ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Multi-label Complementary Labels Learning with Hard Logical Constraints

Yi Gao[1,2], Davide Anguita[2], Fabio Roli[2,3],
Min-Ling Zhang[1], Fulvio Mastrogiovanni[2], and Luca Oneto[2] *

1 - Southeast University, Nanjing, China
2 - University of Genoa, Genova, Italy
3 - University of Cagliari, Cagliari, Italy

Abstract. Two of the main challenges in multi-label classification are the need to collect labeled data, which can be costly or impractical, and the need to satisfy hard logical constraints between labels, which is often computationally expensive. In some applications, complementary labels - that is, labels specifying a class to which a sample does not belong - are available and much less costly to obtain. Researchers have therefore developed methods to learn from such labels efficiently and effectively. Similar efforts have been made to address the problem of learning with hard logical constraints. Nevertheless, to the best of our knowledge, no prior work has investigated the problem of learning from complementary labels with hard logical constraints. In this work, we propose and compare methods to address this problem, showing that hard logical constraints, besides representing restrictions to be satisfied, can also serve as an additional source of weak supervision. The relationships between labels can help bridge the information gap between relevant and complementary labels. Experimental results on different datasets and scenarios support our claims.

1 Introduction

Multi-label classification (MLC) aims to assign multiple relevant labels to each sample [1, 2]. It has been widely applied in several domains, including medical diagnosis [3, 4], image recognition [5, 6], and e-commerce product categorization [7]. Despite its broad applicability, MLC faces two key challenges: (i) collecting precisely annotated multi-label data is labor-intensive and costly [8–10], and (ii) incorporating hard logical constraints (HLC) between labels, such as hierarchical or relational dependencies, is often computationally expensive [8–10].

To mitigate the annotation burden, multi-label complementary label learning (MLCLL) has recently emerged as a promising alternative [8–10]. Unlike MLC, MLCLL learns from complementary labels rather than relevant ones. Each complementary label indicates a class that a sample does not belong to, which substantially reduces annotation costs [11, 12]. Existing MLCLL methods primarily focus on constructing unbiased risk estimators [8, 10] or designing robust loss functions [9]. One of the state-of-the-art approaches in this area is the gradient-descent-friendly (GDF) method, which facilitates stable gradient updates for unbiased risk estimation [10]. These methods, however, overlook HLC between

*This work is partially supported by (i) project ELSA – European Lighthouse on Secure and Safe AI funded by the European Union's Horizon Europe under the grant agreement No. 101070617, (ii) project SERICS (PE00000014) and project FAIR (PE00000013) under the NRRP MUR program funded by the EU - NGEU, (iii) project FISA-2023-00128 funded by the MUR program "Fondo italiano per le scienze applicate", and (iv) project 624B2042 funded by the National Science Foundation of China.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

labels. Such constraints are inherent in multi-label data and are crucial for improving prediction accuracy [13].

Parallel to these efforts, another research line in MLC incorporates HLC to encode domain knowledge and enforce prediction consistency [13–15]. For example, the constraint "if an image is labeled as *Shark*, it must also be labeled as *Fish*" (i.e., $Shark \rightarrow Fish$) provides additional semantic structure that guides model learning. A representative state-of-the-art method in this direction is the coherent-by-construction network (CCN) [13]. Current HLC-based approaches, however, assume access to fully observed relevant labels, which makes them unsuitable for MLCLL, where only complementary labels are provided and constraint relationships remain difficult to leverage.

To the best of our knowledge, this paper presents the first investigation of MLCLL with HLC by integrating GDF with CCN. We propose a novel framework that enforces consistency between model outputs and logical constraints while learning from complementary labels. We argue that logical constraints, beyond serving as rules to satisfy, can act as an additional form of weak supervision. By exploiting these constraints, we bridge the information gap between complementary and relevant labels, which improves the learning efficiency of MLCLL models. Extensive experiments on multiple benchmark datasets demonstrate the effectiveness of the proposed approach.

The remainder of the paper is organized as follows. Section 2 introduces the notations. Section 3 presents the proposed method. Section 4 reports the experimental results and analysis. Section 5 concludes the paper.

