International Journal of Emerging Scientific Research 4 (2022) 15 – 23

International Journal of Emerging Scientific Research (IJESR)

Journal homepage: www.sciengtexopen.org/index.php/ijesr

A Rough Set Approach to Dimensionality Reduction for Performance Enhancement in Machine Learning

M. O. Odighi [a,*], J. O. Olaniyan [b], R. E. Imhanlahimi [c]

[a] *Department of Computer Science, Auchi Polytechnic, Auchi, Nigeria*
[b] *Department of Computer Science, Landmark University, Omu- Aran, Nigeria*
[c] *Department of Computer Science, Ambrose Alli University, Ekpoma, Nigeria*

Authors' Contributions
This work was carried out in collaboration among all authors. The authors have read and approved the revised manuscript.

ARTICLE INFO

Article history:
Received 19 May 2022

Received in revised form
25 August 2022

Accepted 28 September 2022

Available online
2 October 2022

Keywords:
Core
Dimensional space
Dimensionality reduction
Intrinsic dimension
Machine learning
Reducts

ABSTRACT

Machine learning uses complex mathematical algorithms to turn data set into a model for a problem domain. Analysing high dimensional data in their raw form usually causes computational overhead because the higher the size of the data, the higher the time it takes to process it. Therefore, there is a need for a more robust dimensionality reduction approach, among other existing methods, for feature projection (extraction) and selection from data set, which can be passed to a machine learning algorithm for optimal performance. This paper presents a generic mathematical approach for transforming data from a high dimensional space to low dimensional space in such a manner that the intrinsic dimension of the original data is preserved using the concept of indiscernibility, reducts, and the core of the rough set theory. The flue detection dataset available on the Kaggle website was used in this research for demonstration purposes. The original and reduced datasets were tested using a logistic regression machine learning algorithm yielding the same accuracy of 97% with a training time of 25 min and 11 min respectively.

1. Introduction

In this era of big data in which, nearly every twenty months, the magnitude of data and information from various sources in the world gets doubled [1]. Due to this data explosion, there is a need for this immense data to be put to use judiciously, which is why a lot of research scholars are working tirelessly to come up with sophisticated ways for efficient information retrieval and processing. Machine learning is a field of artificial intelligence that deals with building applications, which learn from data and enhance their accuracy over time without being programmed to do so. The main function of a machine learning algorithm is to transform datasets into useful models depending on the type of data and problem domain. In most cases, raw data is always very noisy, making it difficult to use it directly by machine learning algorithms. Hence, modelling requires the selection of input parameter vectors and the corresponding output vectors that sufficiently characterize the component to be modelled [2].

* Corresponding author
E-mail address: odighimatthew@gmail.com
https://doi.org/10.37121/ijesr.v4.192

For optimal performance of any chosen machine learning algorithm, the input sample must be selected in a manner that will weave a domain to cover the minimum model parameter of interest excluding redundancy [1]. This is what is referred to as dimensionality reduction [3]; an essential pre-processing step in preparing data for the machine learning algorithm. It reduces computational intricacy and amends statistical ill-conditioning by rejecting redundant features that may be contained in the raw dataset [4]. Dimensionality reduction also referred to as dimension reduction can transform data from a high-dimensional to a low-dimensional space [5]; the low-dimensional representation can retain some important attributes (ideally close to its intrinsic dimension) of the original data. Working in high-dimensional spaces can be cumbersome and unyielding; for example, raw data are often scarce as a result of the curse of dimensionality [6], which implies that with the increase in the number of features, the machine learning model tends to yield more errors. Also, analysing the data is generally computationally difficult. Dimensionality reduction is commonly employed in fields which deal with large numbers of observations or variables, such as speech recognition, signal processing, bioinformatics, and neuroinformatics, among others [7].

In general, dimensionality reduction techniques strive to locate a reduced number of new dimensions to account for the original data. Some methods are available such as Autoencoder, Generalized discriminant analysis (GDA), Kernel Principal Component Analysis (PCA), Linear discriminant analysis (LDA), Non-negative matrix factorization (NMF), T-distributed Stochastic Neighbour Embedding (t-SNE), and Uniform manifold approximation and projection (UMAP), which are variants of factors analysis to find a smaller set of representative dimensions [8,9]. The PCA practically proves to be the best known of these methods in which the new dimensions, linear combinations of the original features, are given by the eigenvectors (ordered by decreasing eigenvalue) of the covariance matrix of input data. The new features, known as the principal components, are uncorrelated and of maximum variance; thus, the new representation is now minimal. Successive components are of declining importance, and the variance in the input data is accounted for by the first principal components (with higher eigenvalue). However, the size of the covariance matrix is very large for high-dimension data vectors, as input vectors of dimension n give rise to a matrix of size n x n, thus, standard PCA methods cannot deal with data with a huge number of features, because of the computational costs associated with limited space and time [10].

