

REVIEWS

A Review of Label Disambiguation and Feature Selection in Partial Multilabel Learning

Yusong Wang¹

¹ School of Computer and Information Engineering, Jiangxi Agricultural University, Nanchang 330045, Jiangxi, China

Corresponding author: Yusong Wang (654050507@qq.com)

Received 20 October 2026; accepted 20 October 2026; published 20 October 2026.

Abstract. Partial multilabel learning addresses ambiguous supervision in which each instance is assigned a candidate label set containing its relevant labels and additional false positives. Feature selection in this setting faces a circular dependency: reliable labels are needed to identify informative features, while informative features are needed to distinguish relevant labels from misleading candidates. This review examines that dependency through a focused synthesis of recent partial multilabel feature selection methods and related work on disambiguation, partial label learning, and label enhancement. We organize the literature along two complementary dimensions: the evidence used to resolve label ambiguity and the mechanism through which the resulting supervision controls feature selection. Eight representative feature selection methods illustrate latent alignment, mutual information reconstruction, dual-space fusion, multiscale label distributions, sparse regression, fuzzy relevance, and neighborhood rough sets. The comparison shows that similar terminology can conceal different supervision constraints, feature scoring rules, and optimization dependencies. We therefore distinguish label confidence from label description distributions and original feature selection from representation learning. We also examine differences in noise generation, feature budgets, downstream classifiers, and validation protocols that limit comparisons across published results. The review identifies uncertainty calibration, minority-label preservation, reliable structural assumptions, and reproducible evaluation as priorities for developing feature selection methods under ambiguous multilabel supervision.

Keywords: Partial multilabel learning; feature selection; label disambiguation; label enhancement; sparse learning; granular computing

1 Introduction

Many learning problems require several labels to describe one object. An image may contain multiple objects, a document may discuss several topics, and a biological instance may have several functional annotations. Multilabel learning formalizes this situation by associating each instance with a set of relevant labels [1]. Its usual supervised formulation assumes that the observed labels accurately distinguish relevant from irrelevant categories. In practice, producing such annotations can require substantial effort, particularly when labels overlap semantically or annotators have limited expertise.

Partial multilabel learning (PML) relaxes annotation precision. Each training instance has a candidate set that contains its ground-truth labels but may also contain irrelevant labels [2]. The learner must recover a multilabel predictor from this ambiguous supervision. The number of relevant labels is unknown and may exceed one. This distinguishes PML from standard partial label learning (PLL), where exactly one candidate is correct, and from learning with missing positive labels, where relevant labels may be absent from

the observed set. These differences determine which entries can be treated as reliable negatives and which normalization constraints are appropriate.

High feature dimensionality introduces a second source of difficulty. Irrelevant features can distort distances and weaken neighborhood-based disambiguation. Redundant features can cause the same evidence to be counted repeatedly, while features associated with frequent labels can overshadow signals useful for rare labels. Conversely, treating every candidate label as correct can promote features that explain annotation noise. Feature selection and label disambiguation therefore influence each other. A useful selector must identify a subset of original variables while accounting for the uncertain meaning of its supervision.

Existing reviews already provide broad accounts of multilabel learning [1] and multilabel feature selection. In particular, Qian et al. survey feature selection through label fusion, including label correlation, label-specific features, noisy and missing labels, and label enhancement [3]. The present review takes a narrower analytical focus: how disambiguation evidence is converted into the selection or ranking of original features under the candidate-superset assumption.

It does not aim to replace these broader surveys or establish an exhaustive history of PML feature selection.

Two organizing questions guide the discussion. First, what evidence makes a candidate label credible: neighboring instances, label relationships, negative labels, a latent representation, or agreement across granularities? Second, how does that evidence enter feature selection: as a fixed target, an iteratively updated variable, a relevance criterion, or a dependency measure? These questions help compare methods whose names emphasize different tools even when their underlying assumptions overlap.

The review makes three analytical contributions. It defines a common notation and separates the supervision settings most likely to be confused. It develops a taxonomy that distinguishes the source of disambiguation evidence from the selector that consumes it. It then uses this taxonomy to compare eight direct feature selection methods and to examine the evaluation choices that constrain conclusions about their relative merits. The resulting perspective emphasizes the credibility of the supervisory signal and the traceability of original feature scores.

Section 2 describes the scope and formal setting. Section 3 analyzes the principal sources of disambiguation evidence. Section 4 compares representative feature selection mechanisms. Section 5 draws connections across method families. Section 6 discusses experimental evidence and evaluation design. Section 7 develops research directions, followed by conclusions in Section 8.

2 Scope and Problem Formulation

2.1 Review Scope

This is a focused narrative review. Its primary corpus comprises 27 full-text studies: eight direct PML feature selection studies, eight PML formulation or prediction studies, six PLL studies, three studies of label enhancement or inaccurate label distributions, one rough-set reduction study, and one study of feature selection with partially labeled hybrid data. The core comparison concentrates on recent methods for which the feature scoring mechanism and supervision assumptions can be examined directly. Broader reviews supply context [1,3]. The corpus is concentrated in recent work and does not constitute a reproducible, exhaustive database search or a meta-analysis.

The inclusion boundary is functional. A direct method must address ambiguous candidate multilabel supervision and return a ranking or subset of original features. Studies that learn a discriminative subspace, attention representation, or classifier without explicitly producing such a subset serve as methodological background. Their predictive performance is not interpreted as feature selection performance. Likewise, PLL algorithms inform the discussion of mechanisms, but their single-correct-label experiments are not pooled with PML results.

2.2 Candidate Labels and Hidden Relevance

Let X be an n by d feature matrix and let the label vocabulary contain q labels. For instance i , C_i denotes the observed candidate label set and T_i denotes its hidden relevant label set. The standard PML assumption is

$$\emptyset \neq T_i \subseteq C_i, \quad Y^* \leq Y \quad (1)$$

Here Y is the binary candidate matrix and Y^* is the unobserved binary relevance matrix; the matrix inequality is elementwise. A zero in Y means that the corresponding label is not a candidate. Under the superset assumption, it is a reliable negative. A one indicates possible relevance, rather than confirmed relevance. The goal is to learn a predictor and identify an original feature subset A with a specified size or an explicit stopping rule.

Fig. 1 illustrates the boundary between PML and neighboring tasks. In PLL, T_i contains one label. In PML, it may contain several. With missing-label supervision, an observed zero may instead hide an unrecorded positive. A method that uses non-candidates as definite negatives cannot automatically be transferred to this latter setting. If both omissions and false-positive candidates occur, the supervision model must explicitly account for both error types.

2.3 Confidence Is Different From a Label Distribution

A confidence matrix C can assign a score to each candidate, often with zero scores outside the candidate set. A support-preserving confidence model may impose

$$0 \leq c_{ij} \leq y_{ij}, \quad \forall i, j \quad (2)$$

The entries need not sum to one. Several labels can simultaneously receive high confidence because relevance is multilabel. In contrast, a label description distribution D is nonnegative and has unit row sum. It expresses how descriptive degrees are allocated across labels. PML-LD uses label enhancement to recover such a distribution before fitting a predictor [6]. Its normalized outputs should not be interpreted automatically as independent probabilities of relevance.

For example, an instance with two equally relevant labels can have relevance confidences near one for both, while a normalized description distribution allocates about one half to each. Applying a fixed probability threshold of one half has different meanings for these representations. A review must therefore record support constraints, normalization, and the final multilabel decision rule rather than using confidence, probability, membership, and distribution interchangeably.

