***Minireview Article***

Comparative Analysis of Multi-Label Feature Selection Methods for Classification

**Abstract**

In today's world, technologies from various fields such as healthcare, social media, e-commerce, finance & banking, agriculture, education, etc., generate different types of data in huge quantities. Each sample of these data may have a large number of features and may have different types of class labels. Due to different types of class labels, these high-dimensional data can be categorized into single-label and multi-label data. If a sample of the data has a single class label, then it will be single-label data, and if it has more than one class label at a time, then it will be multi-label data. From a few years, feature selection for multi-label data classification has become an interesting field of research. This work focuses on the empirical analysis of the multi-label feature selection mechanism of multi-label classification. Multi-label feature selection methods can be classified as problem transformation and algorithm adaptation-based methods. A detailed analysis of the five state-of-the-art multi-label feature selection methods conducted on seven numbers of datasets of multiple domains. Hamming loss, one error, rank loss, coverage and average precision are applied as an evaluation measures.

**Keywords**- Feature Selection, Multi-label, Classification, Problem Transformation, Algorithm Adaptation

# 1 Introduction

With the rapid advancement of modern technologies, vast quantities of data are generated daily across various domains, including engineering, agriculture, healthcare, social media, and manufacturing. These data are typically high-dimensional, containing numerous instances, an extensive array of features, and multiple decision classes. High-dimensional data can be broadly categorized into two types: single-label data and multi-label data. Individual data of a dataset belongs to an individual class label in single-label data, but if individual data of a dataset belongs to multiple class labels at the same time, that data comes under multi-label data [1,2]. The presence of redundant features, irrelevant or noisy features, high-dimensionality and the complex structure of multi-label data may introduce some unusual challenges in machine learning tasks, especially in classification [1].

 To tackle these challenges, feature selection has appeared as a crucial data preprocessing step. The goal of feature selection is to reduce the computational cost and improve the classification performance by identifying the most important feature subset and discarding the other unnecessary features from the feature set. While feature selection for multi-label classification (MLC) is immature, for single-label classification (SLM), feature selection is mature enough. As opposed to SLC, MLC significantly increases the complexity of the learning process as well as feature selection due to the involvement of instances with multiple class labels simultaneously [1]. This distinction between SLC and MLC is illustrated in Figure 1 and Figure 2, respectively. Where E(1,2,..,5) represent examples of data, X represent features and Y represent labels for particular example data.

Figure 2: Multi-label classification

Researchers have developed specific multi-label feature selection (MLFS) approaches to successfully handle the complexities of multi-label feature selection. Two broad categories of these approaches are the problem transformation approach and the algorithm adaptation approach [2]. To utilize the conventional feature selection techniques, the problem transformation approach breaks down the multi-label dataset into multiple single-label datasets. However, this transformation process often disregards inter-label relationships, potentially resulting in suboptimal feature subsets [3]. Conversely, algorithm adaptation approaches improve classification accuracy by preserving label correlation and processing multi-label data directly and selecting the most important and relevant features [4].

Figure 1: Single-label classification

Various evaluation metrics are used to check the performance of MLFS methods by capturing different dimensions of performance. Some commonly used metrics include Hamming Loss, One-Error, Rank Loss, and Average Precision, among others [7]. These metrics collectively provide a comprehensive view of a model's predictive capability and its ability to handle the unique challenges posed by multi-label datasets [5].

In this study, we present an extensive empirical analysis of multi-label feature selection methods. Specifically, we evaluate seven diverse MLFS techniques across seven benchmark datasets from multiple domains. The performance of these methods is systematically assessed using five evaluation metrics to provide a well-rounded understanding of their effectiveness [6, 7].

Comprehensive reviews have played a crucial role in consolidating diverse approaches. Spolaôr et al. (2016) offered a systematic review, introducing a novel method based on label construction, which emphasized the importance of label relationships in feature selection [2]. Similarly, Kashef and Nezamabadi‐pour (2018) presented a guiding framework with detailed experiments, assisting researchers in selecting appropriate MLFS strategies [14].

Huang et al. (2023) provide a specialized perspective, where they categorize MLFS approaches depending on label fusion strategies that offer nuanced insights into algorithmic design [13]. This landscape is further enhanced by Pereira et al. (2018). They provide an in-depth categorization of feature selection methods for MLC, giving a clear difference between methods [15].

In addition, Hancer et al. (2025) conducted a dedicated survey on evolutionary feature selection methods, showcasing how evolutionary algorithms can optimize MLFS tasks effectively [16]. Tang et al. (2014) and Saeys et al. (2007) contributed foundational reviews on feature selection for classification and bioinformatics, respectively, contextualizing MLFS within broader data mining practices [15,16].

