Swarm Intelligence-Based Feature Selection for Multi-Label Classification: A Review

Adnan Mohsin Abdulazeez1, Dathar A. Hasan2*, Awder Mohammed Ahmed3 and Omar S. Kareem2

1Research Center of Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
2Shekhan Technical Institute, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
3IT Department of Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.

Authors' contributions
This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information
DOI: 10.9734/AJRCOS/2021/v9i430230
Editor(s):
(1) Dr. Hasibun Naher, BRAC University, Bangladesh.
Reviewers:
(1) Lenin Kanagasabai, Prasad V. Potluri Siddhartha Institute of Technology, India.
(2) Ansam Nazar Younis, University of Mosul, Iraq.
Complete Peer review History: http://www.sdiarticle4.com/review-history/69824

Received 10 April 2021
Accepted 15 June 2021
Published 18 June 2021

ABSTRACT

Multi-label classification is the process of specifying more than one class label for each instance. The high-dimensional data in various multi-label classification tasks have a direct impact on reducing the efficiency of traditional multi-label classifiers. To tackle this problem, feature selection is used as an effective approach to retain relevant features and eliminating redundant ones to reduce dimensionality. Multi-label classification has a wide range of real-world applications such as image classification, emotion analysis, text mining and bioinformatics. Moreover, in recent years researchers have focused on applying swarm intelligence methods in selecting prominent features of multi-label data. After reviewing various researches, it seems there are no researches that provide a review of swarm intelligence-based methods for multi-label feature selection. Thus, in this study, a comprehensive review of different swarm intelligence and evolutionary computing methods of feature selection presented for the tasks of multi-label classification. To this end, in this review, we have investigated most of the well-known and state-of-the-art methods and categorize them based on different perspectives. We then provided the main characteristics of the existing multi-label feature selection techniques and compared them analytically. We also introduce benchmarks, evaluation measures and standard datasets to facilitate research in this field. Moreover, we...
performed some experiments to compare existing works and at the end of this survey, some challenges, issues and open problems of this field are introduced to be considered by researchers in future.

Keywords: Multi-Label classification; feature selection; wrapper methods; filter methods; curse of the dimensionality.

1. INTRODUCTION

Data mining refers to the operation of extracting valuable information from large data sets using machine learning and statistical methods. The objective of data mining is to extract useful patterns from data [1-5]. Recently, big data became an interesting topic because it consists of numerous features of high-dimensional data [6]. High dimensionality in the datasets causes a reduction of the classification model performance and consuming more time in the learning process [7]. Therefore, Feature Selection (FS) techniques are used to eliminate redundant and irrelevant features [8-10]. FS techniques are wrapper, filter, and embedded. The wrapper technique first uses a learning model to search through the search space, to evaluate sub-sets of features; filter techniques use information-theoretical methods to assess a subset of features [11-15]. In comparison to filter methods, wrapper methods have better performance, but they rely on a classifier that requires a high computational cost by combining the choice of functions as part of a learning model, embedded forms use the advantage of both wrapper and filter approaches.

Filter techniques do not require any model of learning; therefore, it is very faster than the wrapper techniques. This is why they have recently been used for many applications in the real world. In the other direction, in real-world applications such as text classification, each instance [16,17], image annotation [18,19], gene selection [20-23], information retrieval [24], face recognition [25-28] and cancer classification [29-35], can be associated with multiple class labels. In these tasks, more than one class label can be assigned to each instance.

There are some multivariate methods, such as a MUMI [36], MDMR [37], IGMF [38], SCLS [39], and MLACO [40]. Most of the feature selection approaches are utilized in a single-label application in which only one label is used to assign instances. There are some review papers such as [11,41-44] which provides an introduction and comparison for existing methods for selecting single-label features. In most real-world applications, multi-label classification tasks are used where each instance is associated with multiple class labels. These applications applied feature selection methods applied to the multi-label classification tasks. When selecting a certain task in multi-label feature selection, we have to consider the complicated relations between the features and each class label. There is no more review paper in this area of subject. In spite of there are some review papers, it is better to provide a review of existing works in an analytical manner.

Fig. 1. Multi-label classification example

Swarm intelligence (SI) and Evolutionary Algorithms (EA) are the main two fields of metaheuristic algorithms. Comparing with EA, SI has considered a modern and more effective other single solution-based methods fall under the evolutionary computation section. In SI, the near-optimal solution is determined using approximate and non-deterministic based algorithms. Different applications employ metaheuristic algorithms to minimize processing time and computational complexity. SI is included as one of the nature-inspired artificial intelligence techniques that simulate individual behaviours in self-organized and decentralized systems [45,46].