2.4 Feature Selection and a Common Analytical Objective

Table 1 summarizes the notation. An embedded selector often learns a d by q coefficient matrix W and scores original feature j using the norm of its corresponding row. Row sparsity is commonly encouraged through an $\ell_{2,1}$ penalty, which sums the Euclidean norms of the rows. Other methods form an effective coefficient matrix from several factors, or select features through changes in an explicit relevance or dependency criterion.

A schematic objective makes the common dependencies visible:

$$J(W, C, S) = L(XW, C) + \lambda \Omega(W) + \alpha R(C; Y) + \beta G(W, C, S) \quad (3)$$

The meaning of observed labels depends on the task

One illustrative instance; labels are A, B, C, D and E.

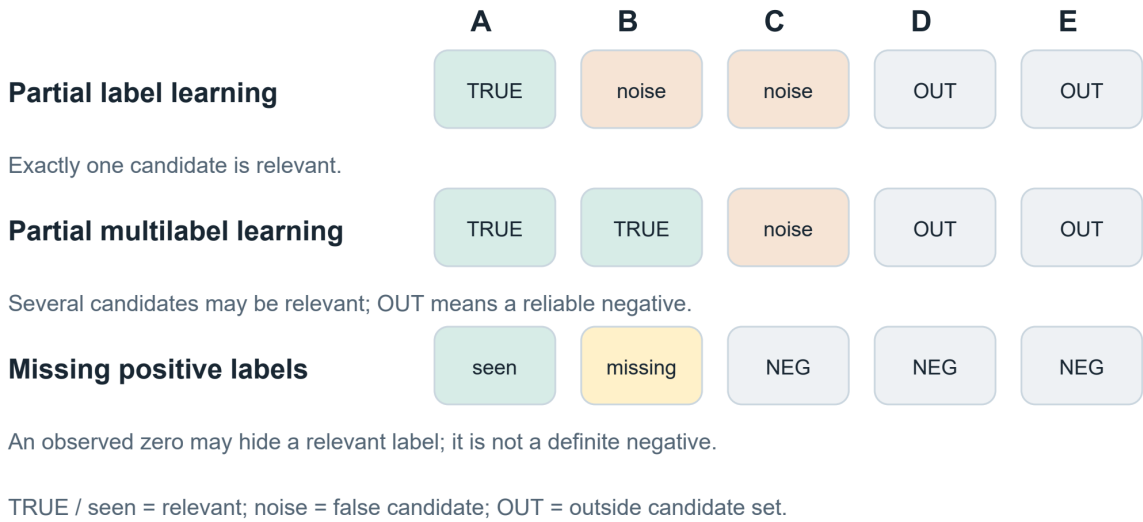


Figure 1. Supervision boundaries. The examples show which labels are actually relevant and which observations are ambiguous. Candidate-superset PML permits several true labels and treats non-candidates as negatives; missing-positive supervision does not justify that inference. Original schematic based on the task definitions [2, 4, 5].

Table 1. Common notation used in this review.

Symbol	Dimension or type	Meaning
n, d, q	Positive integers	Numbers of instances, original features, and labels
X	$n \times d$	Observed feature matrix
Y, Y^*	$n \times q$, binary	Candidate matrix and hidden true relevance matrix
C_i, T_i	Label subsets	Candidate set and true relevant set of instance i
C, D	$n \times q$, real-valued	Label confidence and normalized description distribution
W	$d \times q$	Original feature coefficient matrix, when applicable
$\mathcal{A}$	Subset of $\{1, \dots, d\}$	Selected original feature indices
S	Usually $n \times n$	Instance similarity graph; label graphs have a separate $q \times q$ size

The loss L connects selected or weighted features with estimated supervision C . Ω controls feature complexity, R expresses how C relates to the observed candidates, and G represents structural terms, possibly involving a learned graph S . This expression is an analytical template, not an objective attributed to every reviewed method. Fixed-target methods estimate C before optimizing W . Coupled methods update supervision and feature-related variables within a shared procedure. Filter methods may avoid W entirely and score a candidate feature using relevance, redundancy, and complementarity.

3 Sources of Label Disambiguation Evidence

3.1 Instance Geometry and Neighborhood Agreement

Neighborhood-based methods assume that sufficiently similar instances have related label patterns. This converts feature geometry into evidence about candidate credibility. A label repeatedly supported by nearby instances can receive more confidence than a candidate that is inconsistent with its local context. PML-LD uses feature-space topology together with label correlations to recover soft supervision [6]. PMFS-LDSR extends propagation by considering near and far neighbors as well as negative labels [7].

The assumption depends on the representation used to define similarity. Original feature distances can be unreliable when irrelevant dimensions dominate. A neighborhood may also cross a class boundary or mix several valid label combinations. A globally useful radius can be inappropriate for a sparse region, while a very small neighborhood can provide little support for a rare label. These are reasons to examine the construction and update of geometry separately from the mere presence of a graph term.

Adaptive graph learning provides one response. In PLL, PL-AGGD jointly learns labeling confidences, a similarity graph, and a predictive model [4]. Its relevance here is the idea that the graph can be revised using learned supervision. Its single-label constraint, however, cannot be carried unchanged into PML. In PML-WFGALC, semantic information influences a feature-derived graph while an adaptive label graph captures evolving label relationships [8]. Such prediction frameworks motivate feature selectors that revisit topology, although graph adaptation alone does not demonstrate that the resulting subset is stable or interpretable.

3.2 Label Relationships and Negative Evidence

Label relationships offer evidence beyond one instance. Correlated labels can reinforce each other, but candidate noise can also create spurious co-occurrence. A frequently added false-positive label may appear highly connected precisely because the annotation process is biased. The interpretation of a label graph therefore depends on whether it is computed directly from candidates, reconstructed supervision, or a learned representation.

PML-FSMIR reconstructs candidate supervision using mutual information between labels and later uses label connectivity to revise feature weights [9]. PML-DLAR instead emphasizes removing distorted associations before recovering instance-level and label-level structures through alignment and class prototypes [10]. These approaches share an interest in structural reliability, but they place the correction at different stages. Static candidate-derived statistics risk retaining annotation bias; learned statistics risk reinforcing early model errors.

Non-candidate labels supply another channel under the standard PML assumption. Zhong et al. use negative labels and noise information to guide disambiguation and prediction [11]. The approach highlights that excluded labels contain information about what an instance does not depict. Such evidence can constrain propagation instead of relying exclusively on positive-candidate agreement. Its reliability is conditional on complete candidate coverage of true labels. Missing relevant labels would turn some supposedly reliable negatives into incorrect supervision.

3.3 Cross-Space Consistency and Representation Learning

Feature and label spaces can provide complementary views. FBD-PML connects them through fuzzy label confidence and preserves manifold consistency [12]. DFSDLC learns a discriminative feature subspace and a dynamic label correlation structure [13]. These studies demonstrate different ways to revise ambiguous supervision with feature-related structure. They also expose a recurring assumption: the portions of feature structure retained by the model must be informative about the labels that matter.

Agreement between spaces is not itself proof of correctness. Both views may reflect the same nuisance factor, such as background in an image or a frequent word in a document. Strong alignment can then remove the disagreement that would reveal noise. In a feature selection context, alignment should be assessed together with preservation of rare labels and the recovery of meaningful original variables. A compact subspace is useful for prediction but does not necessarily identify an interpretable feature subset.

PML-TF uses a transformer decoder with partial K-means cross-attention to learn discriminative label embeddings and a class-aware correction loss [14]. Its label-conditioned local representation addresses information that global pooling may miss. This is relevant to disambiguation, especially when evidence for different labels occupies different regions. Attention weights and embeddings, however, should not be treated as rankings of original tabular features without a defined mapping and a dedicated evaluation.

