MCF-Spouse: A Multi-Label Causal Feature Selection Method with Optimal Spouses Discovery

Lin Ma[1] , **Liang Hu**[1] , **Qiang Huang**[2] , **Pingting Hao**[3] and **Juncheng Hu**[1*]
[1]College of Computer Science and Technology, Jilin University
[2]Department of Machine Learning, Mohamed Bin Zayed University of Artificial Intelligence
[3]College of Computer Science and Information Technology, Northeast Normal University
malin23@mails.jlu.edu.cn, hul@jlu.edu.cn, qiang.huang@mbzuai.ac.ae, haopingting@nenu.edu.cn,
jchu@jlu.edu.cn*

Abstract

Multi-label causal feature selection has garnered considerable attention for its ability to identify the most informative features while accounting for the causal dependencies between labels and features. However, previous work often overlooks the unique contributions of labels to the target variables in multi-label settings, focusing instead on prioritizing feature variables. Moreover, existing methods typically rely on traditional Markov Blanket (MB) discovery to construct an initial MB, which often fails to explore the most valuable form of spouse variables to feature selection in multi-label scenarios, leading to significant computational overhead due to redundant Conditional Independence (CI) tests required for spouse search. To address these challenges, we propose the Multi-label Causal Feature Selection Method with Optimal Spouses Discovery, MCF-Spouse, which leverages mutual information to quantify the contributions of both labels and features, ensuring the retention of the most informative variables in multi-label settings. Moreover, we systematically analyze all potential forms of spouse variables to identify the optimal spouse case, significantly reducing the spouse search space and alleviating the time overhead associated with CI tests. Experiments conducted on diverse real-world datasets demonstrate that MCF-Spouse consistently outperforms state-of-the-art methods across multiple metrics, offering a scalable and interpretable solution for multi-label causal feature selection. The code is available at https://github.com/malinjlu/MCF-Spouse.

1 Introduction

Causal inference is crucial in machine learning, enhancing model efficiency, interpretability, and predictive accuracy in complex domains [Chu *et al.*, 2023; Ma *et al.*, 2025]. One of the fundamental methods to causal inference involves Bayesian networks (BNs) [Ben-Gal, 2008; Pearl, 2014], which utilize directed acyclic graphs (DAGs) to model dependencies among variables [Aliferis *et al.*, 2010a; Aliferis *et al.*, 2010b]. Within a BN, the Markov blanket (MB) [Tsamardinos *et al.*, 2003b] of a target variable captures its essential local causal structure, comprising direct causes (parents), direct effects (children), and other causes of its children (spouses). Identifying MBs across multiple variables imposes valuable constraints, effectively reducing the search space and facilitating more scalable approaches for inferring large-scale BNs [Tsamardinos *et al.*, 2006; Pellet and Elisseeff, 2008; Gao *et al.*, 2017]. More importantly, prior studies have demonstrated that the MB represents the optimal feature subset, as all other features become independent of the target when conditioned on its MB, which underscores its significance in feature selection applications [Masegosa and Moral, 2012; Statnikov *et al.*, 2013].

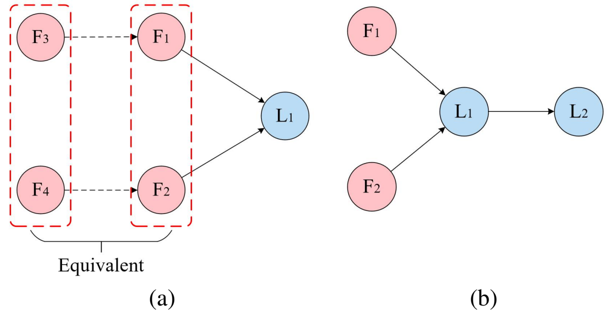


Figure 1: (a) Illustration of equivalent information in multi-label scenario. In this case, $\{F_1, F_2\}$ and $\{F_3, F_4\}$ have the equivalent information to L_1. (b) Strong label correlation can block the MB path from feature to label. In this case, L_1 prevents $\{F_1, F_2\}$ from becoming the MB of L_2.

Single-label causal feature selection methods focus on identifying relevant features for a single target variable [Ling *et al.*, 2022b; Guo *et al.*, 2023; Ling *et al.*, 2024a]. These methods fail to consider label dependencies, which are essential in multi-label scenarios [Ling *et al.*, 2024b; Guo *et al.*, 2024]. For example, when predicting diabetes and hypertension, features such as BMI and cholesterol may influence both conditions, highlighting the interdependencies between labels. Multi-label causal feature selection addresses this issue by simultaneously capturing feature-feature, feature-

[*] Corresponding author

label, and label-label relationships [Wu *et al.*, 2020; Yu *et al.*, 2021; Wu *et al.*, 2022]. By considering causal relationships across multiple labels, multi-label causal feature selection tries to capture the underlying structure among feature-feature, feature-label and label-label simultaneously, enhancing model performance and interpretability in multi-label settings [Yu *et al.*, 2021].

However, due to the unique characteristics of multi-label scenarios, several challenges arise in identifying the MB for target labels. First, complex interdependencies often challenge the faithfulness assumption [Pearl, 2014], particularly due to the presence of equivalent information [Wu *et al.*, 2022]. As illustrated in Fig. 1.(a), $\{F_1, F_2\}$ directly convey information to L_1 through the paths $F_1 \rightarrow L_1$ and $F_2 \rightarrow L_1$. $\{F_3, F_4\}$ have the paths to $\{F_1, F_2\}$ as $F_3 \rightarrow F_1$ and $F_4 \rightarrow F_2$. In this case, $\{F_3, F_4\}$ provide equivalent information to L_1 as $\{F_1, F_2\}$, raising the question of which variables should be retained in L_1's MB. Furthermore, as shown in Fig. 1.(b), the equivalent information provided by L_1 and $\{F_1, F_2\}$ presents additional challenges in the presence of strong correlations between labels. Specifically, L_1, which encapsulates the information from $\{F_1, F_2\}$, can act as a blocker, obstructing the path from $\{F_1, F_2\}$ to L_2 according to the definition of D-separation [Pearl, 2014]. This complicates the identification of informative variables within the MB, as it requires careful consideration of the interplay between features and labels to avoid redundancy or loss of crucial information. Existing studies [Wu *et al.*, 2020; Yu *et al.*, 2021; Wu *et al.*, 2022] tend to retain as many features as possible, prioritizing features over labels. These methods frequently recover features blocked by strong label correlations through extensive Conditional Independence (CI) tests. However, in multi-label contexts, the relationships between labels are equally critical. It is possible that labels, as output variables, encapsulate more concentrated and predictive information than input features, which primarily represent attributes. It is essential to develop methods that not only account for the equivalent information provided by features and labels but also leverage the rich interdependencies between labels to enhance the identification of the MB.

