Convex-Area, Extent, Aspect-Ratio, Roundness, Shapefactor_1, Shapefactor_3, Shapefactor_4}

**Reduct-3:** A feature subset consisting of 12 attributes:
{Area, Perimeter, Major-Axis, Eqdiasq, Solidity, Extent, Aspect-Ratio, Roundness, Compactness, Shapefactor_1, Shapefactor_2, Shapefactor_4}

**Reduct-4:** A feature subset consisting of 12 attributes:
{Perimeter, Major-Axis, Eqdiasq, Solidity, Convex-Area, Extent, Aspect -Ratio, Roundness, Compactness, Shapefactor_1, Shapefactor_2, Shapefactor_4}

**Core:**
{Aspect-Ratio, Eqdiasq, Extent, Major Axis, Perimeter, Roundness, Shapefactor_1, Shapefactor-4, Solidity}

## 6. Rules

The decisional rule will be crucial in identifying any hidden patterns in the data as well as in categorizing new data. The dependencies in the characteristics that can be utilized to categorize test data are represented by rules. The Rules are derived from the set of reducts. Decisional rules are straightforward IF-THEN clauses connected by the antecedent. The final decisional attribute is obtained by connecting the pertinent features from the collection of reducts in the decisional rules. The rigid decisional rules serve as a representation of the classifier. This will be employed to forecast the class types of new objects.

There are numerous methods for choosing the best decision rules. The rule with the condition component that differs from the re-feature image's vector by the fewest features is deemed the most appropriate. The polling method is applied to choose the decisional value when there are many applicable rules. All applicable rules will be considered for decisional values, which are equal to the number of objects the rule matched. In summative, all votes are taken into account, and the choice with the most votes is determined to be the appropriate class type. There are several approaches to get rid of classification rules that are not required.

By using rough sets total of 116 decision rules are generated, out of which 22 rules with good Laplace value and support value is considered. A list of 22 rules based on a good Laplace value is presented in Table 2.

**Table 2.** Support size and Laplace value of decision rules

| S. No. | Rule | Support size | Laplace |
|---|---|---|---|
| 1 | If shapefactor_1 is 4, and shapefactor_4 is 3, and major axis is 2 then class is 3 | 18 | 0.8261 |
| 2 | If major axis is 2, and shapefactor_3 is 1 then class is 3 | 37 | 0.9048 |
| 3 | If shapefactor_1 is 4, and shapefactor_2 is 3, and solidity is 2 then class is 3 | 15 | 0.8 |

(Table *Contd.*)

| S. No. | Rule | Support size | Laplace |
|---|---|---|---|
| 4 | If aspect ratio is 3, and convex area is 3, and solidity is 3 then class is 3 | 23 | 0.82571 |
| 5 | If roundness is 2, and shapefactor_4 is 1, and area is 3 then class is 3 | 15 | 0.8 |
| 6 | If shapefactor_3 is 2, and minor axis is 2, and shapefactor_2 is 1 then class is 3 | 19 | 0.8333 |
| 7 | If shapefactor_1 is 4, and eqdiasq is 1 then class is 3 | 54 | 0.9322 |
| 8 | If compactness is 2, and shapefactor_2 is 3, and perimeter is 2, and shapefactor_1 is 3 then class is 3 | 25 | 0.8667 |
| 9 | If perimeter is 3, and aspect-ratio is 2, and solidity is 3 then class is 2 | 19 | 0.8333 |
| 10 | If aspect-ratio is 2, and eqdiasq is 4 then class is 2 | 97 | 0.9608 |
| 11 | If area is 4 and solidity is 2 then class is 2 | 19 | 0.8333 |
| 12 | If roundness is 3, and major-axis is 3 then class is 2 | 68 | 0.9452 |
| 13 | If shapefactor_1 is 1 then class is 2 | 104 | 0.9633 |
| 14 | If convex-area is 2, and shapefactor_3 is 3, and eccentricity is 3 then class is 4 | 14 | 0.7895 |
| 15 | If compactness is 3, and major-axis is 2, and eccentricity is 3 then class is 4 | 16 | 0.8095 |
| 16 | If perimeter is 3, and minor-axis is 1 then class is 5 | 26 | 0.871 |
| 17 | If perimeter is 4, and minor-axis is 1 then class is 5 | 15 | 0.8 |
| 18 | If major-axis is 4, and eqdiasq is 2 then class is 5 | 21 | 0.8462 |
| 19 | If shapefactor_3 is 1, and eqdiasq is 2, and major-axis is 3, and solidity is 2 then class is 5 | 18 | 0.8261 |
| 20 | If aspect-ratio is 5 then class is 5 | 23 | 0.857 |
| 21 | If eqdiasq is 3, and shapefactor_3 is 2, and major-axis is 4 then class is 3 | 20 | 0.84 |
| 22 | If convex-area is 1, and shapefactor_2 is 3 then class is 3 | 25 | 0.8667 |

The support value represents the randomness of the decision rule. Rules with small values show randomness of the rule, and it is not considered a good classifier. In the same way, good values of Laplace reflects the good power of classification.

## 7. Conclusion

In this study, we used rough sets to develop a set of classification rules for rice grains. The dataset has 16 morphological and morph metric characteristics of rice, categorized as conditional attributes in the decision table. For classification, data includes 5 groups of rice. Four reductions are performed using indiscernibility relations, which take advantage of the unique property of rough sets to discover the smallest subset to describe the original data. Eleven qualities make up each reduct. Out of the 116 decision rules that are calculated for the decision support system, 22 essential rules are taken into account based on Laplace and support value. It is possible to classify data on rice grains effectively using these 22 fundamental principles.