3.4 Label Enhancement and Multiscale Granularity

Label enhancement provides continuous supervision where logical labels are too coarse. Generalized label enhancement models sample correlations through low-rank and tensor-based structure [15], while variational label enhancement treats label distributions as latent variables inferred from logical annotations [16]. These studies supply tools for recovering descriptive degrees. They do not by themselves establish that enhancement removes false-positive candidates in every PML setting.

Distribution quality remains an issue after enhancement. Instance-dependent inaccurate LDL explicitly models distributions contaminated by noise that depends on the input [17]. This matters for feature selection because a smooth distribution can still encode systematic errors. Confidence sharpness, distribution smoothness, and semantic correctness are separate properties. A selector can become very stable around a confidently wrong target.

PML-FSMNG uses neighborhood granules at multiple scales and entropy-guided fusion to estimate soft supervision [18]. Multiple scales provide a way to compare fine and coarse structural evidence. Entropy weights prefer more concentrated local label information, but concentration need not imply correctness when a neighborhood is systematically biased. Fig. 2 places this approach alongside the other evidence sources and distinguishes them from the selection mechanisms that consume the evidence.

4 Representative Feature Selection Mechanisms

4.1 Latent Alignment and Effective Feature Coefficients

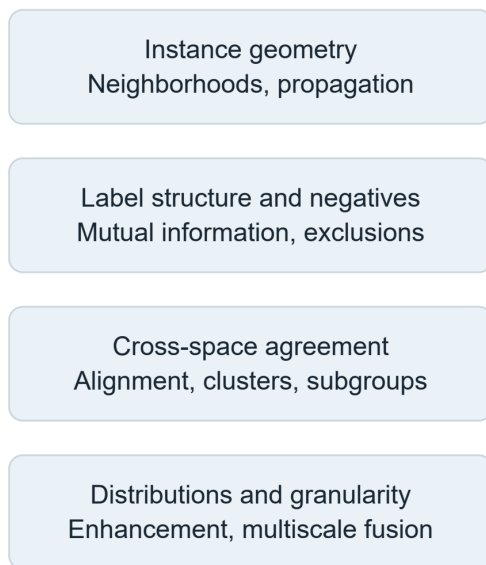
PML-FSLA begins by using OPTICS clustering to determine a latent dimension, factorizes feature and label information, and aligns their instance representations [19]. Its distinctive feature scoring mechanism uses a product of feature-side and label-side projection coefficients. In the paper's notation, the effective coefficient matrix is QR , and original features are ranked by the row norms of that product. Label information therefore enters the score through an explicit cross-space composition rather than only through one reconstruction factor.

This distinction matters when comparing embedded selectors. Ranking rows of a feature reconstruction factor asks which original dimensions reconstruct a latent representation. Ranking rows of a composed feature-to-label matrix asks how those dimensions connect to label-related factors. Neither score should be substituted for the other without examining its dimensions and meaning. PML-FSLA makes this issue visible by incorporating both projections into its sparsity term.

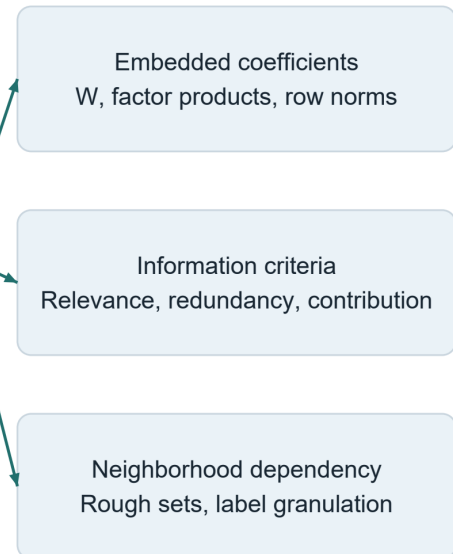
Its benefit is conditional on the feature-derived structure being informative enough to guide label correction. OPTICS also introduces a density and radius choice; deriving a latent dimension from clustering does not remove all model selection. If feature geometry is dominated by nuisance variation, an apparently consistent alignment can preserve misleading structure. This limitation is an analytical consequence of the mechanism, rather than a failure established on every dataset in the original paper.

Two complementary dimensions for comparing methods

Evidence for disambiguation



Mechanism for feature selection



Methods may combine several evidence sources; these are not exclusive families.

Figure 2. An analytical taxonomy with separate axes for disambiguation evidence and feature selection mechanisms. Several evidence sources can feed the same selector, and one method can combine multiple sources. Original synthesis of the reviewed methods [7,9,18–23].

4.2 Mutual Information Reconstruction and Connectivity Weighting

PML-FSMIR is organized into three stages [9]. It first computes label mutual information and reconstructs a candidate-masked supervision matrix. Masking prevents non-candidate labels from receiving nonzero target values through reconstruction. It then learns feature-related coefficients using a reconstructed feature matrix, label-graph regularization, and row sparsity. Finally, it uses label connectivity to reconstruct the learned coefficient matrix before ranking original features.

The method should not be described as eliminating low-rank modeling. It retains a factorization UV of the feature matrix and uses the reconstructed product in its prediction term. Its reformulation preserves the original feature dimensionality of the product while changing how the latent factors enter learning. Preserving the number of columns does not imply that the reconstructed matrix is unconstrained in rank. This distinction prevents dimensionality preservation from being confused with the absence of a low-rank inductive bias.

Connectivity-based reweighting favors features useful for well-connected labels. That can strengthen shared semantic evidence, but connectivity and rarity measure different properties. A sparsely connected, important label may receive little emphasis, whereas a common label may dominate the reconstruction. Consequently, the method raises an evaluation question shared with inverse-frequency approaches: does an aggregate gain reflect better coverage of all labels or mainly improvement on a structurally favored subset?

4.3 Dual-Space and Global-Local Coupling

DGLFS combines local fuzzy clustering with global reconstruction under a shared dual-space structure [20]. Shared memberships connect feature and label clusters, while a learned instance relationship contributes global information. Feature importance is derived from interactions between optimized feature structure and disambiguated label structure. The variables are alternately updated, so feature-related learning can influence subsequent disambiguation rather than merely accepting a fixed preprocessing result.

PMFS_DCDF also combines complementary spaces, but its design differs [21]. Low-dimensional label recovery and feature-label subgroup learning share a coefficient representation. A feature weight matrix is aligned with subgroup-induced importance, and a dynamic label graph is derived from the learned label-side factors. Its final original feature scores use the squared row norms of the learned weight matrix. The formulation includes nonnegativity and Frobenius penalties; it should not be classified automatically as an $\ell_{2,1}$ -sparse method simply because it returns a ranked subset.

Both methods seek to balance compact shared structure with information that may be lost in a single projection. Their comparison is more informative when it names the shared variables, the graph being updated, and the matrix used for ranking. The term dual-space alone does not specify any of these choices. Coupling also introduces several interacting parameters and a nonconvex optimization. An objective that decreases during alternating updates does not establish recovery of the true labels or a globally optimal feature subset.

4.4 Multiscale Distributions and Robust Sparse Regression

PML-FSMNG estimates label distributions from neighborhood granules at several distance scales and fuses them using entropy-based weights [18]. The recovered distribution is then used in a robust regression objective with an $\ell_{2,1}$ residual term, an $\ell_{2,1}$ coefficient penalty, and an entropy-regularized adaptive instance graph. The residual penalty reduces the influence of large instance-level errors, while the coefficient penalty encourages row sparsity and supplies an original feature ranking.

