



(RESEARCH ARTICLE)



## A novel approach to find feature selection using modified fuzzy C-mean

Kumar Siddamallappa U \* and Vinay S \*

*Department of studies in Computer Science, Davangere University, Davangere, India.*

World Journal of Advanced Engineering Technology and Sciences, 2023, 09(02), 359–365

Publication history: Received on 07 July 2023; revised on 23 August 2023; accepted on 25 August 2023

Article DOI: <https://doi.org/10.30574/wjaets.2023.9.2.0224>

### Abstract

Feature selection, also known as dimensionality reduction, is a common preprocessing step in the fields of pattern recognition, data mining, and machine learning. This is a critical issue when mining high-dimensional, massive data sets. Preprocessing the data before analysis to acquire a smaller collection of representative features and keeping the optimal salient properties of the data leads to more compactness of the models trained and better generalization, as well as a decrease in processing time. Therefore, the standard for dimension reduction is to save just the information that is most important from the original data, as determined by certain optimality criteria. In this paper we are going to form a new framework include the combination of MFCM and APSO, to find the HIDS.

**Keywords:** MFCM (modified fuzzy c-mean); APSO (accelerated particle swarm optimization); Feature selection; IDS

### 1. Introduction

An ideal feature subset is always related to a certain criterion in the feature selection process. Generally speaking, various criteria may result in various optimum feature subsets. The capacity of a feature or a combination of characteristics to differentiate between the various class labels is measured by each criterion, though. The dependency or correlation measure is primarily used to determine the correlation between two characteristics or a feature and a class [8], whereas the distance measure is a conventional discrimination or divergence measure. These two measurements are extremely sensitive to noise or outliers in the data set since they depend on the actual values of the training data. On the other hand, information measures that compute the quantity of information or the uncertainty of a feature for categorization include entropy and mutual information [9, 10]. The information measure has been extensively utilised in feature selection since it only depends on the probability distribution of a random variable and not on its actual values [5-7, 9, 10].

The rapid reduct (QR) approach [21, 22], one of the common rough-set-based feature selection techniques, first determines the dependence or quality of approximation of a particular attribute with regard to the class labels or decision attribute. To provide greater quality, more attributes are added to the best attribute once it has been chosen. When the final subset of attributes reaches the highest quality achievable for the data set, or when the quality of the chosen attributes stays the same, attribute additions come to an end. The discernibility-matrix-based technique [23, 24] and dynamic reducts [25] are two further noteworthy methods. Recently, Parthala et al. [26] reported on a distance-measure-based strategy to investigate the rough set border region for feature selection. All of these methods, meanwhile, need a lot of computer power. For feature selection, other heuristic methods built on the rough set theory are also developed [27, 28]. Several techniques have been suggested [29–31] to find an optimal or nearly optimal subset of characteristics by combining rough sets with evolutionary algorithms.

Maji and Pal [13] provided a method for feature selection that uses fuzzy-rough sets to efficiently minimise discrete or real-valued noisy data, or a combination of both, without the requirement for user-specified information. Additionally,

\* Corresponding author: Kumar Siddamallappa U (Orcid Id: [0000-0002-1975-3868](https://orcid.org/0000-0002-1975-3868)); Vinay S. (Orcid Id: [0000-0002-2327-4915](https://orcid.org/0000-0002-2327-4915))

the technique may be used to classify and regression data sets as well as data with continuous or nominal decision criteria. By increasing the relevance and decreasing the redundancy of the chosen features, the approach chooses a subset of features from the entire feature set. The fuzzy approximation spaces' f-information measures are used to determine the relevance and redundancy of the feature sets. The f-information measures for both condition and choice characteristics are derived using the fuzzy equivalence partition matrix (FEPM) idea. As a result, the only data needed for the present feature selection approach is in the form of fuzzy partitions for each characteristic, which may be generated automatically from the provided data set. This removes the requirement for subject matter experts to offer details on the data at hand and links into the benefit of rough sets in that it just needs the data set itself.