This paper provides a systematic review of various swarm intelligence for multi-label classification activities, evolutionary computing methods of feature selection are described. To this end, in this review, we have investigated most of the well-known and state-of-the-art methods and categorize them based on different perspectives. The main properties of the existing multi-label feature selection methods are
introduced. Thereafter, some experiments are performed on these methods to compare them analytically. We also introduce benchmarks, evaluation measures and standard datasets to facilitate research in this field. Moreover, we performed some experiments to compare existing works and at the end of this survey, some challenges, issues and open problems of this field are introduced to be considered by researchers in future.

The remainder of the paper is organized as follows. Section 2 provides the formal definition of multi-label classification. Section 3 reviews existing multi-label feature selection methods. Section 4 discusses the experimental set of multi-label feature selection methods and also reports the comparison results obtained on the performed experiments on the existing works. Finally, the paper is concluded in Section 6.

1.1 Literature Review

Multi-label classification refers to a set of tasks with multi-output classification. Contrarily to supervised learning, each data sample can be assigned multiple labels. Various real-world applications have semantic relations with multiple class label such as text classification, image annotation, protein function classification, music categorization, and semantic scene classification. For example, one can assign more than one label (i.e., Sky, Man, Dog, Shoes, and so on) to the following image.

As another example, in the text classification task, each document can be simultaneously belonging to more than one topic. We first demonstrate the formal definition of the multi-label data and then we provide a taxonomy of the multi-label classification algorithms. Suppose $X = R^M$ denotes an M-dimensional sample space and $L = \{y_1, y_2, ..., y_q\}$ is the label space with $q$ distinct labels. The problem of the multi-label classification is to train a function $h: X \rightarrow 2^q$ from the training set $TS = \{(x_i, y_i), i = 1,2,...,N\}$ with N samples that each sample is a vector in the form of $x_i = \{x_{i1}, x_{i2}, ..., x_{iM}\}$ presented by M features and a set of q labels $Y_i = \{y_{i1}, y_{i2}, ..., y_{iq}\}$. This representation is shown in Fig. 2:

Till now, a variety of solutions were proposed to overcome the problem of multi-label task from various perspectives and have gained valuable results. The training set consists of samples assigned by several labels in the multiple class labels, and the task is predicting the labels for test set data by analysing the training dataset with known label sets. The classification methods of multi-label are divided into (1) problem conversion and (2) adaptive methods. Additional details on these methods are provided in the subsections that follow.

| $X$ | $Y$ |
|-----|-----|
| $x_{11}$ $x_{12}$ ... $x_{1M}$ | 0 1 1 0 |
| $x_{21}$ $x_{22}$ ... $x_{2M}$ | 1 1 0 1 |
| ... ... ... ... | ... ... ... ... |
| $x_{N1}$ $x_{N2}$ ... $x_{NM}$ | 1 0 1 0 |

Fig. 2. Multi-Label Instance Representation

1.2 Transformation Methods

The objective of the methods of problem transformation is to transform the task into a single-label issue and then apply traditional classifiers such as Neural Networks or Naïve Bayes. This approach can be carried out through three primary methods: Binary relevance, Classifier chains, and Label power set. On the other hand, methods of transformation aim to map the multi-label classification of the problem into several single-label tasks. Afterwards, to decrease the dimensionality, any single-label feature selection techniques can be used. Binary Relevance (BR) (Boutell et al., 2004) aims to split multi-label classification task into many binary subtasks, as shown in Fig 3. The label power set (LP) method [47-49] deals with each uniform label’s subset to be single-label and uses in a single-label classifier. Using a single-label classifier to classify test samples, this method predicts its label set and then maps it back to the power set. In this method, the number of classes grows by increasing the labels, although it considers the correlation of the labels. However, predictive performance and the computational efficiency of LP are reduced with growing the number of labels and training samples. The PPT (Pruned Problem Transformation) method is an extension of the LP process. The concept of PPT is to prune those sets of labels that have appeared below a predefined threshold.Instances belonging to the pruned labels can either be assigned or eliminated to another class. This method eliminates the difficulty and grouping error associated with a large number of infrequent sets. The entropy-based label assignment method (ELA) uses labels' entropy to assign labels' weights in a multi-label document.
Adaptive Methods directly perform multi-label classification without transforming it to single-label classification tasks. To define the final set of features, adaptive methods scan directly through the entire solution space of features and labels. For example, MUMI [36] maximizes the correlation between chosen features and the class labels. This method uses the concept of mutual information in an incremental way to compute the correlation between variables. [37] incorporates the parameters for max-dependency and min-redundancy with reciprocal knowledge for the gradual estimation of class label features. The dependence shows how a function relates to each label, and the redundancy shows the connection between the candidate features and the entire labels considered by the selected features. [50] proposed a method that considers the correlation between features and labels without converting the task into a single-label problem. [51] proposed three measures based on mutual neighborhood information to compute the quality of candidate features, as shown in Fig 4. These measures use the margin of the instance to granulate all instances under various labels.