Two uses of entropy must be distinguished. The first weights scale-specific supervisory information during distribution construction. The second regularizes the learned similarity matrix in the feature selection stage. The adaptive graph has instances as its nodes; it is not a q by q label graph. In addition, the distribution is supplied to the feature selection optimization after its construction. Updating W and the graph does not imply that the distribution itself is jointly re-estimated at every optimization step.

Fig. 4 explains this sequence. The framework is attractive when a single neighborhood scale is unreliable, but several scales still share the same feature representation and candidate annotations. Agreement across those scales may therefore preserve a common bias. Dense pairwise relationships and feature-space matrix operations also require attention when sample or feature counts increase. The original study's ablations support its multiscale and fusion components on the datasets examined; they do not establish a universal scale configuration.

4.5 Propagation Followed by Double Regularization

PMFS-LDSR separates instance-level disambiguation from feature coefficient learning [7]. Its first stage combines negative labels with near and far neighbors to derive more reliable supervision. Its second stage uses $\ell_{2,1}$ regularization to suppress irrelevant dimensions and an inner-product-based penalty to address redundancy. Label-specific and label-group-specific information further refines feature weights. This design makes the pathway from corrected labels to selected original variables relatively explicit.

A staged architecture is useful when disambiguation can be checked independently and reused with several selectors. Its principal risk is that a systematic first-stage error remains embedded in the target supplied to the second stage. A redundancy penalty can discourage repeated feature information, but it cannot guarantee correction of a mistaken supervisory target. Evaluation should therefore isolate whether gains arise from improved disambiguation, the selection criterion, or their combination.

4.6 Fuzzy Relevance and Neighborhood Rough Sets

PMFS-F2Deur estimates fuzzy candidate affiliation from sample similarity and then applies a conservative refinement based on an infimum operator [22]. It introduces inverse-frequency label weighting, weighted pure relevance to reduce repeated counting of overlapping label information, and a weighted individual contribution ratio. A feature importance criterion combines these terms with redundancy

information. Selection proceeds through feature evaluation rather than through the row sparsity of a regression matrix.

This family brings the meaning of relevance and redundancy into the foreground. A feature correlated with several overlapping labels should not necessarily receive the sum of several nearly identical rewards. Likewise, a feature that appears redundant marginally may provide conditional information once selected features are considered. The inverse-frequency weighting also makes minority-label protection explicit. However, rarity alone does not imply credibility; rare false-positive candidates can receive undue influence unless disambiguation is reliable. The conservative fuzzy refinement may also penalize a valid label if a highly similar instance provides weak support.

PMFS-RE preserves an intermediate uncertain region during confidence propagation instead of immediately forcing every candidate into a binary decision [23]. It aggregates labels through a graph into coarser representations and uses a neighborhood rough-set dependency criterion for feature selection. Its feature significance is tied to how well selected attributes determine the resulting decision structure. This is different from minimizing a regression residual, even when both approaches exploit neighborhoods.

The rough-set perspective connects to earlier work on positive approximation for accelerating attribute reduction [24]. Such acceleration results should not be transferred automatically to PML: uncertain labels and evolving granules change the underlying decision system. PMFS-RE's coarse label aggregation may reduce uncertainty but can also conceal distinctions needed by individual labels. Assessing performance on original labels, particularly minority labels, remains necessary even when the granulated dependency score improves.

Table 2 summarizes the eight direct methods. The final column states assumptions or questions arising from this review's analysis. These are not claims that every original study demonstrated the corresponding failure.

5 Connections and Tradeoffs Across Method Families

5.1 Information Source and Selection Mechanism Are Separate Dimensions

A method can be graph-based, fuzzy, and sparse at the same time. These terms describe different parts of its construction. Graphs specify relationships, fuzzy memberships express graded evidence, and sparsity controls the representation of feature importance. Treating them as exclusive categories hides meaningful similarities. The taxonomy in Fig. 2 instead asks what produces credible supervision and what turns that supervision into a subset of original variables.

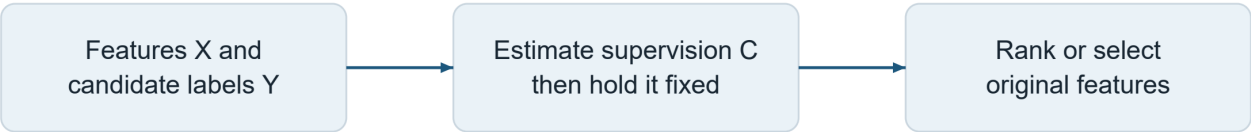
This separation clarifies the comparison of PMFS-LDSR and PML-FSMNG. Both use estimated supervision followed by coefficient learning, but their targets come from different evidence: propagation with negative and neighborhood information versus multiscale distributions. Likewise, PMFS-F2Deur and PMFS-RE both use graded disambiguation but select features through different criteria: information-based relevance and contribution in one case, neighborhood dependency after label aggregation in the other. A hybrid method

Table 2. Eight direct feature selection methods and their principal dependencies. The final column presents analytical questions raised by this review.

Method	Disambiguation evidence	Original feature selection	Assumption or question
PML-FSLA [19]	Feature and label factorization; latent alignment; OPTICS-derived dimension	Row norms of composed projection QR ; $\ell_{2,1}$ penalty	Does feature-derived density structure retain label-relevant variation?
PML-FSMIR [9]	Candidate-masked mutual information reconstruction; label connectivity	Reconstructed coefficient row norms after factor learning	Can candidate-derived connectivity favor frequent or spuriously connected labels?
DGLFS [20]	Shared fuzzy clusters and global instance reconstruction in two spaces	Cross-space feature-label interactions; alternating learning	Can local and global structures reinforce the same nuisance pattern?
PMFS_DCDF [21]	Shared latent factors, subgroup learning, dynamic label graph	Squared row norms of W after subgroup alignment	Does factor sharing preserve label-specific structure at the chosen dimension?
PML-FSMNG [18]	Entropy-weighted multiscale neighborhood distributions	$\ell_{2,1}$ -regularized robust regression; adaptive instance graph	Can concentrated but biased neighborhoods yield misleading targets?
PMFS-LDSR [7]	Negative labels; near and far neighbor propagation	Double-regularized sparse regression; label and group refinement	How much first-stage target error survives downstream selection?
PMFS-F2Deur [22]	Fuzzy two-stage refinement; inverse-frequency label weighting	Weighted pure relevance, individual contribution, and redundancy	Can rare false candidates be overweighted, or conservative refinement suppress valid labels?
PMFS-RE [23]	Confidence propagation; three reliability regions; label aggregation	Neighborhood rough-set dependency and heuristic selection	Does coarse aggregation retain distinctions needed by individual rare labels?

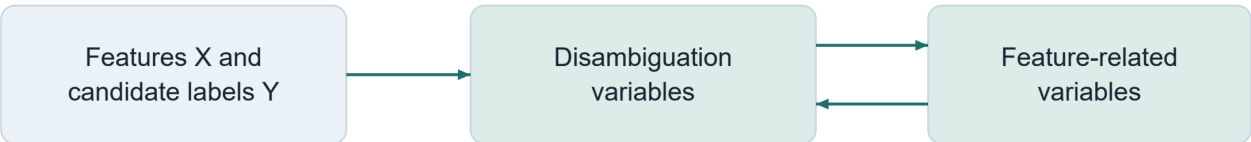
Where information can flow during feature selection

A Staged learning



Errors in the fixed target can pass into the selector.

B Coupled learning



Alternating W and a graph with fixed C is still fixed-target supervision.