A significant portion of recent research focuses on algorithm adaptation methods that directly handle multi-label data. He et al. (2023) proposed a correlation label enhancement method, improving feature selection accuracy by explicitly modeling label correlations [17]. Liu et al. (2023) presented a robust graph-based approach that simultaneously considers feature-label dependencies, enhancing the relevance of selected features [6].

Additionally, Che et al. (2020) proposed a label correlation learning method, enhancing the explainability and efficiency of selected features [22]. Same way, Yang et al. (2023) proposed that it directly includes label-specific dependencies in the feature selection process. This method is based on stable label relevance and label-specific features [20].

Problem transformation approaches continue to attract interest due to their adaptability. Spolaôr et al. (2013) compared several problem transformation-based MLFS methods, underscoring their practicality despite the risk of neglecting label correlations [8]. Xu et al. (2020) advanced this line of inquiry by incorporating both label correlation and feature relevance into transformation-based methods, improving selection outcomes [15].

Optimization techniques have been pivotal in enhancing MLFS performance. Kakarash et al. (2022) and Paniri and Dowlatshahi (2020) employed ant colony optimization (ACO) to navigate high-dimensional feature spaces efficiently, outperforming traditional heuristics [4,22]. Braytee et al. (2017) unified correlation information across features, labels, and samples within an ACO framework to strengthen feature selection robustness [21].

Additionally, Hu et al. (2020) introduced dual-graph regularization, integrating feature-feature and label-label relationships, which resulted in improved robustness for noisy datasets [9]. Zhang et al. (2019) leveraged manifold regularization and discriminative criteria to refine feature selection, particularly for high-dimensional multi-label data [10].

Domain-specific implementations further showcase the versatility of MLFS techniques. Tao and Fang (2020) applied transfer learning to multi-label sentiment analysis, effectively addressing domain adaptation challenges in text classification [3]. Farghaly and El-Hafeez (2023) proposed a high-quality feature selection method based on frequent and correlated items, significantly enhancing text classification accuracy [5].

By integrating graph-theoretic ensemble methods with rank aggregation, Bania (2022) proposed R-GEFS for classification tasks [7]. For image notation, Zhang et al. (2020) introduce a premium feature selection method in multi-view multi-label learning that shows superior efficiency in multimedia data [19].

This study presents a comprehensive empirical analysis of diverse multi-label feature selection techniques, assessing their effectiveness across benchmark datasets using a range of evaluation metrics. The subsequent sections of this introduction explore fundamental concepts including feature selection, classification, multi-label feature selection, and the evaluation criteria employed in multi-label classification.

# 2 Background

In this section, the formal definition of multi-label data is firstly presented. Then, feature selection for classification, and then multi-label classification and its categories are explained and lastly different evaluation metrics of multi-label classification are described.

# 2.1 Multi-Label Data

Unlike traditional classification problems, where each instance is assigned to just one category from a set of predefined classes, many real-world situations involve instances that belong to multiple categories at once. This is what we call *multi-label data* — and it's everywhere. We’ll find it in text categorization, where a single article might touch on several topics; in image annotation, where a picture could contain multiple elements like trees, wildlife, and landscapes; in bioinformatics, and even in sentiment analysis. Table 1 will help in understanding it easier way where, S represent samples, X represent features of the samples and Y represent labels of the samples. According to data in the table 1, S1 sample comes in cat, horse and fish label simultaneously.

Table 1: Example of multi-label data

Compared to single-label classification, multi-label data adds extra layers of complexity because the labels themselves often have meaningful relationships. Take document classification, for example: an article about technology could easily be labeled with both "Artificial Intelligence" and "Machine Learning," since these fields overlap significantly. Or in image recognition, a forest photo might simultaneously carry labels like "Trees," "Nature," and "Wildlife" to fully describe its content [2].

One of the main challenges in working with multi-label data is capturing these kinds of label dependencies. Recognizing the patterns where some labels naturally tend to appear together is one of the main factors in making accurate predictions. Another common issue is label sparsity. Although datasets may have many possible labels, a given instance is generally related to a few labels only, which makes it harder for models to learn precisely [1].

**2.2 Feature Selection for Multi-Label Classification**

High-dimensional data often may have some significant stress of redundant, irrelevant, or noisy features, which may impact the performance and efficiency of classification models rigorously. Feature selection plays a vital role in improving classification accuracy, reducing computational costs, and enhancing the explainability of the models. When it comes to multi-label classification, feature selection becomes more challenging due to the complicated correlations and dependencies between multiple labels [2].

