Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 5 Iss. 2 April 2025, pp: 578-586
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Impact of Cover Parameter Value on Rule Generation in Rough Set Classification

Farhatul Fityah[1*], Pramudya Rakhmadyansyah Sofyan[2]

[1,2]Master of Biomedical Engineering Department,
Graduate School of Universitas Gadjah Mada, Indonesia

E-Mail: [1]farhatulfityah@mail.ugm.ac.id, [2]pramudyarsofyan@mail.ugm.ac.id

Received Dec 15th 2024; Revised Feb 15th 2025; Accepted Feb 24th 2025; Available Online Mar 21th 2025, Published Jan 21th 2025
Corresponding Author: Farhatul Fityah

Abstract

Machine learning plays a crucial role in healthcare classification, with Rough Set Theory (RST) offering effective tools for managing data uncertainty. Within RST, the RSES2 tool supports algorithms like LEM2 and Covering, yet the influence of cover parameter values on rule generalization and specificity remains underexplored. This study investigates these effects using the Differentiated Thyroid Cancer dataset. The research investigates the trade-offs between rule generalization and specificity by adjusting cover parameter settings, which dictate the minimum and maximum cases a rule must cover. The comparison reveals that the LEM2 algorithm maintains high accuracy across various cover parameter values, with only a slight decline as the parameter increases, and shows improved coverage with higher cover values. In contrast, the Covering algorithm displays greater fluctuations in accuracy, peaking at lower cover parameter values and decreasing significantly as the parameter rises. Coverage for the Covering algorithm is highest at lower cover parameters but decreases sharply at higher values. This indicates that LEM2 is more robust in maintaining accuracy and coverage, while the Covering algorithm performs better at lower cover parameters but struggles with stability as the parameter increases.

Keyword: Classification, Covering, LEM2, RSES2, RST

1. INTRODUCTION

Rough Set Theory (RST) has been widely used in data mining and knowledge discovery due to its capability to handle vagueness and uncertainty in data classification [1]. RST operates by approximating a set of data using two definable boundaries: the lower approximation, which contains all elements that definitely belong to the set, and the upper approximation, which includes all elements that possibly belong to the set. The difference between these two approximations is known as the boundary region, representing the uncertainty or vagueness in the classification process [2]. By utilizing these approximations, RST can generate decision rules that are both robust and interpretable, making it particularly effective in domains with incomplete or noisy data [3][4]. One of the key advantages of RST in classification tasks is its ability to work without requiring any preliminary or additional information about data, such as probability distributions or membership grades, which are often necessary in other methods like fuzzy sets or probabilistic models [5]. This makes RST particularly useful in situations where such information is difficult or impossible to obtain. Additionally, RST is capable of identifying and eliminating redundant or irrelevant attributes during the classification process, thereby reducing the dimensionality of the data and improving the efficiency and accuracy of the classifier [6]. This feature selection capability is crucial in handling high-dimensional datasets, as it helps in simplifying models without compromising on performance. Furthermore, the decision rules generated by RST are inherently interpretable, providing clear and concise explanations that can be easily understood by domain experts, thus enhancing the transparency and trustworthiness of the classification results [3].

Recent studies over the past years have continued to explore and extend the application of RST in various domains, and several research efforts have applied RST to diverse classification problems. For instance, a novel rough set-based method was developed to improve feature selection and classification, demonstrating superior performance in handling high-dimensional data [7]. Another approach, known as GBNRS, was designed to offer fast and adaptive attribute reduction, particularly suitable for real-time and streaming data scenarios [8]. Additionally, the integration of Fisher Score with Multilabel Neighborhood

Rough Sets has been proposed to address the complexities inherent in multilabel classification, significantly improving classification accuracy [9]. Furthermore, the concept of multi-scale covering rough sets has been introduced, which enhances data classification by operating across multiple scales and handling hierarchical data structures effectively [10]. Lastly, incremental feature selection using fuzzy rough sets has shown promise in maintaining robust classification performance in dynamic and large-scale datasets [11]. These studies collectively demonstrate the ongoing relevance and applicability of RST in various data mining applications, particularly in handling complex, uncertain, and high-dimensional data. The integration of RST with other methods, such as fuzzy logic and genetic algorithms, further extends its utility in modern data-driven environments.