The rough set approach presented by [2] is another technique to extract patterns from data for classification [11]. The rough set theory (RST) is a recognized approximation of a crisp set in terms of a pair of sets that give the lower and the upper approximation of the original set [12] and can search large, multifaceted databases for deep decision rules [13]. This theory deals with the classificatory analysis of data tables. It has been widely employed in several areas comprising expert systems, machine learning, and knowledge discovery [14]. The overall objective of the rough set analysis is to synthesize an approximation of concepts from the acquired and reduced data to a minimal representation [15]. This concept of set approximation has led to practical forms of information granularity, which has become an essential part of computational intelligence [16].

This paper aims to present a generic mathematical model for data reduction using a flue detection dataset on Kaggle and apply indiscernibility, reducts, and core concepts of the rough set approach to both the original and the reduced data. Section 2 presents some reviewed literature related to the rough set approach to dimensionality reduction. Then, sections 3 and 4 discuss the information and decision systems respectively. Moreover, section 5 explains the concept of indiscernibility, while section 6 presents the mathematical representation of reducts and the core of an information/decision system, and section 7 concludes the paper.

2. Related Literature

The study of a rough set-based dimensionality reduction was carried out by Qiang and Alexios [17]. This study investigated the efficacy of the rough set attribute reduction (RSAR) method in supervised and unsupervised learning to probe RSAR's generality. The study described the three methods in question, discussed how RSAR could be used with a supervised or an unsupervised system and used experimental results to conclude the relative success of the two integration efforts.

The RST appears to be a valuable tool for inductive learning and useful aid for the development of robust expert systems [18]. The authors in [15], proposed a network intrusion detection system based on a rough set and k-nearest neighbours (kNN). In their paper, two machine learning methods (kNN and rough set (LEM2 algorithm)) were used for intrusion detection while an experimental study was conducted on the international knowledge discovery and data mining tools competition (KDD) dataset for benchmarking intrusion detection systems, and their performances were compared. The results revealed that kNN has a better performance in terms of accuracy but consumes more memory and computational time while rough sets classify for a relatively short time and employ simple explainable rules.

Srivastava [14], proposed a hybrid approach to data classification by combining both the rough set technique with the support vector machine (SVM) approach. Their paper reported the introduction of a rough-SVM approach based on the integration of the rough set exploration system (RSES) and SVM. The RSES was utilized to capture reducts while SVM was applied to obtain better classification results.

The authors in [8] proposed a novel feature selection approach based on the fuzzy forward and backward reduct in the year 2012. The feature selection mechanism proposed also involved an entropy-based modification of the original rough set-based method, which was employed to the challenges associated with finding minimal rough set reducts and evaluated experimentally. And the result of the experimental evaluation proved to be effective as the reduced dataset would produce the same results as the original dataset.

A sentiment classification exploiting rough set-based hybrid feature selection was proposed by [19]. The hybrid feature selection technique based on RST and information gain was introduced for sentiment classification. The introduced techniques were evaluated on four standard datasets comprising review datasets, movie reviews, and products. Results suggest that the hybrid feature selection approach performs better than other feature selection techniques for sentiment classification with accuracy of 98% and 92%, precision of 97% and 88%, Recall of 98% and 91%, and F1 score of 96% and 83% respectively.

Medical domain-based features selection was proposed by [4] in which a rough set reduct algorithm was proposed. In their paper, they proposed the application of this algorithm to real-time data that tend to increase dynamically in size believing that it would effectively and efficiently handle the problem associated with humongous input features most of which do not impact the classification task at hand.

Aboul and Jafar [10], provided a rough set mechanism for producing classification rules from a set of observed 360 breast cancer data samples. The rough set dependency rules were generated directly from the real value attribute vector after all the data attributes had been chosen, normalized, and standardized. A minimal subset of features linked with a class label for classification was found in all reductions of the data using the rough set reduction method. Experimental results from the application of the rough set analysis to the set of data samples were given and evaluated. Also, the generated rules were compared to the well-known IDS classifier algorithm.

