

Online Incremental Rough Set Learning in Intelligent Traffic System

Amal Bentaher^{1,2}, Yasser Fouad², Khaled Maher³

¹Faculty of Science, Hadhramout University., Hadhramout, Yemen

²Computer Science Department, Faculty of Science, Alexandria University, Alexandria, Egypt

³College of Information Technology, Arab Academy for Science and Technology, Alexandria, Egypt

Abstract—In the last few years, vehicle to vehicle communication (V2V) technology has been developed to improve the efficiency of traffic communication and road accident avoidance. In this paper, we have proposed a model for online rough sets learning vehicle to vehicle communication algorithm. This model is an incremental learning method, which can learn data object-by-object or class-by-class. This paper proposed a new rules generation for vehicle data classifying in collaborative environments. ROSETTA tool is applied to verify the reliability of the generated results. The experiments show that the online rough sets based algorithm for vehicle data classifying is suitable to be executed in the communication of traffic environments. The implementation of this model on the objectives' (cars') rules that define parameters for the determination of the value of communication, and for reducing the decision rules that leads to the estimation of their optimal value. The confusion matrix is used to assess the performance of the chosen model and classes (Yes or No). The experimental results show the overall accuracy (predicted and actual) of the proposed model. The results show the strength of the online learning model against offline models and demonstrate the importance of the accuracy and adaptability of the incremental learning in improving the prediction ability.

Keywords—Vehicle to vehicle communication; online learning; rough sets theory; intelligent traffic system

I. INTRODUCTION

Nowadays, road accidents are one of the major problems in modern societies that lead to death. The increase of travel time is a main reason for increasing traffic accidents, fuel consumption and increased pollution [1], [2]. Road safety field is on focus by researchers to detect traffic congestions and, thereby, to offer solutions.

The Intelligent Transport System (ITS) is a technology to achieve safe roads and comfortable driving, by reducing accidents and delay [3]. In recent years, a research area in the road safety called vehicular network offers a possible solution that allows a communication and information exchange between vehicles, which is called vehicle to vehicle (V2V) communication, or between vehicles and road infrastructure, which is called vehicle to infrastructure (V2I) communication, [4] as shown in Fig. 1.

Since its development, Rough sets theory has been able to devise a computationally efficient and mathematically sound techniques handling imprecision in decision making [5]. The optimal solutions without losing any information can be found

by using the rough sets theory which find reducts the rules for training the dataset and classifying the test set [6].

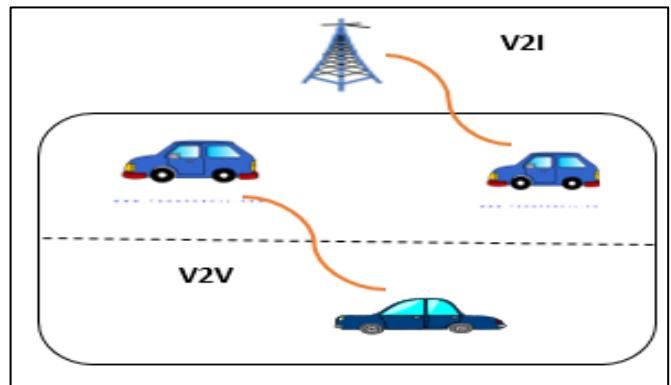


Fig. 1. V2V and V2I communication.

The remainder of this paper is organized as follows. In Section 2, we describe basic concepts of rough sets; in Section 3 we describe architecture and the feasibility decision table of our model. In Section 4, we present the implementation of our proposed model and show the results. Finally, we conclude our paper at Section 5.

II. BASIC CONCEPTS OF ROUGH SETS

Rough sets theory (RST) is a mathematical tool that is developed by Pawlak in 1982 [7]. In this theory, the data is collected in a table, called a decision table. Rows of concepts on rough sets theory are reviewed as follows:

Definition 1 (Information system): Is the 4-tuple [8], [9] (U, A, C, D) where U consists of objects and A consists of attributes, the subsets C and D are called condition attribute set and decision attribute set, respectively. Every $a \in A$ corresponds to the function $a: U \rightarrow V_a$ where V_a is the value set of a .