More importantly, in multi-label causal feature selection, spouse sets are crucial for capturing the complete Markov blanket of a target label by addressing indirect causal relationships [Ling *et al.*, 2022a]. Unlike parent and children variables, which represent direct causal influences, spouse variables affect the target variables through their impact on shared children. This makes spouse variables essential for modeling the intricate interactions and dependencies among multiple labels. Ignoring spouse variables can lead to incomplete MBs, since indirect effects can be overlooked, potentially resulting in biased outcomes. The inclusion of spouse variables is particularly important in multi-label scenarios, where labels often share common causes or are connected through indirect pathways. However, researchers often overlook the spouse variables, focusing instead on PC variables, under the assumption that PC variables contain the most relevant information for the target variables [Wu *et al.*, 2020; Yu *et al.*, 2021]. Alternatively, some methods rely on existing MB construction such as HITON-MB [Aliferis *et al.*, 2003], MMMB [Tsamardinos *et al.*, 2003a], and GetPC [Pena *et al.*, 2007] to establish the initial MB, followed by further optimization [Wu *et al.*, 2022; Wu *et al.*, 2023a; Wu *et al.*, 2023b]. These methods may either risk compromising the integrity of the MB or introduce significant CI tests due to the high dimensionality of features and labels in multi-label datasets when searching for spouse sets.

In this article, we propose the Multi-label Causal Feature Selection Method with Optimal Spouse Discovery (MCF-Spouse), which addresses key challenges in multi-label causal feature selection through several innovations. First, to tackle the issues of equivalent information and the blockage of paths from features to labels caused by strong label correlations, we utilize mutual information (MI) [Zhang *et al.*, 2019b; Pereira *et al.*, 2018] to quantify the contributions of both labels and features to the target variable. This approach ensures that only the most informative variables are retained, prioritizing labels when they provide greater predictive value. Second, to efficiently identify spouse variables without incurring significant computational overhead from CI tests, we systematically analyze all possible configurations of spouse variables in the multi-label domain. By leveraging target-specific guidance, we pinpoint the optimal spouse variables to include in the target's Markov blanket. This targeted strategy significantly narrows the search space for spouse variables, reducing computational complexity and minimizing time overhead. The main contributions of this paper are summarized as follows:

1. We discuss all the possible spouse relationships in the context of multi-label causal feature selection, and analyze the optimal spouse set through the guidance of target-specific variables.

2. To solve the equivalent information and the blockage of paths from features to labels caused by strong label correlations, we use mutual information to compare the contribution of features and labels, and maintain the most informative variables.

3. We conduct extensive experiments to validate our MCF-Spouse method, comparing it with nine state-of-the-art multi-label feature selection methods on four metrics across eight real-world datasets. The results consistently show that our MCF-Spouse method outperforms the compared methods across all four metrics.

2 Related Work

Non-BN-based methods. These methods can be categorized into three categories [Pereira *et al.*, 2018]. *Mutual information-based methods*, such as D2F [Lee and Kim, 2015], MCMFS [Zhang *et al.*, 2020], ENM [Gonzalez-Lopez *et al.*, 2020], and FSSL [Liu *et al.*, 2020], select features by evaluating mutual information between features and labels. *Regularization-based methods*, including SFUS [Ma *et al.*, 2012], MIFS [Jian *et al.*, 2016], MLFS-GLOCAL [Faraji *et al.*, 2024], and ESRFS [Li *et al.*, 2024], incorporate regularization terms to enhance feature selection. *Manifold learning-based methods*, such as MCLS [Huang *et al.*, 2018], MSSL [Cai and Zhu, 2018], and MDFS [Zhang *et* *al.*, 2019a], exploit geometric structures within data. While these methods improve classification performance and mitigate the curse of dimensionality, they can not select the optimal number of features automatically and explicitly model feature-feature and label-label dependencies.

BN-based methods. BN-based methods offer significant advantages by simultaneously modeling feature-feature, feature-label, and label-label relationships [Yu *et al.*, 2021]. For instance, Wu *et al.* explore causal mechanisms in multi-label feature selection by distinguishing common features and label-specific features in MB-MCF [Wu *et al.*, 2020]. They further extend this work with CLFS [Wu *et al.*, 2022], focusing on the interplay between common MB variables and equivalent information. Similarly, Yu *et al.* [Yu *et al.*, 2021] address the issue of false discoveries—where A appears in the MB set of B, but B is absent from the MB set of A in M2LC. This method reduces computational complexity by learning local BN structures rather than entire BN structure.

However, existing BN-based methods often suffer from high computational overhead due to redundant CI tests during spouse discovery. Furthermore, the complex dependencies in multi-label feature selection often lead to equivalent information, making it challenging to determine the most informative MB variables. To address these issues, we propose an efficient spouse discovery method and leverage mutual information to accurately evaluate the contributions of both features and labels, thereby reducing complexity and improving feature selection performance.

3 Preliminaries

In this section, we elaborate on the fundamental definitions and theorems relevant to our discussion. Based on these, we present the theorems and provide their proofs in this article.

Definition 1 (Faithfulness [Pearl, 2014]): a Directed Acyclic Graph (DAG) G is considered faithful to the probability distribution $P(V)$ in the Bayesian Network $< V, G, P(V) >$ *iff* every conditional independence relationship encoded in P is implied by G and satisfies the Markov condition. Furthermore, $P(V)$ is deemed faithful *iff* G accurately represents $P(V)$.

Definition 2 (Equivalent information [Statnikov *et al.*, 2013]): as illustrated in Fig. 1.(a), the features $\{F_3, F_4\}$ provide the same information to the label L_1 as $\{F_1, F_2\}$. This equivalence implies that $\{F_1, F_2\}$ can be replaced by $\{F_3, F_4\}$ without any loss of information. In this scenario, $\{F_1, F_2\}$ and $\{F_3, F_4\}$ are considered equivalent information in their contribution to L_1.