Chaturvedi et al. [11] discussed in detail, the application of the RST in medical health care data analytics. In their article, an attempt to summarize the basic concepts, characteristics of RST, some evolutionary extensions of RST, and applications limited to medical data analysis was made. In keeping the view of learners, a survey on RST-based software tools and packages was outlined with their exhaustive functionalities. It also identifies the importance of RST in the domain of medical or clinical data analytics and also exhibits the strengths and limitations of the respective underlying approaches.

A rough set-based feature selection method for the prediction of learning disabilities was proposed by [20]. In the study, as a pre-processing step for classification, significant symptoms were selected from the learning disabilities dataset with the help of an RST-based feature selection technique. The effect of these selected symptoms in the prediction of learning disability was studied using two common classifiers multilayer perceptron and sequential minimal optimizer available in the Weka data mining tool kit and experiments were conducted on the learning disabilities dataset, which demonstrated the effectiveness of the rough set theory in selecting significant symptoms of learning disabilities. The comparison of the result of the training on the original and the reduced data set established the efficiency of the approach to remove non-significant symptoms from the dataset without negatively affecting the classification performance.

Authors in [21], proposed a technique for Hepatitis disease diagnosis using rough set-based feature selection and a kNN classifier. The proposed system included two modules: the feature extraction module and the predictor module. In the feature extraction module, rough set theory was used to pre-process the attributes in such a way that preserved the important information while redundant attributes were removed. Then, the classification was done using the kNN classifier and the experimental results showed that the proposed system could improve the rate of correct and timely diagnosis of Hepatitis.

Al-Shalabi [22] proposed a rough set-based reduction of incomplete medical datasets in 2019. In this work, the dimensionality reduction concept was used to reduce the number of missing values. This was achieved by splitting the original data set into two subsets; the first one contains complete records and the other one contains imputed records that previously have missing values. The reducts of the two subsets based on rough set theory were merged. The reduct of the merged attributes was constructed and tested using rule-based and decomposition tree classifiers.

A rough set theory-based approach to feature selection for incremental data was proposed by Shampa [3]. The technique generated multiple reducts from the incremental dataset for accurate classification of objects and the generated reducts preserved the property of the entire dataset. The method was utilized in the benchmark dataset collected from the UCI repository and experimental results proved the effectiveness of the model being proposed in the work.

A parallel rough set theory-based attribute reduction approach for attribute reduction in big data was also proposed in 2019 by [9]. In this work, rough set lower and upper approximations were constructed. Then a reduct was detected using both inner and outer importance measures. Then, the MapReduce framework was used to achieve the parallelism for attribute reduction in big data.

In 2020, attributes reduction in big data was an article presented by [16] that analysed a large amount of data from a different perspective. A viewpoint was the processing of reduced collections of big data with fewer computing resources, in which 40 GB of data was analysed to test various strategies to reduce data processing. The objective was to reduce this data while retaining the detection and model learning in machine learning. Other alternatives were analysed, and it was found that in many cases and types of settings, data can be reduced to some extent without compromising detection efficiency.

Also in 2020, the dimensionality reduction technique for hyperspectral image analysis based on rough set theory was proposed by [23]. In this work, the hyperspectral image was represented as a decision system, in which some properties were extracted as decision attributes, and based on information entropy, effective features were selected. The performance of the system was then evaluated using two different data sets, which yielded 94% accuracy by the SVM classifier and reduced computing time by 85%.

3. Information System

An information system can be defined as a tabular representation of a data set, in which each row in the data set represents an object called an entity with its associated attributes that define the properties of the object. Such representation is termed an information system or table. Also, an information system is a relation defined as a pair

$$R = (O, A) \tag{1}$$

Where, O is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes such that

$$a : O \rightarrow V\, a \text{ for every } a \in A. \tag{2}$$

The set V a is called the value set of a. Thus, the information system represents input data, gathered from any domain such as the military, finance, geography, education, or medicine. Table 1 presents an information system sampling 12 patients called object or entity denoted by P1, P2, …, P12, and 3 conditional attributes, X1, X2, and X3 of each patient. X1, X2, and X3 represent Headache, Muscle pain, and Temperature respectively. The attribute values of each entity are also captured, with 1 =Yes, 0 = No, 5 = Normal (N), 7 = High (H) and 9 = Very High (VH).