Within the RST framework, the Rough Set Exploration System (RSES) has emerged as a powerful software tool that facilitates the analysis and implementation of various rough set-based algorithms [12]. RSES2 supports a range of algorithms that can be used for rule generation, including LEM2, covering, genetic, and exhaustive algorithms. Each of these algorithms provides unique approaches to decision rule induction, allowing for flexible and tailored analyses depending on the nature of the data and the specific objectives of the study. The LEM2 algorithm is particularly well-regarded for its efficiency in handling data with missing attributes and its ability to produce concise and accurate decision rules. It constructs rules by iteratively finding minimal sets of conditions that distinguish between different decision classes [13]. On the other hand, covering algorithms offer flexibility by approximating rough sets through various neighborhood-based approaches, which can result in different types of rules depending on how the coverage values are set [14]. Genetic algorithms in RSES2, known for their optimization capabilities, evolve rules by simulating natural selection processes, whereas exhaustive algorithms perform a comprehensive search to identify all possible rules, though often at the cost of higher computational complexity [15].

Sengupta and Sil [16] compared the classification accuracy of rules generated by genetic algorithms, covering algorithms, and LEM2 within the context of network traffic data. Their findings indicated that covering algorithms provided the highest total accuracy, suggesting a strong potential for this approach in certain types of data. A more recent study by Srimani and Koti [17] further corroborates these findings within the medical domain, where the LEM2 algorithm demonstrated the highest accuracy (76%) compared to exhaustive, genetic, and covering algorithms, despite having lower coverage values. This suggests that while covering algorithms excel in accuracy, the balance between coverage and rule generalization remains crucial, particularly in applications where the specificity of rules is as important as their generalization. Sulaiman et al. [18] investigated the application of Rough Set (RS) techniques, including Johnson's and Genetic algorithms, in generating classification rules from AIDS and e-learning datasets. Their results revealed that although both algorithms produced a similar number of rules, the accuracy varied significantly depending on the dataset and the method of cross-validation used. These findings emphasize the importance of considering both rule coverage and accuracy, as the relationship between these factors can significantly influence the overall performance of classification models. Sulaiman et al. highlighted that while higher accuracy might be achieved with certain algorithms, the generalization of rules—particularly in datasets with complex structures—remains a critical aspect that warrants further investigation. This underscores the need for a more nuanced understanding of how coverage values and rule generalization impact the efficacy of covering and LEM2 algorithms across different domains [19]. However, the specific influence of cover parameter values on the performance of covering and LEM2 algorithms, especially in terms of rule generalization and specificity using RSES2, has not been thoroughly investigated, indicating a need for further research in this area. Most studies have focused on the overall effectiveness of individual algorithms without delving into how varying cover parameter values might alter the balance between rule accuracy and complexity.

This research aims to address this gap by systematically comparing the impact of cover parameter values on the rule generation capabilities of both LEM2 and covering algorithms within the RSES2 environment. The study seeks to provide deeper insights into the trade-offs between generalization and specificity in decision rules, offering valuable guidance for practitioners in selecting appropriate algorithms and parameters for various data mining tasks. Using the Differentiated Thyroid Cancer dataset from the UCI Machine Learning Repository, which includes 16 features related to patient demographics, clinical history, and pathological findings, the research will predict the recurrence of well-differentiated thyroid cancer. By applying both LEM2 and covering algorithms within the RSES2 environment, the study will systematically compare classification outputs by adjusting cover parameter settings, which determine the minimum and maximum number of cases a rule must cover. Specifically, the research will evaluate how variations in these settings impact prediction accuracy and the comprehensiveness of the generated rules (coverage value). Multiple experiments with different coverage thresholds will be conducted to observe the trade-offs between more generalized rules that cover a larger dataset portion and more specific rules that may offer higher accuracy but apply to fewer cases. The findings are expected to provide critical insights into optimizing rule generation parameters, aiding practitioners in balancing accuracy and coverage in classification tasks.

This research is limited to testing the cover parameter values of the LEM2 algorithm and the Covering algorithm, to see the relationship between the cover parameter values and the total coverage, on the

RSES2 software as an interface that supports Rough Set Theory-based classification. The implementation is limited to the use of the Thyroid data mentioned earlier.

2. LITERATURE REVIEW

This research was conducted after finding research gaps from the literature review process, especially about the implementation of RST in classification. As mentioned in the introduction, Sengupta and Sil [16] classified network traffic data to improve Intrusion Detection Systems (IDS) based on RST. The data used in the study was KDD network traffic data consisting of 11850 objects with 42 attributes. Of the total data, 2,133 objects were used for training and 1,185 objects for classification testing. Before applying RST, the data obtained was first discretized using supervised learning techniques with the help of WEKA software. This discretization process is important because it can affect classification accuracy. This research applies three algorithms for rule generation in RSES2 namely: Genetic Algorithm, Covering Algorithm, and LEM2 Algorithm. In addition, this research also finds reducts, which are the minimum subset of attributes that can replace the entire data without losing important information, by using extensive algorithms. Rules are then generated from these reducts. The results showed that the classification using the covering algorithm gave the best accuracy, with a total accuracy of 99.1%. The classifications generated from the reduct also showed similar accuracy. This research concludes that the application of RST in the classification of network traffic data can reduce complexity.