Definition 2 (Indiscernibility relation): Let $S=(U,A)$ be an information system, and $B \subseteq A$. we define the B -indiscernibility relation as [8], [10]:

$$INDs(B) = \{ (x,y) \in U^2 \mid \forall b \in B \quad (b(x) = b(y)) \} \quad (1)$$

If $(x,y) \in INDs(B)$, then x and y are indiscernible by attributes from B . The equivalence classes of the B -indiscernibility relation of an object x is denoted by $[x]_{IND(B)}$.

Definition 3 (Lower and upper approximation): Two fundamental concepts of rough sets are the lower and upper approximations of sets (which are a classification of the domain of interest into disjoint categories) in Fig. 2 [8], [9]. Given a set $B \subseteq A$, the lower and upper approximations of a set $X \subseteq U$ are defined by, respectively,

$$\underline{B}Y = \{x \mid [x] B \subseteq X\} \quad (2)$$

$$\bar{B}Y = \{x \mid [x] B \cap X \neq \emptyset\} \quad (3)$$

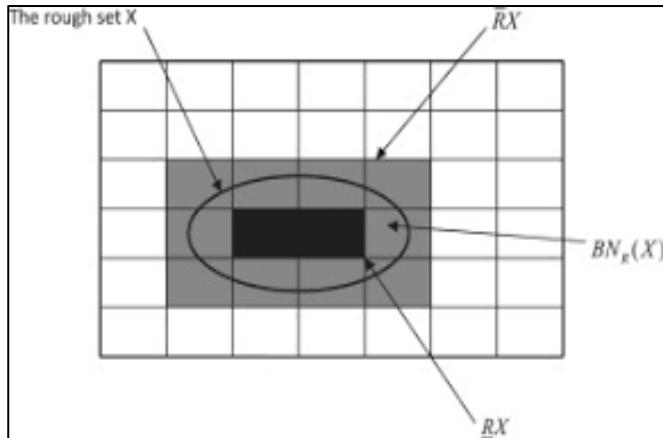


Fig. 2. Depiction of rough set.

Definition 4 (Lower approximation and positive region): [8]-[10]. The positive region $POS_B(X)$ is defined by:

$$POS_B(X) = BX \quad (4)$$

If an object $x \in POS_B(X)$, then it belongs to target set X certainly.

Definition 5 (Upper approximation and negative region): The negative region $BND_B(X)$ is defined by [8]-[10]:

$$NEG_B(X) = U - \bar{B}X \quad (5)$$

If an object $x \in NEG_B(X)$, then it cannot be determined whether the object x belongs to target set X or not.

Definition 6 (Boundary region): The boundary region is the difference between upper and lower approximations of a set X [8]-[10]:

$$BND_B(X) = \bar{B}X - BX \quad (6)$$

If an object $x \in BND_B(X)$ it doesn't belong to target set X certainly.

III. PROPOSED MODEL

Continuous data streams reflect continuous environmental changes, raising the need for online learning to adapt to changing conditions [11]. In this section we present the proposed model that is online rough sets learning vehicle to vehicle communication algorithm. The proposed algorithm is a methodology which uses rough sets theory to compute accurate objects for new (rules) data streams from online traffic environments. Vehicles can detect potential issues on the road and alert nearby users about incoming dangers and reduce the risk of accidents as shown in Fig. 3.

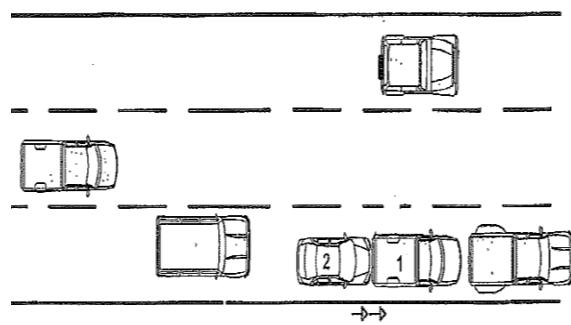


Fig. 3. Vehicles rough set learning communication in online environment.

A. Model Algorithm

In our method online learning rough sets theory in vehicle to vehicle (V2V) communication environment determines objects (cars) one by one and satisfying any object to any rule or addition new rules. Our model use GPS and a wireless LAN module to allow cars to communicate within a range of 100-to-300 meters.

The proposed model is designed to work based on the following algorithm:

Algorithm: Proposed model Algorithm

Input: Incrementally arriving objects in traffic environment.

Output: Optimal decision of communication.

Step 1. Initial data have one object and set of rules R.