## Competing Interests

The authors declare that they have no competing interests.

## Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

## References


[1] H. Chen, T. Li, C. Luo, S.-J. Horng and G. Wang, A rough set-based method for updating decision rules on attribute values' coarsening and refining, *IEEE Transactions on Knowledge and Data Engineering* **26**(12) (2014), 2886 – 2899, DOI: 10.1109/TKDE.2014.2320740.

[2] S. Chen, Z. Jiang, C. Duan, X. Dai and X. Mi, A decision analysis method based on rough set, in: *Proceedings of the 2018 International Conference on Transportation & Logistics, Information & Communication*, Smart City (TLICSC 2018), *Advances in Intelligent Systems Research series*, Vol. **161**, 50 – 56, DOI: 10.2991/tlicsc-18.2018.8.

[3] T. Dimitrovski, D. Andreevska, D. Andov, E. Simeonovska and J. Ibraim, Some seed quality properties of newly introduced Turkish rice varieties (Oryza sativa L.) grown under the environmental conditions of republic of Macedonia, *Congress Book*, 265 – 270 (2017).

[4] S. Mitra and H. Banka, Application of rough sets in pattern recognition, in: *Transactions on Rough Sets VII*, J. F. Peters, A. Skowron, V. W. Marek, E. Orłowska, R. Słowiński and W. Ziarko (editors), Lecture Notes in Computer Science, Vol. **4400**, Springer, Berlin — Heidelberg (2007), DOI: 10.1007/978-3-540-71663-1_10.

[5] H. Patel and D. Patel, Crop prediction framework using rough set theory, *International Journal of Engineering and Technology* **9**(3) (2017), 2505 – 2513, DOI: 10.21817/ijet/2017/v9i3/1709030266.

[6] Z. Pawlak, Rough sets, *International Journal of Computer & Information Sciences* **11** (1982), 341 – 356, DOI: 10.1007/BF01001956.

[7] Z. Pawlak and A. Skowron, Rudiments of rough sets, *Information Sciences* **177**(1) (2007), 3 – 27, DOI: 10.1016/j.ins.2006.06.003.

[8] X. Peng, J. Wen, Z. Li, G. Yang, C. Zhou, A. Reid, D. M. Hepburn, M. D. Judd and W. H. Siew, Rough set theory applied to pattern recognition of partial discharge in noise affected cable data, *IEEE Transactions on Dielectrics and Electrical Insulation* **24**(1) (2017), 147 – 156, DOI: 10.1109/TDEI.2016.006060.

[9] A. W. Przybyszewski, S. Szlufik, P. Habela and D. M. Koziorowski, Rough set rules determine disease progressions in different groups of parkinson's patients, in: *Pattern Recognition and Machine Intelligence* (PReMI 2017), B. Shankar, K. Ghosh, D. Mandal, S. Ray, D. Zhang and S. Pal (eds), Lecture Notes in Computer Science, Vol. **10597**, Springer, Cham. (2017), DOI: 10.1007/978-3-319-69900-4_34.

[10] M. K. Sabu and G. Raju, Rule induction using rough set theory - An application in agriculture, in: *2011 International Conference on Computer, Communication and Electrical Technology* (ICCCET), Tirunelveli, India, pp. 45 – 49, (2011), DOI: 10.1109/ICCCET.2011.5762519.

[11] A. Saxena, L. K. Gavel and M. M. Shrivas, Rough sets for feature selection and classification: An overview with applications, *International Journal of Recent Technology and Engineering* **3**(5) (2014), 62 – 70, URL: https://www.ijrte.org/wp-content/uploads/papers/v3i5/E1269113514.pdf.

**[12]** L. Shi, Q. Duan, J. Zhang, L. Xi, H. Qiao and X. Ma, Rough set based ensemble learning algorithm for agricultural data classification, *Filomat* **32**(5) (2018), 1917 – 1930, DOI: 10.2298/FIL1805917S.

**[13]** L. Sumalatha, P. U. Sankar and B. Sujatha, Rough set based decision rule generation to find behavioural patterns of customers, *Sādhanā* **41** (2016), 985 – 991, DOI: 10.1007/s12046-016-0528-1.

**[14]** R. W. Swiniarski and A. Skowron, Rough set methods in feature selection and recognition, *Pattern Recognition Letters* **24**(6) (2003), 833 – 849, DOI: 10.1016/S0167-8655(02)00196-4.

**[15]** S. Xia, X. Bai, G. Wang, Y. Cheng, D. Meng, X. Gao, Y. Zhai and E. Giem, An efficient and accurate rough set for feature selection, classification, and knowledge representation, *IEEE Transactions on Knowledge and Data Engineering* **35**(8) (2023), 7724 – 7735, DOI: 10.1109/TKDE.2022.3220200.

**[16]** J. Xie, B. Q. Hu and H. Jiang, A novel method to attribute reduction based on weighted neighborhood probabilistic rough sets, *International Journal of Approximate Reasoning* **144** (2022), 1 – 17, DOI: 10.1016/j.ijar.2022.01.010.

**[17]** Y. Yao, Three-way decision: an interpretation of rules in rough set theory, in: *Rough Sets and Knowledge Technology* (RSKT 2009), P. Wen, Y. Li, L. Polkowski, Y. Yao, S. Tsumoto and G. Wang (editors), *Lecture Notes in Computer Science*, Vol. **5589**, Springer, Berlin — Heidelberg (2009), DOI: 10.1007/978-3-642-02962-2_81.