Furthermore, research by Srimani and Koti [17] applied the Rough Set approach to analyze medical data with the aim of generating classification rules and improving decision making. Using the Pima data set, which consists of 768 data samples related to diabetes risk. Each sample has 8 attributes that are considered as major risk factors. This data was divided into a training set and a testing set. This study used Rough Set reduction techniques to select the most relevant subset of attributes for classification. Several induction algorithms were used including Exhaustive, Covering, LEM2, and Genetic Algorithms (GA). Exhaustive algorithm produces more reducts than Genetic Algorithm. In the experiment, the Exhaustive algorithm produced 32 reducts, while the Genetic Algorithm produced 10 reducts. Although the Exhaustive algorithm produced more reducts, the highest accuracy result was achieved by Genetic Algorithm with an accuracy of 78.16%. In contrast, LEM2 had a high accuracy of 76%, albeit with lower coverage than the other algorithms. The importance of the results lies in the discovery that the results obtained from the GA implementation are non-deterministic, meaning that more accurate results can be obtained by calculating the average of multiple trials for the same dataset. This indicates that this approach still requires further exploration to be optimized.

Further research was conducted by Sulaiman et al [18] who explored the application of Rough Set Theory (RST) in generating classification rules for datasets related to AIDS and e-learning. By utilizing web mining techniques, researchers attempt to extract hidden information from large and complex data. RST was chosen for its ability to handle uncertainty and inconsistent information, which is often encountered in the analysis of medical and e-learning data. A discretization process is applied to convert continuous data into categories, which allows for more efficient modeling and analysis. The results showed that for the AIDS dataset, the 5-fold cross validation method showed varied prediction accuracy, with the highest value reaching 81.08%. Meanwhile, 10-fold cross validation resulted in higher accuracy for the e-learning dataset, with the Genetic algorithm (GA) achieving an average accuracy of 97.86%, slightly higher than the Johnson algorithm which recorded 97.74%. Overall, GA and Johnson produced the same number of rules, but the accuracy results obtained from 10-fold cross validation were better than 5-fold for both datasets. In conclusion, this study successfully demonstrated that the use of RST, together with effective reduction and discretization techniques, can generate valid classification rules and improve accuracy in the analysis of complex and incomplete data.

3. MATERIALS AND METHOD

The investigation of trade-offs between rule generalization and specificity by adjusting cover parameter settings using LEM2 and Covering Algorithm is shown in Figure 1. The work was implemented in stages that start from preparing the data, generating rules from two algorithms, cover parameter setting, testing, and finally evaluation based on the metrics to find the best cover parameter value.

2.1. Dataset

The dataset used in this research is sourced from secondary data available in the UCI machine learning repository, specifically focusing on Differentiated Thyroid Cancer Recurrence. The study involved 383 patients who were diagnosed with various forms of thyroid cancer, including papillary, micropapillary, follicular, and Hürthle cell carcinoma, at a single medical center. These patients participated in a retrospective cohort study with a minimum follow-up period of 10 years, beginning from the time of their surgery and initial diagnosis. The research was conducted in strict adherence to local ethical guidelines and the principles outlined in the Declaration of Helsinki. Additionally, the study protocol was thoroughly reviewed and approved by a dedicated board of experts at Hamedan University of Medical Sciences. The dataset includes comprehensive information on each patient's age at diagnosis, biological sex, smoking status (current and past), history of radiation therapy to the head and neck region, thyroid function, presence of adenopathy on physical examination, pathological subtype of cancer, focality, risk assessment according to ATA guidelines, TNM staging, initial treatment response, and recurrence status, which serves as the target variable for the study [20].