Step 2. $F = \emptyset$ {List all best of objects}

Step 3. $f \leftarrow$ get new object (car)

Step 4: Determine object and rule r

Step 5. If $f \subset F$ then the rule r is satisfying to any exited rules

Step 6. Otherwise compute a decision table, generate reduct and generate rules R.

Step 7. Update online data streaming

Step 8. Repeat steps 3-7 until finish all objects.

B. Decision Table of the Model

The rough sets theory has been developed for knowledge discovery in databases and experimental data sets. An attribute-oriented rough sets technique reduces the computational complexity of learning processes and eliminates the unimportant or irrelevant attributes so that the knowledge discovery in database or in experimental data sets can be efficiently learned [12].

The Rough Sets analysis of data involved calculation of reducts from data, derivation of rules from reducts, rule evaluation and prediction processes. The rosetta rough sets toolkit was employed to carry out reducts and generate decision rules. The reducts were created from our selected data are revealed in Fig. 4. We used the Johnhon's reducer algorithm and the equal binning decrettized method.

The screenshot shows the Rosetta software interface with a menu bar (File, Edit, View, Window, Help) and a toolbar with various icons. The main window displays a decision table titled "Sheet1\$". The table has columns: Id, Speed, Range, direction, vehicle color, vehicle type, and Decision. The data consists of 27 rows, each representing a car object with its attributes and a final decision value.

	Id	Speed	Range	direction	vehicle color	vehicle type	Decision
1	1	2	13	-1	green	Truck	Yes
2	2	5	2	-1	indigo	lorry	No
3	3	3	5	-1	orange	lorry	No
4	4	2	6	1	blue	van	Yes
5	5	10	15	1	indigo	bus	No
6	6	4	15	0	green	bus	No
7	7	9	5	-1	blue	lorry	No
8	8	10	18	1	red	car	No
9	9	5	2	1	blue	lorry	Yes
10	10	3	11	1	violet	Truck	No
11	11	6	7	-1	yellow	van	No
12	12	8	3	-1	red	lorry	Yes
13	13	10	9	0	indigo	lorry	Yes
14	14	3	19	1	green	lorry	Yes
15	15	2	11	-1	red	Truck	Yes
16	16	9	8	0	indigo	bus	Yes
17	17	1	9	-1	green	lorry	Yes
18	18	7	5	-1	blue	lorry	No
19	19	10	12	0	red	car	Yes
20	20	8	4	1	green	Truck	No
21	21	4	5	1	violet	car	No
22	22	6	15	-1	red	lorry	No
23	23	10	7	-1	orange	van	No
24	24	10	14	1	violet	car	No
25	25	2	9	-1	blue	Truck	Yes
26	26	8	6	0	yellow	lorry	Yes
27	27	4	4	-1	violet	van	Yes

Fig. 4. Car dataset after use decision table.

A data set can be represented as a decision table, which is used to specify what conditions lead to decisions. A decision table is defined as $T = (U, C, D)$ where U is the set of objects in the table, C is the set of the condition attributes and D is the set of the decision attributes as shown in Fig. 4. The features are: decision, the speed of cars, the range between cars, the directions of the cars, the color and the type of cars. Fig. 4 can be used to decide whether a car has a Yes or No decision according to its features (e.g., the speed, the range and the directions). For example, the first row of this table specifies that the speed of the car is 2, with 13 range, -1 direction, green color and a truck type. The rows in this table are called the objects, and the columns in this table are called attributes. Condition attributes are the features of a car related to its decision communication; therefore, $C = \{Speed, Range, direction, vehicle color, vehicle type\}$. Decision communication is the decision attribute; therefore, $D = \{Decision\}$.

To evaluate the ability of our model to learn incrementally, we conducted experiments using 10 objects in different behaviors. Data are divided into two parts: training and testing sets. The very first task is to find reducts and rules. This Data set contains 7 attributes including the decision attribute which may be Yes or No, and there are 100 objects (or) records in this data and with no missing attribute values. The value and meaning of condition and decision attributes is shown in Fig. 4 as true (*Yes*) class, or false (*NO*) class.

IV. EXPERIMENT, RESULTS AND DISCUSSION

In the experiment, we have evaluated the data with ROSETTA software. ROSETTA is a toolkit application which allows the analysis of tabular data using the rough sets methodology to implement Johnson's algorithm rough sets for attribute selection. Rosetta is an open source collection of C++ classes and routines used for data mining and machine learning in general and particularly for rough sets theory [13].