### 1.1. Fuzzy C-Means

FCM is a technique for unsupervised learning that categorises data samples without using class label information. The fundamental benefit of fuzzy c-means clustering is that it enables data points to gradually join clusters that are quantified as degrees in [0,1]. The ability to describe that data points might be a part of several clusters is made possible by this. FCM is a data clustering approach in which a data set is divided into N clusters, each of which contains some of the data points in the dataset. Performance of the fuzzy c-means algorithm is superior than that of k-means. We decide on fuzzy c-means because of this.

### 1.2. Particle swarm optimization

PSO, an optimization strategy based on research on the behaviour of animals and birds. PSO is a strategy for global search that is effective and efficient [4, 5]. Due to superior representation, the capacity to explore huge areas, being less computationally expensive, being simpler to construct, and needing fewer parameters, it is a suitable technique to handle feature selection difficulties.

### 1.3. Intrusion Detection system

This is why we've settled on the IDS as the target for our unique feature selection planning. When an IDS scans a network or system, it looks for suspicious behaviour or violations of software policies. It's a traffic system that scans network transmissions for signs of malicious behaviour and notifies administrators immediately. Preparing IDSs to distinguish between benign and malicious network activity requires careful traffic.

### 1.4. Host Intrusion Detection System (HIDS):

A host intrusion detection system (HIDS) is a type of IDS that operates on its own computers or other network nodes. The packets entering and leaving the device are the only ones a HIDS watches, and if anything unusual is identified, the administrator will be notified. It creates a copy of the current system files and then checks for changes since the last backup. As soon as any changes are made to, or files are removed from, the analytical system, an alert is issued to the administrator.

---

## 2. Methodology

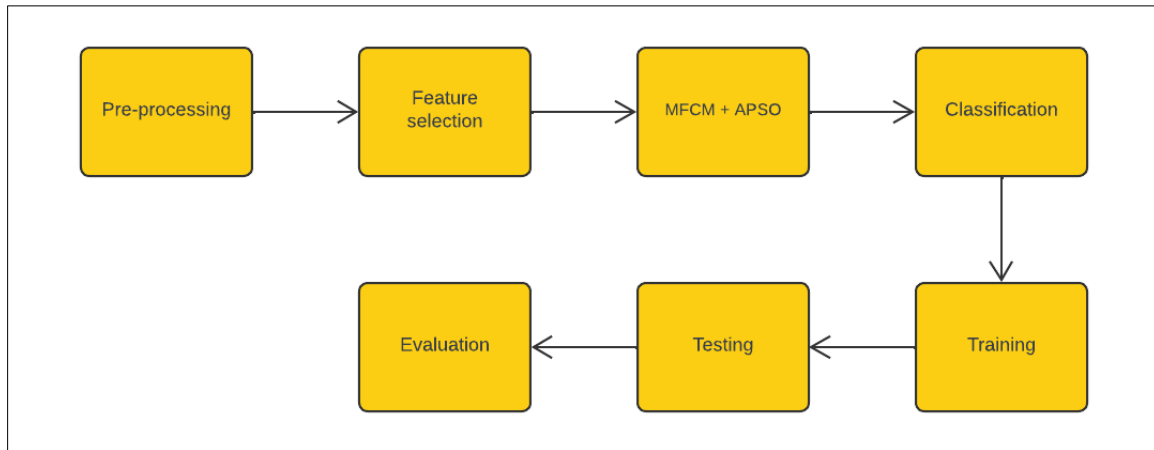
We propose a seven-step approach that includes pre-processing, feature selection, MFCM and APSO, classification, training, testing, and assessment. To test this hypothesis, the NSL-KDD dataset was employed in the suggested model. The NSL-KDD data set is an updated version of the KDD'99 dataset. This data set serves as a useful standard against which researchers can evaluate other HIDS (Host intrusion detection methods). As the raw features confuse the classifier, leading to false alarms, preprocessing is required. In addition, the computational and memory requirements of the classification methods are not fully explored, and the few symbolic characteristics that augment them are underutilised. An expression for the NSL KDD dataset's raw feature set is

$$A_f = \{ax_1, ax_2, ax_3, \dots, ax_n\} \dots \dots \dots (1)$$

where  $n=21$  denotes that the raw dataset has forty-one characteristics. In order to reduce the amount of unnecessary overhead in the learning process, the symbolic features are removed from the raw feature set. The unprocessed feature set obtained looks like this:

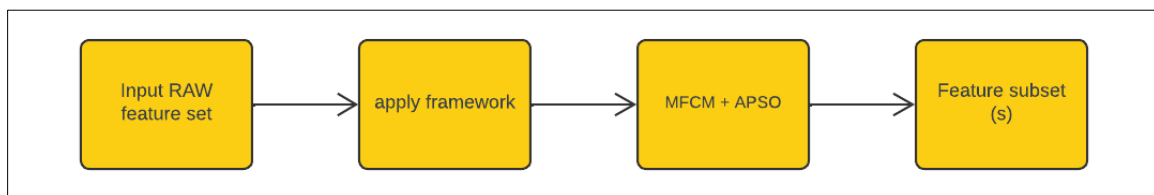
$$A_f' = \{ax_1, ax_2, ax_3, \dots, ax_m\} \dots \dots \dots (2)$$

where the raw feature count ( $m$ ) is set to 18. This raw feature collection needs further preprocessing to choose significant characteristics based on their sensitivity. In Feature Selection, we explore the many methods that may be used to change characteristics and pick the most salient ones for this purpose.



**Figure 1** Proposed framework for feature selection and evaluation

Feature selection's goal is to establish how few feature subsets are required to adequately train a classifier to distinguish between benign and malicious associations or activities. Always,  $p$  is a smaller feature than  $m$ , the initial number of features in the feature set from which it was derived. Figure 2 illustrates the selection of selecting feature subsets. In what follows, we'll go through each of the steps in Figure 2 in step.



**Figure 2** Feature selection flow

In the context of data analysis, the statistical selection known as accelerated particle swarm optimization (APSO) can be a powerful feature for determining which features to prioritise. To further highlight the significance of the data and make them more readily apparent, the APSO is used to turn the raw features into primary features. The last several years have seen this method employed in a variety of fields [21]. Higher eigenvalue traits are chosen in this method, whereas lower eigenvalue features are disregarded. Due to the possibility of losing certain crucial traits, this way of feature selection is not optimum. What feature is crucial, and what is not? How do you know which feature is picked and which isn't in main space? Both APSO and MFCM have been shown to be effective in the past in solving similar optimization issues, making them the best choices here. Therefore, in this study, APSO and MFCM were employed for feature selection while the suggested new framework was used just for feature transformation. The used swarm algorithm is outlined.

### 2.1. Accelerated particle swarm optimization

In [14], a hybrid algorithm is suggested as a means of improving optimization. It doesn't matter what you put in as a beginning value or how you adjust the parameters, APSO and MFCM will still work. However, the APSO matures slowly because of the variable step length and unpredictable behaviour. As a result, the optimization accuracy is rarely good enough, leading to blind spots in the search and a lack of capacity to strike a good balance between discovery and exploitation. The inclusion of crossover grants the particle-swarm the ability to hop to and investigate a more varied region. Particle swarms can skip past the blind search phase and acquire their parents' benefits with a little help from the crossover operator. It was the MFCM concept that motivated them to include the APSO's belief-space-stored normative information and situational knowledge into the MFCM. In this model, the swarm serves as a population space from which subject expertise may be mined. Accordingly, we model and affect the population's evolution iteratively by creating and storing domain knowledge in belief space.

The following equation is used to decide if the algorithm is stuck in local minima:

$$\left\| \frac{f(x_i^t) - f(x_i^{t-1})}{f(x_i^{t-1})} \right\| < 0.1 \dots\dots\dots (3)$$

when the above criterion is satisfied crossover is applied on the *i*th artificial fish as inequation 3.