Table 1. Information system.

Patients	Conditional Attributes		
	X1	X2	X3
P1	1	1	5
P2	1	1	7
P3	1	1	9
P4	0	1	5
P5	0	0	7
P6	0	1	9
P7	1	1	5
P8	1	1	7
P9	1	1	9
P10	0	1	5
P11	0	0	7
P12	0	1	9

4. Decision System

A decision system is an information system (e.g., Table 1) with one or more decision attributes. Table 2 shows a decision system sampling Table 1 and an additional attribute (Y) called the decision attribute. It includes a decision attribute Y (Flue), which indicates whether a patient has Flue or not with binary values 0 and 1, where 0 denotes no and 1 is yes.

A formal representation of a decision system/table is given as DS: T= (U, A $\cup${d}) where U is a non-empty finite set of objects termed the universe and A is a non-empty finite set of attributes, and $d \not\in A$ is the decision attribute, which may be more than one while the elements contained in A are referred to as condition attributes.

5. Indiscernibility Relation

Indiscernibility (IND) is a relation between two or more objects in which all the values are identical to a subset of attributes being considered. A decision system can express all the knowledge about the model in context, which at some point makes it unnecessarily large as a result of redundancy. The same or indiscernible objects can be represented several times, or some of the attributes may be superfluous. For example; let $S = (U, A)$ be an information system and $B \subseteq A$. A binary relation *INDs* (B) defined in equation (3) is called the B-indiscernibility relation.

$$IND\ (B) = \{(x, x^1) \in U^2 \mid \forall a \in B\ a(x) = a(x')\} \tag{3}$$

It is easy to see that $IND_S(B)$ is an equivalence relation. If $(x, x^1) \in IND_S(B)$, then objects x and x^1 are indiscernible from each other by attributes from B. The equivalence classes of the B-indiscernibility relation are denoted $[x]_B$. The subscript S in the indiscernibility relation is usually omitted if it is clear which information system is meant. For a better understanding of indiscernibility in a decision system, consider the decision table presented in Table 2. The non-empty subsets of the condition attributes are: {X1}, {X2}, {X3}, {X1,X2}, {X1,X3}, {X2,X3},{X1,X2,X3}.

The indiscernibility of the above partitions can be derived using the concept of intersection in set theory as follows:

IND({X1}) = {{P1,P2,P3,P7,P8,P9}, {P4,P5,P6,P10,P11,P12}} (4)

IND({X2}) = {{P1,P2,P3,P4,P6,P7,P8,P9,P10,P12},{P5,P11}} (5)

IND({X3}) = {{P1,P4,P7,P10},{P2,P5,P8,P11},{P3,P6,P9,P12}} (6)

IND({X1,X2}) = {{P1,P2,P3,P7,P8,P9},{P4,P6,P10,P12},{P5,P11}} (7)

IND({X1,X3}) = {{P1,P7},{P2,P8},{P3,P9},{P4,P10},{P5,P11},{P6,P12}} (8)

IND({X2,X3}) = {{P1,P4,P7,P10},{P2,P8},{P3,P6,P9,P12},{P5,P11}} (9)

IND({X1,X2,X3}) = {{P1,P7},{P2,P8},{P3,P9},{P4,P10},{P5,P11},{P6,P12}} (10)

Table 2. Decision system.

Patients	Conditional Attributes			Decision
	X1	X2	X3	Y
P1	1	1	5	0
P2	1	1	7	1
P3	1	1	9	1
P4	0	1	5	0
P5	0	0	7	0
P6	0	1	9	1
P7	1	1	5	0
P8	1	1	7	1
P9	1	1	9	1
P10	0	1	5	0
P11	0	0	7	0
P12	0	1	9	1

6. Reducts and Core

Data reduction is concerned with the identification of equivalence classes (indiscernible objects), in which one element of the equivalence class is used to denote the entire class [23]. Another aspect of data reduction is keeping only those attributes, which preserve the indiscernibility relation and, as a result, set approximation. The rejected attributes are redundant since their removal will not have any adverse effect on the classification.