2 Preliminaries

Let $\mathcal{X} = \mathbb{R}^d$ denote the input space and let $\mathcal{Y} = \{l_1, l_2, \ldots, l_c\}$ denote the label space with c classes. A multi-label complementary sample is denoted as $(\boldsymbol{x}, \bar{Y})$, where $\boldsymbol{x} \in \mathcal{X}$ is the feature vector and $\bar{Y} \subset \mathcal{Y}$ is the set of complementary labels. In MLCLL, the relevant label set $Y \subset \mathcal{Y}$ (with $Y \cap \bar{Y} = \emptyset$) is not available to the learning algorithm. Given a dataset of n complementary samples $\bar{\mathcal{D}}_n = \{(\boldsymbol{x}_i, \bar{Y}_i)\}_{i=1}^n$, the goal is to learn a multi-label classifier $\boldsymbol{g} : \mathcal{X} \rightarrow [0,1]^c$ that maps each input $\boldsymbol{x}$ to a vector of confidence scores, where $g_j(\boldsymbol{x})$ denotes the predicted probability that $\boldsymbol{x}$ belongs to label l_j. A label l_j is predicted as relevant for $\boldsymbol{x}$ if $g_j(\boldsymbol{x}) > \theta$, where $\theta \in [0,1]$ is a predefined threshold. Accordingly, the predicted label set is $Y = \{l_j \mid g_j(\boldsymbol{x}) > \theta, 1 \leq j \leq c\}$.

Let Π denote a finite set of logical constraints defined over $\mathcal{Y}$. Each constraint $\pi \in \Pi$ is expressed in normal form as $l_1 \wedge \cdots \wedge l_j \wedge \neg l_{j+1} \wedge \cdots \wedge \neg l_k \rightarrow l_o$, where $0 < j \leq k$. This rule states that if labels $l_1, \ldots, l_j$ are predicted and labels $l_{j+1}, \ldots, l_k$ are not, then label l_o must also be predicted. For each constraint π, we denote its head as $H(\pi) = l_o$ and its body as $B(\pi) = B^+(\pi) \cup B^-(\pi)$, where $B^+(\pi) = \{l_1, \ldots, l_j\}$ and $B^-(\pi) = \{l_{j+1}, \ldots, l_k\}$. For a given input $\boldsymbol{x} \in \mathcal{X}$, the prediction $\boldsymbol{g}(\boldsymbol{x})$ satisfies a constraint π if its binary decision vector renders π true under propositional logic. The classifier $\boldsymbol{g}$ is said to be coherent with Π if all constraints in Π are satisfied for every prediction.

3 Methodology

In this section, we present our proposal to address the MLCLL problem with HLC. To this end, we build upon two state-of-the-art approaches: GDF [10] for

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

handling MLCLL, and CCN [10] for managing HLC. Our framework comprises two main components: (i) a base prediction model $\boldsymbol{f}:\mathcal{X}\to[0,1]^c$ that produces raw class probabilities and is trained using the GDF loss under complementary supervision; and (ii) a logical constraint module $\boldsymbol{h}:[0,1]^c\to[0,1]^c$ built atop $\boldsymbol{f}$, which refines the predictions into logic-consistent outputs by enforcing a set of logical constraints Π. This refinement step ensures that all predictions remain coherent with the predefined logical relations. We now introduce the standard GDF loss [10], defined for each sample $(\boldsymbol{x},\bar{Y})$ as:

$$\bar{\ell}(\boldsymbol{f}(\boldsymbol{x}),\bar{Y}) = -\sum_{j=1}^{c}\left[\mathbb{I}(l_j\notin\bar{Y})\log(f_j(\boldsymbol{x}))+\mathbb{I}(l_j\in\bar{Y})\log(1-f_j(\boldsymbol{x}))\right], \quad (1)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, returning 1 if the condition holds and 0 otherwise, and f_j is the j-th component of $\boldsymbol{f}$. The refined predictions produced by the constraint module $\boldsymbol{h}$ are subsequently passed into the GDF loss, replacing the raw prediction probabilities. This integration allows the learning process to leverage the additional information provided by the HLC. Accordingly, the overall learning objective is formulated as:

$$\bar{L} = \frac{1}{n}\sum_{i=1}^{n}\bar{\ell}(\boldsymbol{h}(\boldsymbol{f}(\boldsymbol{x}_i);\Pi),\bar{Y}_i). \quad (2)$$