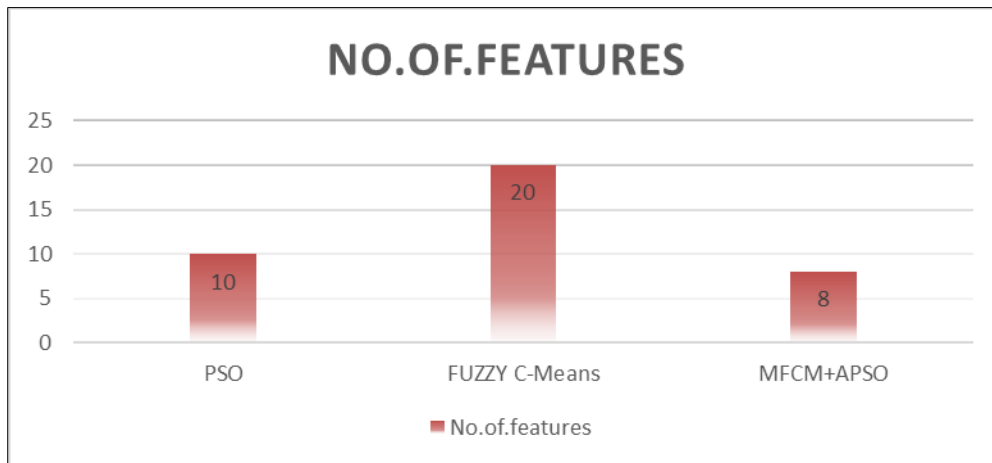
$$x_i^{new} = x_{r1} + \alpha(x_{r2} - x_{r1}) \dots\dots\dots (4)$$

The main steps for the proposed one:

- Initialize the N particle-swarm in the search ranges with random places and set all the parameter values.
- Start with a blank belief space and evaluate all particle-swarms with the fitness function *y*.
- To find the best offspring, we must replicate each particle swarm's preying behaviour, swarming pattern, and chasing pattern independently. If the offspring is superior, it should be substituted for the *i*th particle-swarm.
- Revise the assumption set.
- If the crossover condition holds, the *i*th particle-swarm from Step 3 will be subjected to the crossover operator.
- In order to keep going until the end requirement is met, go back to Step 3.

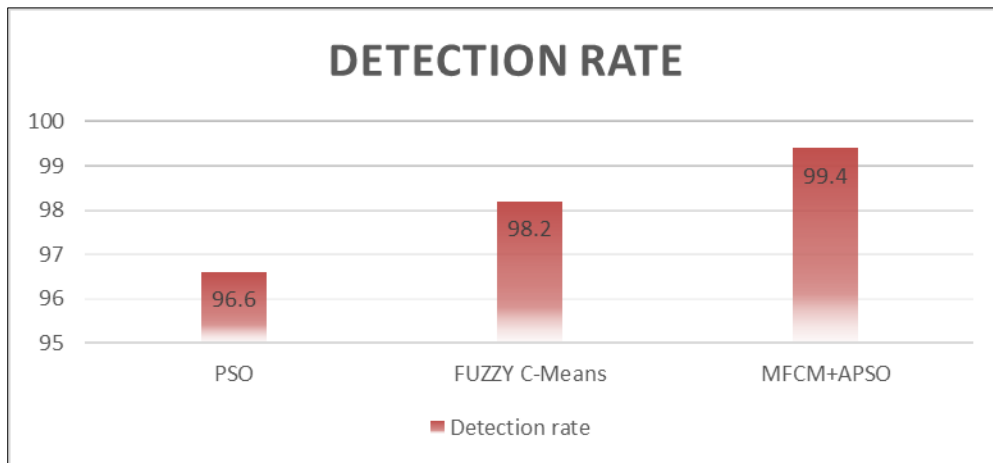
### 3. Result and discussion

There is a direct comparison of feature selection methods that uses different numbers of characteristics, as shown in Figure 3. The raw datasets contain 38 features, the PCA conventional based method dataset has 20, the PCA + GA based selection method dataset contains 10, and the PCA + PSO based method dataset contains 8 features. As a result, the final approach delivered a more manageable collection of characteristics that enhanced the classifier's efficiency and reduced the complexity of its underlying architecture.