The faithfulness condition in Definition 1 is always invalid due to the presence of equivalent information in the context of multi-label learning and raises the question of which should be retained in the target's Markov blanket. We introduce Definition 3 and 4 and Theorem 1 to solve this problem.

Definition 3 (Common causal variable): as illustrated in Fig. 2, C is identified as the common variable, since it is present in the Markov blankets of both X and Y.

Definition 4 (Target-specific variable): as illustrated in Fig. 2, A is the target-specific variable for X, and B is the target-specific variable for Y, which means that A uniquely captures the local dependencies relevant to X, and B uniquely captures the local dependencies relevant to Y.

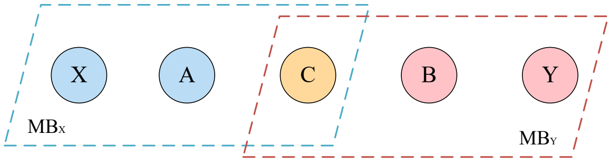


Figure 2: Illustration of common and target-specific variables: C is identified as the common variable, since it is present in the Markov blankets of both X and Y. A serves as the target-specific variable for X, and B serves as the target-specific variable for Y.

Target-specific MB variables capture the unique local relationships specific to individual targets, enabling more accurate predictions or inferences tailored to each target. In contrast, common MB variables represent shared dependencies across multiple targets, efficiently summarizing the information relevant to all targets using a minimal set of variables.

Theorem 1. *Equivalent information can be resolved by comparing the contributions of a feature F and a label L_2 to the target label L_1.*

Proof. Mutual information [Kraskov *et al.*, 2004] is a widely used metric for evaluating the contribution of variables. In the context of feature F and labels L_1 and L_2, we compare the mutual information between feature F and target label L_1 ($I(L_1; F)$) with the mutual information between label L_2 and target label L_1 ($I(L_1; L_2)$).

If $I(L_1; F) \leq I(L_1; L_2)$, this implies that feature F provides information similar to that of label L_2 or is redundant. In such cases, we choose not to retain F, thereby reducing redundancy from equivalent information. Conversely, if $I(L_1; F) > I(L_1; L_2)$, the feature F offers additional information beyond what label L_2 provides, and we therefore retain F. □

Definition 5 (V-structure [Pearl, 2014]): three variables X, Y and T form a V-structure (i.e., $X \rightarrow T \leftarrow Y$), *iff* X and Y have a directed edge pointing to T, regardless of whether X and Y are directly connected.

The V-structure is a critical concept for identifying the spouse set of a variable. For example, in a V-structure $T \rightarrow X \leftarrow Y$, Y is a spouse of T, and X is a collider. However, in the context of multi-label causal feature selection, directly applying Definition 5 for spouse search can lead to several issues: first, as the number of features and labels increases, the time overhead grows exponentially due to the need of numerous conditional independence tests; second, in the multi-label domain, the identification of spouses is more complex, as illustrated in Fig. 3, which presents four cases of spouse relationships in a multi-label DAG. It is not clear which spouse should be retained. To address these challenges, we introduce Theorem 2 for efficient and precise spouse search.

In Fig. 3, we illustrate the four possible relationships between the target label and the spouse. Based on this figure, we conclude that Fig. 3.(a) is the case we want to retain, as the spouse of L_1 is the feature F_3. In the case of Fig. 3.(b), we do not need to retain it since the spouse of L_1 here is label L_2.

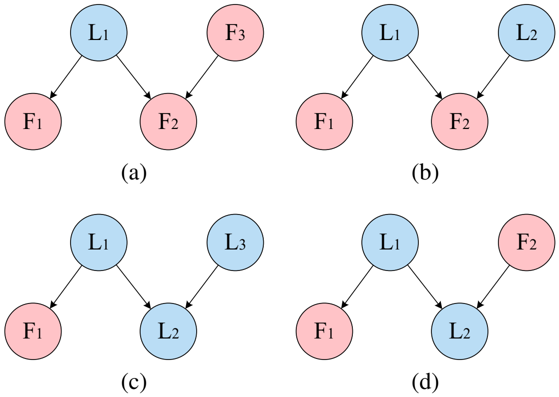


Figure 3: Illustration of four spouse relationships in multi-label DAG. (a) $L_1 \rightarrow F_2 \leftarrow F_3$. (b) $L_1 \rightarrow F_2 \leftarrow L_2$. (c) $L_1 \rightarrow L_2 \leftarrow L_3$. (d) $L_1 \rightarrow L_2 \leftarrow F_2$.

If we designate label L_2 as the spouse of L_1, we will conduct feature selection using only the feature set from L_1's Markov blanket, making it unnecessary to retain label L_2. Similarly, in Fig. 3.(c), we also do not need to retain it, since the spouse of L_1 in this case is label L_3. Finally, for Fig. 3.(d), we conclude that it should not be retained, as proven in Theorem 2.

Theorem 2. *The label-specific variable is not contained in other label's MB.*

Proof. According to Definition 4, in Fig. 3.(d), feature F_2 is the label-specific variable of label L_2. Therefore, the following conditions hold:

- $F_2 \not\perp\!\!\!\perp L_2 \mid \{\mathcal{L} \backslash L_2\}$, which indicates that F_2 is dependent on L_2 conditioned on the rest of the label set $\mathcal{L} \backslash L_2$.
- $\forall L \in \mathcal{L}, L \neq L_2, F_2 \perp\!\!\!\perp L \mid \{\mathcal{L} \backslash L_2\}$, meaning F_2 is independent of any other label L (where $L \neq L_2$) conditioned on $\mathcal{L} \backslash L_2$.

Assume that F_2 is part of the Markov blanket of label L_1. This implies that there exists a path between F_2 and L_1, such as $L_1 \rightarrow L_2 \leftarrow F_2$. According to this path, $F_2 \not\perp\!\!\!\perp L_1 \mid \{\mathcal{L} \backslash L_1\}$, indicating that F_2 is dependent on L_1 given the rest of the labels. However, this contradicts the earlier condition that F_2 is independent of all other labels except L_2.