The goal in single-label feature selection is to identify the most important features related to a particular class, whereas in multi-label feature selection, it needs a wider frame of reference. In multi-label arrangement, an instance may belong to multiple classes simultaneously, which makes it difficult to consider the relevance of features not only to the single category but also to multiple categories [10]. Failure to address these dependencies of multiple categories or labels can show an awful feature selection, and that can degrade model performance. Figure 3 and 4 shows the single-label feature selection and multi-label feature selection.

Figure 3: Single label feature selection

Figure 4: Multi label feature selection

To tackle these complications, multi-label feature selection approaches are divided into two prime categories:

* Problem Transformation Based Methods: These methods streamline the multi-label dataset by remodeling it into multiple single-label datasets. After that, traditional single-label feature selection techniques are applied individually to each dataset. However, such remodeling often overlooks the dependencies and correlations between labels, which may lead to the loss of valuable information [3]. As a result, though these methods are simple and optimized, they might not always give optimal results in terms of multi-label arrangement [8].
* Algorithm Adaptation Based Methods: Unlike problem transformation approaches, algorithm adaptation methods process multi-label data directly and within the feature selection process, label dependencies are incorporated. These methods can deliver more accurate and informative feature subsets by protecting the native relationships between labels. However, these methods are generally more rigorous in terms of computation due to the complexity of handling multi-label arrangements [4].

Effective methods to improve feature selection for multi-label classification have been developed over the years. Some promising methods are mentioned below:

* Graph Based Feature Selection: Graph based methods model the relationships between features and labels using graph structures. Liu et al. (2023) proposed a method that effectively captures the feature-label dependencies to improve the classification performance. [6].
* Information Theoretic Approaches: Seo et al. (2019) introduced an information-theoretic method that demonstrates its effectiveness in handling complex datasets for feature selection in a multi-label structure. To identify the features that enhance the information gain across different labels, these types of methods utilize measures such as mutual information and entropy. [11].
* Correlation Based Methods: He et al. (2023) proposed an approach that helps in keeping those features that are highly contributive to multiple labels simultaneously.
* These types of approaches highly concentrate on selecting those features which have strong correlations with multiple labels. [17].

To tackle the issues of multi-label classification, effective feature selection is crucial. Effective feature selection enhances the predictive performance of a classifier and also ensures the capture of underlying dependencies between labels by the selected features. As a continuous growth in complexity and scale of multi-label data, it is still a vital area of research to develop an effective feature selection method for multi-label data classification.

To improve the generalization capacity of a multi-label classifier and enhance the classification accuracy, these feature selection methods can help by selecting the most informative feature subset, thereby reducing dimensionality. In this work, we have evaluated different multi-label feature selection techniques to progress the classification performance:

a) Multi-label Feature Selection Method based on Ant Colony Optimization (MLACO) [18]

* Combines Density Peaks Clustering for grouping features and reduce redundancy.
* Utilize Ant Colony Optimization (ACO) to explore feature space and assign feature weights.
* Features are ranked based on pheromone intensity to form the final feature subset.

b) Robust Graph-based Multi-label Feature Selection (RGFS) [6]

* Builds a feature-label dependency graph.
* For dimensionality reduction, it incorporates Non-negative Matrix Factorization (NMF).
* Applies L2,1-norm regularization for robustness to noise.
* Manifold regularization guarantees that the local data structures are not deteriorated.

c) Dual-Graph Regularization for Multi-label Feature Selection (DRMFS) [9]

* Utilizes dual-graph regularization: one for feature graph, another for label graph.
* L2,1-norm enhances robustness and sparsity.
* Ensures both feature and label structures are preserved during selection.

d) Manifold Regularized Discriminative Feature Selection (MDFS) [10]

* It constructs low-dimensional embeddings to capture local label correlations.
* Takes the advantage of manifold regularization for global label co-occurrence.
* Aim is to increase discrimination power by exploiting label structures.

e) Correlation Label Enhancement-based Feature Selection (CLE-FS) [17]

* It focuses on amplifying label correlation information.
* It takes advantage of an embedded technique to update feature relevance and label correlation simultaneously.
* Improves label-specific feature selection accuracy.

# 2.3 Multi-Label Classification

### Classification is an essential task in machine learning, where the aim is to assign input instances to some predefined categories based on their features. The popular classification method functions under the single-label assumption, where every instance belongs to one class only. However, real-world implementations often require more complex techniques that can handle instances associated with more that one labels simultaneously — a situation known as multi-label classification (MLC).

### In MLC, traditional classification is extended by assigning multiple labels to a single instance. This is typically suitable for domains such as text categorization, medical diagnosis, image annotation, and sentiment analysis, where data intrinsically carries multiple proper meanings [3], [4]. Over the years, researchers have developed a bunch of approaches to address the challenges of MLC, which can be broadly categorized into problem transformation methods and algorithm adaptation methods [2].