Note that $\boldsymbol{h}$ must be a differentiable module ensuring that the composite function $\boldsymbol{h}\circ\boldsymbol{f}$ satisfies all constraints in Π. This differentiability allows gradients to flow through $\boldsymbol{h}$, enabling the base model to learn representations that are both logic-consistent and complementary-aware. To construct $\boldsymbol{h}$, we adopt a stratified forward-chaining scheme [13] over the constraint set Π, under the assumption that Π is stratified, i.e., it contains no cyclic dependencies involving negation. Given this assumption, Π can be partitioned into m strata $\Pi_1,\ldots,\Pi_m$ using the CompStrata(Π) algorithm [13], where each constraint in Π_i depends only on labels that are either input features or derived from constraints in preceding strata $\Pi_{<i}$ $(i\in\{1,2,\ldots,m\})$. The strata are mutually exclusive, i.e., $\Pi_r\cap\Pi_s=\emptyset$ for $r\neq s$, and their number is minimized to reduce the number of propagation steps. This stratification eliminates circular dependencies and guarantees the existence of a unique minimal coherent extension, thereby ensuring both correctness and computational efficiency in constraint propagation. We now describe the updating process of the constraint module $\boldsymbol{h}$, which operates layer by layer over the stratified constraint set $\Pi_1,\ldots,\Pi_m$. At each step i, let $\boldsymbol{h}(\boldsymbol{x};\Pi_{i-1})$ denote the refined predictions obtained from the previous stratum, which are then used to compute $\boldsymbol{h}(\boldsymbol{x};\Pi_i)$. For each input $\boldsymbol{x}$ and constraint $\pi\in\Pi_i$, the module updates the confidence score of the head $H(\pi)$ based on the satisfaction degree of the body of π, which quantifies how strongly the antecedent holds for $\boldsymbol{x}$. Formally, the confidence score of $H(\pi)$ is refined as:

$$h_{H(\pi)}(\boldsymbol{x};\Pi_i)=\max_{\pi\in\Pi_i}\left\{h_{H(\pi)}(\boldsymbol{x};\Pi_{i-1}),\min\left\{\begin{matrix}\min_{l_j\in B^+(\pi)}h_{l_j}(\boldsymbol{x};\Pi_{i-1}),\\ \min_{l_j\in B^-(\pi)}[1-h_{l_j}(\boldsymbol{x};\Pi_{i-1})]\end{matrix}\right\}\right\}, \quad (3)$$

where $h_{l_j}(\cdot)$ denotes the j-th component of $\boldsymbol{h}(\cdot)$. The predictions are initialized with the raw outputs of the base model, that is, $\boldsymbol{h}(\boldsymbol{x};\Pi_0)=\boldsymbol{f}(\boldsymbol{x})$. Eq. (3) ensures that when the body of a constraint is strongly satisfied, the corresponding head prediction is increased accordingly, while all other label predictions remain unchanged:

$$h_{l_j}(\boldsymbol{x};\Pi_i)=h_{l_j}(\boldsymbol{x};\Pi_{i-1}),\quad \text{if } l_j\notin\{H(\pi)\mid\pi\in\Pi_i\}. \quad (4)$$

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

After processing all strata, the final logic-consistent outputs are obtained as $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}; \Pi) = \boldsymbol{h}(\boldsymbol{x}; \Pi_m)$. This design guarantees that $\boldsymbol{g}$ satisfies all constraints in Π, thus eliminating the need for any post-hoc projection. Consequently, our approach can be trained in an end-to-end manner, allowing logical consistency to directly influence both feature learning and gradient propagation. Furthermore, the incorporated HLC serve as an additional form of weak supervision, effectively bridging the gap between complementary and relevant labels.

4 Experimental Results

In this work we employ thirteen widely used MLC datasets[1] that include relevant labels and constraints[2]. For each dataset, complementary labels are randomly sampled from the label set, excluding the relevant labels, until the number of complementary labels reaches $q\%$ of c. We vary q within $\{10, 50\}$ to simulate different degrees of weak supervision. The training data include only complementary labels, while validation and test data include relevant labels. Following [13], 30% of the samples in each dataset are reserved for testing. From the remaining data, 15% are randomly selected for validation and the rest are used for training. We also assess our method in two scenario where just $s\%$ of the dataset is available. We vary s within $\{10, 100\}$: full dataset and low-data scenario where only 10% of the training data are used. We compare our approach with the base GDF [10] MLCLL method. We adopt four standard MLC metrics [1]: *ranking loss*, *one-error*, *coverage*, and *average precision*. Higher average precision and lower values for the remaining metrics indicate better performance. For each dataset, we repeat the random data split ten times and report the average performance across all trials. Both linear and MLP architectures are used as base models, with MLP hidden-layer dimensions following the configuration in [13]. For fairness, all baseline methods use the same architecture as our model. Training is performed with the Adam optimizer using a batch size of 256 for 200 epochs [16]. Learning rate and weight decay are selected from $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ based on validation results. All experiments are implemented in PyTorch [17] and executed on an NVIDIA RTX 3090 Ti GPU.

