An Efficient Elastic Net-based High Sparse Personalized and Low Redundancy Feature Selection Method with Missing Label for Multi-label Learning

Yichun Liu, Qing Ai*, Yuting Xu and Bo Cui

***Abstract*—Multi-label learning, which involves training models on data annotated with more than one label, has garnered considerable attention. However, there are still several challenges in this field. A) In multi-label learning, traditional algorithms have not fully exploited samples with missing labels, despite the fact that such samples are prevalent in real-world applications. B) Traditional methods typically assume that strongly correlated labels share similar label-specific features. However, this assumption does not hold in some scenarios, as even though there is a correlation between two labels, their corresponding specific features may not be identical. C) The sparsity and redundancy of features hinder the selection of high-quality features. To address these limitations, we design an Efficient Elastic net based high Sparse personalized and low Redundancy Feature Selection method with Missing Label for multi-label data named EESRFSML. First, we introduce the label correlation matrix to construct an enhanced label matrix, which recovers as many missing labels as possible. Second, we leverage a label-level regularizer to capture both global and local label correlations from the label outputs, rather than relying on the coefficient matrix. Meanwhile, We also fully consider both label-specific features and common features. Finally, the classification model is efficiently optimized using the accelerated proximal gradient algorithm. Extensive experiments demonstrate that the proposed method outperforms existing methods in multi-label learning tasks.**

***Index Terms*—Multi-label learning; Missing labels; Label-specific features; Common features;**

I. Introduction

MULTI-LABEL learning aims to address classification tasks where each sample can be associated with multiple labels simultaneously and is widely applied across various fields, such as text categorization [1], social media [2], image search and partitioning [3], protein function prediction [4], and so on.

Multi-label learning algorithms are generally categorized into two main groups: problem transformation (PT) methods and algorithm adaptation (AA) methods. PT methods reformulate a multi-label classification task into either a multi-class classification task or a set of binary classification tasks. Representative PT algorithms include Binary Relevance (BR) [5], Label Powerset (LP) [6], and Classifier Chains (CC) [7]. In contrast, AA methods modify existing classification algorithms to handle multi-label data directly. Typical AA algorithms include Multi-label Decision Trees [8], CML [9], Rank-SVM [10], and ML-KNN [11].

The traditional multi-label algorithms generally assume that all labels share the same feature set [12], [13]. However, in practical applications, this assumption may not always hold. For example, in multi-label image classification tasks, labels such as "sunset" and "ocean" may rely on entirely different visual features, with the former focusing on color gradients and atmospheric hues, and the latter emphasizing textures and wave patterns. Such label-specific dependencies underscore the limitations of shared feature sets in capturing nuanced label-specific characteristics. In contrast, label-specific feature learning algorithms extract distinct features for each individual label, which effectively improves classification performance [14]. The typical algorithms of label-specific features include LITF [15], LSML [16], CLML [17], and LRLSF [18].

In many cases, different labels may still share underlying patterns or relationships. For instance, labels like "beach" and "ocean" may both rely on features capturing sand textures or water dynamics. By leveraging these shared patterns, common features can complement label-specific features, enabling the model to better generalize across all labels. This complementary relationship underscores the importance of simultaneously learning common and label-specific features to achieve robust performance. Therefore, common features play a crucial role in providing shared discrimination for all labels in multi-label learning. The typical algorithms of common features include SCMFS [19], MLMLFS [20], FSLCLC [21], CLML [17], and ESRFS [22].

In multi-label classification, labels are often interrelated. Based on the level of label correlation considered, multi-label learning algorithms can be categorized into three types. First-order strategies ignore label correlations entirely. Representative algorithms in this category include Binary Relevance (BR) [5], ML-KNN [11], LIFT [15], and LSFML-MLTSVM [23]. Second-order strategies take into account pairwise label correlations. Typical algorithms include LLSF [24], LSML [16], CLML [17], NMDG [25], and LRLSF [18]. Higher-order strategies aim to capture correlations among all labels simultaneously. Notable

Manuscript received May 12, 2025; revised Aug 09, 2025.

Yichun Liu is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (e-mail: ustlliuyichun@163.com).

Qing Ai is a professor at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (Corresponding author to provide phone: +86-155-0492-8887; e-mail: lyaiqing@126.com).

Yuting Xu is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (e-mail: ustlxuyuting@163.com).

Bo Cui is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (e-mail: ustlcuibo@163.com).

examples include LLSF-DL [24], LFFS [26], GLMAM [27], and GLFS [28].

Multi-label learning, which involves training models on data annotated with more than one label, has garnered considerable attention. However, there are still several challenges in this field. A) In multi-label learning, traditional algorithms have not fully exploited samples with missing labels, despite the fact that such samples are prevalent in real-world applications. B) Traditional methods typically assume that strongly correlated labels share similar specific label features. However, this assumption does not hold in some scenarios, as even though there is a correlation between two labels, their corresponding specific features may not be identical. C) The sparsity and redundancy of features hinder the selection of high-quality features. To address these limitations, we design an Efficient Elastic net based high Sparse personalized and low Redundancy Feature Selection Method with Missing Label for multi-label data named EESRFSML. First, we introduce the label correlation matrix to construct an enhanced label matrix, which recovers as many missing labels as possible. Second, we leverage a label-level regularizer to capture both global and local label correlations from the label outputs, rather than relying on the coefficient matrix. Meanwhile, We also fully consider both label-specific features and common features. Finally, the classification model is efficiently optimized using the accelerated proximal gradient algorithm. Extensive experiments demonstrate that the proposed method outperforms existing methods in multi-label learning tasks.

The structure of the paper is as follows: Section 2 provides an overview of related works. Section 3 introduces our EESRFSML algorithm, including the classification model, optimization and analysis of time complexity. Section 4 presents the experimental results and examines parameter sensitivity. Section 5 concludes the paper.

II. Related works

A. Label-specific features and common features learning

Zhang et al. propose the LIFT algorithm [15], which first utilizes label-specific features to address the multi-label learning problem. It clusters instances according to their labels and leverages the separation between instances to acquire label-specific features. Huang et al. propose the LLSF algorithm [24], which assumes that each label is correlated to a subset of features from the original feature set, and uses a linear regression model to obtain the relevant label-specific features. Huang et al. also propose the LLSF-DL [24] algorithm by extending the LLSF algorithm, which learns label-specific features using a sparse stacking method. However, the LLSF and LLSF-DL algorithms ignore common feature learning. To address this issue, Li et al. proposed the CLML [17] algorithm, which introduces the $l_{2,1}$-norm regularizer to learn common features to enhance classification performance. Hu et al. propose the SCMFS [29] algorithm, which employs coupled matrix factorization (CMF) to identify the common modes between the feature and label matrices, effectively capturing the underlying information from both data sources. Additionally, the algorithm extracts common features from these modes using the $l_{2,1}$-norm. However, the SCMFS algorithm can only utilize samples with complete labels and cannot make use of samples with missing labels. Li et al. propose the CLSML [30] algorithm, which employs a two-stage, second-order label correlation learning method based on cosine similarity to better acquire label-specific features and common features. However, the $l_{2,1}$-norm ignores the redundant correlations among features, leading to the selection of shared common features with considerable redundancy. Li et al. propose the ESRFS [22] algorithm, which identifies common features with low redundancy and strong discriminative power by introducing a novel regularization term.

B. Missing labels

Huang et al. propose the LSML [16] algorithm, which leverages the label correlation coefficient matrix to recover missing labels and learn label-specific features from the new supplementary label matrix. However, the LSML algorithm ignores the potential correlations between the feature space and multidimensional label data. Deng et al. propose the NMFR [31] algorithm, which utilizes non-negative matrix factorization to learn a low-rank approximation of the feature and label matrices, enhancing the accuracy of multi-label classification tasks by exploiting both feature-label relationships and latent patterns. However, the NMFR algorithm does not address the issue of label imbalance. Faraji et al. propose the MLFS-GLOCAL [32] algorithm, which integrates both global and local label correlations to enhance feature selection, aiming to improve classification performance by identifying discriminative features that are relevant across multiple labels.

C. Label correlation learning

Zhu et al. propose the GLOCAL [33] algorithm, which addresses multi-label learning with missing labels by modeling both global and local label correlations, learning a latent label representation, and optimizing label manifolds. However, the GLOCAL algorithm ignores the potential issue of redundant features in high-dimensional spaces. Fan et al. propose the LCIFS [34] algorithm, which employs a manifold-based regression model to capture the relationship between the feature space and label distribution. Additionally, an adaptive spectral graph is used to enhance the accuracy of label correlations. However, the LCIFS algorithm only considers samples with complete labels and does not sufficiently address the issue of missing labels. Xu et al. propose the IncomLDL-LCD [35] algorithm, which decomposes label correlation into sparse local label correlation and low-rank global label correlation using a soft-thresholding operator and a singular value thresholding operator, respectively.

III. EESRFSML

Let $X = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^{n\times d}$ represent the training data with n samples and d features. The corresponding label matrix $Y = [y_1, y_2, \ldots, y_n]^T \in \{-1, 0, 1\}^{d\times l}$ contains the class labels. For the i-th sample, $y_{ij} = 1$ denotes the i-th sample has the j-th label y_j, $y_{ij} = -1$ denotes the i-th sample has not the j-th label y_j, and $y_{ij} = 0$ denotes the label of the i-th sample is unobserved.