Figure 3. Staged and coupled information flow. A fixed recovered target limits downstream feedback. Coupled updates can revise both supervision and feature-related variables, but iteration with a fixed target does not make disambiguation joint. Original schematic.

can therefore change either the supervisory estimator or the selector, and ablations should identify which change is responsible for a gain.

5.2 Staged and Coupled Learning Have Different Error Pathways

Fig. 3 contrasts the principal dependencies. In a staged pipeline, confidence estimation precedes selection. This makes the modules easier to replace and audit, but errors can be inherited by every downstream selector. In a coupled framework, feature-related updates can improve dis-

ambiguation, while updated supervision can refine feature importance. The same feedback can amplify a mistaken early neighborhood or a spurious label correlation.

Iteration alone does not define joint disambiguation and selection. An algorithm may alternate several matrix variables while keeping its supervisory target fixed. Conversely, a filter may update scores as the selected feature set grows without revising its confidence estimates. Reports should state which variables change, which annotations anchor those updates, and when supervision becomes fixed. This is more informative than labeling every iterative procedure a unified framework.

Multiscale supervision and adaptive feature selection

Conceptual sequence of PML-FSMNG; no empirical values are shown.

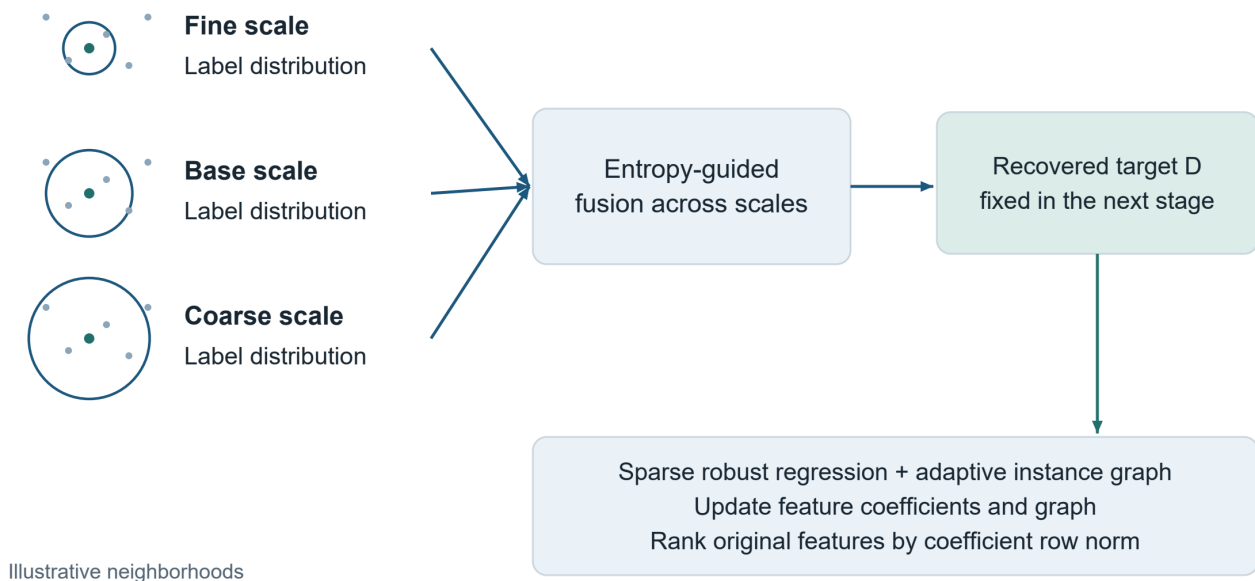


Figure 4. Multiscale supervision construction followed by feature selection in PML-FSMNG [18]. The distribution supplied to the selector is distinguished from the adaptive instance graph updated during regression. The neighborhoods are illustrative. Original conceptual redrawing.

5.3 What Can Be Transferred From Related Tasks

PLL provides several useful mechanisms. PL-AGGD illustrates adaptive topology [4]. Fuzzy neighborhood selection combines iterative disambiguation with fuzzy dependency, information gain, and local similarity [25]. Neighborhood mutual information selection measures continuous-feature relevance and redundancy using confidence and label consistency [26]. Granular-ball selection derives confidence from candidate and non-candidate structures and evaluates fuzzy correlation and redundancy [27]. Weighted centroid disambiguation uses local similarity and class-level information [28], while PLOD formulates disambiguation through constrained clustering [29].

These connections support methodological transfer, not direct equivalence. A constraint that assigns exactly one cluster label per instance must be relaxed or redesigned for multiple relevant labels. Selecting the maximum confidence candidate is appropriate for a single-label decision but can delete valid positives in PML. Neighborhood and granular representations also need to represent partial overlap between label sets, rather than assuming that instances in one class share an exclusive decision label.

Feature selection with partially labeled hybrid data represents a different boundary. Pan and Lin combine self-information, neighbor-based label prediction, and a Student-t kernel rough set [5]. Its partially labeled setting involves unobserved annotation information rather than the same candidate-superset ambiguity as the core PML methods. It is relevant to broader weak supervision, but it is not additional evidence that a selector handles false-positive candi-

date labels. The distinction prevents an apparent expansion of coverage through papers that solve different problems.

6 Experimental Evidence and Evaluation Practice

6.1 What Published Comparisons Establish

The reviewed studies report improvements over their chosen baselines, often accompanied by ablations and statistical comparisons. These findings support particular mechanisms under the corresponding experimental conditions. They do not produce a common ranking of the eight methods. The studies differ in datasets, synthetic ambiguity, downstream learners, feature budgets, and access to clean training labels during evaluation. Table 3 gives several directly documented examples of this heterogeneity.

PML-FSLA and PML-FSMIR evaluate selected features with SVM and ten-fold cross-validation [9, 19]. PML-FSMNG uses ML-KNN with five-fold cross-validation and reports major comparison tables at the top 20 percent of features [18]. PMFS_DCDF uses ML-KNN, five-fold cross-validation, and a top 10 percent rule [21]. PMFS-LDSR also investigates fixed absolute feature counts and several downstream classifiers [7]. A difference in reported average precision across these protocols cannot be assigned to the feature selector alone.

A further distinction concerns what the downstream classifier is allowed to see. DGLFS explicitly discusses an MLkNN-G setting trained with ground-truth label sets to evaluate feature quality [20]. Such a diagnostic can isolate the utility of selected features, but it supplies supervision

Table 3. Selected source-reported protocol differences. These rows describe experimental designs, not a cross-paper performance ranking.

Method	Downstream evaluation	Validation	Feature budget or access note
PML-FSLA [19]	SVM	Ten-fold	1–20% for most datasets; separate handling for Water
PML-FSMIR [9]	SVM	Ten-fold	Feature percentage curves; source reports 20% noise
DGLFS [20]	Includes MLkNN-G	Ten-fold	Clean-label downstream diagnostic is distinguished explicitly
PMFS_DCDF [21]	ML-KNN, $k = 10$	Five-fold	Top 10% rule
PML-FSMNG [18]	ML-KNN, $k = 10$	Five-fold	Major result tables use top 20%
PMFS-LDSR [7]	Several classifiers, including ML-KNN	Five-fold	Absolute counts 5, 10, . . . , 50

unavailable in an end-to-end PML deployment. An oracle evaluation and a weak-supervision evaluation answer different questions and should be reported separately. This distinction does not invalidate the diagnostic; it specifies the conclusion that the diagnostic can support.