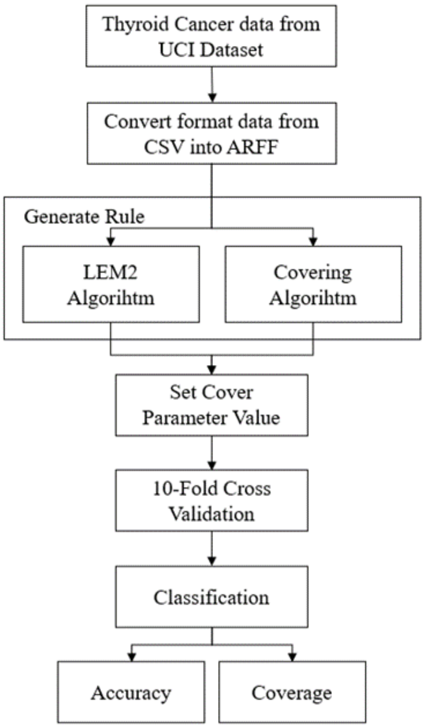

Figure 1. Research Method Flow

2.2 Data Processing

The dataset obtained from the UCI repository, initially provided in CSV format, was first processed for compatibility with the RSES2 software, which necessitates the use of the ARFF format. To achieve this conversion, Python was utilized to effectively transform the data from CSV to ARFF format. This preprocessing step was essential to ensure that the dataset could be efficiently analyzed and processed within the RSES2 environment, facilitating the subsequent stages of the research.

2.3 Rules Generation Method

This research is comparing two rules generation methods which will analyze the classification outcomes using rough set theory. The focus is on evaluating two rule generation approaches that depend on the set values of the cover parameter in RSES2, namely LEM2 and the Covering algorithm. Lower cover parameter value (e.g., 0.1) results in more granular and detailed rules. Conversely, a higher cover parameter value (e.g., 1.0) produces more general rules that are less detailed.

In LEM2, the total coverage of the generated rules is always ensured, as rules are iteratively constructed until all objects in the decision class are covered. The total coverage of the rule set R for a decision class X is given by equation 1 and 2.

$$C_{LEM2} = \frac{|\bigcup_{r \subset R} Covers(r)|}{|X|} \tag{1}$$

where:

- R is the set of generated rules.
- $Covers(r)$ represents the subset of objects covered by rule r, given by:

$$Covers\ (r) = X \cap \bigcap_{(a,v) \subset r} [a = v] \tag{2}$$

- X is the set of all objects in the decision class.

Since LEM2 ensures that all objects are eventually covered by at least one rule, it follows that equation 3.

$$C_{LEM2} = 1 \tag{3}$$

Indicating full coverage, regardless of the chosen cover parameter value.

In contrast, the Covering Algorithm selects attribute-value pairs based on their fraction of coverage, stopping when a predefined threshold γ is met. The total coverage of the generated rules is given by equation 4 and 5.

$$C_{Covering} = \frac{|\bigcup_{r \subset R} Covers(r)|}{|X|} \tag{4}$$

Where:

$$Covers\ (r) = X \cap \bigcap_{(a,v) \subset r} [a = v] \tag{5}$$

Unlike LEM2, the Covering Algorithm does not guarantee full coverage, as it depends on the selected cover parameter value γ. The individual coverage of an attribute-value pair *(a,v)* within a decision class X is given by equation 6 and 7.

$$Coverage\ (a, v) = \frac{|X \cap [a = v]|}{|X|} \tag{6}$$

A condition *(a,v)* is selected if the equation 7:

$$\frac{|X \cap [a = v]|}{|X|} \geq \gamma \tag{7}$$

Since the algorithm may stop before all objects are covered, the total coverage equation 8.

$$C_{Covering} \leq 1 \tag{8}$$

where a higher γ can lead to some objects being left uncovered. These differences highlight the fundamental trade-offs between the two rule generation methods, particularly in terms of coverage completeness and rule generalization.

The research involves applying the Rough Set Theory framework to build decision rules using LEM2 and covering algorithms, with cover parameter values tested at specific intervals: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. These variations in rule granularity and generalization will affect the classification outcomes, influencing metrics such as accuracy and coverage.

2.4. Testing and Evaluation

The data is classified based on the generated decision rules using 10-fold cross-validation within the RSES2 environment. The outcomes of each test are evaluated and compared in terms of total accuracy, total coverage, and total test number across the different coverage settings for both LEM2 and covering algorithms. This comparative analysis aims to identify the optimal coverage settings for rule generation and assess the performance differences between the two algorithms. The outcomes of each test are evaluated using the default evaluation metrics provided by RSES2, specifically total accuracy, total coverage, and total test number across the different cover parameter settings for both LEM2 and covering algorithms. Total accuracy refers to the percentage of correctly classified instances out of the total number of instances in the dataset. Meanwhile, total coverage represents the proportion of instances that are covered by the generated decision rules. In rough set theory, due to its inherent approach to handling uncertainty and incomplete information, not all data may be classified. This explains why total coverage is a crucial metric—some instances may remain unclassified, and total coverage reflects the extent to which the rules can cover the dataset.