A. Learning label-specific features and common features

We utilize the elastic net to overcome the limitations of the l_1-norm and introduce an inner-product-based regularization term to capture sparse personalized features for each label while considering common features across labels, as in [22]. The original problem is formulated as follows.

$$\begin{aligned}\min_{W} \quad & \frac{1}{2}\|XW - Y\|_F^2 + \lambda_1\|W\|_1 \\ & + \lambda_2\|W\|_F^2 + \lambda_3(\|WW^T\|_1 - \|W\|_F^2)\end{aligned} \tag{1}$$

where $W = [w^1, w^2, \ldots, w^d] \in \mathbb{R}^{d\times l}$ represents the weighted matrix, and $\lambda_1, \lambda_2, \lambda_3$ serve as the trade-off parameters.

B. Recovering the missing labels

Traditional algorithms typically assume that all labels for training samples are fully available. However, in real-world applications, only a subset of the label set is often observable [36]. To address the problem of missing labels, it is commonly assumed that incomplete labels can be reconstructed based on their correlations with other labels. Accordingly, the original label matrix with missing entries is augmented into a new, more complete label matrix by leveraging a label correlation coefficient matrix. We then define:

$$\begin{aligned}\min_{W,C} \frac{1}{2}\|XW - YC\|_F^2 + \frac{\alpha}{2}\|YC - Y\|_F^2 + \lambda_1\|W\|_1 \\ +\lambda_2\|W\|_F^2 + \lambda_3(\|WW^T\|_1 - \|W\|_F^2) + \lambda_4\|C\|_1\end{aligned} \tag{2}$$

By minimizing $\frac{\alpha}{2}\|YC - Y\|_F^2$, we aim for the new label matrix YC to closely approximate the true label matrix Y. The model enables the redefinition of each label by taking into account high-order correlations among labels. Besides, the label correlations coefficient matrix C is constrained by l_1-norm that can select the most important label correlation information, thereby improving the model of generalization ability and prediction accuracy [37].

C. Incorporating global and local label correlations

In traditional multi-label learning, it is generally assumed that correlated labels share similar label-specific feature representations. However, this assumption may not always hold in practice. Therefore, we propose an additional assumption: when two labels are strongly correlated, their corresponding outputs are likely to be similar [17]. To address this issue, we utilize a new label matrix XW instead of the coefficient matrix W. Finally, we employ a graph Laplacian regularization to improve model stability. Then we defined:

$$\begin{aligned}\sum_{p,q=1}^{l} S_{pq}\|(XW)_p - (XW)_q\| &= \operatorname{tr}\left((XW)(S^{\Delta} - S)(XW)^T\right) \\ &= \operatorname{tr}\left((XW)^T L_1(XW)\right)\end{aligned} \tag{3}$$

where $S_{ij} \in S$ represents the cosine similarity-based correlation between labels y_i and y_j. $L_1 = S^{\Delta} - S$ denotes the $l \times l$ label laplacian matrix of S. Meanwhile, S^{Δ} is the diagonal matrix, where $S_{ii}^{\Delta} = \sum_{j=1}^{l} S_{ij}$. The objective function is subsequently defined.

$$\begin{aligned}\min_{W,C} \quad & \frac{1}{2}\|XW - YC\|_F^2 + \frac{\alpha}{2}\|YC - Y\|_F^2 \\ & + \frac{\beta}{2}\operatorname{tr}((XW)L_1(XW)^T) + \lambda_1\|W\|_1 + \lambda_2\|W\|_F^2 \\ & + \lambda_3(\|WW^T\|_1 - \|W\|_F^2) + \lambda_4\|C\|_1\end{aligned} \tag{4}$$

In our EESRFSML algorithm, we also analyze the local correlations among labels. However, the effectiveness of the cosine similarity is limited by redundant features and noisy samples. Thus, we introduce k-nearest neighbors method to analyze the instance similarity matrix G. For example, the similarity of the p-th and q-th samples is defined:

$$G_{pq} = \begin{cases} 1, & \text{if } x_p \in \text{KNN}(x_q) \text{ or } x_q \in \text{KNN}(x_p) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

If two instances are strongly correlated, their predicted labels should be similar. To capture this relationship, we introduce a graph laplacian regularization term, defined as follows:

$$\sum_{p,q=1}^{n} G_{pq}\|(x_pW) - (x_qW)\| = \operatorname{tr}((XW)^T L_2(XW)) \tag{6}$$

Finally, the objective function of our EESRFSML algorithm can be expressed as an equation.

$$\begin{aligned}\min_{W,C} \quad & \frac{1}{2}\|XW - YC\|_F^2 + \frac{\alpha}{2}\|YC - Y\|_F^2 \\ & + \frac{\beta}{2}\operatorname{tr}((XW)L_1(XW)^T) + \frac{\gamma}{2}\operatorname{tr}((XW)^T L_2(XW)) \\ & + \lambda_1\|W\|_1 + \lambda_2\|W\|_F^2 \\ & + \lambda_3(\|WW^T\|_1 - \|W\|_F^2) + \lambda_4\|C\|_1\end{aligned} \tag{7}$$

where $\alpha, \beta, \gamma, \lambda_1, \lambda_2, \lambda_3$ and λ_4 are constant coefficients.

D. Optimization

Due to the non-smooth nature of the l_1-norm regularization term, the optimization problem (8) is convex but overall non-smooth. Therefore, this paper uses the accelerated proximal gradient descent algorithm to solve problem (8). The model has two solution variables, we denote ϕ as W and C. Simplifying the objective framework to:

$$\min_{\phi} \Gamma(\phi) = h(\phi) + \psi(\phi) \tag{8}$$

where

$$\begin{aligned}h(\phi) = & \\ & \frac{1}{2}\|XW - YC\|_F^2 + \frac{\alpha}{2}\|YC - Y\|_F^2 \\ & + \frac{\beta}{2}\operatorname{tr}((XW)L_1(XW)^T) + \frac{\gamma}{2}\operatorname{tr}((XW)^T L_2(XW)) \\ & + \lambda_2\|W\|_F^2 - \lambda_3\|W\|_F^2.\end{aligned} \tag{9}$$

$$\psi(\phi) = \lambda_1\|W\|_1 + \lambda_4\|C\|_1 + \lambda_3\|WW^T\|_1. \tag{10}$$

Both $h(\phi)$ and $\psi(\phi)$ are convex, but $\psi(\phi)$ is non-smooth. For any $L > 0$, we define the second-order approximation of the former:

$$Q_L(\phi, \phi^{(t)}) = h(\phi^{(t)}) + \langle \nabla h(\phi^{(t)}), \phi - \phi^{(t)} \rangle + \frac{L}{2}\|\phi - \phi^{(t)}\|_F^2 + \psi(\phi) \tag{11}$$

For any $L \geq L_f$, it can be stated that $Q_L(\phi, \phi^t) \geq \Gamma(\phi)$, where L_f is Lipschitz constant. We do not directly minimize $\Gamma(\phi)$, the proximal gradient algorithm approximates the objective function $\Gamma(\phi)$ by minimizing many separable quadratic approximations. By defining $G^{(t)} = \phi^{(t)} - \frac{1}{L}\nabla h(\phi^{(t)})$, this solution for ϕ be acquired by minimizing $Q_L(\phi, \phi^t)$.

$$\phi^* = \arg\min_{\phi} Q_L(\phi, \phi^{(t)}) = \arg\min_{\phi} g(\phi) + \frac{L}{2}\|\phi - G^{(t)}\|_F^2 \tag{12}$$

where $G^{(t)} = \phi^{(t)} - \frac{1}{L}\nabla h(\phi^{(t)})$, by setting b_t in a sequence, the convergence of the model can be accelerated, with $\phi^{(t)} = \phi_t + \frac{b_{t-1}-1}{b_t}(\phi_t - \phi_{t-1})$, where $b_{t+1}^2 - b_{t+1} \leq b_t^2$, ϕ_t is the t-th iteration of ϕ.

1) Updating W : First, C is fixed to update W, and the partial derivative W is computed.

$$\nabla_W h(\phi) = X^T XW - X^T YC + \beta X^T XWL_1 + \gamma X^T L_2 XW + 2\lambda_2 W - 2\lambda_3 W \tag{13}$$

The update process for W can be obtained through equation (13).

$$W^{(t)} = W_t + \frac{b_{t-1}-1}{b_t}(W_t - W_{t-1})$$
$$W^{(t+1)} = \text{prox}_{\varepsilon}\left(W^t - \frac{1}{L}\nabla f(W^{(t)}, C)\right) \tag{14}$$

where τ represents the step size, regarding $\psi(\phi)$, The l_1-norm of W is obtained using the element-wise soft-thresholding operator.

$$\text{prox}_{\varepsilon}(W_{ij}) = (|W_{ij}| - \tau)_+ \text{sign}(W_{ij}) \tag{15}$$

where $(\cdot)_+ = \max(\cdot, 0)$.

2) Updating C : First, W is fixed to update C, and the partial derivative C is computed.

$$\nabla_S f(\phi) = (1+\alpha)Y^T YC - Y^T XW - \alpha Y^T Y \tag{16}$$

Similarly, we can obtain the update process for C,

$$C^{(t)} = C_t + \frac{b_{t-1}-1}{b_t}(C_t - C_{t-1})$$
$$C^{(t+1)} = \text{prox}_{\varepsilon}\left(C^{(t)} - \frac{1}{L}\nabla f(W, C^{(t)})\right) \tag{17}$$

where ε represents the step size, regarding $\psi(\phi)$, the representation of the l_1-norm of C can be derived from the definition of the element-wise soft-thresholding operator.

$$\text{prox}_{\varepsilon}(S_{ij}) = (|S_{ij}| - \varepsilon)_+ \text{sign}(S_{ij}) \tag{18}$$

where $(\cdot)_+ = \max(\cdot, 0)$.