2. RELATED WORKS

The main objective of the selection of multi-label features is to identify a set of prominent characteristics with minimum redundancies that are most relevant to a set of target classes. In general, these methods are divided into two main categories: data transformation and adaptive techniques. Methods of transformation aim to convert the multi-label task to a single-label study and then use a traditional method of feature selection to decrease dimensionality. In contrast, adoption methods are applied directly to the multi-label space to reduce dimensionality. A majority part of single-table and multi-label feature selection methods employ swarm intelligence methods. To this end, in this section, we have a comprehensive survey on swarm intelligence-based feature selection methods, and then we have reviewed existing multi-label transformation and adoption methods. In their corresponding sections, the details of these methods are described.
For feature selection, swarm intelligence such as Particle Swarm Optimisation (PSO) [52,53], Artificial Bee Colony (ABC) [54-56], and ACO and evolutionary methods such as Genetic Algorithm (GA) [50] have been used successfully. Among them, ACO has gained better results than the others. In general, ACO-based methods are categorized into Wrapper and Filter methods. To evaluate a set of characteristics selected by each ant through its traverse, wrapper techniques use a learning model. Filter methods, by contrast, use an information-theoretical measure to evaluate the results of ants. For example, the authors of [57,58] proposed an ACO-based unsupervised feature selection method called UFSACO. This method uses a measure to compute the redundancy of features. The relevance of features to the target class is not considered by this method, however. A multivariate filter-based relevance-redundancy feature selection method called RRFSACO was suggested by the same authors in [57]. This method uses ACO and considers in its search process both the concepts of relevance and redundancy. In [26], an unsupervised probabilistic feature selection method called UPFS is proposed. To reduce the redundancy between functions within the iterative search process of ACO, this method uses ACO and looks for the optimal feature subset by considering the inter-feature data. In [59], the authors proposed an unsupervised filter method called GCACO, which combines ACO with graph clustering for feature selection. GCACO first divides similar features into clusters and then uses the search strategy of ACO to rank features. This method reduces the redundancy by force the method to prevent choosing similar features. To this aim, GCACO assigns a penalty for ants to remain in the same cluster the authors of [60] suggested a method called MGCACO, which extends GCACO by proposing a measure to evaluate both the relevance and redundancy of characteristics using multiple discriminant analysis (MDA) [61], proposed an ACO-based feature selection ABACO which combines Artificial Bee Colony (ABC) with ACO. This method’s features are mapped to a fully connected graph where each node has two sub-nodes for determining the selection or deselection of features. Ant colony optimization algorithms have been used to select nodes [30] proposed a modified binary-coded ACO integrated with the genetic algorithm. [62] proposed the method called MBACO uses the pheromone density model to initialize pheromones. Recently [63] proposed a specific update rule for ACO, which prevents the algorithm from falling into the local optima. To this aim, this method improves the path transfer probability method by adding pheromones to more paths. The feature selection task is regarded as a multi-objective task by some ACO-based methods. For example, [64] suggested a method of multi-objective ACO-based feature selection to classify disturbances in power quality. This method evaluates the solutions by using two contradictory objectives: one for reducing the feature subset's size and the other for improving the classifier performance. The authors of [65] introduced a memory to keep the best ants and heuristic desirability in ACO is computed using a specific strategy to make the algorithm computationally efficient than its ancestors. The authors of [66] proposed the key principle component analysis (PCA) is used to classify redundant characteristics and then to pick the final function subset using the genetic algorithm(GA). This method uses the multi-label naïve Bayes (MLNB) classified to evaluate each solution of GA. The authors of [67] proposed a classification algorithm for multi-label features (through the integration of shared knowledge with GA). This method uses GA to search through the solution space and employ mutual information to determine the importance of the features for each label. The authors of [51] use mutual information to compute the label granularity to identify the relationships between labels and features. In [68] the constrained convex optimization was used to maximize relevance and minimizing the redundancy simultaneously proposed a measure called “label frequency difference” (LFD) to compute the conditional frequencies of labels to recognize the discrimination power of features. LRFs [69] analysed the differences between labels and them into two independent and dependent groups on computing the label redundancy. A PSO-based multi-objective multi-label feature choice algorithm called MPSOFS was suggested in [70]. This method first transforms the feature selection task into an ongoing problem. It employs non-dominated comparison, the crowding distance, and probability-based encoding concepts to convert a separated features problem into a suitable continuous one. The authors employed the idea to prune the archive. IGMF [38] computes the label correlations by using the Information gain between features and labels. Recently, the authors of [71] presents a many-objective optimization based multi-label feature selection algorithm (MMFS). Moreover, in [53] a PSO based multi-label learning and the arrival of features in an online fashion was proposed. The
3. RESULTS AND DISCUSSION

In section two, each transformation method and an adaptive method has been viewed. Recently, researchers have focused on adaptive and Swarm intelligence-based methods due to their success in gaining performance. This is why, in this section, we have only compared these methods. Table 1 summarizes the main properties of the existing works.