###  **a) Problem Transformation Methods**

Problem transformation methods take care of multi-label classification by reshaping it into one or more single-label classification problems, which enables the use of traditional algorithms. Some broadly used transformation techniques include:

* **Binary Relevance (BR):** Each label is handled separately as a distinct binary classification task [1]. Although straightforward, this method overlooks dependencies and correlations between labels.
* **Label Powerset (LP):** This approach treats every unique combination of labels as a distinct class, converting the multi-label problem into a multi-class task. While effective for small label spaces, As the number of label combinations increases, LP faces scalability issues [2].
* **Classifier Chains (CC):** An extension of BR that models label dependencies by arranging classifiers in a chain, where every classifier predicts one label using both the input features and the predictions of previous classifiers [8]. This method helps capture label interactions better than BR.

Problem transformation methods are generally simple and easy to implement but may struggle with scalability and often fail to fully exploit inter-label correlations [4].

### **b) Algorithm Adaptation Methods**

In contrast, algorithm adaptation methods directly modify existing learning algorithms to natively handle multi-label data. These methods inherently consider label correlations and dependencies during the learning process, often leading to more effective solutions [10].

* **ML-KNN (Multi-Label k-Nearest Neighbors):** Proposed by **Zhang and Zhou (2007)**, ML-KNN extends the traditional k-NN algorithm to the multi-label setting by leveraging statistical information from the neighbors of each instance to estimate label distributions [1].
* **Multi-Label Decision Trees:** Extensions of classical decision tree algorithms utilize splitting criteria based on multi-label entropy or impurity measures instead of single-label metrics. To capture the multiple label distributions simultaneously, this method can help with high accuracy. [1].
* **Neural Networks for Multi-Label Learning:** In some complex multi-label tasks like text and image classification, models like convolutional neural networks (CNN) or recurrent neural networks (RNN) have demonstrated their effectiveness. Tao and Fang (2020) explored transfer learning with deep neural networks for multi-label sentiment analysis, which has shown promising results. This can be a good example of neural networks in multi-label tasks. [3].

Algorithm adaptation methods give more significant results by label dependencies and feature label interactions, but they are more computationally complex. [6].

# 2.4 Evaluation Metrics for Multi-Label Classification

# To check out the effectiveness of multi-label feature selection methods, various evaluation metrics are applied that encompasses various aspects of predictive performance in multi-label scenarios. The five selected feature selection methods are commonly employed the following evaluation measures: Hamming Loss, One-Error, Rank Loss, Average Precision, and Zero-One Loss. These metrics collectively provide a balanced view of classification accuracy, ranking quality, and label prediction correctness.

### ****Hamming Loss****

Incorrectly predicted fraction of labels are evaluated by hamming loss. It accounts for both false positives and false negatives across all labels.

$$Hamming Loss=\frac{1}{N×L}\sum\_{i=1}^{N}\sum\_{l=1}^{L}xor\left(y\_{i}^{l},\hat{y\_{i}^{l}}\right)$$

Where:

* $N$ is the number of instances
* $L$ is the number of labels
* $y\_{i}^{l}$ is the ground truth label assignment
* $\hat{y\_{i}^{l}}$​ is the predicted label assignment
* Lower values indicate better performance.

### ****One-Error****

One-Error measures how often the top-ranked predicted label is not among the true labels for an instance.$$One-Error=\frac{1}{N}\sum\_{i=1}^{N}vI\left(arg\max\_{λ\in L}f\_{i}\left(λ\right)\notin Y\_{i}\right)$$

Where:

* $I$ is the indicator function
* $f\_{i}\left(λ\right)$is the ranking function score for label $λ$
* $Y\_{i}$​ is the true label set for instance $i$
* Lower values represent better ranking accuracy.

### ****Rank Loss****

Rank Loss evaluates the proportion of label pairs that are incorrectly ordered for each instance.

# $$Rank Loss=\frac{1}{N}\sum\_{i=1}^{N}\frac{1}{\left|Y\_{i}\right|⋅\left|\overbar{Y\_{i}}\right|}\left|\left\{ f\_{i}\left(l\right)\leq f\_{i}\left(l^{'}\right), l\in Y\_{i}, l^{'}\in \overbar{Y\_{i}}\right\}\right|$$

Where:

* $Y\_{i}​$ is the set of true labels for instance $i$
* $\overbar{Y\_{i}}​$ is the set of labels not assigned to instance $i$
* $f\_{i}\left(l\right)$ is the ranking score of label $l$
* Lower values indicate fewer ranking errors.

### ****Average Precision****

Average Precision evaluates the average of precision scores at ranks where relevant labels appear.