E. Proof of Lipschitz continuity

Lipschitz is essential in accelerated proximal gradient algorithms. We give $\phi_1 = (W_1, C_1)$ and $\phi_2 = (W_2, C_2)$. Based on equations (14) and (17), we can obtain the following expressions.

$$\begin{aligned}
&\|\nabla f(\phi_1) - \nabla f(\phi_2)\|_F^2 \\
&= \|X^T X\Delta W + \beta X^T X\Delta WL_1 \\
&+ \gamma X^T L_2 X\Delta W + 2\lambda_2 \Delta W \\
&- 2\lambda_3 \Delta W + (1+\alpha)Y^T Y\Delta C\|_F^2 \\
&= \|X^T X\Delta W + \beta X^T X\Delta CL_1 \\
&+ \gamma X^T L_2 X\Delta W + 2\lambda_2 \Delta W \\
&- 2\lambda_3 \Delta W\|_F^2 + \|(1+\alpha)Y^T Y\Delta C\|_F^2 \\
&\leq 5(\|X^T X\|_F^2 + \|\beta X^T X\|_2^2\|L_1\|_2^2 \\
&+ \|\gamma X^T L_2 X\|_F^2 + \|2\lambda_2\|_2^2 \\
&- \|2\lambda_3\|_2^2)\|\Delta W\|_F^2 \\
&\quad + 2(\|Y^T Y\|_2^2 + \|\alpha Y^T Y\|_2^2)\|\Delta C\|_F^2 \\
&\leq 5(\|X^T X\|_F^2 + \|\beta X^T X\|_2^2\|L_1\|_2^2 \\
&+ \|\gamma X^T L_2 X\|_F^2 + \|2\lambda_2\|_2^2 + \|2\lambda_3\|_2^2) \\
&+ 2(\|Y^T Y\|_2^2 + \|\alpha Y^T Y\|_2^2)\|\frac{\Delta W}{\Delta C}\|_F^2
\end{aligned} \tag{19}$$

where $\Delta W = W_1 - W_2, \Delta C = C_1 - C_2$.

Therefore, the objective function for the Lipschitz constant is expressed as

$$L_f = \sqrt{\begin{array}{l} 5\left(\|X^T X\|_2^2 + \|\beta X^T X\|_2^2\|L_1\|_2^2 \right. \\ \left. + \|\gamma X^T L_2 X\|_2^2 + \|2\lambda_2\|_2^2 + \|2\lambda_3\|_2^2\right) \\ + 2\left(\|Y^T Y\|_2^2 + \|\alpha Y^T Y\|_2^2\right) \end{array}} \tag{20}$$

Algorithm LSMLLC.

Input: Train data matrix $X \in \mathbb{R}^{n\times d}$, train label matrix $Y \in \mathbb{R}^{n\times l}$, and weighting parameters $\alpha, \beta, \gamma, \lambda_1$, and λ_2.

Output: Coefficient matrix $W \in \mathbb{R}^{d\times l}$.

1) **Initialization:** $W_0, W_1 \leftarrow \text{rand}(d, l)$; $C_0, C_1 \leftarrow \text{rand}(l, l)$; $\phi^{(1)} = \{W_1, C_1\}$; $t_0, t_1 \leftarrow 1$; $k \leftarrow 1$.
2) Calculate label correlation matrix S by calculating cosine similarity on Y; calculate H by using k-nearest neighbors; calculate the Lipschitz constant L_f.
3) **repeat:**
   a) $W^{(k)} \leftarrow W_k + \frac{t_{k-1}-1}{t_k}(W_k - W_{k-1})$.
   b) $F_W^{(k)} \leftarrow W^{(k)} - \frac{1}{L}\nabla_W f(W^{(k)}, C_k)$.
   c) $W_{k+1} \leftarrow \text{prox}_{\mathcal{E}}\left(\frac{\lambda_1}{L}F_W^{(k)}\right)$ (by according to (14)).
   d) $W^{(k+1)} \leftarrow W_{k+1}$.
   e) $C^{(k)} \leftarrow C_k + \frac{t_{k-1}-1}{t_k}(C_k - C_{k-1})$.
   f) $F_C^{(k)} \leftarrow C^{(k)} - \frac{1}{L}\nabla_C f(W_k, C^{(k)})$.
   g) $C_{k+1} \leftarrow \text{prox}_{\varepsilon}\left(\frac{\lambda_2}{L}F_C^{(k)}\right)$ (by according to (17)).
   h) $C^{(k+1)} \leftarrow C_{k+1}$.
   i) $t_{k+1} \leftarrow \frac{1+\sqrt{4t_k^2+1}}{2}$.
   j) $k \leftarrow k + 1$.
4) **until convergence**.
5) **Return** W.

TABLE I
THE DETAILS OF THE EXPERIMENTAL DATASETS

Dataset	Domain	# Label-Sample	#Unlabel-Sample	#Sample	#Feature	#Label
emotions	music	296	297	593	72	6
medical	text	489	489	978	1449	45
enron	text	851	851	1702	1001	52
yeast	biology	1208	1208	2416	103	14
education	text(web)	2500	2500	5000	550	333
arts	text(web)	2500	2500	5000	462	26
science	text(web)	2500	2500	5000	743	40
bibtex	text	3697	3698	7395	1836	159
delicious	text(web)	8052	8053	16105	500	983

F. Analysis of time complexity

The time complexity of the algorithm derived from the Lipschitz constant can be analyzed based on the main operations involved in the computation. In the preprocessing phase, the computation of terms results in a complexity of $O(nd^2 + d^2k + ndl + dl)$, where n is the number of data points, d is the feature dimension, k is the output dimension, and l is the number of labels. During iterative optimization, each iteration involves matrix multiplications and norm calculations, with a per-iteration complexity of $O(nd^2 + nl^2 + d^2k + dl + dk + l^2)$. Assuming the number of iterations is bounded by maxIter, the total iterative optimization complexity is $O(\text{maxIter} \times (nd^2 + nl^2 + d^2k + dl + dk + l^2))$. Combining both phases, the overall time complexity of the algorithm is $O(nd^2 + d^2k + ndl + dl + \text{maxIter} \times (nd^2 + nl^2 + d^2k + dl + dk + l^2))$.

IV. EXPERIMENTS AND EVALUATION

A. Datasets

We perform experiments on nine multi-label benchmark datasets to assess the performance of our EESRFSML algorithm. Table 1 summarizes the characteristics of the experimental datasets, where #Sample, #Label-Sample, #Unlabel-Sample, #Feature, and #Label denote the total number of samples, labeled samples, unlabeled samples, features and labels, respectively [38].

B. Comparative algorithms

In this experiment, we perform the comparative analysis of our EESRFSML algorithm and eight state-of-the-art algorithms. Meanwhile, we select six evaluation metrics: average precision, coverage, one error, ranking loss, Hamming loss and AUC. The following section provides the descriptions and parameter settings for these algorithms:

LIFT [15] is a multi-label classification algorithm that first leverages label-specific features and employs LIBSVM for classification. It has one parameter r, which is set to 0.1 in our experiment [39].

LLSF [24] is a multi-label classification algorithm that leverages label-specific features using the l_1-norm and incorporates label correlation learning. The parameters of LLSF include λ_1 and λ_2, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

LLSF-DL [24] extends LLSF by introducing class-dependent labels through sparse superposition and incorporating higher-order label correlations. The parameters of LLSF-DL include λ_1, λ_2, and λ_3, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

LSML [16] is a multi-label classification algorithm designed for missing labels, which learns label-specific features from a newly completed label matrix and incorporates label correlations. The parameters of LSML include $\lambda_1, \lambda_2, \lambda_3$ and λ_4, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

LSLC-ML [37] is a multi-label classification algorithm designed for missing labels, which recovers missing labels using label correlations and leverages label-specific features. The parameters of LSLC-ML include $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ and λ_6, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

CLML [17] employs both label-specific features and common features for multi-label classification, directly exploring label correlations through their labels, rather than relying on coefficient matrices. The parameters of CLML include $\lambda_1, \lambda_2, \lambda_3$ and λ_4, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

LRLSF [18] introduces the robust global label correlation method using the self-expression matrix and includes a manifold regularization term to capture local label correlations. The parameters of LRLSF include $\lambda_1, \lambda_2, \lambda_3$ and λ_4, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

ESRFS [22] is a multi-label classification algorithm that proposes an efficient Elastic Net-based method for multi-label data, designed to tackle sparsity and redundancy by selecting high-sparsity, low-redundancy features. The parameters of ESRFS include $\lambda_1, \lambda_2, \lambda_3$ and λ_4, these parameters are selected from $[2^{-5}, 2^5]$ in our experiment.

C. Experimental results

In this paper, we evaluate the classification performance of each algorithm on five benchmark datasets using five-fold cross-validation. The detailed results are presented in Tables 2 to 7, with the corresponding rankings summarized in Tables 8 to 13. For each evaluation metric, '↓' denotes that lower values indicate better performance, while '↑' indicates that higher values are preferred. As shown in Tables 8 through 13, our EESRFSML algorithm consistently surpasses LIFT [15], LLSF [24], LLSF-DL [24], LSML [16], LSLC-ML [37], CLML [17], LRLSF [18] and ESRFS [22] across nine datasets in terms of each evaluation metric. To evaluate the significant performance differences between our EESRFSML algorithm and the competing algorithms, we performed a Friedman test [40].