6.2 Ambiguity Generation and Dataset Identity

Synthetic PML datasets are commonly obtained by adding irrelevant labels to fully annotated multilabel data. At least three quantities should be recorded: the probability of adding each negative label, the number of added candidates per instance, and the realized number of candidates relative to true positives. These measures are not interchangeable. For example, a ratio of added false candidates to true labels can exceed 100 percent, whereas a probability cannot. A statement such as 20 percent noise is incomplete without its denominator and sampling procedure.

Candidate corruption should also be separated from corruption of test relevance. For a standard generalization experiment, the outer test set retains reference labels for scoring, while only permissible training annotations are given to the learner. Structured ambiguity deserves additional evaluation: false candidates can depend on label co-occurrence, instance appearance, or annotator behavior instead of being sampled uniformly. Strong performance under uniform additions does not establish robustness to these dependencies.

Dataset names alone are insufficient identifiers. Studies may use different feature representations, sample subsets, label vocabularies, or preprocessing versions of a nominally identical benchmark. A reproducible report needs the exact source, dimensions, preprocessing, candidate construction seed, and split indices. Label cardinality, label density, and candidate cardinality should be distinguished. In particular, several corrupted versions of one base dataset are related experimental conditions, not independent evidence from several unrelated applications.

6.3 Validation Without Information Leakage

Fig. 5 gives a recommended inductive evaluation workflow. The outer split precedes learned normalization, confidence recovery, graph construction, feature selection, and threshold tuning. These operations are fitted using training information. Hyperparameters and feature budgets are chosen within an inner validation procedure, after which the chosen pipeline is refitted on the outer training set and assessed on the untouched test set. Candidate corruption and random initialization must be controlled so compared methods receive equivalent training conditions.

Graph construction requires particular care. Using test features without test labels is not the same as using test labels, but it changes an inductive experiment into a transductive one when the learner depends on those features. Either design can be meaningful if its information access is explicit and baselines receive the same access. Building a graph on the complete dataset and then reporting an inductive score obscures that difference. The need for such disclosure is a recommendation; this review does not infer leakage merely because a paper omits implementation detail.

Clean-label use should be disclosed at every stage. An oracle feature-quality experiment can train a downstream classifier with clean training labels after feature selection. A deployable PML pipeline must instead train from candidates or the recovered supervision available to the learner. Hyperparameters selected using hidden training truth should also be identified as oracle-assisted. Without this separation, disambiguation quality, feature quality, and classifier supervision are confounded.

6.4 Metrics Measure Different Aspects of Success

Table 4 summarizes complementary metrics. Hamming loss averages binary errors across instance-label pairs and can appear small when most labels are negative. Micro-F1 emphasizes aggregate positive prediction, while macro-F1 exposes uneven performance across labels. Average precision and ranking loss evaluate the order of relevant and irrelevant labels; they do not require the same threshold decisions as classification measures. Subset accuracy requires an entire predicted label set to match and is correspondingly strict.

A concrete within-paper example shows why several metrics are needed. In PML-FSMNG’s Table 4, the reported Yeast micro-F1 is 0.6104, compared with 0.6119 for SCLS, while its average precision is 0.7490, compared with 0.7443 for SCLS [18]. The values are source-reported means, not results reproduced here, and they do not by themselves establish statistical significance. They show that a method can improve ranking quality without improving a particular thresholded score on the same dataset. A single count of best results would hide this distinction.

Feature selection also needs measurements beyond predictive accuracy. Reports should give the feature budget, variation over budgets, runtime, peak memory where possible, and stability of selected sets under changes in sampling or corruption. A high ranking score does not establish calibrated confidence, and an $\ell_{2,1}$ penalty does not guarantee a stable subset when features are interchangeable. Minority-label recall and macro-averaged measures are especially rel-

A recommended inductive evaluation protocol

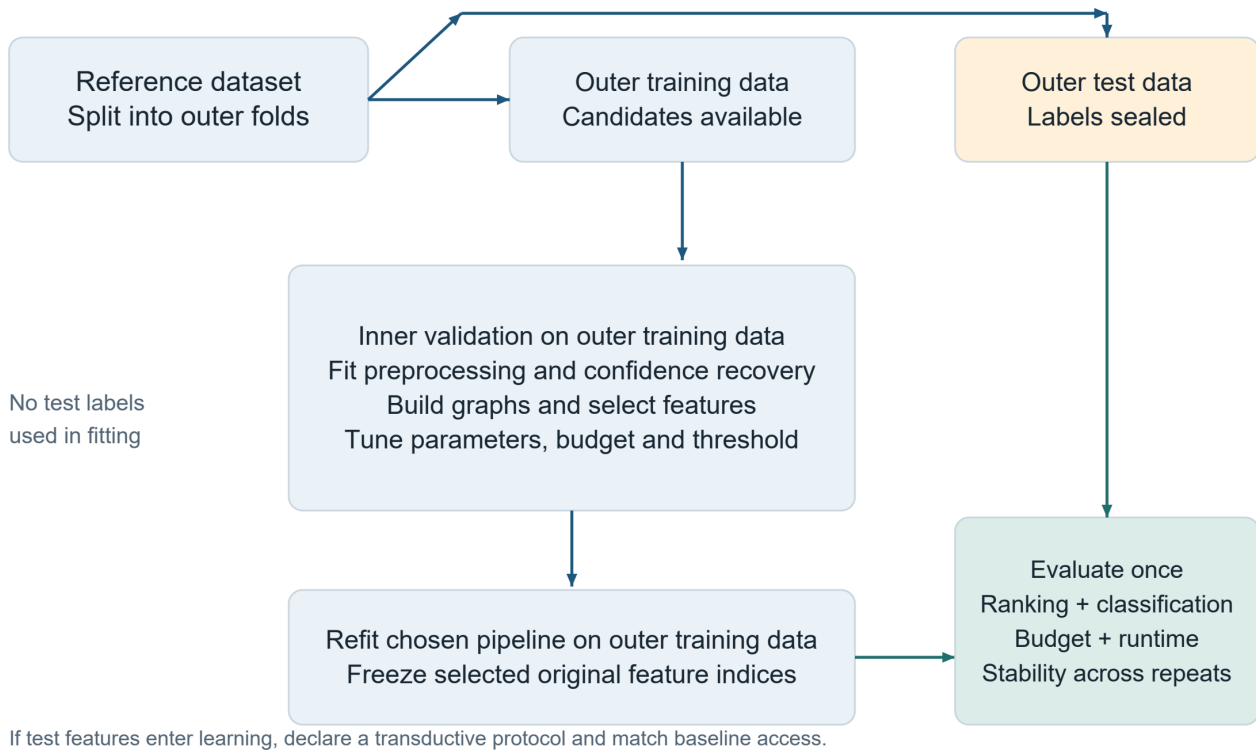


Figure 5. Recommended information flow for an inductive evaluation. Learned preprocessing, disambiguation, graph construction, feature selection, and tuning occur inside the training process. Test-feature access must be declared if a transductive design is used. Original protocol schematic.

Table 4. Complementary evaluation measures and the questions they answer.

Measure	What it measures	Interpretation and limitation
Hamming loss	Binary instance–label errors; lower is better	Can be dominated by numerous negative labels
Micro-F1	F1 after pooling label decisions; higher is better	Reflects aggregate positive prediction; frequent labels have more influence
Macro-F1	Average F1 across labels; higher is better	Shows uneven label performance; report treatment of labels with no positives
Average precision	Ranking of relevant labels; higher is better	Does not establish threshold calibration
Ranking loss	Inverted relevant–irrelevant pairs; lower is better	Not equivalent to the quality of the predicted label set
Subset accuracy	Exact set match; higher is better	Strict when many labels or label combinations are possible

evant when the proposed contribution concerns label weighting or positive-label identification.