Methods for selecting multi-label features may be classified based on the label and search technique. Training samples may be labelled from a labelling standpoint, unlabelled and semi-labelled. Therefore, this property leads to the prescience of three categories of multi-label feature selection methods include supervised, unsupervised and semi-supervised feature selection methods, respectively. It is assumed that training samples contain class labels in supervised methods and are expected to have high accuracy. Multi-label ranker (MLFR) [76], multi-label feature selection based on the information gain (IGMF) [77], and multi-label correlation-based feature selection (ML-CFS) [78] are some well-known examples in this category. The presence of the labels in the training samples leads to enhance the accuracy. On the other hand, all wrapper methods such as MR2PSO [10], ABACO [13], OFS [20], and MPSOFS [26] require the class labels to train the multi-label classifiers. Moreover, the wrapper methods in each iteration use a learning model such as MLKNN or MLNB to evaluate the feature subset. Training a learning model is a time-consuming process, and thus, we can claim that wrapper methods can be scalable for large-scale and real-world applications. On the other hand, the filter methods employ some information-theoretic criteria to evaluate feature subsets instead of running a learning model. Therefore, these methods are much faster than the wrapper methods. For example, MGCACO [16], MDMR [3], MUMI [5], MICO [68], and FACO [19] are some well-known supervised filter methods. All these methods require the class labels to search through the solution space. While in many real-world applications, it is hard or time-consuming to provide labels to the training samples.

Unsupervised feature selection methods are categorized as filter methods, which do not require any class labels. Among them, MR2PSO [10] are well-known and fast unsupervised filter methods whose performance is comparable with many wrapper and supervised-filter methods. This method uses the ACO in its search process and employs a specific metric to evaluate the ants’ features. Moreover, in many real-world applications, there are many samples without class labels, and only a few samples consist of class labels. In this situation, supervised learning methods cannot be utilized because of a few samples containing class labels. Therefore, it is crucial to develop some semi-supervised feature selection methods, which considers both labelled and unlabelled training samples in their processes. A majority of semi-supervised feature selection methods convert the solution space into a graph and then propagate the labels of those labelled samples to unlabelled ones.

3.1 Experimental Evaluation

The performance of the existing works is evaluated using six well-known real-world and diverse datasets. These datasets include Arts, Birds, CAL500, Computer, Corel5K, Education, and Yeast. All these datasets can be accessed from Mulan Library. Table 2 summarizes the properties of these datasets. This table for each dataset shows the dataset name (Name), the
### Table 1. Result of all evaluation metrics by using Naïve Bayes

| Methods          | Year | - Single label - Multi-label | Search Process | Approach type | Relevancy-Redundancy | Filter/Wrapper |
|------------------|------|------------------------------|----------------|--------------|----------------------|----------------|
| MMLACO[72]       | 2021 | ML                           | ACO            | AD           | MV                   | Filter         |
| MMOFS[53]        | 2021 | ML                           | PSO            | AD           | MV                   | Filter         |
| MMFS[71]         | 2020 | ML                           | GA             | AD           | MV                   | Wrapper        |
| FMABC-FS[56]     | 2020 | SL                           | ABC            | -            | -                    | Filter         |
| MICO [68]        | 2019 | ML                           | MI             | AD           | MV                   | Filter         |
| LRFS [29]        | 2019 | ML                           | MI             | AD           | MV                   | Filter         |
| MLACO[40]        | 2019 | ML                           | ACO            | AD           | MV                   | Filter         |
| OFS [20]         | 2019 | SL                           | ACO            | -            | -                    | Wrapper        |
| WFACOFS [65]     | 2019 | SL                           | ACO            | -            | -                    | Wrapper - Filter|
| MGCACO [16]      | 2018 | SL                           | ACO            | -            | -                    | Filter         |
| FACO [19]        | 2018 | SL                           | ACO            | -            | -                    | Filter         |
| SCLS [25]        | 2017 | ML                           | MI             | AD           | MV                   | Filter         |
| MPSOFS [26]      | 2017 | ML                           | PSO            | AD           | UV                   | Wrapper        |
| GMFS[27]         | 2017 | ML                           | MI             | AD           | MV                   | Wrapper        |
| MFNMI[24]        | 2016 | ML                           | MI             | AD           | MV                   | Filter         |
| MBACO[14]        | 2016 | SL                           | ACO + GA       | -            | -                    | Wrapper        |
| MDMR[3]          | 2015 | ML                           | MI             | AD           | MV                   | Filter         |
| GA-ML-CFS[23]    | 2015 | ML                           | GA             | AD           | UV                   | Filter         |
| ABACO[13]        | 2015 | SL                           | ACO            | -            | -                    | Wrapper        |
| RRRFSACO[12]     | 2015 | SL                           | ACO            | -            | -                    | Filter         |
| GCACO [8]        | 2015 | SL                           | ACO            | -            | -                    | Filter         |
| PSO-RR[11]       | 2014 | SL                           | PSO            | -            | -                    | Wrapper        |
| UFSACO[21]       | 2014 | SL                           | ACO            | -            | -                    | Filter         |
| MUMI[5]          | 2013 | ML                           | MI             | AD           | MV                   | Filter         |
| MR2PSO[10]       | 2011 | SL                           | PSO            | -            | -                    | Wrapper        |