Thus, we reach a contradiction. Therefore, F_2 cannot be part of the Markov blanket of L_1, and more generally, a label-specific variable cannot be contained in the Markov blanket of any other label. □

Theorem 3. *For three variables X, Y and T, given condition set Z if $T \not\perp\!\!\!\perp X \mid Z, T \perp\!\!\!\perp Y \mid Z, T \not\perp\!\!\!\perp Y \mid Z \cup X$, then T, X and Y form a V-structure ($T \rightarrow X \leftarrow Y$), variable Y is T's spouse* [Spirtes *et al.*, 2001; Pearl, 2014].

Theorem 4. *Only children with multiple parents can make target variable T and variables in $non-PC_T$ from independence to dependence* [Ling *et al.*, 2022a].

According to Theorems 2, 3, and 4, the spouse search can be conducted within a multi-label scenario. First, we identify children with multiple parents, which can serve as colliders. Then, we determine the optimal spouse set, which is illustrated in Fig. 3.(a), indicating that spouses can only take the form of feature-feature pairs. Finally, using Theorem 3, we can estimate the existence of a spouse variable based on the identification of the V-structure.

4 MCF-Spouse Algorithm

In this section, we present the Multi-label Causal Feature Selection with Optimal Spouse Discovery (MCF-Spouse, Algorithm 1). MCF-Spouse consists of three phases: Phase 1 (lines 1-3) mines the PC sets using the HITON-PC method [Aliferis *et al.*, 2003], Phase 2 (lines 4-17) recovers features ignored due to strong label correlation, and Phase 3 (lines 18-28) conducts the spouse search.

4.1 Learning the Local Causal Structure of Each Class Label

In phase 1, we employ the HITON-PC [Aliferis *et al.*, 2003] as our primary method for PC discovery. HITON-PC is known for its efficiency in identifying relevant features by minimizing false positives and false negatives, making it a robust choice for PC discovery in high-dimensional data. Phase 2 refers to the feature recovery. After the PC discovery in Phase 1, some features are not selected due to strong correlations between labels and the issue of equivalent information. These ignored features may still hold significant value but are overshadowed by the influence of other labels in the dataset. To address this, we re-evaluate these features to determine whether they should be included in the candidate PC set.

In lines 5-7, the feature F_j is examined. This feature does not belong to the current candidate PC set of label L_i, but it could have been blocked by strong correlations with other labels, making it a candidate for recovery. Then, in lines 8-10, if F_j is dependent on Y_i, F_j will be included in $Fe_{re}(Y_i)$ in descending order, and the top $k_1\%$ features in $Fe_{re}(Y_i)$ are considered as the most possible ignored features. Lines 11-14 address the key condition: when label L_k is not in the condition set, feature F_j remains dependent on label L_i; once label L_k is included in the condition set, feature F_j and label L_i become independent. We conclude that label L_k blocks the path from F_j to L_i, and this path-blocking mechanism shows how labels can obscure feature relationships, potentially leading to the omission of valuable features.

In lines 15-17, we compare the MI between feature F_j and the target label L_i ($I(L_i; F_j)$) with that between label L_k and the target label L_i ($I(L_i; L_k)$). This comparison allows us to quantify the contribution of feature F_j relative to the label L_k. If the feature F_j contributes more information to L_i than label L_k, we conclude that F_j is more significant, and therefore, it is included in the candidate PC set (CPC_{L_i}). Conversely, if label L_k provides more information than feature F_j, L_k is included in the candidate PC set instead.

4.2 Spouse Discovery

Phase 3 is spouse discovery. As discussed in Fig. 3, the optimal spouse variables can only exist in the configuration shown in Fig. 3.(a), which forms a collider structure: target → feature ← feature. In lines 19-21, we begin by traversing

Algorithm 1 MCF-Spouse Algorithm

```
Require: feature set $\mathcal{F}$=$\{F_1, F_2, \ldots, F_m\}$, label set $\mathcal{L}$=$\{L_1, L_2, \ldots, L_l\}$, PC discovery algorithm $\mathcal{A}$, $V$=$\mathcal{F} \cup \mathcal{L}$.
Ensure: $MB_{L_i}$
 1: {Phase 1: PC discovery using $\mathcal{A}$}
 2: for $i = 1, ..., l$ do
 3:     $CPC_{L_i}$ : discover the $PC$ of $L_i$ from $V \backslash L_i$
 4: {Phase 2: recover the ignored features caused by strong label correlation}
 5: for $i = 1, ..., l$ do
 6:     if $\exists k \neq i, Y_k \in CMB_{Y_i}$ then
 7:         for $\forall F_j \in F \backslash CMB_{Y_i}$ do
 8:             if $F_j \not\perp\!\!\!\perp Y_i$ then
 9:                 $Fe_{re}(Y_i) = Fe_{re}(Y_i) \cup \{F_j\}$ in descending order
10:                 Keep top $k_1\%$ features in $Fe_{re}(Y_i)$
11: for $i = 1, ..., l$ do
12:     if $\exists k \neq i, L_k \in CPC_{L_i}$ then
13:         for $\forall F_j \in F \backslash CPC_{L_i}$ do
14:             if $F_j \not\perp\!\!\!\perp L_i | \{V \backslash CPC_{L_i}\}$ $and$ $F_j \perp\!\!\!\perp L_i | \{V \backslash CPC_{L_i}\} \cup \{L_k\}$ then
15:                 if $I(L_i; F_j) > I(L_i; L_k)$ then
16:                     $CPC_{L_i} = CPC_{L_i} \cup \{F_j\}$
17:                     $CPC_{L_i} = CPC_{L_i} \backslash \{L_k\}$
18: {Phase 3: spouse discovery}
19: for $i = 1, ..., l$ do
20:     $CSP_{L_i} = \emptyset$
21:     for $\forall F_j \in CPC_{L_i}$ do
22:         for $p = 1, ..., m$ do
23:             if $\exists p$ $s.t.$ $F_j \in CPC_p$ then
24:                 for $\forall F_k \in \{F \backslash CPC_{L_i}\}$ do
25:                     if $F_k \not\perp\!\!\!\perp L_i | \{V \backslash CPC_{L_i} \cup F_j\}$ then
26:                         $CSP_{L_i} = CSP_{L_i} \cup \{F_k\}$
27:     $MB_{Y_i} = CPC_{Y_i} \cup CSP_{Y_i}$
28: return $MB_{Y_i}$
```

each feature F_j in the candidate PC set (CPC_{L_i}). In lines 22-23, we estimate whether F_j belongs to the PC set of other features, since only features with multiple parent variables can serve as colliders. This step is crucial because identifying colliders is essential for detecting indirect dependencies between features and labels.