Table 1 reports for the different datasets, metrics, models, s, and q the results of classical GDF MLCLL method against our proposal. From Table 1, it can be observed that our method consistently outperforms the baseline in which the constraints are not exploited. The performance gains are particularly pronounced when the available information or the model capacity is limited. Specifically, the improvements are more evident: (i) for linear models compared to the non-linear one, (ii) when s is smaller, i.e., when fewer training samples are available, and (iii) when p is smaller, meaning that the complementary labels provide less informative supervision to the algorithm.

5 Conclusions

This work explored the previously unaddressed problem of learning from complementary labels with hard logical constraints in multi-label classification. We

[1] `https://mulan.sourceforge.net/datasets-mlc.html`
[2] `https://github.com/atatomir/CCN`

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 1: Results of each comparing approaches (mean±std. deviation), where ↑ / ↓ indicates the larger/smaller the value, the better the performance. Best results are shown in boldface.

Dataset	s%	q%	Linear GDF	Linear Ours	MLP GDF	MLP Ours	Linear GDF	Linear Ours	MLP GDF	MLP Ours
			one error↓				coverage↓			
arts	10	10	0.929±0.007	**0.895±0.015**	0.866±0.037	**0.842±0.032**	0.547±0.012	**0.475±0.012**	0.458±0.038	**0.378±0.018**
		50	**0.877±0.011**	0.890±0.010	0.768±0.022	**0.757±0.017**	0.477±0.010	**0.435±0.011**	0.356±0.018	**0.321±0.017**
	100	10	**0.816±0.015**	0.852±0.015	**0.759±0.019**	0.811±0.039	0.459±0.009	**0.433±0.011**	0.378±0.016	**0.359±0.025**
		50	0.698±0.011	**0.653±0.011**	**0.564±0.008**	0.618±0.009	0.374±0.008	**0.309±0.006**	0.274±0.008	**0.258±0.006**
business	10	10	0.947±0.009	**0.823±0.010**	**0.532±0.098**	0.619±0.131	0.571±0.013	**0.297±0.009**	**0.258±0.024**	0.271±0.039
		50	0.912±0.007	**0.808±0.024**	0.133±0.006	**0.132±0.006**	0.507±0.008	**0.279±0.009**	0.170±0.019	**0.151±0.019**
	100	10	0.941±0.010	**0.696±0.025**	**0.132±0.006**	0.569±0.289	0.562±0.012	**0.235±0.008**	**0.174±0.011**	0.200±0.025
		50	**0.221±0.009**	0.366±0.021	0.133±0.006	0.133±0.006	0.206±0.009	**0.152±0.007**	0.125±0.007	**0.113±0.006**
cal500	10	10	0.850±0.023	**0.758±0.083**	0.699±0.092	**0.689±0.053**	0.959±0.004	**0.945±0.006**	0.936±0.006	**0.933±0.006**
		50	0.793±0.031	**0.783±0.037**	0.193±0.055	**0.185±0.053**	0.944±0.003	**0.938±0.004**	0.895±0.013	0.895±0.013
	100	10	0.850±0.027	**0.553±0.061**	**0.150±0.064**	0.320±0.174	0.952±0.008	**0.940±0.007**	0.910±0.011	0.910±0.012
		50	0.507±0.049	**0.440±0.028**	0.164±0.056	**0.148±0.037**	0.919±0.006	**0.911±0.008**	0.844±0.014	0.844±0.014