Let $S = (U,A)$ be an information system, $B \subseteq A$, and $a \in B$, therefore, it can be said that a is *dispensable* in B if $IND(\{B\}) = IND(\{B\}-\{a\})$; otherwise a is indispensable in B. A set 'B' is called independent if all its attributes are indispensable. Any subset B' of B is called a reduct of B if B' is independent, that is $IND(\{B'\}) = IND(\{B\})$. Hence, a reduct is a set of attributes that preserves partition. It means that a reduct is the minimal subset of attributes that enables the same classification of elements of the universe as the whole set of attributes. In other words, attributes that do not belong to a reduct are redundant concerning the classification of elements of the universe.

Let B be a subset of A. The *core* of B is the set of all indispensable attributes of B. The following is an important property, connecting the concept of the core and reducts:

$$Core(B) = \bigcap Red(B), \tag{11}$$

where, *Red* (B) is the set of all reducts of B.

Since the core is the intersection of all reducts, it is incorporated in every reduct, in other words, each element of the core belongs to some reducts. Thus, in a sense, the core is the most important subset of attributes, for none of its elements can be removed without affecting the classification efficacy of attributes. The following properties are very important for dimensionality reduction:

(a) $S = (U,A)$
(b) $B \subseteq A$
(c) $a \in B$
(d) $IND\ (B) = \{(x, x^1) \in U^2 \mid \forall a \in B\ a(x) = a(x')\}$
(e) $IND(\{B\}) = IND(\{B\}-\{a\}$
(f) $IND(\{B'\}) = IND(\{B\})$
(g) $Core(\{B\}) = n\ Red(\{B\})$

Using the information system presented (Table 2), let us consider the following partitions (to produce Table 3):

$$\text{IND}(\{X1,X2\}) = \{\{P1,P2,P3,P7,P8,P9\},\{P4,P6,P10,P12\},\{P5,P11\}\} \tag{12}$$

$$\text{IND}(\{X1,X3\}) = \{\{P1,P7\},\{P2,P8\},\{P3,P9\},\{P4,P10\},\{P5,P11\},\{P6,P12\}\} \tag{13}$$

$$\text{IND}(\{X2,X3\}) = \{\{P1,P4,P7,P10\},\{P2,P8\},\{P3,P6,P9,P12\},\{P5,P11\}\} \tag{14}$$

$$\text{IND}(\{X1,X2,X3\}) = \{\{P1,P7\},\{P2,P8\},\{P3,P9\},\{P4,P10\},\{P5,P11\},\{P6,P12\}\} \tag{15}$$

Fig. 1 is a representation of the partition of the sample data set presented as a decision system in Table 2.

Table 3. Data information system.

Patients	Conditional Attributes			Decision
	X1	X2	X3	Y
P1	1	1	5	0
P2	1	1	7	1
P3	1	1	9	1
P4	0	1	5	0
P5	0	0	7	0
P6	0	1	9	1
P7	1	1	5	0
P8	1	1	7	1
P9	1	1	9	1
P10	0	1	5	0
P11	0	0	7	0
P12	0	1	9	1