Then, in line 24, we define the candidate spouse set as the set of features that are not already included in CPC_{L_i}. These features have the potential to be spouses, as they are not directly connected to the target label L_i but could form indirect relationships via colliders. In line 25, we determine the spouse relationship by examining the dependency between feature F_k and the target label L_i when feature F_j is conditioned upon. Specifically, when F_j is part of the condition set, F_k and L_i become dependent, indicating that F_k, F_j, and L_i form a V-structure represented as $L_i \rightarrow F_j \leftarrow F_k$. In this V-structure, feature F_j serves as the collider, while feature F_k serves as the spouse of label L_i.

Theorem 5. *In a multi-label causal feature selection algorithm, employing a phased method—performing only PC set discovery in the first phase and deferring spouse discovery to the third phase, is more efficient and accurate than performing a complete MB search in the first phase.*

Datasets	Domains	Instances	Features	Labels
Flags	Image	194	19	7
VirusGO	Biology	207	749	6
CHD_49	Medicine	555	49	6
PlantGO	Biology	978	3091	12
Enron	Text	1702	1001	53
Image	Image	2000	294	5
Yeast	Biology	2417	103	14
HumanGO	Biology	3106	9844	14

Table 1: Description of Datasets

Theorem 5 indicates that after feature recovery, spouse discovery reduces computational complexity while improving the accuracy of multi-label causal feature selection.

4.3 Complexity Analysis

The MCF-Spouse method consists of three main phases: (1) PC discovery, (2) feature recovery, and (3) spouse discovery.

Phase 1: PC Discovery: in this phase, the algorithm identifies the PC set for each label L_i using a PC discovery method (e.g., HITON-PC, MMPC). The time complexity of this step is $O(m \cdot logm)$ for each label, where m is the number of features. Therefore, the total time complexity for this phase is $O(l \cdot m \cdot logm)$, where l is the number of labels.

Phase 2: Feature Recovery: in this phase, the algorithm recovers features ignored due to strong label correlation. For each label L_i, the algorithm compares the mutual information of features outside the PC set with other labels. The time complexity for this step is $O(m^2 l)$, since the MI for each feature is computed with every label.

Phase 3: Spouse Discovery: the algorithm identifies spouse features by examining V-structures in the PC set of each label. The complexity of checking dependencies and performing CI tests for each feature pair is $O(m^2)$ per label, resulting in a total time complexity of $O(m^2 l)$.

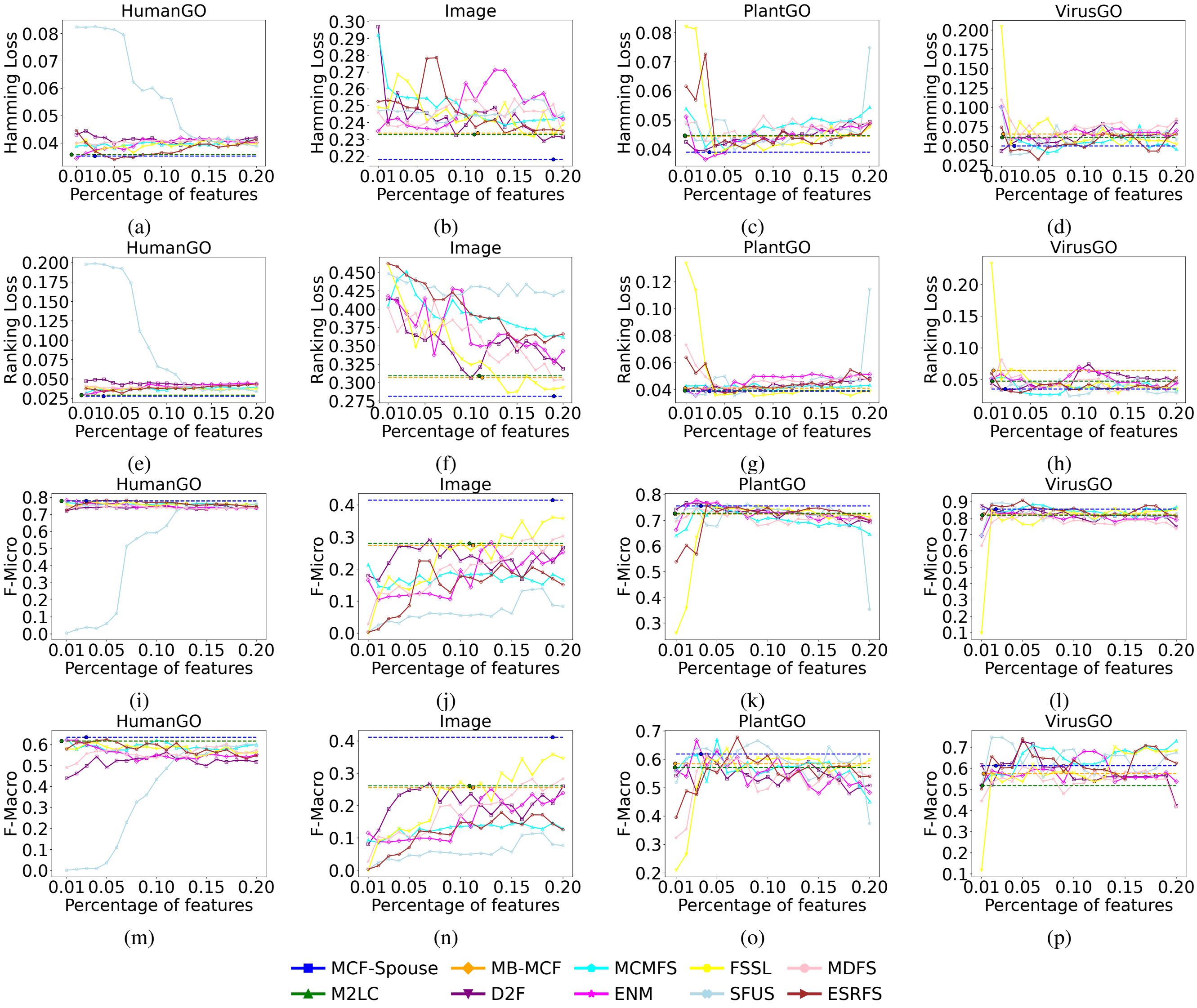
The overall time complexity of the MCF-Spouse algorithm is dominated by the most computationally expensive phase, which is consistent across all three phases. Consequently, the total time complexity is $O(m^2 l)$.