*Where: DT is Data Transformation, AD is Adaptive Methods, IG is Information Gain, ACO is Ant Colony Optimization, GA is Genetic Algorithm, NB is Naïve Bayes, MI is Mutual Information, PR is Page Rank UV is Univariate, MV is Multivariate, SL is Single Label, ML is Multi-Label, DT is Data Transformation, MI is Mutual Information, PR is Page Rank, UV is Univariate, and MV is Multivariate.*
number of samples (Instance), the number of features (Features), the number of labels (Labels), dataset density (Density), feature type (Type), and dataset domain (Domain).

Moreover, our method is compared to six state-of-art multi-label feature selection, and filter-based methods include MDMR [37], SCLS [39], MGFS [79], MLACO [40], AMI [80], and IGMF [38].

In the experiments, to evaluate the performance of the proposed method, the ML-KNN classifier has been employed. The ML-KNN is a multi-label version of the traditional and famous KNN (k-nearest neighbour classifier) [81] with k=10 neighbors. ML-KNN works by detecting the k nearest neighbors for each sample at first in training data, and then the test samples will be assigned to the labels set, which is most popular among its neighbor. It identifies k nearest neighbors for each unseen instance in the training set. Then, the maximum posterior (MAP) which is statistical information gained from the labels set has been applied to determine the set of labels for the unseen instances. In single-label feature selection, an instance can be classified incorrectly or correctly. While, in multi-label feature selection, the problem is much harder and difficult than the single-label feature selection, as the predicted label subset has very different from the ones in the actual label subset. Thus, there are different criteria evaluation for multi-label methods than those used in single-labels. To show the power of prediction of the used classifier ML-KNN and then show our method's performance, we have used five well-known multi-label evaluation measures: Hamming loss, Average precision, Ranking loss, F1-micro, and F1-macro [82].

- **Hamming loss**: This measure computes the average difference between the predicted and ground-truth labels, and it is defined as:

$$HL = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{m} y_{ij} \oplus y_{ij}'}{m}$$  \hspace{1cm} (1)

Hamming loss will be achieved to the best performance when it is approaching 0.

- **Ranking Loss**: This measure evaluates the number of occurrences that relevant labels are ranked lower than irrelevant labels.

$$RL = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\|y_i\|} \sum_{\lambda_i \in L, \lambda_i \neq y_i} \lambda_i \leq \lambda_i', (\lambda_i, \lambda_i', y_i^*, \bar{y}_i)$$  \hspace{1cm} (2)

where $\lambda_i$ shows a likelihood between $x_j$ and each label $l_i \in L$, and $\bar{y}_i$ is the complementary set of $y_i$. The smaller value of it shows better performance.

The methods are evaluated using a various number of features, and the results are reported in Figs. (3)-(4). The vertical axis of these figures indicates the classification performance value in terms of Hamming loss and Ranking loss measures, respectively. The results are reported over five independent runs. It is clear from Fig. (3) that ACO based methods achieved smaller Hamming loss criteria. This value is significantly increased while a higher number of features are chosen. Moreover, most multi-label filter-based feature selection methods do not consider both relevancy and redundancy and some of them only consider the count of features without paying attention to the ants that choose the feature. Thus, these methods are not capable of finding and removing redundant features as well. To tackle this issue, some methods such as MMLACO can find and eliminate redundant and irrelevant features.