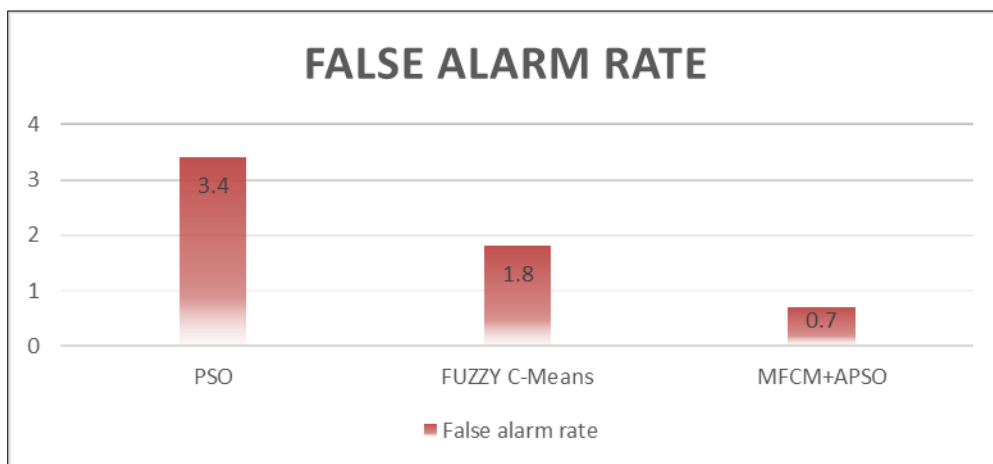
**Figure 3** Comparison of feature selection techniques based on number of features

In Figure 4 we see a comparison of several feature selection methods in terms of their detection rates. A raw feature set is outperformed by the obtained feature subsets. In addition, the MFCM+APSO-based feature selection indicates the optimal performance with the fewest characteristics.



**Figure 4** Comparison of feature selection techniques based on detection rate

The false alarm rate comparison among the feature selection strategies is shown in Figure 5. When compared to alternative approaches, the MFCM+APSO-based technique of feature selection has a lower false-alarm rate.



**Figure 5** Comparison of feature selection techniques based on false alarm rate

#### 4. Conclusion

There is no shortage of intrusion detection methods, however the most of them suffer from poor performance. By suggesting a good approach of feature selection and categorization, we can boost performance. This study proposes a technique for feature selection in Host-intrusion detection for wireless sensor networks that, using APSO and MFCM, can pick the best features to use from either the primary space or the sample KNN space. The suggested method is evaluated using the NSL KDD dataset, which is widely recognised as the gold standard for testing intrusion detection approaches. After validating the chosen feature subset (based on APSO) on a modular neural network, and contrasting it with another feature subset (based on MFCM) and other approaches, the former is deemed to be the superior choice (KNN based, raw). The outcomes show that the feature selection approach based on APSO and MFCM works better than the current methods.