TABLE II
AVERAGE PRECISION OF ALL COMPARED ALGORITHMS ACROSS NINE PUBLIC AVAILABLE DATASETS

	Average Precision ↑								
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious
LIFT	0.5528 ± 0.0159	0.7599 ± 0.0055	0.5586 ± 0.0021	0.7311 ± 0.0017	0.5872 ± 0.0.0169	0.5628 ± 0.0029	0.5203 ± 0.0051	0.4235 ± 0.0029	0.2894 ± 0.0006
LLSF	0.6019 ± 0.0014	0.6775 ± 0.0145	0.5311 ± 0.0161	0.6884 ± 0.0073	0.5928 ± 0.0036	0.5980 ± 0.0112	0.5309 ± 0.0048	0.4106 ± 0.0029	0.2782 ± 0.0021
LLSF-DL	0.6006 ± 0.0028	0.8497 ± 0.0178	0.5511 ± 0.0026	0.7323 ± 0.0079	0.6070 ± 0.0137	0.5939 ± 0.0034	0.5281 ± 0.0108	0.4309 ± 0.0044	0.3416 ± 0.0006
LSML	0.5913 ± 0.0248	0.8235 ± 0.0041	0.5911 ± 0.0032	0.7377 ± 0.0007	0.5489 ± 0.0023	0.5542 ± 0.0019	0.5322 ± 0.0040	0.4144 ± 0.0035	0.3209 ± 0.0003
LSLC-ML	0.5970 ± 0.0.0116	0.8451 ± 0.0100	0.6111 ± 0.0042	0.7393 ± 0.0029	0.6089 ± 0.0071	0.5964 ± 0.0084	0.5341 ± 0.0072	0.4295 ± 0.0039	0.3316 ± 0.0008
CLML	0.5923 ± 0.0043	0.8496 ± 0.0112	0.6246 ± 0.0060	0.7408 ± 0.0030	0.5579 ± 0.0364	0.5790 ± 0.0094	0.5055 ± 0.0055	0.4340 ± 0.0079	0.2792 ± 0.0005
LRSLF	0.6039 ± 0.0139	0.8540 ± 0.0126	0.6012 ± 0.0106	0.6977 ± 0.0073	0.5919 ± 0.0174	0.5918 ± 0.0033	0.5373 ± 0.0044	0.4252 ± 0.0096	0.2926 ± 0.0009
ESRFS	0.5728 ± 0.0120	0.8247 ± 0.0123	0.5981 ± 0.0145	0.7195 ± 0.0081	0.5877 ± 0.0138	0.5826 ± 0.0092	0.5332 ± 0.0090	0.4318 ± 0.0100	0.3031 ± 0.0006
EESRFSML	0.6037 ± 0.0050	0.8620 ± 0.0019	0.6228 ± 0.0045	0.7418 ± 0.0013	0.6235 ± 0.0013	0.6146 ± 0.0014	0.5494 ± 0.0028	0.4562 ± 0.0031	0.3385 ± 0.0005

TABLE III
COVERAGE OF ALL COMPARED ALGORITHMS ACROSS NINE PUBLIC AVAILABLE DATASETS

	Coverage ↓								
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious
LIFT	0.6042 ± 0.00359	0.0739 ± 0.0046	0.4115 ± 0.0098	0.5354 ± 0.0019	0.2511 ± 0.0101	0.2017 ± 0.0036	0.1972 ± 0.0041	0.2308 ± 0.0094	0.6696 ± 0.0047
LLSF	0.5245 ± 0.0035	0.1453 ± 0.0084	0.3936 ± 0.0174	0.5544 ± 0.0164	0.2324 ± 0.0057	0.1954 ± 0.0059	0.1723 ± 0.0031	0.2486 ± 0.0045	0.6685 ± 0.0035
LLSF-DL	0.5161 ± 0.0034	0.0361 ± 0.0031	0.3371 ± 0.0018	0.5104 ± 0.0071	0.2246 ± 0.0057	0.1939 ± 0.0034	0.1769 ± 0.0006	0.1996 ± 0.0006	0.7281 ± 0.0044
LSML	0.5404 ± 0.0339	0.0680 ± 0.0034	0.3381 ± 0.0102	0.4874 ± 0.0028	0.2427 ± 0.0018	0.2881 ± 0.0031	0.2250 ± 0.0035	0.3136 ± 0.0035	0.7598 ± 0.0009
LSLC-ML	0.5479 ± 0.0082	0.0397 ± 0.0016	0.3471 ± 0.0092	0.4886 ± 0.0048	0.1944 ± 0.0046	0.1959 ± 0.0037	0.1723 ± 0.0020	0.1915 ± 0.0042	0.7276 ± 0.0023
CLML	0.5432 ± 0.0088	0.0361 ± 0.0094	0.3572 ± 0.0102	0.4813 ± 0.0005	0.2466 ± 0.0111	0.2678 ± 0.0134	0.2878 ± 0.0034	0.2850 ± 0.0094	0.7131 ± 0.0028
LRSLF	0.5240 ± 0.0035	0.0293 ± 0.0070	0.3153 ± 0.0066	0.4872 ± 0.0059	0.2271 ± 0.0080	0.1937 ± 0.0039	0.1531 ± 0.0028	0.2156 ± 0.0041	0.7601 ± 0.0037
ESRFS	0.5183 ± 0.0132	0.0496 ± 0.0129	0.3414 ± 0.0167	0.5053 ± 0.0141	0.2085 ± 0.0039	0.2132 ± 0.0097	0.1513 ± 0.0041	0.2235 ± 0.0091	0.6734 ± 0.0019
EESRFSML	0.5208 ± 0.0121	0.0266 ± 0.0045	0.3262 ± 0.0093	0.4632 ± 0.0020	0.2184 ± 0.0011	0.1797 ± 0.0016	0.1702 ± 0.0019	0.2224 ± 0.0030	0.6917 ± 0.0013

TABLE IV
ONE ERROR OF ALL COMPARED ALGORITHMS ACROSS NINE PUBLIC AVAILABLE DATASETS

	One Error ↓								
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious
LIFT	0.5781 ± 0.0211	0.2955 ± 0.0039	0.2444 ± 0.0121	0.2318 ± 0.0031	0.5256 ± 0.0324	0.5047 ± 0.0068	0.5969 ± 0.0092	0.5153 ± 0.0171	0.3962 ± 0.0056
LLSF	0.5625 ± 0.0000	0.3932 ± 0.0211	0.3333 ± 0.0271	0.3128 ± 0.0216	0.5204 ± 0.0053	0.4778 ± 0.0186	0.5769 ± 0.0070	0.5023 ± 0.0179	0.4745 ± 0.0078
LLSF-DL	0.5656 ± 0.0117	0.2259 ± 0.0249	0.3229 ± 0.0078	0.2477 ± 0.0199	0.4987 ± 0.0277	0.5022 ± 0.0058	0.5924 ± 0.0204	0.5030 ± 0.0088	0.3655 ± 0.0063
LSML	0.5427 ± 0.0211	0.2089 ± 0.0045	0.2939 ± 0.0163	0.2474 ± 0.0027	0.5167 ± 0.0045	0.5047 ± 0.0061	0.5361 ± 0.0085	0.4876 ± 0.0054	0.3496 ± 0.0018
LSLC-ML	0.5550 ± 0.0.0286	0.2250 ± 0.0151	0.2889 ± 0.0144	0.2440 ± 0.0079	0.4929 ± 0.0126	0.4978 ± 0.0149	0.5693 ± 0.0094	0.4672 0.0080	0.3665 ± 0.0019
CLML	0.5656 ± 0.0063	0.2205 ± 0.0223	0.2301 ± 0.0177	0.2147 ± 0.0018	0.5360 ± 0.0454	0.4929 ± 0.0101	0.5698 ± 0.0096	0.5350 ± 0.0181	0.4679 ± 0.0054
LRSLF	0.5250 ± 0.0249	0.2205 ± 0.0211	0.2458 ± 0.0263	0.2596 ± 0.0214	0.5625 ± 0.0550	0.4844 ± 0.0074	0.5689 ± 0.0051	0.5386 ± 0.0092	0.4375 ± 0.0066
ESRFS	0.5565 ± 0.0132	0.2868 ± 0.0249	0.3035 ± 0.0131	0.2619 ± 0.0021	0.5310 ± 0.0176	0.4992 ± 0.0099	0.5700 ± 0.0123	0.4852 ± 0.0081	0.3602 ± 0.0028
EESRFSML	0.5225 ± 0.0000	0.2217 ± 0.0155	0.2705 ± 0.0092	0.2428 ± 0.0033	0.4944 ± 0.0025	0.4907 ± 0.0044	0.5540 ± 0.0019	0.4748 ± 0.0047	0.3472 ± 0.0017