Moreover, similar results were achieved where the experiments are reported using the Ranking loss measure. It is clear from Fig. 4 that those methods in their search methods, they take into consideration both relevancy and redundancy principles., could achieve better Ranking loss values. Moreover, from the results, it can be

| Dataset     | Instances | Features | Labels | Density | Type    | Domain   |
|-------------|-----------|----------|--------|---------|---------|----------|
| Arts        | 5000      | 462      | 26     | 0.063   | Numeric | Text     |
| Birds       | 645       | 260      | 19     | 0.053   | Numeric | Audio    |
| CAL500      | 502       | 68       | 174    | 0.150   | Numeric | Music    |
| Corel5K     | 5000      | 499      | 374    | 0.009   | Nominal | Image    |
| Education   | 5000      | 550      | 33     | 0.044   | Numeric | Text     |
| Yeast       | 2417      | 103      | 14     | 0.303   | Numeric | Biology  |
concluded that swarm intelligence-based methods such as MMLACO and MLACO could better search through the search space and find more relevant features in comparison with the heuristic-based methods such as MGFS and MDMR.

![Graphs showing Hamming loss values for different datasets](image)

**Fig. 5.** The Hamming loss values of the multi-label feature selection methods with the various number of features
4. CONCLUSION

In this study, we proposed a detailed survey on multi-label feature selection methods based on swarm intelligence. Such techniques are categorized into wrapper and filter methods. To test a feature set, wrapper methods use a multi-label classifier, such as MLKNN or MLNB. These methods yield precise results; however, they require a high computing cost to classify a range of prominent features due to the classifier’s use. Simultaneously, filter methods use an information-theoretical criterion to measure the similarities among various features and calculations the significance of each feature with labels, such as shared information. They are, thus, much quicker than the wrapper techniques. On the other hand, most of these approaches use a swarm intelligence method such as PSO, ACO, ABC, and Genetic Algorithm to search for a feature set across the solution space. We have analyzed this paper to compare the current methods of selection of multi-label features. The findings show that compared to other swarm-intelligence approaches, ACO-based methods yield better results. A potential path for enhancing
the swarm-intelligence method search process by clustering the search space and related community features into a single cluster.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Boutell MR, Luo J, Shen X, Brown CM. Learning multi-label scene classification. Pattern Recognition. 2004;37(9):1757-1771.
2. Ibrahim I, Abdulazeez A. The role of machine learning algorithms for diagnosing diseases. Journal of Applied Science and Technology Trends. 2021;2(01):10–19.
3. Abas Hasan D, Mohsin Abdulazeez A. A modified convolutional neural networks model for medical image segmentation. Test Engineering and Management. 2020;83:16798–16808.
4. Rashid H, Mohsin Abdulazeez A, Zebari D. Data mining classification techniques for diabetes prediction. Qubahan Academic Journal. 2021;1.
5. Mohsin Abdulazeez A. Journal of Soft Computing and Data Mining Evaluating Data Mining Classification Methods Performance in Internet of Things Applications; 2020.
6. Shi et al. F. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. IEEE Reviews in Biomedical Engineering; 2020.
7. Hasan B, Mohsin Abdulazeez A. Journal of Soft Computing and Data Mining A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. Applied Soft Computing. 2021;2:20-30.
8. Abdulqader D, Abdulazeez A, Zeebaree D. Machine learning supervised algorithms of gene selection: A review; 2020.
9. Mohammed M, Mohsin Abdulazeez A, Saeed J, Zebari D. A review on region of interest segmentation based on clustering techniques for breast cancer ultrasound images. Journal of Applied Science and Technology Trends. 2020;01:78-91.
10. Zebari D, Abdulazeez A, Zeebaree D, Salih M. a fusion scheme of texture features for covid-19 detection of ct scan images. 2020;1-6.
11. Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. Journal of Applied Science and Technology Trends. 2020;1(2):56 – 70.
12. Abdulkareem NM, Mohsin Abdulazeez A, Qader Zeebaree D, Hasan DA. COVID-19 world vaccination progress using machine learning classification algorithms. Qubahan Academic Journal. 2021;1(2):100-105.
13. Zeebaree DQ, Haron H, Abdulazeez AM, Zebari DA. Trainable model based on new uniform lbp feature to identify the risk of the breast cancer. in 2019 International Conference on Advanced Science and Engineering (ICOASE). 2019;106-111.
14. Eesa A, Mohsin Abdulazeez A, Orman Z. A new dids design based on a combination feature selection approach; 2015.
15. Mohsin Abdulazeez A, Hajy D, Zebari D. Robust watermarking scheme based LWT and SVD using artificial bee colony optimization. Indonesian Journal of Electrical Engineering and Computer Science. 2021;21:1218-1229.
16. Sinoara RA, Camacho-Collados J, Rossi RG, Navigli R, Rezende SO. Knowledge-enhanced document embeddings for text classification. Knowledge-Based Systems. 2019;163:955-971.
17. Mohsin Abdulazeez A, Zebari D, Jijo B. Machine learning classifiers based classification for IRIS recognition. Qubahan Academic Journal. 2021;1.
18. Zhu P, Xu Q, Hu Q, Zhang C, Zhao H. Multi-label feature selection with missing labels. Pattern Recognition. 2018;74:488-502.
19. Zhang Y, Wu J, Cai Z, Philip SY. Multi-view multi-label learning with sparse feature selection for image annotation. IEEE Transactions on Multimedia; 2020.
20. Tabakhi S, Najafi A, Ranjbar R, Moradi P. Gene selection for microarray data classification using a novel ant colony optimization. Neurocomputing. 2015;168:1024-1036.
21. Zeebaree DQ, Haron H, Abdulazeez AM. Gene selection and classification of microarray data using convolutional neural network. in 2018 International Conference on Advanced Science and Engineering (ICOASE). 2018;145-150.
22. Ahmed O, Brificani A. Gene expression classification based on deep learning. in
23. Ahmed N. Swarm intelligence algorithms in gene selection profile based on classification of microarray data: A review. Journal of Applied Science and Technology Trends. 2021;2:1-09.