The toolkit follows some important procedures for producing the accurate result.

A. Implementation Process

The steps are: importing data from any valid data source (e.g. Excel format), applying the binary splitting algorithm in the imported data to split the original dataset into training and test data, removing the missing values, and finally applying the reduction and classification algorithms. The reduction algorithm is used to compute the reduct set and the classification algorithm is used to reduct rules and compute the classification result.

The input data set is divided into two parts with the 0.9, 0.8, 0.7, 0.6 and 0.5 split factor. The first part is known as training data set and the other one is known as test data set. The training set was reduced by using Johnson's reduction algorithm [14], which uses greedy search to find one reduct. In Table I, numbers of reduct sets produced through the application of Johnson's reduction algorithm are illustrated. The Johnson's reduction algorithm produced 9 combinations of reduct sets. An example of a rule obtained from reducts is shown in Table I. A full training dataset of each dataset object is used to train the classifiers to build the classification models that were evaluated on the test data of the same objects.

TABLE I. REDUCTS OF THE CAR DATASETS

	Reduct	Support	Length
1	{Speed, Range}	100	2
2	{Speed, vehicle type}	100	2
3	{Speed, direction}	100	2
4	{Range}	100	1
5	{Speed, vehicle color}	100	2
6	{Range, direction}	100	2
7	{direction, vehicle color}	100	2
8	{Speed, vehicle color}	100	2
9	{Range, vehicle color}	100	2

The table shows the average accuracy for the predicted accuracy and the accuracy sensitivity as 50.59% and 65.87%, respectively.

The testing1 is training-testing of 90-10, testing 2 is training-testing of 80-20, testing 3 is training-testing of 70-30, testing 4 is training-testing of 60-40 and testing 5 is training-testing of 50-50. The classification results of original object are shown in Table II.

The reduction rule explains the rule support, stability, length, coverage and accuracy. Each row of the reduction rule is called descriptors (Attribute → value). The left hand side of the rule is called the antecedent and right hand side of the rule is called consequent. This reduction rule result used in the classification process. This rule is used to make the confusion matrix.

A. Results and Discussion

Table II exhibits the classification accuracy of original object. To find the percentage of accuracy, dataset has been changed as training set and testing set according to the mentioned ratio.

For more evaluation for the model's capability to learn incrementally, we conducted experiments using different testing types this process was repeated 10 times in different (10 objects) cars. The ten different testing results (classification accuracy) for our model are shown in Table III.

Our model has the capability to learn new objects (cars) from data streams in online environments and can accurately detect the appropriate car to communicate in road traffic. Fig. 5 and 6 show the predicted accuracy and the actual accuracy of our proposed model, respectively. The results suggest that our proposed model can handle the concept of vehicle to vehicle communication in online environments.

TABLE II. CLASSIFICATION RESULTS OF ORIGINAL OBJECT

Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
Testing 1	66.67%	40.00%	40.00%
Testing 2	42.86%	50.00%	45.00%
Testing 3	46.15%	66.67%	46.67%
Testing 4	48.48%	85.71%	50.00%
Testing 5	48.78%	86.96%	52.00%
Average	50.59%	65.87%	46.37%

TABLE III. CLASSIFICATION RESULTS OF ORIGINAL TEN OBJECT

	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
Testing Object1	Testing 1	75.00%	50.00%	60.00%
	Testing 2	50.00%	41.67%	40.00%
	Testing 3	50.00%	58.82%	45.16%
	Testing 4	47.37%	39.13%	41.46%
	Testing 5	33.33%	21.43%	33.33%
	Average	51.14%	42.21%	43.99%

CLASSIFICATION'S PREDICTED ACCURACY

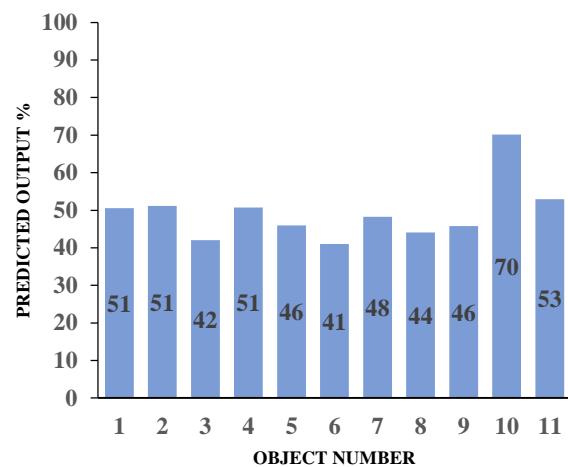


Fig. 5. Classification's prediction accuracy.