6.5 Ablations and Statistical Interpretation

The most useful ablations separate the supervisory estimator, the feature criterion, and their interaction. A staged method can compare its selector with and without disambiguation and can hold the recovered target fixed while changing selectors. A coupled method can freeze one direction of feedback or replace a learned graph with an otherwise comparable fixed graph. Multiscale fusion should be compared with a single scale and uniform fusion under the same computational budget, as investigated in PML-FSMNG [18]. Tests should also address the assumptions that motivate each component.

Comparisons should use paired data splits and corruption realizations, report dispersion, and identify the unit of

statistical analysis. Repeated folds, feature budgets, and corrupted variants of one dataset are dependent observations. Treating all of them as independent datasets exaggerates evidence. Statistical significance across a benchmark collection does not demonstrate superiority on every dataset, and overlapping or nonoverlapping standard deviations are not a substitute for a suitable paired analysis. This review therefore synthesizes mechanisms and protocol differences without pooling heterogeneous scores into a numerical meta-analysis.

7 Open Research Directions

7.1 Calibrated Uncertainty and Selective Disambiguation

Most methods must decide how strongly each uncertain candidate should influence selection. A useful next step is to

assess whether confidence corresponds to actual correctness, rather than only whether it is concentrated or smoothly propagated. Calibration is particularly important when thresholds determine which candidate labels are retained or which instance-label pairs contribute to feature scores. A small trusted validation subset could support reliability assessment, provided its annotation cost is reported and equivalent supervision is available to baselines.

The uncertain region in PMFS-RE provides one starting point [23]. More generally, a selector could retain unresolved candidates with bounded influence and postpone hard decisions when neighborhood, graph, and model evidence disagree. The research question is whether this reduces error propagation at a fixed feature budget. Relevant outcomes include disambiguation precision and recall, downstream macro-F1, and variation across candidate corruptions. Confidence disagreement should be treated as an observable diagnostic, rather than automatically averaged away.

7.2 Rare Labels and Conditional Feature Utility

The contrast between connectivity weighting [9] and inverse-frequency weighting [22] exposes a broader issue: structural prominence and semantic importance need not coincide. Frequent, highly connected labels provide many observations, whereas rare labels can require features with little global relevance. A selector that optimizes only aggregate reconstruction may discard these variables. Conversely, aggressively weighting rare labels may promote rare annotation errors.

Future work should distinguish label rarity from label reliability and report their interaction. One direction is to estimate relevance conditionally on already selected features and on a label's confidence, allowing a variable with modest marginal relevance to survive if it contributes unique evidence for a minority label. Evaluation should stratify performance by training frequency and ambiguity, and compare the selected feature sets at matched budgets. Such tests would reveal whether an apparent minority benefit comes from better disambiguation or simply a larger effective model.

7.3 Structural Assumptions That Can Be Checked

Low rank, neighborhood smoothness, shared latent spaces, and label correlation all offer useful inductive biases. Their value depends on the data. Shared nuisance structure can make feature-label alignment persuasive but incorrect, while a label graph derived from ambiguous candidates can encode annotation habits. Adding more structural terms can strengthen the same mistaken agreement instead of supplying independent evidence.

Methods should therefore expose diagnostics for their assumptions. Possible tests include perturbing irrelevant features, withholding common co-occurrence patterns, evaluating label combinations not seen in training, and varying the neighborhood definition. Structure learned from feature geometry can be compared with structure inferred from a small trusted annotation sample. The purpose is to identify conditions under which an assumption helps, and the point at which it becomes a source of bias. Consistent objective descent should be reported as an optimization property,

distinct from these semantic checks.

7.4 Mixed Annotation Errors and Changing Data

The reliable-negative interpretation of non-candidates is central to many PML methods. Practical annotations may also omit true labels, making a richer supervision model necessary. A mixed-error extension must state which observations are trusted and how false-positive candidates can be distinguished from unobserved positives. Without additional assumptions or trusted information, the two sources of ambiguity may not be identifiable from the candidate matrix alone.

Changing label vocabularies, streaming features, and domain shifts introduce related difficulties. Confidence calibrated on one annotation process may become misleading under another. Future selectors should be evaluated under explicitly defined shifts rather than assuming that robustness to random candidate additions transfers automatically. The partially labeled hybrid setting [5] offers a reminder that neighboring weak-supervision problems can inform extensions only after their observation mechanisms are reconciled.

7.5 Scalable and Auditable Original Feature Selection

Several reviewed methods construct pairwise instance or label relationships and solve matrix optimization problems. Scaling requires more than timing one benchmark. Memory use, graph construction, confidence estimation, iterative updates, and downstream fitting should be reported separately. Sparse graphs, sampled neighborhoods, structured matrix operations, and granular summaries may reduce costs, but their effect on minority labels and boundary instances must be measured.

Interpretability also requires an explicit chain from original variables to scores. Factor products, reconstructed coefficient matrices, and rough-set dependency increments should be available for inspection. A feature ranking is more useful when its variability is reported and the reason for selecting a variable can be traced to relevant labels or instances. Shared splits, candidate corruption scripts, exact feature indices, and executable evaluation code would make these claims easier to assess. A common benchmark implementing both oracle feature-quality diagnostics and end-to-end weak-supervision evaluation would address a major source of ambiguity in current comparisons.

8 Conclusions

Feature selection under partial multilabel supervision depends on how ambiguous labels become credible evidence. Neighborhoods, label relationships, negative labels, cross-space agreement, and multiscale granularity provide different routes to that evidence. Sparse coefficients, composed latent mappings, information-based relevance, and rough-set dependency then translate it into original feature rankings or subsets. These components should be compared separately before their combinations are judged.

The eight direct methods reviewed here illustrate a move toward richer interactions between supervision and feature

structure. They also retain distinct assumptions and error pathways. Staged pipelines can preserve target errors, while coupled learning can reinforce incorrect agreement. Normalized label distributions require a different interpretation from multilabel relevance confidence, and representation learning needs an explicit original-variable mapping before it qualifies as feature selection. Finally, published scores support conclusions within their own protocols; differences in corruption, classifiers, feature budgets, and clean-label access prevent a defensible global ranking. Progress will depend on calibrated supervision, preservation of minority-label evidence, and evaluations that make both information access and feature utility explicit.

Originality

The author confirms that this manuscript is original, has not been published previously, and is not under consideration by any other journal.

Data and code availability

This review analyses published literature and reports no new data. The comparison tables are reproduced from the cited sources.

Conflicts of interest

The author declares no competing financial or non-financial interests.

Use of generative AI

A generative AI tool was used for language editing of the manuscript. The author checked the technical content, the references and the figures, and takes full responsibility for the published version.

Funding

This work received no specific grant from any funding agency.

Ethics approval

Not applicable. This review involves no human participants and no animals.