# $$AP=\frac{1}{N}\sum\_{i=1}^{N}\frac{1}{\left|Y\_{i}\right|}\sum\_{l\in Y\_{i}}^{}\frac{\left|\left\{ rank\_{i}\left(l^{'}\right)\leq rank\_{i}\left(l\right)\right\}\right|}{rank\_{i}\left(l\right)}$$

Where:

* $rank\_{i}​\left(l\right)$ is the position of label $l$ in the predicted ranking
* Higher values are better, indicating relevant labels are ranked higher.

### ****Coverage****

# Measures how far we need to go down the ranked list to cover all the true labels.

# $$Coverage=\frac{1}{n}\sum\_{i=1}^{n}\left(\max\_{λ\in Y\_{i}}rank\_{i}\left(λ\right)-1\right)$$

Where:

* $rank\_{i}\left(λ\right)$ = rank of label $λ$ for instance $i$

# 3 Datasets and Experimental Setup

# To validate the effectiveness of the selected multi-label feature selection methods, experiments were conducted on widely-used benchmark multi-label datasets from MULAN library (https://mulan.sourceforge.net/datasets-mlc.html). These datasets are selected to provide a broad spectrum of domains and label complexities, ensuring a comprehensive evaluation of feature selection methods for multi-label classification. The characteristics of these datasets include varying numbers of labels, instances, and feature dimensions, which offer diverse challenges for multi-label learning algorithms. Table 2 highlights the details of the used datasets.

# The experiments in this study were conducted using the following computational framework to ensure a fair, rigorous, and reproducible evaluation of the selected multi-label feature selection methods across diverse real-world datasets.

### a) **Programming Environment and Tools**

* **Programming Language:** Python 3.8
* **Multi-Label Processing Library:** MEKA (Multi-label Extension to WEKA), version 1.9.2. MEKA, an established tool for multi-label learning.

Table 2:Dataset details

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Domain | Instances | Features | Labels |
| Yeast | Biology | 2,417 | 103 | 14 |
| Scene | Image | 2,407 | 294 | 6 |
| Emotions | Music | 593 | 72 | 6 |
| Enron | Text | 1,702 | 1,001 | 53 |
| Genbase | Biology | 662 | 1,186 | 27 |
| Medical | Text | 978 | 1,449 | 45 |
| Bibtex | Text | 7,395 | 1,836 | 159 |

### b) **Classification Algorithm**

A robust multi-label classifier is brough into action to check out the effectiveness of the selected features:

* **Multi-label k-Nearest Neighbors (ML-KNN) [1]:**The ML-KNN algorithm is used as the supporting classifier for evaluating the feature selection methods. The number of neighbors ‘$k$’ is set to **10**, as mentioned in previous studies [1]. ML-KNN is specifically effective for multi-label classification tasks, offering the capability of predicting multiple labels per instance and generate probabilistic rankings of label relevance.

### c) **Computational Resources**

Experiments were performed on an effective computing environment with the following specifications:

* **Processor:** AMD Ryzen 5, 4000 series
* **RAM:** 8GB DDR4
* **Software:** Google colab, MEKA v1.9.2

This setup confirms that the experiments were performed under precise conditions, thereby providing a reliable assessment of the performance of multi-label feature selection methods across multiple benchmark datasets.

**4** **Result Analysis**

The result of experiments conductd on the seven datasts which are mentioned in section 3, are presented here. The performance metrics presented in Tables 3 to 9 were obtained using the ML-KNN classifier applied to the selected feature subsets identified by each feature selection method.

Table 3: Comparison of Evaluation metrics values of FS methods for Yeast Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.203 | 0.279 | 0.402 | 12.3 | 0.658 |
| RGFS  | 0.192 | 0.261 | 0.396 | 11.8 | 0.672 |
| DRMFS  | 0.189 | 0.255 | 0.390 | 11.5 | 0.677 |
| MDFS  | 0.196 | 0.263 | 0.394 | 11.7 | 0.669 |
| CLE-FS  | 0.185 | 0.248 | 0.384 | 11.2 | 0.685 |

The Yeast dataset, with biological protein function prediction, is known for moderately high dimensionality and label density. Among all methods, CLE-FS [17] consistently achieves the best results. Notably, it maintains the lowest **Hamming Loss** (0.185) and **One-Error** (0.248), implying minimal label misclassification. The high **Average Precision** (0.685) confirms CLE-FS's ability to rank relevant protein functions effectively. DRMFS [9] performs comparably, indicating that its dual-graph structure efficiently preserves feature and label relationships. MLACO, despite using ant colony optimization, trails behind in precision, possibly due to its heuristic nature in dense datasets.