3. RESULTS AND DISCUSSION

Thyroid Cancer data which was originally in CSV file format, was converted into Atribute Relation File Format (ARFF) format which follows the RSES2 input system. Afterwards, according to the roughset way of working, rules are built using LEM2 and Covering algorithms as previously designed. With the experimental treatment of the cover parameter value set as designed. The following are the classification results of each algorithm based on each treatment of the cover parameter set, with a focus on results comparing total accuracy and total coverage.

3.1. LEM2 Algorithm

Table 1 presents the impact of varying cover parameters on the LEM2 algorithm's performance in terms of accuracy and coverage. The LEM2 algorithm consistently achieves high accuracy across a range of cover parameters, with values fluctuating slightly between 0.979 and 1.000. Notably, perfect accuracy (1.000) is observed when the cover parameter is set at 0.2, 0.3, and 0.4, indicating that at these settings, the algorithm is highly precise in classifying instances without any errors. As the cover parameter increases beyond 0.4, a slight decline in accuracy is observed, with the lowest accuracy (0.973) occurring at the maximum cover parameter value of 1.0. Despite this minor reduction, the algorithm still maintains a relatively high level of accuracy, demonstrating its robustness even with higher cover parameters.

Table 1. Cover Parameter Variation Impact in LEM2 Algorithm

Cover Parameter	LEM2 Algorithm		
	Total Accuracy	Total Coverage	Total Tested Object
0.1	0.989	0.411	38
0.2	1.000	0.408	38
0.3	1.000	0.403	38
0.4	1.000	0.418	38
0.5	0.996	0.495	38
0.6	0.987	0.579	38
0.7	0.984	0.634	38
0.8	0.979	0.726	38
0.9	0.979	0.768	38
1.0	0.973	0.824	38

In contrast, the coverage of the LEM2 algorithm displays more variability as the cover parameter changes. At lower cover parameters, such as 0.1, 0.2, and 0.3, the coverage is relatively low, with the minimum value (0.403) occurring at a cover parameter of 0.3. However, as the cover parameter increases, the coverage improves steadily, peaking at 0.824 when the cover parameter is set to 1.0. This trend suggests that while lower cover parameters yield perfect accuracy, they do so at the expense of coverage, which remains limited. Conversely, higher cover parameters enhance the algorithm's ability to generalize across a broader set of instances, thereby increasing coverage, although with a slight trade-off in accuracy. Therefore, to achieve a balanced performance, particularly around 1.0 is recommended, as it optimizes the algorithm’s performance in both accuracy and coverage.

The LEM2 algorithm is designed to identify minimal sets of conditions that can accurately classify instances while balancing the trade-off between generalization and specificity of rules generated. As the cover parameter increases, LEM2 is capable of generating rules that generalize better, meaning they apply to a larger number of instances without sacrificing accuracy. This occurs because LEM2 prioritizes creating the most general rule possible that still effectively distinguishes between different decision classes. The cover parameter in LEM2 influences the granularity of the rule sets; higher cover parameters allow the algorithm to create more generalized rules that cover a greater number of instances, leading to increased coverage. However, the accuracy remains high because LEM2's rule induction process inherently seeks to balance generalization with specificity, avoiding overfitting by not over-complicating the rules.

Moreover, LEM2 is particularly effective in maintaining a minimal and efficient set of rules that ensures both high accuracy and adequate coverage across different cover parameter settings. The algorithm's flexibility in adjusting the rules according to the cover parameter allows it to maintain strong performance, preventing overfitting by focusing on the essential conditions needed to distinguish between classes. As the cover parameter increases, the algorithm broadens the scope of its rules, thereby improving coverage without a significant loss in accuracy. This demonstrates the LEM2 algorithm's capability to adapt to varying conditions while maintaining a balance between accuracy and coverage, making it a versatile tool for classification tasks.

3.2. Covering Algorithm

The impact of varying cover parameters on the performance of the Covering algorithm in terms of accuracy and coverage is illustrated by Table 2. The algorithm’s accuracy fluctuates significantly across different cover parameter values, ranging from 0.868 to 0.937. The highest accuracy, 0.937, is observed at a cover parameter of 0.2, indicating that at this specific setting, the algorithm is particularly precise in classifying instances. Accuracy remains relatively high, above 0.92, at cover parameters 0.1, 0.4, and 0.5, but begins to decline noticeably as the cover parameter increases beyond 0.5. The lowest accuracy, 0.868, occurs at cover parameters of 0.8 and 0.9, showing that the algorithm's performance in terms of accuracy deteriorates at higher cover parameters.