CLASSIFICATION'S ACTUAL ACCURACY

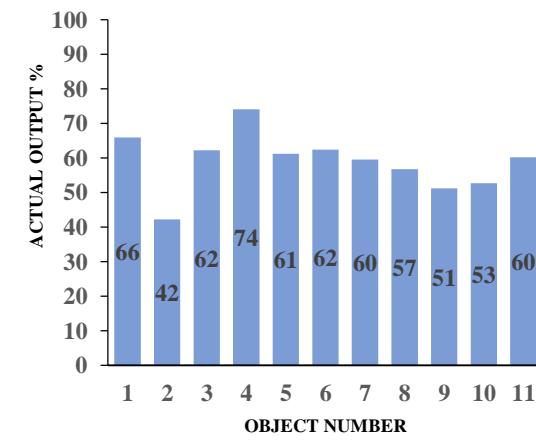


Fig. 6. Classification's actual accuracy.

Rough sets have been employed here to remove redundant conditional attributes from discrete-valued datasets, while retaining their information content. This approach has been applied to aid classification of online traffic environment, with very promising results.

Using online rough sets in our model allows efficient updates and avoids the process of retaining the whole data, a major disadvantage of offline models, when new data are coming, and it is most helpful when dealing with big data.

	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
Testing Object2	Testing 1	45.95%	70.83%	47.06%
	Testing 2	40.00%	40.00%	40.00%
	Testing 3	42.86%	54.55%	38.10%
	Testing 4	40.91%	69.23%	45.16%
	Testing 5	40.63%	76.47%	43.90%
	Average	42.07%	62.22%	42.84%
Testing Object3	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	66.67%	80.00%	70.00%
	Testing 2	42.86%	66.67%	47.62%
	Testing 3	47.37%	69.23%	54.84%
	Testing 4	48.15%	76.47%	57.14%
	Testing 5	48.65%	78.26%	53.85%
Testing Object4	Average	50.74%	74.13%	56.69%
	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	50.00%	66.67%	45.45%
	Testing 2	42.86%	66.67%	47.62%
	Testing 3	40.00%	53.33%	40.63%
	Testing 4	50.00%	61.90%	50.00%
Testing Object5	Testing 5	46.88%	57.69%	46.15%
	Average	45.95%	61.25%	45.97%
	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	28.57%	40.00%	27.27%
	Testing 2	41.67%	62.50%	52.38%
	Testing 3	47.62%	66.67%	50.00%
Testing Object6	Testing 4	44.83%	72.22%	50.00%
	Testing 5	42.50%	70.83%	43.40%
	Average	41.04%	62.44%	44.61%
	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	50.00%	33.33%	45.45%
	Testing 2	50.00%	70.00%	52.38%
Testing Object7	Testing 3	50.00%	73.33%	53.13%
	Testing 4	48.00%	57.14%	48.84%
	Testing 5	43.24%	64.00%	43.40%
	Average	48.25%	59.56%	48.64%
	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	50.00%	50.00%	45.45%
Testing Object8	Testing 2	46.15%	54.55%	45.45%
	Testing 3	45.00%	64.29%	50.00%
	Testing 4	37.50%	50.00%	44.19%
	Testing 5	41.67%	65.22%	46.30%
	Average	44.06%	56.81%	46.28%
	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
Testing Object9	Testing 1	40.00%	28.57%	27.27%
	Testing 2	50.00%	42.86%	36.36%
	Testing 3	45.00%	50.00%	39.39%
	Testing 4	46.43%	61.90%	47.73%
	Testing 5	47.50%	73.08%	48.15%
	Average	45.79%	51.28%	39.78%
Testing Object10	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	100.00%	55.56%	63.64%
	Testing 2	75.00%	60.00%	59.09%
	Testing 3	66.67%	60.00%	57.58%
	Testing 4	57.14%	48.00%	50.00%
	Testing 5	52.17%	40.00%	47.27%
	Average	70.20%	52.71%	55.52%
Testing Object10	Data Allocation (%)	The predicted accuracy	The actual accuracy	The overall accuracy
	Testing 1	57.14%	66.67%	50.00%
	Testing 2	60.00%	69.23%	50.00%
	Testing 3	50.00%	68.75%	46.67%
	Testing 4	56.00%	60.87%	50.00%
	Testing 5	41.67%	35.71%	36.00%
	Average	52.96%	60.25%	46.53%