enron	10	10	0.898±0.049	**0.744±0.092**	0.694±0.111	**0.637±0.076**	0.655±0.032	**0.618±0.045**	0.543±0.043	**0.535±0.036**
		50	0.801±0.014	**0.723±0.052**	**0.389±0.034**	0.479±0.025	0.594±0.014	**0.557±0.016**	0.437±0.026	**0.425±0.024**
	100	10	0.868±0.044	**0.623±0.048**	**0.382±0.040**	0.494±0.020	0.637±0.048	**0.586±0.026**	0.464±0.018	**0.456±0.016**
		50	0.636±0.049	**0.505±0.043**	**0.237±0.016**	0.475±0.020	0.531±0.014	**0.447±0.012**	**0.363±0.017**	0.364±0.013
genbase	10	10	0.870±0.068	**0.858±0.041**	**0.830±0.087**	0.916±0.050	**0.307±0.044**	0.419±0.031	**0.347±0.039**	0.463±0.047
		50	**0.531±0.069**	0.640±0.100	**0.468±0.084**	0.619±0.093	**0.150±0.022**	0.190±0.024	**0.126±0.021**	0.196±0.027
	100	10	**0.514±0.033**	0.870±0.032	**0.341±0.073**	0.891±0.044	**0.135±0.020**	0.361±0.025	**0.082±0.013**	0.437±0.021
		50	0.083±0.030	**0.106±0.026**	**0.025±0.015**	0.211±0.028	**0.031±0.008**	0.042±0.010	**0.023±0.007**	0.059±0.007
image	10	10	**0.617±0.038**	0.622±0.039	0.576±0.034	**0.565±0.033**	**0.344±0.020**	0.354±0.024	0.337±0.025	**0.335±0.022**
		50	0.449±0.030	**0.446±0.027**	0.426±0.027	**0.424±0.025**	0.258±0.013	**0.254±0.012**	0.243±0.017	0.243±0.016
	100	10	0.497±0.015	**0.473±0.012**	0.381±0.016	**0.375±0.016**	0.271±0.009	**0.267±0.009**	0.226±0.009	**0.221±0.006**
		50	0.345±0.019	**0.343±0.019**	0.286±0.022	**0.283±0.025**	**0.202±0.008**	0.205±0.009	**0.178±0.010**	0.179±0.009
rcv1subset1	10	10	0.963±0.005	**0.919±0.022**	0.942±0.024	**0.820±0.037**	0.663±0.009	**0.570±0.011**	0.618±0.029	**0.545±0.011**
		50	0.907±0.009	**0.897±0.014**	**0.790±0.021**	0.793±0.023	0.566±0.013	**0.492±0.013**	0.489±0.020	**0.467±0.014**
	100	10	0.946±0.004	**0.858±0.027**	0.804±0.012	**0.759±0.014**	0.622±0.005	**0.520±0.027**	0.484±0.007	**0.410±0.015**
		50	0.725±0.012	**0.633±0.021**	**0.556±0.006**	0.674±0.011	0.422±0.007	**0.308±0.007**	**0.330±0.009**	0.341±0.011
rcv1subset2	10	10	0.967±0.006	**0.929±0.019**	**0.905±0.042**	0.936±0.018	0.630±0.016	**0.543±0.015**	0.573±0.024	**0.527±0.018**
		50	0.920±0.007	**0.911±0.007**	**0.767±0.019**	0.840±0.018	0.542±0.011	**0.462±0.013**	0.463±0.014	**0.450±0.030**
	100	10	0.926±0.006	**0.904±0.020**	**0.777±0.013**	0.821±0.029	0.571±0.003	**0.488±0.024**	**0.450±0.013**	0.510±0.010
		50	0.765±0.016	**0.724±0.023**	**0.545±0.012**	0.684±0.014	0.439±0.024	**0.307±0.011**	**0.317±0.009**	0.336±0.009
rcv1subset3	10	10	0.966±0.004	**0.919±0.016**	0.927±0.034	**0.882±0.030**	0.632±0.009	**0.558±0.013**	0.580±0.023	**0.515±0.027**
		50	0.913±0.011	**0.901±0.017**	**0.757±0.026**	0.822±0.040	0.518±0.012	**0.455±0.013**	0.446±0.015	**0.418±0.016**
	100	10	0.922±0.005	**0.893±0.014**	**0.756±0.025**	0.862±0.011	0.567±0.005	**0.516±0.027**	**0.443±0.007**	0.466±0.006