24. Kadhe S, Garcia B, Heidarzadeh A, El Rouayheb S, Sprintson A. Private information retrieval with side information. IEEE Transactions on Information Theory; 2019.

25. Kute RS, Vyas V, Anuse A. Component-based face recognition under transfer learning for forensic applications. Information Sciences. 2019;476:176-191.

26. Du H, Ma L, Li G, Wang S. Low-rank graph preserving discriminative dictionary learning for image recognition. Knowledge-Based Systems. 2020;187:104823.

27. Hassan R, Mohsin Abdulazeez A. Deep learning convolutional neural network for face recognition: A review; 2021.

28. Mohammed S, Mohsin Abdulazeez A. Deep convolutional neural network for facial expression recognition deep convolution neural network for facial expression recognition deep convolution neural network for facial expression recognition. Palarch’s Journal Of Archaeology Of Egypt/Egyptology. ISSN 1567 -214x Keywords-facial expression recognition, principle component analysis, convolutional neural network; 2021.

29. Jain I, Jain VK, Jain R. Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification. Applied Soft Computing. 2018;62:203-215.

30. Ting FF, Tan YJ, Sim KS. Convolutional neural network improvement for breast cancer classification. Expert Systems with Applications. 2019;120:103-115.

31. Almugren N, Al shamlian H. A survey on hybrid feature selection methods in microarray gene expression data for cancer classification. IEEE Access. 2019;7:78533-78548.

32. Zeebaree DQ, Haron H, Abdulazeez AM, Zebari DA. Aichne learning and region growing for breast cancer segmentation. in 2019 International Conference on Advanced Science and Engineering (ICOASE). 2019;88-93.

33. Rashid Ismael H, Mohsin Abdulazeez A, Hasan DA. Comparative study for classification algorithms performance in crop yields prediction systems. Qubahan Academic Journal. 2021;1(2):19-24.

34. Kareem O, Mohsin Abdulazeez A, Zeebaree D. Skin lesions classification using deep learning techniques: Review. 2021;1-22.

35. Zebari DA, Zeebaree DQ, Abdulazeez AM, Haron H, Hamed HNA. Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. IEEE Access. 2020;8:203097-203116.

36. Lee J, Kim DW. Feature selection for multi-label classification using multivariate mutual information. Pattern Recognition Letters. 2013;34(3):349-357.

37. Lin Y, Hu Q, Liu J, Duan J. Multi-label feature selection based on max-dependency and min-redundancy. Neurocomputing. 2015;168:92-103.

38. Li L, et al. Multi-label feature selection via information gain. in International Conference on Advanced Data Mining and Applications. 2014;345-355: Springer.

39. Lee J, Kim DW. SCLS: Multi-label feature selection based on scalable criterion for large label set. Pattern Recognition. 2017;66:342-352.

40. Paniri M, Dowlatshahi MB, Nezamabadi-pour H. MLACO: A multi-label feature selection algorithm based on ant colony optimization. Knowledge-Based Systems. 2019;105285.

41. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Computers in Biology and Medicine. 2019;112:103375.

42. Rostami M, Berahmand K, Nasiri E, Forouzande S. Review of swarm intelligence-based feature selection methods. Engineering Applications of Artificial Intelligence. 2021;104210.

43. Di Mauro M, Galatro G, Fortino G, Liotta A. Supervised feature selection techniques in network intrusion detection: A critical review. Engineering Applications of Artificial Intelligence. 2021;104216.