#### Compliance with ethical standards

##### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

- [1] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, Englewood Cliffs, NJ, 1982.
- [2] D. Koller and M. Sahami. Toward Optimal Feature Selection In Proceedings of the International Conference on Machine Learning, Bari, Italy, pages 284–292, 1996.
- [3] R. Kohavi and G. H. John. Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(1–2):273–324, 1997.
- [4] I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
- [5] J. R. Quinlan. Induction of Decision Trees. *Machine Learning*, 1:81–106, 1986.
- [6] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco, CA, 1993.
- [7] J. R. Quinlan. Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research*, 4:77–90, 1996.
- [8] M. Dash and H. Liu. Consistency Based Search in Feature Selection. *Artificial Intelligence*, 151(1–2):155–176, 2003.
- [9] R. Battiti. Using Mutual Information for Selecting Features in Supervised Neural Net Learning. *IEEE Transactions on Neural Network*, 5(4):537–550, 1994.
- [10] D. Huang and T. W. S. Chow. Effective Feature Selection Scheme Using Mutual Information. *Neurocomputing*, 63:325–343, 2004.
- [11] I. Vajda. *Theory of Statistical Inference and Information*. Kluwer Academic, Dordrecht, The Netherlands, 1989.
- [12] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. *f-Information Measures in Medical Image Registration*. *IEEE Transactions on Medical Imaging*, 23(12):1508–1516, 2004.
- [13] P. Maji and S. K. Pal. Feature Selection Using *f-Information Measures in Fuzzy Approximation Spaces*. *IEEE Transactions on Knowledge and Data Engineering*, 22(6):854–867, 2010.
- [14] P. Maji. *f-Information Measures for Efficient Selection of Discriminative Genes from Microarray Data*. *IEEE Transactions on Biomedical Engineering*, 56(4):1063–1069, 2009.
- [15] P. Maji and S. K. Pal. Fuzzy-Rough Sets for Information Measures and Selection of Relevant Genes from Microarray Data. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 40(3):741–752, 2010.
- [16] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer, Dordrecht, The Netherlands, 1991.
- [17] Q. Shen and A. Chouchoulas. Combining Rough Sets and Data-Driven Fuzzy Learning for Generation of Classification Rules. *Pattern Recognition*, 32(12):2073–2076, 1999.
- [18] A. Skowron, R. W. Swiniarski, and P. Synak. Approximation Spaces and Information Granulation. *LNCS Transactions on Rough Sets*, 3:175–189, 2005.
- [19] K. H. Chen, Z. W. Ras, and A. Skowron. Attributes and Rough Properties in Information Systems. *International Journal of Approximate Reasoning*, 2:365–376, 1988.
- [20] D. Yamaguchi. Attribute Dependency Functions Considering Data Efficiency. *International Journal of Approximate Reasoning*, 51:89–98, 2009.
- [21] A. Chouchoulas and Q. Shen. Rough Set-Aided Keyword Reduction for Text Categorisation. *Applied Artificial Intelligence*, 15(9):843–873, 2001.
- [22] C. Cornelis, R. Jensen, G. H. Martin, and D. Slezak. Attribute Selection with Fuzzy Decision Reducts. *Information Sciences*, 180:209–224, 2010.
- [23] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron. Rough Sets: A Tutorial. In S. K. Pal and A. Skowron, editors, *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, pages 3–98. Springer-Verlag, Singapore, 1999.
- [24] A. Skowron and C. Rauszer. The Discernibility Matrices and Functions in Information Systems. In R. Slowinski, editor, *Intelligent Decision Support*, pages 331–362. Kluwer Academic, The Netherlands, Dordrecht, 1992.

- [25] J. Bazan, A. Skowron, and P. Synak. Dynamic Reducts as a Tool for Extracting Laws from Decision Tables. In Z. W. Ras and M. Zemankova, editors, *Proceedings of the 8th Symposium on Methodologies for Intelligent Systems*, volume 869, pages 346–355. *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Charlotte, North Carolina, 1994.
- [26] N. Parthalaïn, Q. Shen, and R. Jensen. A Distance Measure Approach to Exploring the Rough Set Boundary Region for Attribute Reduction. *IEEE Transactions on Knowledge and Data Engineering*, 22(3):305–317, 2010.
- [27] M. Modrzejewski. Feature Selection Using Rough Sets Theory. In *Proceedings of the 11th International Conference on Machine Learning*, Amherst, MA, pages 213–226, 1993.
- [28] N. Zhong, J. Dong, and S. Ohsuga. Using Rough Sets with Heuristics for Feature Selection. *Journal of Intelligent Information Systems*, 16:199–214, 2001.
- [29] A. T. Bjorvand and J. Komorowski. Practical Applications of Genetic Algorithms for Efficient Reduct Computation. In *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modeling and Applied Mathematics*, Berlin, volume 4, pages 601–606, 1997.
- [30] D. Slezak. Approximate Reducts in Decision Tables. In *Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Granada, pages 1159–1164, 1996.
- [31] J. Wroblewski. Finding Minimal Reducts Using Genetic Algorithms. In *Proceedings of the 2nd Annual Joint Conference on Information Sciences*, North Carolina, pages 186–189, 1995.