		50	0.721±0.014	**0.625±0.043**	**0.539±0.014**	0.657±0.018	0.468±0.005	**0.301±0.006**	**0.302±0.006**	0.333±0.006
rcv1subset4	10	10	**0.968±0.006**	0.977±0.006	**0.896±0.054**	0.972±0.007	**0.587±0.012**	0.616±0.011	**0.513±0.033**	0.625±0.023
		50	0.917±0.012	**0.901±0.013**	**0.711±0.015**	0.931±0.022	0.485±0.016	**0.471±0.012**	**0.397±0.011**	0.458±0.015
	100	10	**0.912±0.007**	0.918±0.007	**0.753±0.012**	0.969±0.012	**0.518±0.002**	0.536±0.003	**0.390±0.010**	0.595±0.025
		50	**0.608±0.005**	0.727±0.031	**0.470±0.009**	0.959±0.007	0.393±0.007	**0.292±0.009**	**0.254±0.006**	0.303±0.003
rcv1subset5	10	10	0.968±0.005	**0.967±0.011**	**0.908±0.030**	0.957±0.005	**0.627±0.009**	0.634±0.013	**0.550±0.024**	0.578±0.022
		50	0.915±0.012	**0.911±0.018**	**0.761±0.037**	0.912±0.029	0.524±0.016	**0.507±0.018**	**0.452±0.013**	0.486±0.024
	100	10	**0.929±0.009**	0.936±0.007	**0.713±0.024**	0.962±0.005	**0.578±0.007**	0.582±0.008	0.438±0.010	**0.415±0.008**
		50	**0.709±0.007**	0.747±0.032	**0.498±0.007**	0.817±0.005	0.461±0.007	**0.327±0.014**	**0.292±0.007**	0.430±0.008
scene	10	10	0.718±0.034	**0.698±0.027**	**0.664±0.072**	0.661±0.063	0.340±0.027	**0.316±0.028**	0.294±0.043	**0.293±0.043**
		50	0.553±0.038	**0.512±0.030**	0.440±0.033	**0.438±0.033**	0.225±0.022	**0.205±0.018**	0.171±0.017	**0.170±0.017**
	100	10	0.623±0.026	**0.582±0.017**	0.389±0.019	**0.378±0.015**	0.276±0.013	**0.247±0.012**	0.163±0.007	**0.160±0.008**
		50	0.394±0.019	**0.381±0.019**	0.287±0.010	0.287±0.010	0.156±0.010	**0.151±0.009**	0.111±0.006	0.111±0.006
yeast	10	10	0.683±0.031	**0.524±0.032**	**0.251±0.002**	0.294±0.087	0.746±0.016	**0.664±0.011**	**0.621±0.029**	0.633±0.018
		50	0.559±0.023	**0.436±0.025**	0.252±0.002	0.252±0.003	0.665±0.008	**0.591±0.009**	**0.521±0.013**	0.524±0.016
	100	10	**0.327±0.011**	0.391±0.030	**0.248±0.004**	0.251±0.001	**0.616±0.004**	0.628±0.004	**0.536±0.008**	0.605±0.015
		50	0.263±0.006	**0.242±0.004**	**0.240±0.004**	0.241±0.002	0.521±0.002	**0.489±0.002**	0.479±0.004	**0.473±0.002**
			ranking loss↓				average precision↑			
arts	10	10	0.476±0.012	**0.409±0.014**	0.382±0.039	**0.305±0.019**	0.182±0.007	**0.216±0.013**	0.252±0.035	**0.269±0.023**
		50	0.403±0.010	**0.372±0.010**	0.278±0.015	**0.247±0.016**	**0.240±0.009**	0.239±0.007	0.360±0.022	**0.374±0.019**
	100	10	0.380±0.008	**0.358±0.012**	0.299±0.016	**0.280±0.022**	**0.289±0.010**	0.265±0.013	**0.351±0.018**	0.322±0.025
		50	0.295±0.006	**0.239±0.004**	0.197±0.006	**0.189±0.005**	0.390±0.008	**0.438±0.009**	**0.518±0.007**	0.485±0.006
business	10	10	0.502±0.014	**0.236±0.007**	**0.173±0.020**	0.218±0.043	0.149±0.009	**0.293±0.010**	**0.540±0.060**	0.317±0.049
		50	0.441±0.011	**0.222±0.008**	0.097±0.011	**0.083±0.010**	0.191±0.008	**0.314±0.020**	0.818±0.018	**0.836±0.014**