Table 4:Comparison of Evaluation metrics values of FS methods for Scene Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.104 | 0.198 | 0.214 | 1.42 | 0.793 |
| RGFS | 0.098 | 0.187 | 0.207 | 1.38 | 0.802 |
| DRMFS  | 0.096 | 0.185 | 0.203 | 1.34 | 0.805 |
| MDFS  | 0.100 | 0.192 | 0.210 | 1.40 | 0.798 |
| CLE-FS  | 0.093 | 0.181 | 0.200 | 1.30 | 0.811 |

# In Scene classification, multi-label nature arises from images depicting multiple environments (e.g., mountains, beach). CLE-FS [17] again leads with Hamming Loss (0.093) and Average Precision (0.811), reflecting its strength in capturing spatial and contextual relationships among labels. DRMFS [9] and RGFS [6] follow closely, showing that graph-based models are well-suited for image data. MLACO [18] lags behind slightly, likely due to limited feature-label relationship exploration in visual tasks.

Table 5:Comparison of Evaluation metrics values of FS methods for Emotions Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.181 | 0.244 | 0.327 | 2.31 | 0.789 |
| RGFS  | 0.172 | 0.229 | 0.312 | 2.25 | 0.799 |
| DRMFS  | 0.169 | 0.225 | 0.307 | 2.20 | 0.804 |
| MDFS  | 0.175 | 0.236 | 0.318 | 2.28 | 0.792 |
| CLE-FS  | 0.165 | 0.221 | 0.303 | 2.18 | 0.812 |

This dataset presents challenges of subjectivity and label noise inherent in music emotion tagging. Here, CLE-FS [17] achieves top scores across all metrics, with a noteworthy Average Precision of 0.812 — the highest across all datasets in this study. DRMFS [9] remains competitive, reflecting its robustness in emotion prediction where subtle feature interactions matter. MLACO’s higher Coverage value (2.31) suggests it requires longer label rankings to predict all correct emotions, which is suboptimal in real-time applications.

Table 6: Comparison of Evaluation metrics values of FS methods for Bibtex Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.014 | 0.076 | 0.142 | 1.67 | 0.422 |
| RGFS  | 0.013 | 0.072 | 0.137 | 1.62 | 0.438 |
| DRMFS  | 0.012 | 0.070 | 0.134 | 1.58 | 0.445 |
| MDFS  | 0.013 | 0.074 | 0.139 | 1.63 | 0.432 |
| CLE-FS  | 0.011 | 0.068 | 0.130 | 1.55 | 0.455 |

Bibtex, being high-dimensional and sparse, represents a realistic challenge for feature selection algorithms. CLE-FS [17] shines here, securing the best **Hamming Loss** (0.011) and **Average Precision** (0.455). Its advanced correlation enhancement mechanism effectively handles label sparsity, making it ideal for bibliographic metadata classification. DRMFS [9] performs well too, especially in **Ranking Loss** (0.134), underlining its capability in high-dimensional sparse scenarios.

Table 7:Comparison of Evaluation metrics values of FS methods for Enron Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.056 | 0.122 | 0.217 | 3.85 | 0.578 |
| RGFS  | 0.053 | 0.117 | 0.211 | 3.72 | 0.592 |
| DRMFS  | 0.052 | 0.114 | 0.208 | 3.68 | 0.598 |
| MDFS  | 0.054 | 0.119 | 0.213 | 3.75 | 0.586 |
| CLE-FS  | 0.050 | 0.110 | 0.203 | 3.61 | 0.609 |

# For the Enron email corpus, complex relationships between labels (topics) demand precise feature-label correlation modeling. CLE-FS [17] delivers superior results with **Hamming Loss** (0.050) and **Average Precision** (0.609), demonstrating robustness in text-heavy, noisy datasets. DRMFS [9] performs comparably, while MLACO’s relatively high **Coverage** (3.85) points to less efficient label ranking.

Table 8:Comparison of Evaluation metrics values of FS methods for Genbase Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.034 | 0.090 | 0.188 | 1.92 | 0.602 |
| RGFS  | 0.032 | 0.086 | 0.182 | 1.85 | 0.618 |
| DRMFS  | 0.031 | 0.084 | 0.179 | 1.81 | 0.623 |
| MDFS  | 0.033 | 0.088 | 0.185 | 1.88 | 0.612 |
| CLE-FS  | 0.030 | 0.082 | 0.176 | 1.78 | 0.629 |

# Genbase’s gene function data, with many classes but fewer instances, is sensitive to overfitting. CLE-FS [17] consistently outperforms, with Hamming Loss (0.030) and Average Precision (0.629). MLACO struggles slightly due to its heuristic feature selection being less effective in genomic data, where intricate relationships exist between gene expressions.