TABLE V
RANKING LOSS OF ALL COMPARED ALGORITHMS ACROSS NINE PUBLIC AVAILABLE DATASETS

	Ranking Loss ↓								
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious
LIFT	0.4086 ± 0.0251	0.0505 ± 0.0040	0.1675 ± 0.0022	0.2017 ± 0.0021	0.1648 ± 0.0121	0.1558 ± 0.0027	0.1612 ± 0.0034	0.1553 ± 0.0072	0.1478 ± 0.0019
LLSF	0.4245 ± 0.0049	0.1209 ± 0.0090	0.1740 ± 0.0070	0.2563 ± 0.0064	0.2498 ± 0.0037	0.1327 ± 0.0037	0.1238 ± 0.0031	0.1511 ± 0.0040	0.1779 ± 0.0011
LLSF-DL	0.4373 ± 0.0051	0.0232 ± 0.0074	0.1440 ± 0.0033	0.2077 ± 0.0055	0.1920 ± 0.0097	0.1332 ± 0.0019	0.1172 ± 0.0099	0.1428 ± 0.0015	0.1617 ± 0.0003
LSML	0.3958 ± 0.0361	0.0453 ± 0.0033	0.1426 ± 0.0032	0.1907 ± 0.0019	0.1837 ± 0.0020	0.2058 ± 0.0027	0.1373 ± 0.0029	0.1855 ± 0.0022	0.1915 ± 0.0004
LSLC-ML	0.4175 ± 0.0130	0.0276 ± 0.0067	0.1226 ± 0.0052	0.1895 ± 0.0033	0.1052 ± 0.0027	0.1342 ± 0.0038	0.1275 ± 0.0015	0.1506 ± 0.0025	0.31609 ± 0.0018
CLML	0.4138 ± 0.0054	0.0234 ± 0.0086	0.1357 ± 0.0043	0.1917 ± 0.0022	0.1864 ± 0.0128	0.1940 ± 0.0124	0.2292 ± 0.0039	0.1553 ± 0.0062	0.1583 ± 0.0004
LRSLF	0.4405 ± 0.0084	0.0222 ± 0.0088	0.1236 ± 0.0027	0.2550 ± 0.0068	0.2451 ± 0.0104	0.1432 ± 0.0024	0.1153 ± 0.0017	0.1214 ± 0.0050	0.2199 ± 0.0015
ESRFS	0.4183 ± 0.0144	0.0476 ± 0.0084	0.1573 ± 0.0057	0.2006 ± 0.0082	0.0811 ± 0.0041	0.1452 ± 0.0074	0.1167 ± 0.0032	0.1423 ± 0.0053	0.1534 ± 0.0014
EESRFSML	0.4106 ± 0.0065	0.0231 ± 0.0031	0.1172 ± 0.0040	0.1806 ± 0.0004	0.0855 ± 0.0009	0.1180 ± 0.0014	0.1268 ± 0.0016	0.1402 ± 0.0020	0.1533 ± 0.0003

TABLE VI
HAMMING LOSS OF ALL COMPARED ALGORITHMS ACROSS NINE PUBLIC AVAILABLE DATASETS

	Hamming Loss ↓								
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious
LIFT	0.3561 ± 0.0066	0.0284 ± 0.0005	0.0597 ± 0.0029	0.2325 ± 0.0019	0.0398 ± 0.0019	0.0669 ± 0.0001	0.0381 ± 0.0005	0.0240 ± 0.0004	0.0186 ± 0.0016
LLSF	0.3786 ± 0.0027	0.0275 ± 0.0024	0.0725 ± 0.0046	0.2406 ± 0.0061	0.0392 ± 0.0002	0.0555 ± 0.0006	0.0346 ± 0.0002	0.0149 ± 0.0004	0.0255 ± 0.0001
LLSF-DL	0.4156 ± 0.0080	0.0192 ± 0.0012	0.1115 ± 0.0021	0.2308 ± 0.0052	0.0404 ± 0.0002	0.0573 ± 0.0002	0.0343 ± 0.0002	0.0138 ± 0.0001	0.0182 ± 0.0000
LSML	0.3028 ± 0.0088	0.0159 ± 0.0004	0.0778 ± 0.0108	0.2457 ± 0.0016	0.0498 ± 0.0003	0.0556 ± 0.0001	0.0338 ± 0.0001	0.0132 ± 0.0000	0.0181 ± 0.0000
LSLC-ML	0.3385 ± 0.0023	0.0211 ± 0.0009	0.0578 ± 0.0021	0.2556 ± 0.0066	0.0407 ± 0.0001	0.0584 ± 0.0004	0.0347 ± 0.0001	0.0139 ± 0.0001	0.0182 ± 0.0009
CLML	0.3385 ± 0.0029	0.0207 ± 0.0007	0.0523 ± 0.0006	0.2240 ± 0.0068	0.0424 ± 0.0009	0.0582 ± 0.0007	0.0337 ± 0.0005	0.0211 ± 0.0005	0.0342 ± 0.0001
LRSLF	0.3630 ± 0.0042	0.0337 ± 0.0018	0.0548 ± 0.0049	0.2348 ± 0.0035	0.0589 ± 0.0030	0.0555 ± 0.0028	0.0387 ± 0.0011	0.0369 ± 0.0005	0.0176 ± 0.0012
ESRFS	0.3110 ± 0.0100	0.0277 ± 0.0008	0.0624 ± 0.0069	0.2155 ± 0.0087	0.0576 ± 0.0007	0.0628 ± 0.0011	0.0362 ± 0.0005	0.0151 ± 0.0001	0.0180 ± 0.0005
EESRFSML	0.3249 ± 0.0037	0.0227 ± 0.0044	0.0570 ± 0.0003	0.2260 ± 0.0020	0.0419 ± 0.0002	0.0525 ± 0.0001	0.0343 ± 0.0000	0.0138 ± 0.0000	0.0180 ± 0.0000

TABLE VII
AUC OF ALL COMPARED ALGORITHMS ACROSS NINE PUBLIC AVAILABLE DATASETS

	AUC ↑								
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious
LIFT	0.5052 ± 0.0298	0.9417 ± 0.0029	0.8372 ± 0.0036	0.7817 ± 0.0019	0.7699 ± 0.0095	0.8166 ± 0.0034	0.8210 ± 0.0029	0.8356 ± 0.0061	0.8510 ± 0.0008
LLSF	0.5541 ± 0.0041	0.8672 ± 0.0071	0.8108 ± 0.0054	0.7240 ± 0.0075	0.7852 ± 0.0038	0.8330 ± 0.0046	0.8563 ± 0.0027	0.8515 ± 0.0029	0.8203 ± 0.0014
LLSF-DL	0.5415 ± 0.0052	0.9692 ± 0.0062	0.8585 ± 0.0031	0.7828 ± 0.0058	0.7875 ± 0.0154	0.8289 ± 0.0026	0.8601 ± 0.0119	0.8890 ± 0.0026	0.8203 ± 0.0005
LSML	0.5500 ± 0.0379	0.9427 ± 0.0029	0.8493 ± 0.0032	0.7960 ± 0.0017	0.7860 ± 0.0013	0.7473 ± 0.0023	0.7977 ± 0.0070	0.8808 ± 0.0023	0.8070 ± 0.0004
LSLC-ML	0.5617 ± 0.0108	0.9660 ± 0.0063	0.8593 ± 0.0043	0.7985 ± 0.0030	0.8693 ± 0.0045	0.8276 ± 0.0046	0.8472 ± 0.0026	0.8494 ± 0.0017	0.8376 ± 0.0013
CLML	0.5645 ± 0.0054	0.9692 ± 0.0079	0.8630 ± 0.0032	0.7990 ± 0.0023	0.7799 ± 0.0085	0.7648 ± 0.0117	0.7506 ± 0.0031	0.8293 ± 0.0051	0.8738 ± 0.0004
LRSLF	0.5487 ± 0.0068	0.9750 ± 0.0017	0.8791 ± 0.0027	0.7401 ± 0.0060	0.7459 ± 0.0082	0.8298 ± 0.0032	0.8641 ± 0.0023	0.8830 ± 0.0026	0.7869 ± 0.0006
ESRFS	0.5608 ± 0.0148	0.9751 ± 0.0092	0.8589 ± 0.0047	0.7835 ± 0.0063	0.9077 ± 0.0036	0.8169 ± 0.0075	0.8404 ± 0.0039	0.8249 ± 0.0085	0.8452 ± 0.0014
EESRFSML	0.5620 ± 0.0106	0.9784 ± 0.0041	0.8780 ± 0.0032	0.8063 ± 0.0005	0.8993 ± 0.0011	0.8473 ± 0.0013	0.8480 ± 0.0016	0.8526 ± 0.0014	0.8455 ± 0.0003

TABLE VIII
THE RANKING OF AVERAGE PRECISION ACROSS NINE PUBLIC AVAILABLE DATASETS

	Average Precision									
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious	average
LIFT	9	8	8	6	7	8	7	7	7	7.44444
LLSF	3	9	9	9	4	2	6	9	9	6.66667
LLSF-DL	4	3	7	5	3	4	8	4	1	4.33333
LSML	7	7	6	4	9	9	5	8	4	6.55556
LSLC-ML	5	5	3	3	2	3	3	5	3	3.55556
CLML	6	4	1	2	8	7	9	2	8	5.22222
LRSLF	1	2	4	8	5	5	2	6	6	4.33333
ESRFS	8	6	5	7	6	6	4	3	5	5.55556
EESRFSML	2	1	2	1	1	1	1	1	2	1.33333

TABLE IX
THE RANKING OF COVERAGE ACROSS NINE PUBLIC AVAILABLE DATASETS

	Coverage									
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious	average
LIFT	9	8	9	8	9	6	7	6	2	7.11111
LLSF	5	9	8	9	6	4	5	7	1	6.00000
LLSF-DL	1	3	3	7	4	3	6	2	7	4.00000
LSML	6	7	4	4	8	9	8	9	8	7.00000
LSLC-ML	8	5	6	5	1	5	4	1	6	4.55556
CLML	7	3	7	2	7	8	9	8	5	6.22222
LRLSF	4	2	1	3	5	2	2	3	9	3.44444
ESRFS	2	6	5	6	2	7	1	5	3	4.11111
EESRFSML	3	1	2	1	3	1	3	4	4	2.44444

TABLE X
THE RANKING OF ONE ERROR ACROSS NINE PUBLIC AVAILABLE DATASETS

	One Error									
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious	average
LIFT	9	8	2	2	6	8	9	7	6	6.33333
LLSF	5	9	9	9	5	1	7	5	9	6.55556
LLSF-DL	6	6	8	6	3	7	8	6	4	6.00000
LSML	3	1	7	5	4	8	1	4	2	3.88889
LSLC-ML	4	5	5	4	1	6	4	1	4	3.77778
CLML	6	2	1	1	8	4	5	8	8	4.77778
LRLSF	2	2	3	7	9	2	3	9	7	4.88889
ESRFS	8	7	6	8	7	5	6	3	3	5.88889
EESRFSML	2	4	4	4	2	3	2	2	1	2.44444