	100	10	0.493±0.013	**0.170±0.006**	**0.100±0.007**	0.131±0.029	0.152±0.009	**0.414±0.019**	**0.803±0.010**	0.501±0.187
		50	0.129±0.006	**0.090±0.005**	0.068±0.004	**0.061±0.004**	**0.728±0.009**	0.701±0.012	0.847±0.007	**0.851±0.006**
cal500	10	10	0.494±0.007	**0.457±0.008**	0.396±0.019	**0.393±0.019**	0.174±0.005	**0.195±0.007**	**0.242±0.023**	0.240±0.018
		50	0.454±0.008	**0.420±0.009**	0.284±0.007	**0.281±0.006**	0.202±0.006	**0.221±0.009**	0.392±0.012	**0.401±0.010**
	100	10	0.483±0.007	**0.426±0.005**	0.307±0.015	**0.304±0.011**	0.179±0.004	**0.232±0.007**	**0.368±0.016**	0.367±0.010
		50	0.353±0.005	**0.331±0.005**	0.216±0.006	0.216±0.007	0.298±0.006	**0.327±0.009**	0.472±0.008	0.472±0.009
enron	10	10	0.421±0.030	**0.362±0.034**	0.300±0.033	**0.280±0.028**	0.176±0.043	**0.222±0.019**	0.286±0.049	0.286±0.043
		50	0.352±0.010	**0.306±0.013**	0.187±0.014	**0.180±0.014**	0.252±0.015	**0.263±0.021**	**0.499±0.023**	0.455±0.015
	100	10	0.405±0.040	**0.332±0.023**	**0.206±0.009**	0.214±0.009	0.196±0.030	**0.250±0.014**	**0.487±0.014**	0.346±0.009
		50	0.290±0.009	**0.209±0.007**	**0.136±0.008**	0.140±0.007	0.348±0.025	**0.420±0.019**	**0.633±0.012**	0.538±0.012
genbase	10	10	**0.285±0.040**	0.410±0.030	**0.330±0.046**	0.461±0.049	0.280±0.058	**0.227±0.032**	**0.299±0.063**	0.141±0.025
		50	**0.127±0.021**	0.179±0.023	**0.103±0.020**	0.182±0.029	**0.584±0.051**	0.432±0.050	**0.640±0.055**	0.469±0.070
	100	10	**0.115±0.019**	0.354±0.028	**0.059±0.011**	0.434±0.023	**0.612±0.027**	0.234±0.026	**0.756±0.047**	0.163±0.023
		50	**0.017±0.005**	0.029±0.008	**0.010±0.005**	0.042±0.006	**0.932±0.019**	0.891±0.017	**0.972±0.013**	0.833±0.016
image	10	10	**0.364±0.027**	0.378±0.034	0.356±0.033	**0.351±0.029**	**0.593±0.025**	0.586±0.028	0.613±0.025	**0.618±0.024**
		50	0.255±0.018	**0.250±0.016**	0.238±0.021	**0.236±0.019**	0.705±0.019	**0.709±0.016**	0.722±0.016	**0.723±0.015**
	100	10	0.274±0.008	**0.269±0.008**	0.214±0.010	**0.209±0.007**	0.681±0.010	**0.692±0.009**	0.749±0.010	**0.754±0.009**
		50	**0.185±0.010**	0.188±0.010	**0.156±0.012**	0.157±0.011	**0.777±0.012**	0.776±0.012	0.813±0.014	**0.814±0.015**
rcv1subset1	10	10	0.460±0.010	**0.377±0.010**	0.414±0.027	**0.325±0.012**	0.081±0.004	**0.121±0.012**	0.100±0.014	**0.188±0.015**
		50	0.367±0.010	**0.318±0.010**	0.288±0.016	**0.260±0.009**	0.131±0.006	**0.152±0.010**	0.212±0.015	**0.227±0.009**
	100	10	0.420±0.005	**0.312±0.017**	0.283±0.004	**0.225±0.009**	0.097±0.003	**0.161±0.020**	0.204±0.008	**0.224±0.007**
		50	0.247±0.005	**0.162±0.004**	**0.165±0.005**	0.177±0.006	0.265±0.006	**0.372±0.010**	**0.406±0.006**	0.298±0.006
rcv1subset2	10	10	0.452±0.015	**0.369±0.013**	0.385±0.029	**0.321±0.011**	0.080±0.006	**0.116±0.011**	0.130±0.034	**0.140±0.005**