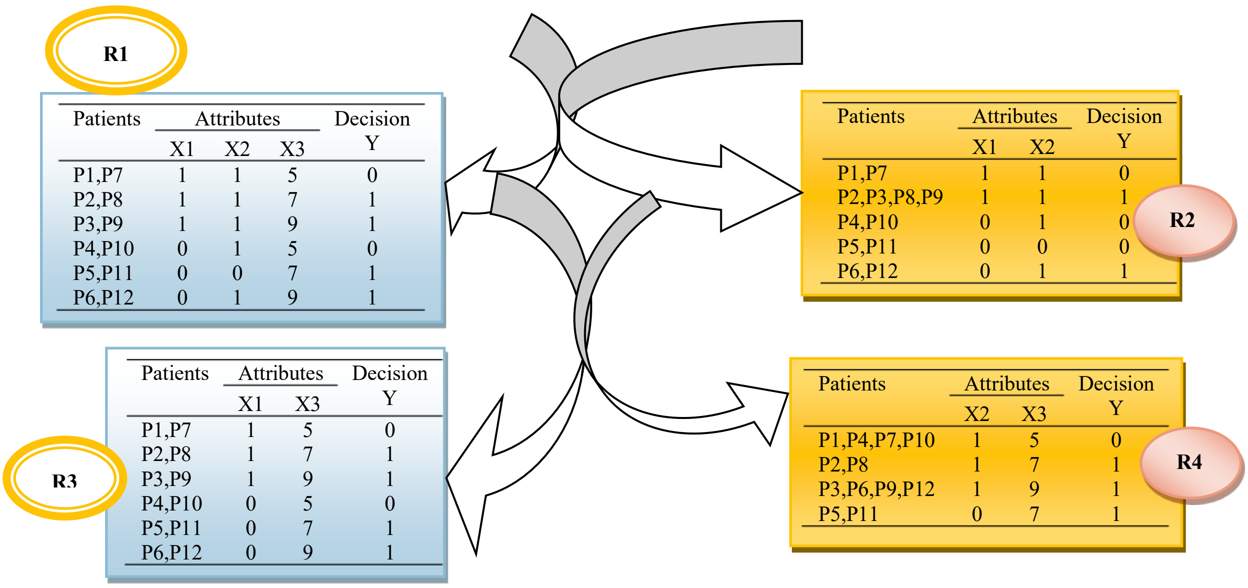

Fig. 1 Data set partitions.

The data set has been partitioned into R1, R2, R3, and R4 by the application of indiscernibility, reducts and core concepts of the RST. From the partitions, the following can be inferred:

(a) R1 removes redundancy but retains all the features in the original data set. That is arrow-level, let R be S, and if R=S, then R1 ⊆ R.

(b) R2 ⊆ R reduces row-wise redundancy and the number of input features but distorts classification.

(c) R3 reduces both row and column-wise redundancies while preserving the ultimate classification as would by using the original data set, thus satisfying the concept of dimensionality reduction.

(d) R4 removes row-wise redundancy but negatively affects the classification.

That is,

$$R1 = (\{P\},(\{X1\},\{X2\}, \{X3\}) \subseteq A) \quad (16)$$

$$R2 = (\{P\},(\{X1\},\{X2\}) \subseteq A) \quad (17)$$

$$R3 = (\{P\},(\{X1\},\{X3\}) \subseteq A) \quad (18)$$

$$R4 = (\{P\},(\{X2\},\{X3\}) \subseteq A) \quad (19)$$

Where P ⊆ U in R if R=S.

Since R1 does not reduce the number of features and R2 worsens the classification, then R3 and R4 will be considered to calculate the reducts and core of S as follows:

R3 =>X1,X3 and R4 => X2,X3

$$Red(R)=(X1,X3) \text{ and } (X2,X3) \quad (20)$$

$$Core(R) = nRed(R) = (X1,X3)\ n\ (X2,X3) = X3 \quad (21)$$

Therefore, it can be inferred that X3 (temperature) is the most important attribute (i.e., core) while X1 (headache) and X3 (temperature) or X2 (muscle pain) and X3 (temperature) are the reducts in this context.

7. Conclusion

The set theory and rough set approaches have been combined and employed in this paper to reduce the number of input features for optimal performance of a chosen machine learning algorithm. After the reduction process, and due to decreasing the complexity of the classification models, the results obtained as presented in Fig. 1 revealed that hybridizing these two approaches has a great improvement in dimensionality reduction as seen in R3. Thus, this certainly reduced the training time and computational overhead tremendously for the chosen machine learning algorithm (logistic regression) compared with using the whole feature vector as in the decision system shown in Table 2. Therefore, mixing set theory with the rough set technique for dimensionality reduction is a very powerful mathematical technique for reducing high dimensional space data to low dimensional space without negatively affecting the underlined classification accuracy, thus, improving the performance of the machine learning model being used. However, the rough set approach can also be combined with neural networks for better results in dimensionality reduction, especially in data pre-processing activities.

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] M. Sammany, T. Medhat, "Dimensionality reduction using rough set approach for two neural networks-based applications," In M. Kryszkiewicz, J. F. Peters, H. Rybinski, A. Skowron, (eds), Rough Sets and Intelligent Systems Paradigms. RSEISP 2007. Lecture Notes in Computer Science, vol. 4585, Springer, Berlin, Heidelberg.

[2] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences,* vol. 11, no. 5, pp. 341-356, 1982.

[3] S. Shampa, "A rough set theory-based approach to feature selection for incremental data," *International Journal for Research in Engineering Application & Management,* vol. 5, no. 4., pp. 247-252, 2019.

[4] T. Keerthika, K. Premalatha, "Medical domain-based feature selection using rough set reduct algorithm," *International Journal of Engineering Research & Technology*, vol. 3, no. 15., pp. 1-6, 2015.