5 Experiment

In this experiment, we compare MCF-Spouse with nine methods, which are introduced in related work. Classification accuracy is computed using the ML-kNN [Zhang and Zhou, 2007], with the number of nearest neighbors k fixed at 10. Each experiment is repeated 10 times.

5.1 Experiment Settings

1) Datasets: we utilize eight real-world datasets sourced from various application domains, including ***Flags*** [Gonçalves *et al.*, 2013], ***VirusGO*** [Xu *et al.*, 2016], ***CHD_49*** [Shao *et al.*, 2013], ***PlantGO*** [Xu *et al.*, 2016], ***Enron*** [Read *et al.*, 2008], ***Image*** [Zhang and Zhou, 2007], ***Yeast*** [Elisseeff and Weston, 2001], and ***HumanGO*** [Xu *et al.*, 2016] detailed in Table 1.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

Figure 4: The Hamming Loss (↓), Ranking Loss (↓), F-Micro (↑), and F-Macro (↑) of MCF-Spouse and nine state-of-the-art methods on the HumanGO, Image, PlantGO, and VirusGO dataset.

These datasets are accessible for download from the Multi-Label Classification Dataset Repository[1].

2) Evaluation Metrics: to assess the performance of selected feature subsets, we utilize the following four metrics in multi-label feature selection: ***Hamming Loss***, ***Ranking Loss***, ***F-micro*** and ***F-macro*** [Wu *et al.*, 2020].

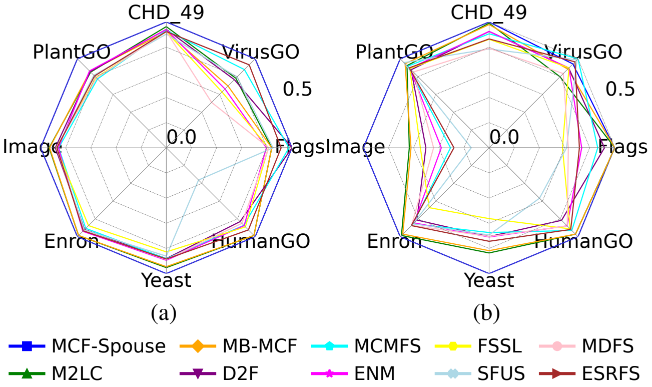
3) Parameter Settings: for comparing methods, their parameters are set according to the suggestions in the corresponding literature. For D2F, MCMFS, ENM, and FSSL, these mutual information-based methods can not determine the optimal number of features. Therefore, we gradually increase the percentage of selected features from 1% to 20% with a step size of 1%. For SFUS, the value of α and β are searched within the range of $\{10^{-3}, 10^{-2},..., 10^{2}, 10^{3}\}$. Similarly, for MDFS, the values of α, β, and γ are searched within the same range. In the case of ESRFS, the values of α, β, γ, and λ are also evaluated within $\{10^{-3}, 10^{-2},..., 10^{2}, 10^{3}\}$. Since SFUS, MDFS and ESRFS are unable to determine the optimal number of features, we incrementally increase the percentage of selected features in the same manner as mutual information-based methods.

For MB-MCF, M2LC, and MCF-Spouse, these BN-based methods do not require the number of selected features to be predetermined. However, MB-MCF and M2LC have parameters: k_1 which determines the number of ignored features to be recovered, and k_2, which selects the features that need to undergo symmetry checks. We perform the grid search within the range of [0.1, 1] to find the values of k_1 and k_2 that yield the best results. For parameter k_1 in MCF-Spouse, we also perform the grid search within the range of [0.1, 1] to find the

[1] http://www.uco.es/kdis/mllresources/#3sourcesDesc

Datasets	*Time(s)*									
	MCF-Spouse	M2LC	MB-MCF	D2F	MCMFS	ENM	FSSL	SFUS	MDFS	ESRFS
Flags	7.828	7.734	7.063	7.075	8.249	8.310	7.947	35.080	82.859	81.761
Image	137.016	278.250	345.828	782.446	658.217	651.289	649.886	656.092	680.634	595.777

Table 2: Run time of ten algorithms on Flags and Image datasets



(a)

(b)

Figure 5: Spider web diagrams demonstrate the stability of the ten methods across eight datasets with Hamming Loss and F-macro.

optimal performance.

5.2 Results of MCF-Spouse Comparing With Nine Multi-Label Feature Selection Methods

As previously discussed, Non-BN-based methods cannot determine the optimal number of features. Therefore, we incrementally increase the percentage of selected features from 1% to 20% in 1% steps to search for the optimal number. In contrast, the BN-based methods discussed in this article (MCF-Spouse, M2LC, and MB-MCF) are heuristic methods that do not require matrix training and can automatically select the optimal number of features (size of MB). These characteristics eliminate variance, further highlighting the superiority of BN-based methods. We evaluate the performance of MCF-Spouse and nine state-of-the-art methods on the *HumanGO*, *Image*, *PlantGO*, and *VirusGO* datasets. The results are shown in Fig. 4.

From Fig. 4, we observe that MCF-Spouse, which incorporates the search for spouse variables, selects more features compared to MB-MCF and M2LC-two BN-based methods that do not consider spouses. This results in significant performance improvements across the four datasets in all metrics. These findings highlight the crucial role of spouse variables in enhancing predictive accuracy and overall model performance. Compared to non-BN-based methods that determine the optimal number of features by exploring in 1% increments, MCF-Spouse consistently delivers superior performance across most datasets, with the exception of the *F-macro* score on *VirusGO*. Non-BN-based methods often exhibit instability and require a time-intensive stepwise search to identify the optimal feature count. In contrast, MCF-Spouse not only achieves more stable results but also effectively selects a smaller subset of features with higher precision, leading to better overall performance.