Table 2. Cover Parameter Variation Impact in Covering Algorithm

Cover Parameter	Covering Algorithm		
	Total Accuracy	Total Coverage	Total Tested Object
0.1	0.927	0.863	38
0.2	0.937	0.863	38
0.3	0.906	0.834	38
0.4	0.932	0.845	38
0.5	0.936	0.847	38
0.6	0.877	0.637	38
0.7	0.895	0.458	38
0.8	0.868	0.471	38
0.9	0.869	0.458	38
1.0	0.891	0.466	38

In terms of coverage, the Covering algorithm performs best at lower cover parameters. The maximum coverage of 0.863 is achieved at cover parameters of 0.1 and 0.2. However, as the cover parameter increases, coverage decreases progressively, with significant drops occurring beyond a cover parameter of 0.6. The lowest coverage, 0.466, is observed at the highest cover parameter of 1.0. This trend indicates that while lower cover parameters yield both high accuracy and maximum coverage, higher cover parameters lead to a reduction in both coverage and accuracy. Therefore, to achieve a balanced performance, maintaining a lower cover parameter, particularly around 0.2, is recommended, as it optimizes the algorithm's performance in both accuracy and coverage.

The mechanics of the Covering algorithm contribute to these observed trends. The algorithm operates by generating rules that "cover" as many instances as possible based on the conditions set by the cover parameter. When the cover parameter is low, the algorithm tends to create more general rules that apply to a broader range of instances, which results in higher coverage. However, these general rules may not be as precise, potentially leading to a slight decrease in accuracy. Conversely, as the cover parameter increases, the algorithm generates rules that are more specific. These specific rules are designed to cover fewer instances but aim to be more accurate for those instances. This increase in specificity reduces coverage because fewer instances meet the stricter rule conditions.

As the cover parameter increases, the algorithm demands that rules must cover more instances to be considered valid, leading to rules that are more restrictive and apply only to very specific cases. Consequently, coverage decreases because fewer instances are included under these more specific rules. The slight drop in accuracy at higher cover parameters may be attributed to the algorithm overfitting to certain data patterns, making it less effective at generalizing across the entire dataset. This overfitting is a result of the algorithm's focus on maximizing the coverage of each rule based on the parameter settings, which can lead to overly specific rules that perform well on the training data but fail to generalize to new, unseen data.

In summary, the Covering algorithm's performance is highly dependent on the cover parameter. Lower cover parameters result in more general rules that provide optimal coverage and relatively high accuracy, while higher cover parameters create more specific rules that reduce coverage and slightly decrease accuracy. The trade-off between rule coverage and specificity is evident, with higher specificity at higher cover parameters leading to a decline in overall coverage and the potential for overfitting, which ultimately affects the algorithm's ability to generalize to new data.

3.3. Performance Comparison Between LEM2 and Covering Algorithm

The comparison between the LEM2 and Covering algorithms, based on the impact of cover parameter variations on accuracy and coverage, reveals distinct performance characteristics for each algorithm. The LEM2 algorithm maintains consistently high accuracy across all cover parameter values, ranging from 0.979 to 1.000, with perfect accuracy achieved at cover parameters of 0.2, 0.3, and 0.4. Even as the cover parameter increases to 1.0, the accuracy only slightly declines to 0.973, demonstrating the robustness of LEM2 in maintaining strong accuracy regardless of the cover parameter. In terms of coverage, the LEM2 algorithm shows a clear upward trend as the cover parameter increases, with coverage starting low at 0.403 for lower cover parameters and peaking at 0.824 when the cover parameter reaches 1.0. This indicates that LEM2 benefits from higher cover parameters in terms of coverage, with only a slight trade-off in accuracy.

On the other hand, the Covering algorithm exhibits more fluctuation in accuracy, ranging from 0.868 to 0.937, with the highest accuracy at a cover parameter of 0.2. As the cover parameter increases beyond 0.5, a noticeable decline in accuracy occurs, highlighting the Covering algorithm's sensitivity to changes in the cover parameter. In terms of coverage, the Covering algorithm performs best at lower cover parameters, particularly 0.1 and 0.2, where coverage is highest at 0.863. However, coverage progressively decreases as the cover parameter increases, with significant drops beyond 0.6, resulting in the lowest coverage of 0.466 at a cover parameter of 1.0. This decline shows that the Covering algorithm is more effective with lower cover parameters but struggles to maintain performance as the cover parameter increases.