TABLE XI
THE RANKING OF RANKING LOSS ACROSS NINE PUBLIC AVAILABLE DATASETS

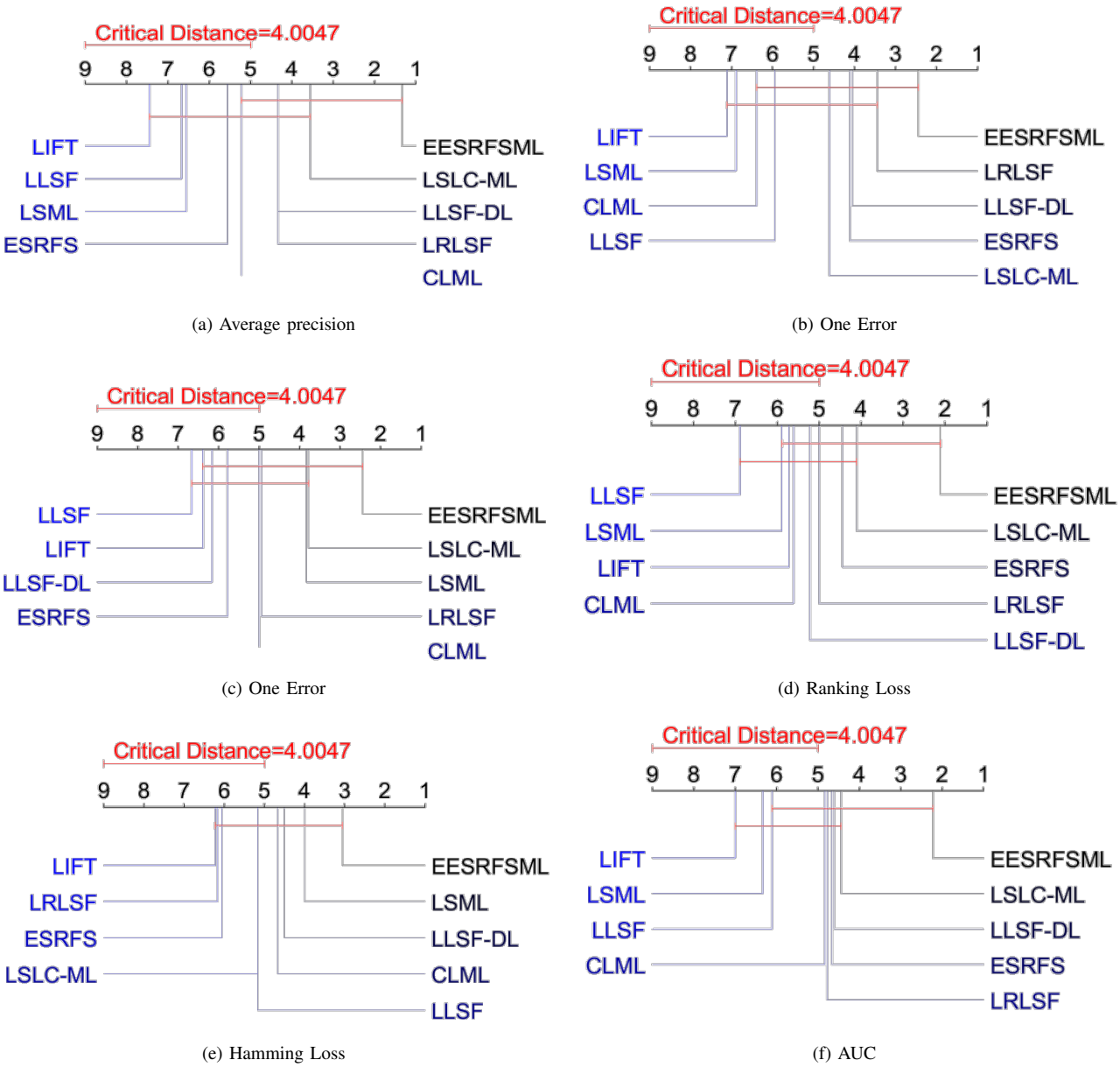
	Ranking Loss									
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious	average
LIFT	2	8	8	6	4	7	8	7	1	5.66667
LLSF	7	9	9	9	9	2	4	6	7	6.88889
LLSF-DL	8	3	6	7	6	3	3	4	5	5.00000
LSML	1	6	5	3	5	9	7	9	8	5.88889
LSLC-ML	5	5	2	2	3	4	6	5	6	4.22222
CLML	4	4	4	4	7	8	9	7	4	5.66667
LRLSF	9	1	3	8	8	5	1	1	9	5.00000
ESRFS	6	7	7	5	1	6	2	3	3	4.44444
EESRFSML	3	2	1	1	2	1	5	2	2	2.11111

TABLE XII
THE RANKING OF HAMMING LOSS ACROSS NINE PUBLIC AVAILABLE DATASETS

	Hamming Loss									
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious	average
LIFT	1	6	9	7	5	2	9	8	8	6.77778
LLSF	2	8	7	8	7	1	2	5	5	5.66667
LLSF-DL	3	9	2	9	4	3	5	3	2	4.66667
LSML	4	1	1	6	8	7	4	2	1	3.77778
LSLC-ML	5	4	4	4	9	4	7	6	4	5.22222
CLML	6	4	3	1	2	6	6	1	7	4.33333
LRLSF	7	7	8	2	6	9	2	9	9	5.88889
ESRFS	8	2	6	5	1	8	8	7	6	5.00000
EESRFSML	9	3	5	3	3	5	1	3	2	3.00000

TABLE XIII
THE RANKING OF AUC ACROSS NINE PUBLIC AVAILABLE DATASETS

	AUC									
Algorithm	emotion	medical	enron	yeast	education	art	science	bibtex	delicious	average
LIFT	9	8	8	7	8	7	7	7	2	7.00000
LLSF	6	9	9	8	6	2	3	5	7	6.11111
LLSF-DL	4	4	6	6	4	4	2	1	6	4.11111
LSML	7	7	7	4	5	9	8	3	8	6.44444
LSLC-ML	3	6	4	3	3	5	5	6	5	4.44444
CLML	1	4	3	2	7	8	9	8	1	4.77778
LRLSF	8	3	1	9	9	3	1	2	9	5.00000
ESRFS	5	2	5	5	1	6	6	9	4	4.77778
EESRFSML	2	1	2	1	2	1	4	4	3	2.22222



(a) Average precision

(b) One Error

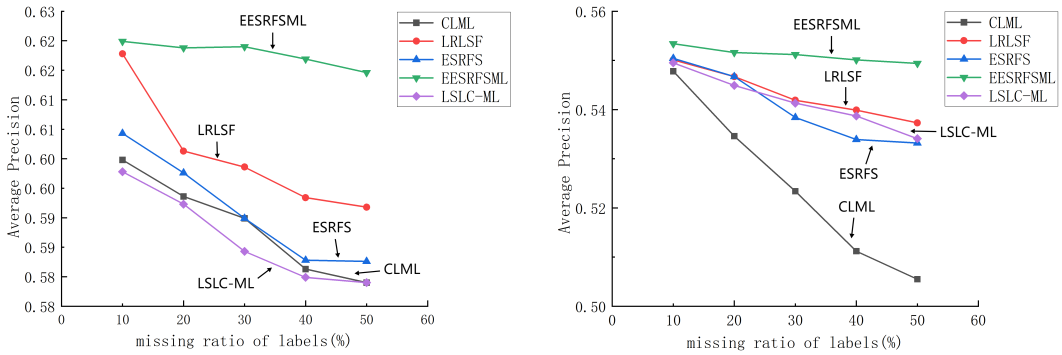
(c) One Error

(d) Ranking Loss

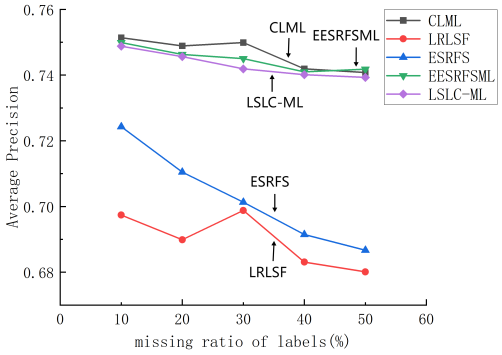
(e) Hamming Loss

(f) AUC

Fig.1. Comparison of the EESRFSML algorithm with other algorithms using the Nemenyi test at a 0.05 significance level.