To demonstrate the stability [Wu *et al.*, 2022] of MCF-Spouse across various metrics and datasets, we normalize the results to the range [0, 0.5], assigning the best-performing method a value of 0.5. For metric *Hamming Loss*, lower values indicate better performance, so the minimum is normalized to 0.5. In contrast, for metric *F-Macro*, higher values are preferable, and the maximum is set to 0.5. As illustrated in Fig. 5, MCF-Spouse consistently demonstrates superior performance, forming a regular octagon.

5.3 Run Time Analysis

In this section, we evaluate the runtime of ten methods on the *Flags* and *Image* datasets. The timing starts at the beginning of the method and ends after classification using ML-kNN. For Non-BN-based methods, runtimes are averaged across feature selections ranging from 1% to 20% of the total features to ensure a fair comparison.

As summarized in Table 2, the proposed MCF-Spouse method demonstrates competitive efficiency. On the *Flags* dataset, its runtime is comparable to methods like MCMFS and ENM. For the larger *Image* dataset, MCF-Spouse significantly reduces runtime compared to MB-MCF, MCMFS, and others. These results highlight MCF-Spouse's ability to maintain a favorable balance between computational efficiency and feature selection quality, making it highly applicable for complex, high-dimensional datasets.

6 Conclusion

In this paper, we propose Multi-label Causal Feature Selection Method with Optimal Spouses Discovery (MCF-Spouse) to address challenges in multi-label causal feature selection. MCF-Spouse utilizes MI to evaluate contributions of features and labels and retain only the most informative variables, resolving issues of equivalent information and strong label correlations. Additionally, MCF-Spouse introduces a novel spouse discovery mechanism, optimizing spouse discovery through reducing search space and alleviating time overhead associated with CI tests. Our results highlight the importance of distinguishing label-label interactions and label-feature contributions in multi-label causal inference. Future work will focus on addressing spurious variables in the MB, specifically by exploring the use of “AND” and “OR” rules to refine MB construction, further improving the accuracy and efficiency of multi-label causal feature selection.

Acknowledgments

This work is supported by the National Key Research and Development Program, Unified Integrated Development technology of scientific computing language and engineering physics modeling language[Grant No.2024YFB3310202], and by the talent project of Department of Science and Technology of Jilin Province of China[Grant No.20240602106RC].

References

[Aliferis *et al.*, 2003] Constantin F Aliferis, Ioannis Tsamardinos, and Alexander Statnikov. Hiton: a novel markov blanket algorithm for optimal variable selection. In *AMIA annual symposium proceedings*, volume 2003, page 21. American Medical Informatics Association, 2003.

[Aliferis *et al.*, 2010a] Constantin F Aliferis, Alexander Statnikov, Ioannis Tsamardinos, Subramani Mani, and Xenofon D Koutsoukos. Local causal and markov blanket induction for causal discovery and feature selection for classification part i: algorithms and empirical evaluation. *Journal of Machine Learning Research*, 11(1), 2010.

[Aliferis *et al.*, 2010b] Constantin F Aliferis, Alexander Statnikov, Ioannis Tsamardinos, Subramani Mani, and Xenofon D Koutsoukos. Local causal and markov blanket induction for causal discovery and feature selection for classification part ii: analysis and extensions. *Journal of Machine Learning Research*, 11(1), 2010.

[Ben-Gal, 2008] Irad Ben-Gal. Bayesian networks. *Encyclopedia of statistics in quality and reliability*, 2008.

[Cai and Zhu, 2018] Zhiling Cai and William Zhu. Multi-label feature selection via feature manifold learning and sparsity regularization. *International journal of machine learning and cybernetics*, 9:1321–1334, 2018.

[Chu *et al.*, 2023] Zhixuan Chu, Ruopeng Li, Stephen Rathbun, and Sheng Li. Continual causal inference with incremental observational data. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 3430–3439. IEEE, 2023.

[Elisseeff and Weston, 2001] André Elisseeff and Jason Weston. A kernel method for multi-labelled classification. *Advances in neural information processing systems*, 14, 2001.

[Faraji *et al.*, 2024] Mohammad Faraji, Seyed Amjad Seyedi, Fardin Akhlaghian Tab, and Reza Mahmoodi. Multi-label feature selection with global and local label correlation. *Expert Systems with Applications*, 246:123198, 2024.

[Gao *et al.*, 2017] Tian Gao, Kshitij Fadnis, and Murray Campbell. Local-to-global bayesian network structure learning. In *International Conference on Machine Learning*, pages 1193–1202. PMLR, 2017.

[Gonçalves *et al.*, 2013] Eduardo Corrêa Gonçalves, Alexandre Plastino, and Alex A Freitas. A genetic algorithm for optimizing the label ordering in multi-label classifier chains. In *2013 IEEE 25th international conference on tools with artificial intelligence*, pages 469–476. IEEE, 2013.

[Gonzalez-Lopez *et al.*, 2020] Jorge Gonzalez-Lopez, Sebastián Ventura, and Alberto Cano. Distributed multi-label feature selection using individual mutual information measures. *Knowledge-Based Systems*, 188:105052, 2020.

[Guo *et al.*, 2023] Xianjie Guo, Kui Yu, Lin Liu, Peipei Li, and Jiuyong Li. Adaptive skeleton construction for accurate dag learning. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[Guo *et al.*, 2024] Xianjie Guo, Kui Yu, Lin Liu, Jiuyong Li, Jiye Liang, Fuyuan Cao, and Xindong Wu. Progressive skeleton learning for effective local-to-global causal structure learning. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Huang *et al.*, 2018] Rui Huang, Weidong Jiang, and Guangling Sun. Manifold-based constraint laplacian score for multi-label feature selection. *Pattern Recognition Letters*, 112:346–352, 2018.

[Jian *et al.*, 2016] Ling Jian, Jundong Li, Kai Shu, and Huan Liu. Multi-label informed feature selection. In *IJCAI*, volume 16, pages 1627–33, 2016.

[Kraskov *et al.*, 2004] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 69(6):066138, 2004.

[Lee and Kim, 2015] Jaesung Lee and Dae-Won Kim. Mutual information-based multi-label feature selection using interaction information. *Expert Systems with Applications*, 42(4):2013–2025, 2015.

[Li *et al.*, 2024] Yonghao Li, Liang Hu, and Wanfu Gao. Multi-label feature selection with high-sparse personalized and low-redundancy shared common features. *Information Processing & Management*, 61(3):103633, 2024.