V. CONCLUSION AND FUTURE WORKS

Vehicle to vehicle communication has offered several solutions for traffic problems that reduce the danger of collision. This research aimed to design a prototype online rough sets learning in traffic vehicular communication. Optimal communication vehicles were decided by using online learning rough sets to evaluate the percentage of accuracy decision (Yes or No). Generally, online learning data requires an updatable model. That is, the model should in some objects “evolve” in response to the streaming data.

This study presents an incremental learning algorithm which learns new classes in online environments, allowing our model to be updatable and evolve to detect new objects (cars). The model attempts to learn the rules of objects (cars) where the following vehicle notifies the cars.

Our model uses ROSETTA tool in rough sets data analysis to emphasize the classification, in prediction of the learning. Several tests were done by changing the training and the testing dataset ratio. The confusion matrix is used to assess performance of chosen model and classes (Yes or No). The experimental results show that overall accuracy (predicted and actual accuracy) of the object is evolved in our proposed online model.

The limitation of our model is that the run time is not perfect which affects the classification accuracy. More experiments are needed in future.

REFERENCES

- [1] G. Bulumulle and L. Böloni, “Reducing Side-Sweep Accidents with Vehicle-to-Vehicle Communication,” *J. Sens. Actuator Netw.*, vol. 5, no. 4, p. 19, Dec. 2016.
- [2] M. Kang, P. Bao, and Y. Cai, “Effect of Residential Quarters Opening on Urban Traffic from the View of Mathematical Modeling,” *Open J. Model. Simul.*, vol. 05, no. 01, pp. 59–69, 2017.
- [3] S.-Z. Huang, C.-L. Hu, S. Chen, and L. Guo, “Inter-vehicle Media Distribution for Driving Safety,” in *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, vol. 81, J.-S. Pan, P.-W. Tsai, J. Watada, and L. C. Jain, Eds. Cham: Springer International Publishing, 2018, pp. 121–128.
- [4] A. Bhawiyuga, R. A. Sabriansyah, W. Yahya, and R. E. Putra, “A Wi-Fi based Electronic Road Sign for Enhancing the Awareness of Vehicle Driver,” *J. Phys. Conf. Ser.*, vol. 801, p. 012085, Jan. 2017.
- [5] M. Aggarwal, “Rough Information Set and Its Applications in Decision Making,” *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 2, pp. 265–276, Apr. 2017.
- [6] M. T. Rezvan, A. Z. Hamadani, and S. R. Hejazi, “An exact feature selection algorithm based on rough set theory,” *Complexity*, vol. 20, no. 5, pp. 50–62, May 2015.
- [7] Z. Pawlak, “Rough sets,” *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
- [8] Q. Zhang, Q. Xie, and G. Wang, “A survey on rough set theory and its applications,” *CAAI Trans. Intell. Technol.*, vol. 1, no. 4, pp. 323–333, Oct. 2016.
- [9] S. Eskandari and M. M. Javidi, “Online streaming feature selection using rough sets,” *Int. J. Approx. Reason.*, vol. 69, pp. 35–57, Feb. 2016.
- [10] J. Bai, K. Xia, Y. Lin, and P. Wu, “Attribute Reduction Based on Consistent Covering Rough Set and Its Application,” *Complexity*, vol. 2017, pp. 1–9, 2017.
- [11] H.-S. Chiang and Z.-W. Wu, “Online incremental learning for sleep quality assessment using associative Petri net,” *Appl. Soft Comput.*, Aug. 2017.
- [12] S. Danjuma, T. Herawan, M. A. Ismail, H. Chiroma, A. I. Abubakar, and A. M. Zeki, “A Review on Soft Set-Based Parameter Reduction and Decision Making,” *IEEE Access*, vol. 5, pp. 4671–4689, 2017.
- [13] A. Øhrn, *The ROSETTA C++ Library: Overview of Files and Classes*. 2000.
- [14] X. Li, “Attribute Selection Methods in Rough Set Theory,” San Jose State University, California, 2014.