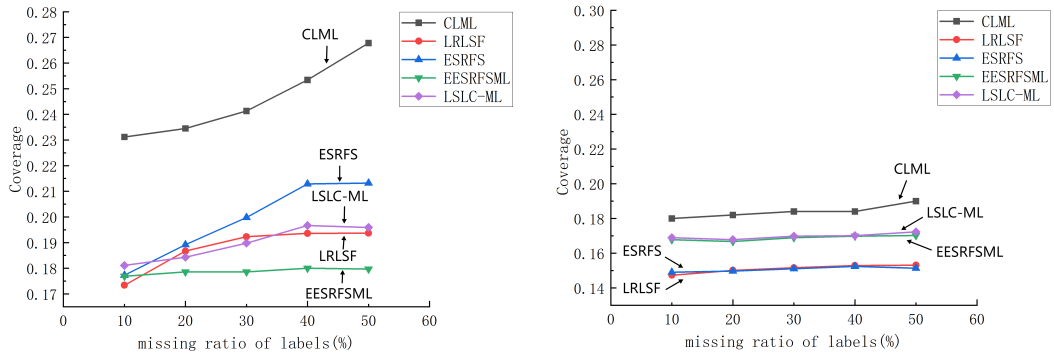


(a) art (b) science

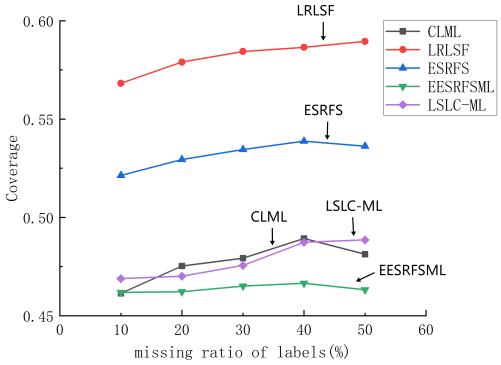


(b) yeast

Fig.2. The impact of the samples of missing labels on average precision.

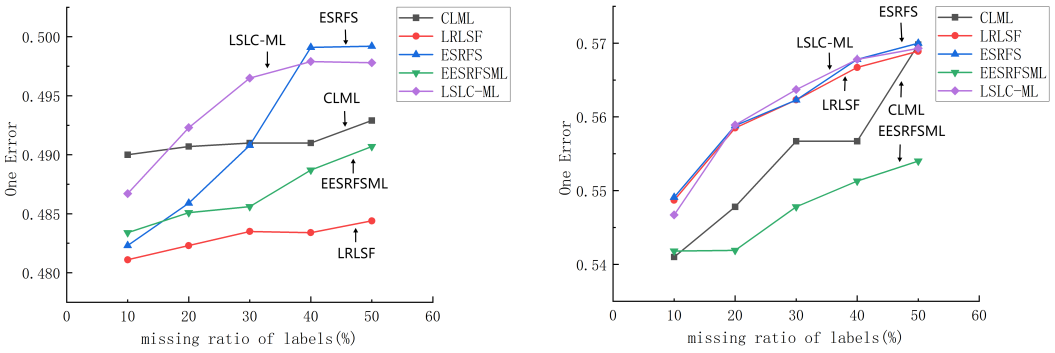


(a) art (b) science

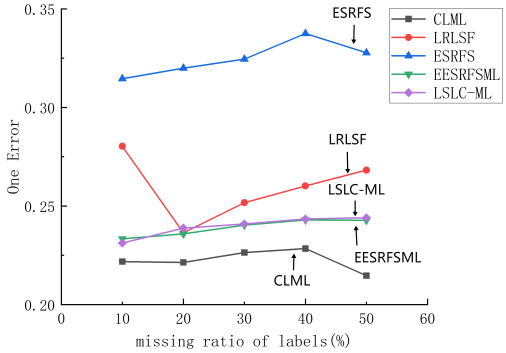


(b) yeast

Fig.3. The impact of the samples of missing labels on coverage.

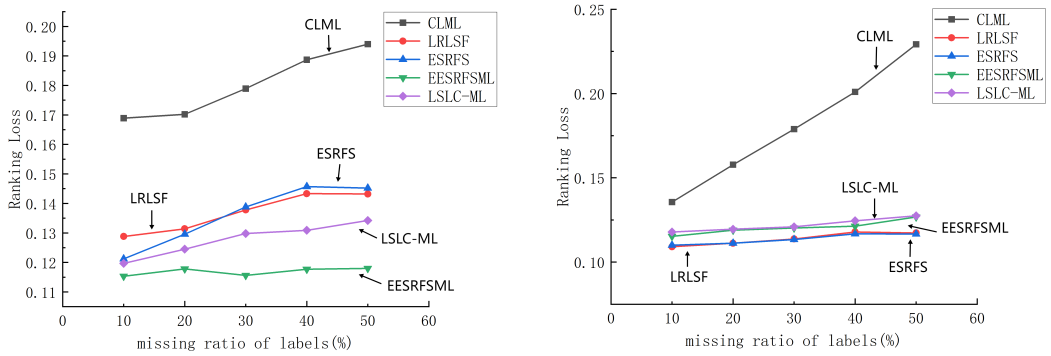


(a) art (b) science

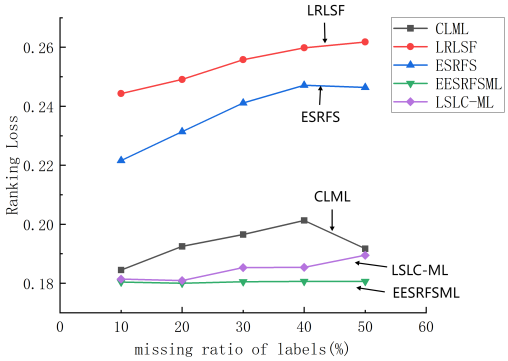


(b) yeast

Fig.4. The impact of the samples of missing labels on one error.

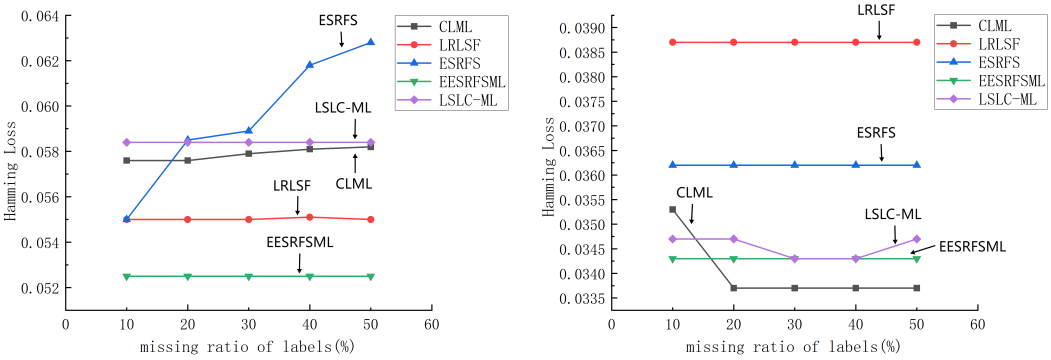


(a) art (b) science

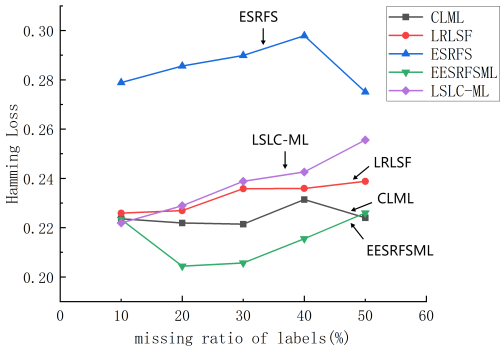


(b) yeast

Fig.5. The impact of the samples of missing labels on ranking loss.

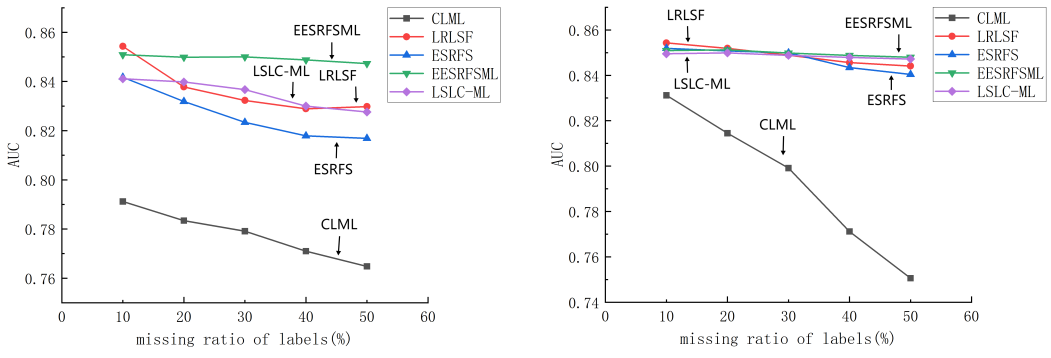


(a) art (b) science

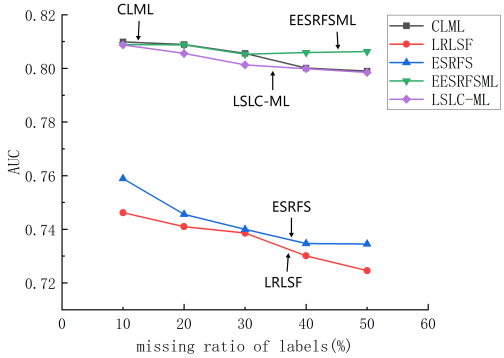


(b) yeast

Fig.6. The impact of the samples of missing labels Hamming loss.



(a) art (b) science



(b) yeast

Fig.7. The impact of the samples of missing labels on AUC.

TABLE XIV
FRIEDMAN TEST RESULTS $F_F(K = 9, N = 6)$ AND CRITICAL VALUES FOR DIFFERENT METRICS

Evaluation Metric	F_F	Critical Value($\alpha = 0.05$)
One Error	19.274074	2.086758
Ranking Loss	17.696296	
Average Precision	33.540741	
Hamming Loss	10.377778	
AUC	18.451852	
Coverage	25.955556	

Table 14 presents each evaluation metric, along with its corresponding F_F value and critical value. At a significance level of $\alpha = 0.05$, we rejected the null hypothesis that all competing algorithms perform equally, as the F_F value surpassed the critical value, indicating significant differences between the algorithms. We applied the Nemenyi test [40] to determine whether EESRFSML outperforms the other algorithms, designating EESRFSML as the reference method. We compared the average rank differences between algorithm pairs using the critical difference (CD): $CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}$. For Nemenyi test, $q_\alpha = 2.08676$, $CD = 4.0047(K = 9, N = 6)$ at a significance level of $\alpha = 0.05$. The CD diagrams for each metric are shown in Fig. 1. It displays the CD diagrams for each evaluation metric. In each subplot, a line connects two algorithms if their average ranks are within one CD of each other. Conversely, two unconnected algorithms are considered to exhibit a significant difference for that evaluation metric. It is evident that our EESRFSML algorithm outperforms the other eight algorithms.