[Ling *et al.*, 2022a] Zhaolong Ling, Bo Li, Yiwen Zhang, Qingren Wang, Kui Yu, and Xindong Wu. Causal feature selection with efficient spouses discovery. *IEEE Transactions on Big Data*, 9(2):555–568, 2022.

[Ling *et al.*, 2022b] Zhaolong Ling, Ying Li, Yiwen Zhang, Kui Yu, Peng Zhou, Bo Li, and Xindong Wu. A light causal feature selection approach to high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[Ling *et al.*, 2024a] Zhaolong Ling, Bo Li, Yiwen Zhang, Peng Zhou, Xingyu Wu, Yuee Huang, Kui Yu, and Xindong Wu. Causal discovery using weight-based conditional independence test. *ACM Transactions on Knowledge Discovery from Data*, 19(1):1–24, 2024.

[Ling *et al.*, 2024b] Zhaolong Ling, Jingxuan Wu, Yiwen Zhang, Peng Zhou, Xingyu Wu, Kui Yu, and Xindong Wu. Label-aware causal feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Liu *et al.*, 2020] Jinghua Liu, Yuwen Li, Wei Weng, Jia Zhang, Baihua Chen, and Shunxiang Wu. Feature selection for multi-label learning with streaming label. *Neurocomputing*, 387:268–278, 2020.

[Ma *et al.*, 2012] Zhigang Ma, Feiping Nie, Yi Yang, Jasper RR Uijlings, and Nicu Sebe. Web image annotation via subspace-sparsity collaborated feature selection. *IEEE Transactions on Multimedia*, 14(4):1021–1030, 2012.

[Ma *et al.*, 2025] Lin Ma, Liang Hu, Yonghao Li, Weiping Ding, and Wanfu Gao. Mi-mcf: A mutual information-based multilabel causal feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):9864–9878, 2025.

[Masegosa and Moral, 2012] Andrés R Masegosa and Serafín Moral. A bayesian stochastic search method for discovering markov boundaries. *Knowledge-Based Systems*, 35:211–223, 2012.

[Pearl, 2014] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[Pellet and Elisseeff, 2008] Jean-Philippe Pellet and André Elisseeff. Using markov blankets for causal structure learning. *Journal of Machine Learning Research*, 9(7), 2008.

[Pena *et al.*, 2007] Jose M Pena, Roland Nilsson, Johan Björkegren, and Jesper Tegnér. Towards scalable and data efficient learning of markov boundaries. *International Journal of Approximate Reasoning*, 45(2):211–232, 2007.

[Pereira *et al.*, 2018] Rafael B Pereira, Alexandre Plastino, Bianca Zadrozny, and Luiz HC Merschmann. Categorizing feature selection methods for multi-label classification. *Artificial intelligence review*, 49:57–78, 2018.

[Read *et al.*, 2008] Jesse Read, Bernhard Pfahringer, and Geoff Holmes. Multi-label classification using ensembles of pruned sets. In *2008 eighth IEEE international conference on data mining*, pages 995–1000. IEEE, 2008.

[Shao *et al.*, 2013] Huan Shao, GuoZheng Li, GuoPing Liu, and YiQin Wang. Symptom selection for multi-label data of inquiry diagnosis in traditional chinese medicine. *Science China Information Sciences*, 56:1–13, 2013.

[Spirtes *et al.*, 2001] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2001.

[Statnikov *et al.*, 2013] Alexander Statnikov, Nikita I Lytkin, Jan Lemeire, and Constantin F Aliferis. Algorithms for discovery of multiple markov boundaries. *Journal of Machine Learning Research*, 14(Feb):499–566, 2013.

[Tsamardinos *et al.*, 2003a] Ioannis Tsamardinos, Constantin F Aliferis, and Alexander Statnikov. Time and sample efficient discovery of markov blankets and direct causal relations. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 673–678, 2003.

[Tsamardinos *et al.*, 2003b] Ioannis Tsamardinos, Constantin F Aliferis, Alexander R Statnikov, and Er Statnikov. Algorithms for large scale markov blanket discovery. In *FLAIRS*, volume 2, pages 376–81, 2003.

[Tsamardinos *et al.*, 2006] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65:31–78, 2006.

[Wu *et al.*, 2020] Xingyu Wu, Bingbing Jiang, Kui Yu, Huanhuan Chen, and Chunyan Miao. Multi-label causal feature selection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 6430–6437, 2020.

[Wu *et al.*, 2022] Xingyu Wu, Bingbing Jiang, Yan Zhong, and Huanhuan Chen. Multi-target markov boundary discovery: Theory, algorithm, and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4964–4980, 2022.

[Wu *et al.*, 2023a] Xingyu Wu, Bingbing Jiang, Xiangyu Wang, Taiyu Ban, and Huanhuan Chen. Feature selection in the data stream based on incremental markov boundary learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10):6740–6754, 2023.

[Wu *et al.*, 2023b] Xingyu Wu, Bingbing Jiang, Tianhao Wu, and Huanhuan Chen. Practical markov boundary learning without strong assumptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10388–10398, 2023.

[Xu *et al.*, 2016] Jianhua Xu, Jiali Liu, Jing Yin, and Chengyu Sun. A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously. *Knowledge-Based Systems*, 98:172–184, 2016.

[Yu *et al.*, 2021] Kui Yu, Mingzhu Cai, Xingyu Wu, Lin Liu, and Jiuyong Li. Multilabel feature selection: a local causal structure learning approach. *IEEE Transactions on Neural Networks and Learning Systems*, 34(6):3044–3057, 2021.

[Zhang and Zhou, 2007] Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048, 2007.

[Zhang *et al.*, 2019a] Jia Zhang, Zhiming Luo, Candong Li, Changen Zhou, and Shaozi Li. Manifold regularized discriminative feature selection for multi-label learning. *Pattern Recognition*, 95:136–150, 2019.

[Zhang *et al.*, 2019b] Ping Zhang, Guixia Liu, and Wanfu Gao. Distinguishing two types of labels for multi-label feature selection. *Pattern recognition*, 95:72–82, 2019.

[Zhang *et al.*, 2020] Ping Zhang, Wanfu Gao, Juncheng Hu, and Yonghao Li. Multi-label feature selection based on high-order label correlation assumption. *Entropy*, 22(7):797, 2020.