The cover parameter specifies the minimum number of instances that a rule must cover to be considered valid. Essentially, it controls the granularity or specificity of the rules generated by the algorithm. While total coverage refers to the proportion of instances in the dataset that are covered by the rules generated by the classifier. It measures how well the rules fit the data. Theoretically, when the cover parameter value is low, the rules generated will be more specific and may cover fewer objects from the dataset. The total coverage is low because only a small percentage of objects are included in the rule. Whereas when the cover parameter value is high, the generated rules will be more general and cover more objects from the dataset. Total coverage tends to increase as more objects are described by the rule. Based on this theory, the LEM2 results show a match between the relationship between the cover parameter value and the total coverage generated. In contrast, the covering algorithm shows the opposite result, which has a lower total coverage value when the cover parameter value is higher. This happens because there is overgeneralization in the rules generated when the cover parameter value gets higher in the covering algorithm. The more generalized rules become too general that they are not effective enough in capturing specific patterns in the data, so the model cannot classify the data well and is eventually chosen not to be classified because it is considered ambiguous [21][22].

4. CONCLUSION

This study highlights the crucial role of cover parameter values in shaping the performance of LEM2 and Covering algorithms within the Rough Set Theory (RST) framework. By systematically adjusting the cover parameter, the research uncovers the distinct trade-offs between rule generalization and specificity for both algorithms. The LEM2 algorithm demonstrates strong robustness, maintaining high accuracy across different cover parameter settings with only minimal decline at higher values, making it a reliable choice for applications requiring consistent classification performance. Conversely, the Covering algorithm exhibits greater sensitivity to parameter variations, with accuracy and coverage decreasing at higher values, potentially leading to overfitting. Lower cover parameters favor broader rule coverage with relatively high accuracy, whereas higher values yield more specific rules at the cost of reduced coverage. The implications of this research extend beyond theoretical analysis, offering practical guidance for optimizing classification models in real-world applications. In healthcare diagnostics, where rule-based models assist in medical decision-making, understanding the impact of cover parameters can help fine-tune classification systems for more interpretable and reliable predictions. Furthermore, this study provides a foundation for improving the adaptability of rough set-based classifiers, enabling better parameter selection based on dataset characteristics. The findings can also inform future research in machine learning, particularly in developing automated methods for optimizing rule-based classification models across various domains, including finance, cybersecurity, and bioinformatics.

Future research could explore strategies to optimize the selection of cover parameters, potentially using heuristic or machine learning-based approaches to dynamically adjust the values based on dataset characteristics. Additionally, applying this methodology to different domains, such as finance, cybersecurity, or other medical datasets, could help assess the generalizability of the findings. Another potential direction is investigating hybrid models that integrate Rough Set Theory with other classification techniques to improve accuracy and robustness. Furthermore, addressing challenges related to imbalanced datasets and enhancing the interpretability of generated rules could increase the practical applicability of Rough Set-based classification. Finally, real-world validation through collaboration with domain experts, particularly in clinical decision-making, would further strengthen the relevance of the findings for practical implementation.

REFERENCES

[1] W. Z. Wu, J. S. Mi, and W. X. Zhang, “A new rough set approach to knowledge discovery in incomplete information systems,” *Int. Conf. Mach. Learn. Cybern.*, vol. 3, pp. 1713–1718, 2003, doi: 10.1109/icmlc.2003.1259773.

[2] Z. Pawlak, “RoughSets,” *Inst. Theor. Appl. Informatics, Polish Acad. Sci. ul. Bałtycka 5, 44 100 Gliwice, Pol.*, vol. 2, pp. 1–51, 1982.

[3] J. Komorowski, L. Polkowski, and A. Skowron, “Rough sets: A tutorial,” *Rough fuzzy Hybrid. A new trend Decis.*, pp. 3–98, 1999, [Online]. Available: http://secs.ceas.uc.edu/~mazlack/dbm.w2011/Komorowski.RoughSets.tutor.pdf.

[4] L. Polkowski and A. Skowron, “Rough mereology: A new paradigm for approximate reasoning,” *Int. J. Approx. Reason.*, vol. 15, no. 4 SPEC. ISS., pp. 333–365, 1996, doi: 10.1016/s0888-613x(96)00072-2.