References

- [1] Min-Ling Zhang, Zhi-Hua Zhou. A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8), 1819-1837, 2014. <https://doi.org/10.1109/tkde.2013.39>
- [2] Ming-Kun Xie, Sheng-Jun Huang. Partial Multi-Label Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 4302-4309, 2018. <https://doi.org/10.1609/aaai.v32i1.11644>
- [3] Wenbin Qian, Jintao Huang, Fankang Xu, Wenhao Shu, Weiping Ding. A survey on multi-label feature selection from perspectives of label fusion. *Information Fusion*, 100, 101948, 2023. <https://doi.org/10.1016/j.inffus.2023.101948>
- [4] Deng-Bao Wang, Li Li, Min-Ling Zhang. Adaptive Graph Guided Disambiguation for Partial Label Learning. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 83-91, 2019. <https://doi.org/10.1145/3292500.3330840>
- [5] Sujuan Pan, Yonghua Lin. Feature selection for partially labeled hybrid data via self information, prediction label using k-nearest neighbor and Student-t kernel rough set. *Engineering Applications of Artificial Intelligence*, 164, 113223, 2026. <https://doi.org/10.1016/j.engappai.2025.113223>
- [6] Ning Xu, Yun-Peng Liu, Xin Geng. Partial Multi-Label Learning with Label Distribution. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04), 6510-6517, 2020. <https://doi.org/10.1609/aaai.v34i04.6124>
- [7] Jing Chai, Yemin Han, Khursheed Aurangzeb, Fa Zhu, Wadood Abdul. Partial multi-label feature selection based on label disambiguation and double-regularized sparse regression. *Journal of King Saud University Computer and Information Sciences*, 38(5), 301, 2026. <https://doi.org/10.1007/s44443-025-00434-3>
- [8] Zhenzhen Sun, Shuhao Shan, Hongbo Zhang, Jin Gou, Jia Zhong, Jinghua Liu. Weighted Feature Graph With Adaptive Label Correlation for Partial Multi-Label Learning. *IEEE Transactions on Multimedia*, 28, 8324-8338, 2026. <https://doi.org/10.1109/tmm.2026.3678814>
- [9] Wanfu Gao, Hanlin Pan, Qingqi Han, Kunpeng Liu. Noise-Resistant Label Reconstruction Feature Selection for Partial Multi-Label Learning. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 5172-5180, 2025. <https://doi.org/10.24963/ijcai.2025/576>
- [10] Xuhuan Zhu, Yu Chen, Xiaozhao Fang, Jie Wen, Jing Zhang, Guoxu Zhou, Shengli Xie. Dual Label Association Recovery for Partial Multi-Label Learning. *IEEE Transactions on Multimedia*, 28, 6251-6265, 2026. <https://doi.org/10.1109/tmm.2026.3668550>
- [11] Jingyu Zhong, Ronghua Shang, Feng Zhao, Weitong Zhang, Songhua Xu. Negative Label and Noise Information Guided Disambiguation for Partial Multi-Label Learning. *IEEE Transactions on Multimedia*, 26, 9920-9935, 2024. <https://doi.org/10.1109/tmm.2024.3402534>
- [12] Xiaozhao Fang, Xi Hu, Yan Hu, Yonghao Chen, Shengli Xie, Na Han. Fuzzy bifocal disambiguation for partial multi-label learning. *Neural Networks*, 185, 107137, 2025. <https://doi.org/10.1016/j.neunet.2025.107137>
- [13] Hao Xie, Jianhua Dai. Partial multi-label learning with discriminative feature subspace and dynamic label correlation. *Pattern Recognition*, 178, 113543, 2026. <https://doi.org/10.1016/j.patcog.2026.113543>
- [14] Jinfu Fan, Jiangnan Li, Wenpeng Lu, Linqing Huang. Partial multi-label learning via transformer to discover discriminative label embeddings. *Neurocomputing*, 684, 133518, 2026. <https://doi.org/10.1016/j.neucom.2026.133518>
- [15] Qinghai Zheng, Jihua Zhu, Haoyu Tang, Xinyuan Liu, Zhongyu Li, Huimin Lu. Generalized Label Enhancement with Sample Correlations. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 482-495, 2023. <https://doi.org/10.1109/tkde.2021.3073157>
- [16] Ning Xu, Jun Shu, Yun-Peng Liu, Xin Geng. Variational Label Enhancement. *Proceedings of the 37th International Conference on Machine Learning*, 119, 10597-10606, 2020. <https://proceedings.mlr.press/v119/xu20g.html>
- [17] Zhiqiang Kou, Jing Wang, Yuheng Jia, Biao Liu, Xin Geng. Instance-Dependent Inaccurate Label Distribution Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1), 1425-1437, 2025. <https://doi.org/10.1109/tnnls.2023.3329870>
- [18] Yifan Cao, Mao Li, Cong Wang, Shuyu Fan, Ziqiao Yin, Binghui Guo. Partial Multi-Label Feature Selection via Entropy-Weighted Multi-Scale Neighborhood Granular Label Distribution Learning. *Entropy*, 28(4), 422, 2026. <https://doi.org/10.3390/e28040422>
- [19] Hanlin Pan, Kunpeng Liu, Wanfu Gao. Reconsidering Feature Structure Information and Latent Space Alignment in Partial Multi-label Feature Selection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(19), 19786-19794, 2025. <https://doi.org/10.1609/aaai.v39i19.34179>

- [20] Cheng Peng, Wanfu Gao, Pingting Hao, Qingqi Han. Dual-space guided global-local disambiguation for partial multi-label feature selection. *Neurocomputing*, 677, 133120, 2026. <https://doi.org/10.1016/j.neucom.2026.133120>
- [21] Hongkun Yang, Hongmei Chen, Biao Xiang, Zhong Yuan, Tianrui Li. Partial multi-label feature selection via dual-space combination and disambiguation strategy fusion. *Pattern Recognition*, 179, 113537, 2026. <https://doi.org/10.1016/j.patcog.2026.113537>
- [22] Cheng Huang, Duo Xu, Yun Kang, Jianhua Dai. Partial Multilabel Feature Selection via Fuzzy Two-Stage Disambiguation and Weighted Pure Relevance. *IEEE Transactions on Fuzzy Systems*, 34(7), 2104-2117, 2026. <https://doi.org/10.1109/tfuzz.2026.3686128>
- [23] Jinghua Liu, Shuqi Huang, Yuling Fan, Hongbo Zhang, Jin Gou, Yaojin Lin. Reliability-enhanced partial multi-label feature selection based on neighborhood rough set. *Pattern Recognition*, 179, 113762, 2026. <https://doi.org/10.1016/j.patcog.2026.113762>
- [24] Yuhua Qian, Jiye Liang, Witold Pedrycz, Chuangyin Dang. Positive approximation: An accelerator for attribute reduction in rough set theory. *Artificial Intelligence*, 174(9-10), 597-618, 2010. <https://doi.org/10.1016/j.artint.2010.04.018>
- [25] Junqi Li, Wenbin Qian, Wenji Yang, Suxuan Liu, Jintao Huang. Fuzzy neighborhood-based partial label feature selection via label iterative disambiguation. *International Journal of Approximate Reasoning*, 179, 109358, 2025. <https://doi.org/10.1016/j.ijar.2024.109358>
- [26] Jinfei Ding, Wenbin Qian, Yihui Li, Wenji Yang, Jintao Huang. Partial label feature selection via label disambiguation and neighborhood mutual information. *Information Sciences*, 680, 121163, 2024. <https://doi.org/10.1016/j.ins.2024.121163>
- [27] Wenbin Qian, Junqi Li, Xinxin Cai, Jintao Huang, Weiping Ding. Granular ball-based partial label feature selection via fuzzy correlation and redundancy. *Information Sciences*, 709, 122047, 2025. <https://doi.org/10.1016/j.ins.2025.122047>
- [28] Yuhang Tian, Xin Niu, Jing Chai. Partial label learning via weighted centroid clustering disambiguation. *Neurocomputing*, 604, 128312, 2024. <https://doi.org/10.1016/j.neucom.2024.128312>
- [29] Jun-Ying Liu, Jian-Ping Sun, Ya-Hong Zhao, Bin-Bin Jia, Min-Ling Zhang. Partial label learning with semi-supervised clustering disambiguation. *Pattern Recognition*, 171, 112138, 2026. <https://doi.org/10.1016/j.patcog.2025.112138>