[5] C. Avian, J. S. Leu, S.W. Prakosa, Muhamad Faisal, "An improved classification of pork adulteration in beef based on the electronic nose using modified deep extreme learning with principal component analysis as feature learning," *Food Analytical Methods*, vol. 21, 2022, DOI: 10.1007/s12161-022-02361-9

[6] S. Akodad, "Ensemble learning methods on the space of covariance matrices: Application to remote sensing scene and multivariate time series classification," Automatic Control Engineering, Université de Bordeaux, NNT: 2021BORD0310ff, tel-03484011v2f, 2021.

[7] R. Silvia, L. Germano, "Rough set theory – Fundamental concepts, principals, data extraction, and applications data mining and knowledge discovery in real life applications, J. Ponce, A. Karahoca (eds)*,* I-Tech, Vienna, Austria*,* 2009.

[8] T. R. JeraldBeno, M. Karnan, "Dimensionality reduction: Rough set-based feature reduction," *International Journal of Scientific and Research Publications,* vol. 2, no. 9, pp. 1-6, 2012.

[9] M. Amsaveni, S. Duraisamy, "A parallel rough set theory for nonlinear dimension-reduction in big data analysis," *International Journal of Intelligent Engineering & Systems,* vol.12, no.5, pp. 170-178, 2019.

[10] E. Aboul, M. H. A. Jafar, "Rough set approach for generation of classification rules of breast cancer data," *Informatica,* vol. 15, no. 1, pp. 23–38, 2004.

[11] P. Chaturvedi, A. K. Daniel, K. Khusboo, "Concept of rough set theory and its applications in decision-making processes," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, special issue 2, pp. 43-46, 2017.

[12] D. Miao, W. Pedrycz, D. Ślęzak, G. Peters, Q. Hu, R. Wang, "Rough Sets and Knowledge Technology," 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014,

[13] N. Sengupta, J. Sil, "Dimension reduction using rough set theory for intrusion detection system," Proceedings of the 4th National Conference, Computing for Nation Development, February 25 – 26, 2010, Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi.

[14] D. K. Srivastava, "Data classification: A rough - SVM approach," *Contemporary Engineering Sciences,* vol. 3, no. 2, pp. 77 – 86, 2010.

[15] A. O. Adetunmbi, S. O. Falaki, O. S. Adewale, B. K. Alese, "Network intrusion detection based on rough set and k-nearest neighbour," *International Journal of Computing & ICT Research,* vol. 2, no. 1, pp. 60-66, 2008.

[16] A. Waleed, U. K. Rehan, K. Khalil, "Attributes reduction in big data," *Applied Sciences,* vol. 10, no. 14, 2020, DOI: 10.3390/app10144901

[17] S. Qiang, C. Alexios, "Rough set-based dimensionality reduction for supervised and unsupervised learning," *International Journal of Applied Mathematics & Computational Science,* vol.11, no. 3, pp. 583-601, 2001.

[18] Y. Li, S. C. K. Shiu, S. K. Pal, J. N. K. Liu, "A rough set-based case-based reasoner for text categorization," *International Journal of Approximate Reasoning*, vol. 41, no. 2, pp. 229-255, 2006.

[19] A. Basant, M. Namita, "Sentiment classification using rough set based hybrid feature selection," Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis, pp. 115–119, 2013.

[20] M. K. Sabu, "A rough set based feature selection approach for the prediction of learning disabilities," *International Journal of Advanced Computational Engineering and Networking,* vol. 2, no. 12, pp. 43-48, 2014.

[21] B. Femina, S. Anto, "Disease diagnosis using rough set based feature selection and K-nearest neighbour classifier," *International Journal of Multidisciplinary Research and Development,* vol. 2, no. 4, pp. 664-668, 2015.

[22] L. Al-Shalabi, "Rough set-based reduction of incomplete medical datasets by reducing the number of missing values," *International Arab Journal of Information Technology,* vol. 16, no. 2, pp. 203-210, 2019

[23] W. Zhenhua, L. Suling, X. Lizhi, S. Wei, W. Dexing, H. Dongmei, "Dimensionality reduction method for hyperspectral image analysis based on rough set theory," *European Journal of Remote Sensing,* vol. 53, no. 1, pp. 192–200, 2020.