D. Discussion on the experimental results

The experimental results lead to the following conclusions:

1.LLSF and LLSF-DL outperform LIFT by effectively leveraging label correlation information. This comprehensive exploitation of label dependencies enhances their robustness and generalization performance, even under conditions of label space redundancy and imbalanced label distributions.

2.CLML and ESRFS outperform LLSF and LLSF-DL by effectively leveraging shared feature representations. By constructing a shared feature strategy within the feature space, they achieve superior classification performance on datasets characterized by high label semantic overlap or strong feature correlations.

3.EESRFSML outperforms CLML, LRSFS, and ESRFS by effectively addressing the challenge of missing label information. In real-world multi-label learning scenarios, missing labels are a common issue. However, many existing methods either disregard these incomplete samples or rely on strong assumptions to infer the missing labels, often resulting in reduced classification performance. EESRFSML overcomes this limitation by employing a label correlation coefficient matrix to complete missing labels at the label level, enabling more accurate recovery of label information and more effective utilization of incomplete data.

E. Analysis of parameter sensitivity

Samples with missing labels can substantially affect an algorithms classification performance. As illustrated in Figures 2 to 7, we evaluate the performance of various algorithms under different missing label ratios. The results demonstrate a general decline in classification accuracy as the proportion of incomplete samples increases. Nevertheless, the proposed EESRFSML algorithm maintains notably stable performance across varying missing rates, highlighting its superior robustness to missing label scenarios compared to other methods.

V. CONCLUSION

In this paper, we present EESRFSML, an effective algorithm for multi-label classification with missing labels that outperforms existing methods. Our approach utilizes the label correlation matrix to reconstruct an enhanced label matrix, enabling the recovery of a substantial portion of missing labels. Additionally, we propose a label-level regularizer that directly captures both global and local label dependencies from the label outputs, avoiding sole reliance on the correlation matrix. Future work will explore integrating this algorithm within a multi-view learning framework and tackling the challenges posed by noisy labels.

REFERENCES

[1] Sun K., He M., Xu Y., et al. Multi-label classification of fundus images with graph convolutional network and lightgbm. *Computers in Biology and Medicine*, 149:105909, 2022.

[2] Ameer I., Bölücü N., Siddiqui M H F., et al. Multi-label emotion classification in texts using transfer learning. *Expert Systems with Applications*, 213:118534, 2023.

[3] Qin Q., Xian L., Xie K., et al. Deep multi-similarity hashing with semantic-aware preservation for multi-label image retrieval. *Expert Systems with Applications*, 205:117674, 2022.

[4] Dhanuka R., Tripathi A., and Singh J P. A semi-supervised autoencoder-based approach for protein function prediction. *IEEE Journal of Biomedical and Health Informatics*, 26(10):4957–4965, 2022.

[5] Boutell M R., Luo J., Shen X., et al. Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771, 2004.

[6] Tsoumakas G. and Vlahavas I. Random k-labelsets: An ensemble method for multilabel classification. In *European Conference on Machine Learning*, pages 406–417. Springer Berlin Heidelberg, 2007.

[7] Read J., Pfahringer B., Holmes G., and Frank E. Classifier chains for multi-label classification. *Machine Learning*, 85:333–359, 2011.

[8] Clare A. and King R D. Knowledge discovery in multi-label phenotype data. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 42–53. Springer, 2001.

[9] Ghamrawi N. and McCallum A. Collective multi-label classification. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pages 195–200, 2005.

[10] Elisseeff A. and Weston J. A kernel method for multi-labelled classification. In *Advances in Neural Information Processing Systems*, volume 14, 2001.

[11] Zhang M. and Zhou Z.-H. Ml-knn: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7):2038–2048, 2007.

[12] Yu Liu, Jie-Sheng Wang, Jia-Yao Wen, Yu-Tong Li, and Peng-Guo Yan. A sparse genetic algorithm to solve feature selection of sparse high-dimensional data and liver totxicity classification. *Engineering Letters*, vol. 33, no. 4, pp1045–1060, 2025.

[13] Qing Ai, Fei Li, Qingyun Gao, and Fei Zhao. An improved lst-ksvc based on energy model. *Engineering Letters*, vol. 30, no. 4, pp1436–1443, 2022.

[14] Li F., Ai Q., Li X., Wang W., Gao Q., and Zhao F. Intuitionistic fuzzy least squares mltsvm for noisy label data using label-specific features and local label correlation. *Expert Systems with Applications*, 260:125351, 2025.

[15] Zhang M. and Wu L. Lift: Multi-label learning with label-specific features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):107–120, 2014.

[16] Huang J., Qin F., Zheng X., Cheng Z., Yuan Z., Zhang W., and Huang Q. Improving multi-label classification with missing labels by learning label-specific features. *Information Sciences*, 492:124–146, 2019.

[17] Li J., Li P., Hu X., and Yu K. Learning common and label-specific features for multi-label classification with correlation information. *Pattern Recognition*, 121:108259, 2022.

[18] Yang Y., Chen H., Mi Y., Luo C., Horng S.-J., and Li T. Multi-label feature selection based on stable label relevance and label-specific features. *Information Sciences*, 648:119525, 2023.

[19] Hu L., Li Y., Gao W., Zhang P., and Hu J. Multi-label feature selection with shared common mode. *Pattern Recognition*, 104:107344, 2020.

[20] Zhu P., Xu Q., Hu Q., Zhang C., and Zhao H. Multi-label feature selection with missing labels. *Pattern Recognition*, 74:488–502, 2018.

[21] Jiang L., Yu G., Guo M., and Wang J. Feature selection with missing labels based on label compression and local feature correlation. *Neurocomputing*, 395:95–106, 2020.

[22] Li Y., Hu L., and Gao W. Multi-label feature selection with high-sparse personalized and low-redundancy shared common features. *Information Processing and Management*, 61(3):103633, 2024.

[23] Ai Q., Li F., Li X., Zhao J., Wang W., Gao Q., and Zhao F. An improved mltsvm using label-specific features with missing labels. *Applied Intelligence*, 53(7):8039–8060, 2023.

[24] Huang J., Li G., Huang Q., and Wu X. Learning label-specific features and class-dependent labels for multi-label classification. *IEEE Transactions on Knowledge and Data Engineering*, 28(12):3309–3323, 2016.

[25] Yu Z.-B. and Zhang M.-L. Multi-label classification with label-specific feature generation: A wrapped approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5199–5210, 2021.

[26] Fan Y., Chen B., Huang W., Liu J., Weng W., and Lan W. Multi-label feature selection based on label correlations and feature redundancy. *Knowledge-Based Systems*, 241:108256, 2022.

[27] Cheng Y., Qian K., and Min F. Global and local attention-based multi-label learning with missing labels. *Information Sciences*, 594:20–42, 2022.

[28] Zhang J., Wu H., Jiang M., Liu J., Li S., Tang Y., and Long J. Group-preserving label-specific feature selection for multi-label learning. *Expert Systems with Applications*, 213:118861, 2023.

[29] Hu L., Li Y., Gao W., et al. Multi-label feature selection with shared common mode. *Pattern Recognition*, 104:107344, 2020.

[30] Li R., Ouyang Z., Shang Z., et al. Learning common and label-specific features for multi-label classification with missing labels. *IEEE Access*, 12:81170–81195, 2024.

[31] Deng J., Chen D., Wang H., et al. Matrix factorization algorithm for multi-label learning with missing labels based on fuzzy rough set. *Fuzzy Sets and Systems*, 498.

[32] Faraji M., Seyedi S. A., Tab F. A., et al. Multi-label feature selection with global and local label correlation. *Expert Systems with Applications*, 246:123198, 2024.

[33] Zhu Y., Kwok J. T., and Zhou Z. Multi-label learning with global and local label correlation. *IEEE Transactions on Knowledge and Data Engineering*, 30(6):1081–1094, 2018.

[34] Fan Y., Liu J., Tang J., et al. Learning correlation information for multi-label feature selection. *Pattern Recognition*, 145:109899, 2024.

[35] Xu S., Shang L., Shen F., et al. Incomplete label distribution learning via label correlation decomposition. *Information Fusion*, 113:102600, 2025.

[36] Rastogi R. and Chowdhury S. Binary-tree based mean-averaging estimation for multi-label classification. In *International Conference on Pattern Recognition*, pages 271–285. Springer, Cham, 2025.

[37] Cheng Z. and Zeng Z. Joint label-specific features and label correlation for multi-label learning with missing label. *Applied Intelligence*, 50(11):4029–4049, 2020.

[38] Guang Yi Chen and Adam Krzyzak. An experimental study for the effects of noise on pattern recognition algorithms. *Engineering Letters*, vol. 33, no. 5, pp1185–1193, 2025.

[39] Guo Zhenwei, Li Haojie, Zhao Ruiqiang, Jiang Yongyan, and Deng Yingcai. Longitudinal protection of line direction based on vmd-ht and svm. *Engineering Letters*, vol. 33, no. 5, pp1429–1436, 2025.

[40] Demsar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.