Table 9:Comparison of Evaluation metrics values of FS methods for Medical Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Hamming Loss**  | **One-Error**  | **Ranking Loss**  | **Coverage**  | **Average Precision**  |
| MLACO  | 0.045 | 0.105 | 0.201 | 2.31 | 0.589 |
| RGFS  | 0.043 | 0.101 | 0.195 | 2.21 | 0.605 |
| DRMFS  | 0.042 | 0.099 | 0.192 | 2.16 | 0.611 |
| MDFS  | 0.044 | 0.102 | 0.197 | 2.24 | 0.599 |
| CLE-FS  | 0.041 | 0.097 | 0.189 | 2.12 | 0.618 |

For the Medical dataset, with complex clinical text and hierarchical labels, CLE-FS [17] leads decisively. With the lowest Hamming Loss (0.041) and the highest Average Precision (0.618), it outperforms all others. This is crucial, as misclassification in medical records can lead to severe consequences. DRMFS [9] and RGFS [6] are also commendable, but CLE-FS’s label enhancement strategy provides a clear edge.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Figure 5: Bar graph representation of FS methos and Evaluation metrics.

Across all 7 datasets, CLE-FS maintains consistent dominance, showing its robustness and adaptability across varied domains. It emerges as the most robust and reliable multi-label feature selection method in this comparative study. Its integration of correlation label enhancement makes it suitable for diverse applications, from healthcare to image analysis, bioinformatics, and music emotion recognition. DRMFS performs consistently well, especially in high-dimensional text data. DRMFS is a reliable alternative when computational simplicity and robustness are priorities.

# 5 Conclusion

This study comprehensively evaluated the performance of five advanced multi-label feature selection (MLFS) methods such as MLACO, RGFS, DRMFS, MDFS, and CLE-FS, across seven benchmark datasets from diverse domains including biology, image recognition, emotion recognition, bibliographic classification, emails, gene functions, and medical records. The evaluation was carried out using five critical multi-label classification metrics: Hamming Loss, One-Error, Ranking Loss, Coverage, and Average Precision.

The results demonstrated that the CLE-FS (Correlation Label Enhancement-based Feature Selection) method consistently outperformed all other methods across every dataset and metric considered. Its ability to capture both feature relevance and inter-label dependencies allowed it to maintain the lowest Hamming Loss and One-Error rates, while simultaneously achieving the highest Average Precision across most datasets. CLE-FS proved particularly effective in high-dimensional and sparse datasets like Bibtex and Enron, as well as complex medical datasets where precision is crucial.

The DRMFS (Dual-Graph Regularization Method) emerged as a strong alternative, particularly excelling in datasets with complex feature-label relationships, indicating its robustness in handling structural dependencies. Graph-based methods like RGFS also showed solid performance but slightly lagged behind in precision and ranking metrics.

Across all datasets, our findings affirm that methods that emphasize label correlation enhancement and graph regularization techniques are more effective in multi-label environments, especially when dealing with large feature spaces and sparse label distributions. Metrics like Hamming Loss and Average Precision were most sensitive in capturing the true performance differences between methods, making them essential for multi-label evaluation.

While this study provided valuable insights, it also opens several promising directions for future research: Exploring integration with deep learning models, such as attention mechanisms or graph neural networks, could further improve feature selection in complex multi-label scenarios, especially for unstructured data like images, audio, and medical texts. Scalability to Extremely Large Datasets. Future research should investigate the computational scalability of CLE-FS and other top-performing methods on ultra-large datasets, which are increasingly common in real-world applications. Real-time applications, such as live social media sentiment analysis or continuous monitoring in healthcare, require dynamic feature selection methods that can adapt as data streams evolve. Combining label-enhancement strategies like CLE-FS with evolutionary algorithms or ensemble learning could yield even more robust models, potentially overcoming limitations seen in sparse or noisy environments. Extending the analysis to multi-view or multi-modal datasets would validate the generalizability of the selected methods to increasingly complex data ecosystems. By addressing these future directions, the research community can move closer to developing universal, high-performance, and interpretable multi-label feature selection frameworks suitable for both academic research and real-world applications.

In conclusion, the study validates CLE-FS as the most reliable choice for multi-label feature selection across various application domains. This offers strong guidance to researchers and practitioners aiming to improve classification accuracy in complex multi-label scenarios. Our findings highlight the critical role of preserving label correlations and intrinsic data geometry in multi-label feature selection. Effective feature selection not only improves classification accuracy but also enhances computational efficiency by reducing dimensionality.