[5] R. Jensen and Q. Shen, “Finding rough set reducts with ant colony optimization,” *Proc. 2003 UK Work.*, vol. 1, no. 2, pp. 15–22, 2003, [Online]. Available: http://users.aber.ac.uk/rkj/pubs/papers/antRoughSets.pdf.

[6] H. Rana and M. Lal, “A Rough Set Theory Approach for Rule Generation and Validation Using RSES,” *Int. J. Rough Sets Data Anal.*, vol. 3, no. 1, pp. 55–70, 2016, doi: 10.4018/ijrsda.2016010104.

[7] S. Xia *et al.*, “An Efficient and Accurate Rough Set for Feature Selection, Classification, and Knowledge Representation,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 7724–7735, 2023, doi: 10.1109/TKDE.2022.3220200.

[8] S. Xia, H. Zhang, W. Li, G. Wang, E. Giem, and Z. Chen, “GBNRS: A Novel Rough Set Algorithm for Fast Adaptive Attribute Reduction in Classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1231–1242, 2022, doi: 10.1109/TKDE.2020.2997039.

[9] L. Sun, T. Wang, W. Ding, J. Xu, and Y. Lin, “Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification,” *Inf. Sci. (Ny).*, vol. 578, pp. 887–912, 2021, doi: 10.1016/j.ins.2021.08.032.

[10] Z. Huang and J. Li, “Multi-scale covering rough sets with applications to data classification,” *Appl. Soft Comput.*, vol. 110, p. 107736, 2021, doi: 10.1016/j.asoc.2021.107736.

[11] W. Huang, Y. She, X. He, and W. Ding, “Fuzzy Rough Sets-Based Incremental Feature Selection for Hierarchical Classification,” *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 10, pp. 3721–3733, 2023, doi: 10.1109/TFUZZ.2023.3300913.

[12] J. G. Bazan and M. Szczuka, “The rough set exploration system,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3400 LNCS, pp. 37–56, 2005, doi: 10.1007/11427834_2.

[13] J. W. Grzymala-Busse, “Data with missing attribute values: Generalization of indiscernibility relation and rule induction,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3100, pp. 78–95, 2004, doi: 10.1007/978-3-540-27794-1_3.

[14] P. Zhu, “Covering rough sets based on neighborhoods: An approach without using neighborhoods,” *Int. J. Approx. Reason.*, vol. 52, no. 3, pp. 461–472, 2011, doi: 10.1016/j.ijar.2010.10.005.

[15] J. G. Bazan, M. S. Szczuka, and J. Wróblewski, “A new version of rough set exploration system,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2475, pp. 397–404, 2002, doi: 10.1007/3-540-45813-1_52.

[16] N. Sengupta and J. Sil, “Comparison of Different Rule Calculation Method forRough Set Theory,” *Int. J. Inf. Electron. Eng.*, vol. 2, no. 3, pp. 464–466, 2012, doi: 10.7763/ijiee.2012.v2.138.

[17] P. K. Srimani and M. S. Koti, “Rough Set Approach for Novel Decision Making in Medical Data for Rule Generation and Cost Sensitiveness,” *Adv. Intell. Syst. Comput.*, vol. 249 VOLUME, pp. 303–311, 2014, doi: 10.1007/978-3-319-03095-1_33.

[18] S. Sulaiman, N. A. A. Rahim, and A. Pranolo, “Generated rules for AIDS and e-learning classifier using rough set approach,” *Int. J. Adv. Intell. Informatics*, vol. 2, no. 2, pp. 103–122, 2016, doi: 10.26555/ijain.v2i2.74.

[19] Grzymala-Busse, *Rule Induction, Missing Attribute Values and Discretization*. New York: Springer NY, 2009.

[20] S. Borzooei, G. Briganti, M. Golparian, J. R. Lechien, and A. Tarokhian, “Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study,” *Eur. Arch. Oto-Rhino-Laryngology*, vol. 281, no. 4, pp. 2095–2104, 2024, doi: 10.1007/s00405-023-08299-w.

[21] M. Ye, X. Wu, X. Hu, and D. Hu, “Multi-level rough set reduction for decision rule mining,” *Appl. Intell.*, vol. 39, no. 3, pp. 642–658, 2013, doi: 10.1007/s10489-013-0439-3.

[22] Y. L. Chen, Y. Y. Wu, and R. I. Chang, “From data to global generalized knowledge,” *Decis. Support Syst.*, vol. 52, no. 2, pp. 295–307, 2012, doi: 10.1016/j.dss.2011.08.005.