44. Mahmoud MR, Abdulazeez AM. Different model for hand gesture recognition with a novel line feature extraction. in 2019 International Conference on Advanced Science and Engineering (ICOASE). 2019;52-57.

45. Rostami M, Berahmand K, Forouzandehe SJ. Review of swarm intelligence-
on artificial bee colony: An acceleration approach with variable sample size. Applied Soft Computing. 2020;88:106041.

57. Tabakh S, Moradi P. Relevance–redundancy feature selection based on ant colony optimization. Pattern Recognition. 2015;48(9):2798-2811.

58. Tabakh S, Moradi P, Akhlaghian F. An unsupervised feature selection algorithm based on ant colony optimization. Engineering Applications of Artificial Intelligence. 2014;32:112-123.

59. Moradi P, Rostami M. Integration of graph clustering with ant colony optimization for feature selection. Knowledge-Based Systems. 2015;84:144-161.

60. Ghimatgar H, Kazemi K, Helfroush MS, Aarabi A. An improved feature selection algorithm based on graph clustering and ant colony optimization. Knowledge-Based Systems. 2018;159:270-285.

61. Kashef S, Nezamabadi-pour H. An advanced ACO algorithm for feature subset selection. Neurocomputing. 2015;147:271-279.

62. Wan Y, Wang M, Ye Z, Lai X. A feature selection method based on modified binary coded ant colony optimization algorithm. Applied Soft Computing. 2016;49:248-258.

63. Peng H, Ying C, Tan S, Hu B, Sun Z. An improved feature selection algorithm based on ant colony optimization. IEEE Access. 2018;6:69203-69209.

64. Singh U, Singh SN. A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework. Applied Soft Computing. 2019;74:216-225.

65. Ghosh M, Guha R, Sarkar R, Abraham A. A wrapper-filter feature selection technique based on ant colony optimization. Neural Computing and Applications. 2019;1-19.

66. Zhang ML, Peña JM, Robles V. Feature selection for multi-label naive Bayes classification. Information Sciences. 2009;179(19):3218-3229.

67. Yu Y, Wang Y. Feature selection for multi-label learning using mutual information and GA. in International Conference on Rough Sets and Knowledge Technology. 2014:454-463:Springer.

68. Sun Z, et al. Mutual information based multi-label feature selection via constrained convex optimization. Neurocomputing. 2019;329:447-456.

69. Zhang P, Liu G, Gao W. Distinguishing two types of labels for multi-label feature
Zhang Y, Gong Dw, Sun Xy, Guo Yn. A PSO-based multi-objective multi-label feature selection method in classification. Scientific reports. 2017;7(1):1-12.

Dong H, Sun J, Sun X, Ding R. A many-objective feature selection for multi-label classification. Knowledge-Based Systems. 2020;208:106456.

Hatami M, Mahmood SR, Moradi P. A graph-based multi-label feature selection using ant colony optimization. in 2020 10th International Symposium on Telecommunications (IST), 2020;175-180.

Spolaôr N, Cherman EA, Monard MC, Lee HD. A comparison of multi-label feature selection methods using the problem transformation approach. Electronic Notes in Theoretical Computer Science. 2013;292:135-151.

Spolaôr N, Monard MC, Tsoumakas G, Lee HD. A systematic review of multi-label feature selection and a new method based on label construction. Neurocomputing. 2016;180:3-15.

Kashef S, Nezamabadi-pour H, Nikpour B. Multilabel feature selection: A comprehensive review and guiding experiments. WIREs Data Mining and Knowledge Discovery. 2018;8(2):e1240.

Lastra G, Luaces O, Quevedo JR, Bahamonde A. Graphical feature selection for multilabel classification tasks. In Advances in Intelligent Data Analysis X, Berlin, Heidelberg. 2011;246-257: Springer Berlin Heidelberg.

Li L, et al. Multi-label Feature Selection via Information Gain. in Advanced Data Mining and Applications, Cham. 2014;345-355: Springer International Publishing.

Jungjit S, Michaelis M, Freitas AA, Cinatl J. Two extensions to multi-label correlation-based feature selection: a case study in bioinformatics. In 2013 IEEE International Conference on Systems, Man, and Cybernetics. 2013;1519-1524.

Hashemi A, Dowlatshahi MB, Nezamabadi-pour H. MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality. Expert Systems with Applications. 2020;142:113024.

Lee J, Lim H, Kim DW. Approximating mutual information for multi-label feature selection. Electronics letters. 2012;48(15):929-930.

Zhang ML, Zhou ZH. ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition. 2007;40(7):2038-2048.

Wu XZ, Zhou ZH. A unified view of multi-label performance measures. In Proceedings of the 34th International Conference on Machine Learning-70. 2017;3780-3788:JMLR. org.

© 2021 Abdulazeez et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.