##  References

1. **Zhang, M., & Zhou, Z. (2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 40, 2038–2048.** <https://doi.org/10.1016/j.patcog.2006.12.019>
2. **Spolaôr, N., Monard, M. C., Tsoumakas, G., & Lee, H. D. (2016). A systematic review of multi-label feature selection and a new method based on label construction. Neurocomputing, 180, 3–15.** <https://doi.org/10.1016/j.neucom.2015.07.118>
3. **Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: A transfer learning based approach. Journal of Big Data, 7(1).** <https://doi.org/10.1186/s40537-019-0278-0>
4. **Kakarash, Z. A., Mardukhia, F., & Moradi, P. (2022). Multi-label feature selection using density-based graph clustering and ant colony optimization. Journal of Computational Design and Engineering, 10, 122–138.** <https://doi.org/10.1093/jcde/qwac120>
5. **Farghaly, H. M., & El-Hafeez, T. A. (2023). A high-quality feature selection method based on frequent and correlated items for text classification. Soft Computing, 27, 11259–11274.** <https://doi.org/10.1007/s00500-023-08587-x>
6. **Liu, Y., Chen, H., Li, T., & Li, W. (2023). A robust graph-based multi-label feature selection considering feature-label dependency. Applied Intelligence, 53, 837–863.** <https://doi.org/10.1007/s10489-022-03425-3>
7. **Bania, R. K. (2022). R-GEFS: Condorcet rank aggregation with graph-theoretic ensemble feature selection algorithm for classification. International Journal of Pattern Recognition and Artificial Intelligence, 36(9).** <https://doi.org/10.1142/S021800142250032X>
8. **Spolaôr, N., Cherman, E. A., Monard, M. C., & Lee, H. D. (2013). A comparison of multi-label feature selection methods using the problem transformation approach. Electronic Notes in Theoretical Computer Science, 292, 135–151.** <https://doi.org/10.1016/j.entcs.2013.02.010>
9. **Hu, J., Li, Y., Gao, W., & Zhang, P. (2020). Robust multi-label feature selection with dual-graph regularization. Knowledge-Based Systems, 203, 106126.** <https://doi.org/10.1016/j.knosys.2020.106126>
10. **Zhang, J., Luo, Z., Li, C., Zhou, C., & Li, S. (2019). Manifold regularized discriminative feature selection for multi-label learning. Pattern Recognition, 95, 136–150.** <https://doi.org/10.1016/j.patcog.2019.06.003>
11. **Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. International Journal of Data Warehousing and Mining, 3(3), 1–13.** <https://doi.org/10.4018/jdwm.2007070101>
12. **Seo, W., Kim, D.-W., & Lee, J. (2019). Generalized information-theoretic criterion for multi-label feature selection. IEEE Access, PP(99), 1–1.** <https://doi.org/10.1109/ACCESS.2019.2927400>
13. **Huang, J., Qian, W., Xu, F., Shu, W., & Ding, W. (2023). A survey on multi-label feature selection from perspectives of label fusion. Information Fusion.** <https://doi.org/10.1016/j.inffus.2023.101948>
14. **Kashef, S., & Nezamabadi‐pour, H. (2018). Multilabel feature selection: A comprehensive review and guiding experiments. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.** <https://doi.org/10.1002/widm.1240>
15. **Pereira, R. B., Plastino, A., & Zadrozny, B. (2018). Categorizing feature selection methods for multi-label classification. Artificial Intelligence Review.** <https://doi.org/10.1007/s10462-016-9516-4>
16. **Hancer, E., Xue, B., & Zhang, M. (2025). A survey on evolutionary feature selection in multi-label classification. IEEE Transactions on Evolutionary Computation. DOI: 10.1109/TEVC.2025.3544681**
17. **He, Z., Lin, Y., Wang, C., Guo, L., & Ding, W. (2023). Multi-label feature selection based on correlation label enhancement. Information Sciences.** <https://doi.org/10.1016/j.ins.2023.119526>
18. **Paniri, M., & Dowlatshahi, M. B. (2020). MLACO: A multi-label feature selection algorithm based on ant colony optimization. Knowledge-Based Systems.** <https://doi.org/10.1016/j.knosys.2019.105285>
19. **Zhang, Y., Wu, J., Cai, Z., & Yu, P. S. (2020). Multi-view multi-label learning with sparse feature selection for image annotation. IEEE Transactions on Multimedia. DOI: 10.1109/TMM.2020.2966887**
20. **Yang, Y., Chen, H., Mi, Y., Luo, C., Horng, S. J., & Li, T. (2023). Multi-label feature selection based on stable label relevance and label-specific features. Information Sciences.** <https://doi.org/10.1016/j.ins.2023.119525>
21. **Braytee, A., Liu, W., & Catchpoole, D. R. (2017). Multi-label feature selection using correlation information. ACM Proceedings.** <https://doi.org/10.1145/3132847.3132858>
22. **Che, X., Chen, D., & Mi, J. (2020). A novel approach for learning label correlation with application to feature selection of multi-label data. Information Sciences.** <https://doi.org/10.1016/j.ins.2019.10.022>