		50	0.367±0.010	**0.309±0.008**	0.278±0.012	**0.259±0.016**	0.126±0.008	**0.153±0.006**	**0.240±0.017**	0.212±0.006
	100	10	0.393±0.003	**0.308±0.019**	**0.269±0.008**	0.305±0.011	0.121±0.004	**0.136±0.007**	**0.240±0.010**	0.173±0.008
		50	0.274±0.018	**0.168±0.006**	**0.164±0.005**	0.175±0.004	0.256±0.007	**0.348±0.014**	**0.425±0.008**	0.332±0.008
rcv1subset3	10	10	0.455±0.008	**0.375±0.011**	0.393±0.025	**0.305±0.020**	0.078±0.003	**0.118±0.006**	0.117±0.028	**0.161±0.015**
		50	0.345±0.010	**0.300±0.010**	0.266±0.012	**0.229±0.009**	0.135±0.010	**0.158±0.013**	**0.258±0.020**	0.237±0.011
	100	10	0.390±0.004	**0.331±0.023**	0.263±0.006	0.271±0.005	0.124±0.005	**0.145±0.006**	**0.254±0.016**	0.170±0.005
		50	0.295±0.005	**0.161±0.003**	**0.153±0.003**	0.170±0.003	0.274±0.009	**0.394±0.020**	**0.443±0.009**	0.352±0.008
rcv1subset4	10	10	**0.445±0.010**	0.479±0.008	**0.359±0.036**	0.484±0.025	0.078±0.005	**0.064±0.004**	**0.152±0.042**	0.067±0.010
		50	0.345±0.015	**0.333±0.011**	**0.237±0.014**	0.319±0.010	0.139±0.011	**0.154±0.012**	**0.326±0.013**	0.144±0.018
	100	10	**0.371±0.002**	0.388±0.003	**0.240±0.007**	0.454±0.026	**0.137±0.005**	0.128±0.005	**0.283±0.011**	0.073±0.014
		50	0.251±0.005	**0.170±0.008**	**0.129±0.004**	0.187±0.002	**0.375±0.005**	0.348±0.020	**0.510±0.006**	0.207±0.003
rcv1subset5	10	10	**0.451±0.011**	0.459±0.013	**0.363±0.031**	0.401±0.026	**0.079±0.004**	0.076±0.005	**0.135±0.032**	0.091±0.005
		50	0.352±0.015	**0.338±0.018**	**0.263±0.011**	0.315±0.025	0.133±0.012	**0.145±0.016**	**0.247±0.023**	0.148±0.013
	100	10	**0.397±0.006**	0.403±0.007	**0.248±0.007**	0.254±0.006	**0.122±0.006**	0.115±0.006	**0.274±0.019**	0.111±0.002
		50	0.288±0.004	**0.183±0.008**	**0.145±0.003**	0.269±0.004	0.280±0.004	**0.329±0.018**	**0.464±0.005**	0.227±0.003
scene	10	10	0.385±0.033	**0.358±0.033**	0.331±0.051	**0.330±0.052**	0.510±0.029	**0.531±0.027**	0.558±0.053	**0.560±0.049**
		50	0.248±0.026	**0.225±0.021**	0.183±0.021	**0.182±0.020**	0.643±0.028	**0.671±0.023**	0.722±0.022	**0.723±0.022**
	100	10	0.308±0.016	**0.273±0.015**	0.171±0.009	**0.168±0.010**	0.585±0.017	**0.618±0.013**	0.749±0.011	**0.755±0.011**
		50	0.165±0.012	**0.159±0.011**	**0.111±0.007**	0.112±0.006	0.749±0.013	**0.758±0.013**	0.822±0.007	0.822±0.006
yeast	10	10	0.469±0.016	**0.372±0.014**	**0.270±0.019**	0.287±0.019	0.443±0.013	**0.539±0.017**	**0.667±0.013**	0.643±0.037
		50	0.378±0.010	**0.301±0.011**	**0.214±0.006**	0.218±0.009	0.529±0.009	**0.610±0.009**	**0.713±0.005**	0.712±0.007
	100	10	**0.283±0.003**	0.317±0.005	**0.211±0.004**	0.254±0.008	**0.641±0.005**	0.607±0.012	**0.719±0.003**	0.678±0.007
		50	0.213±0.001	**0.194±0.001**	**0.182±0.002**	0.183±0.001	0.716±0.002	**0.740±0.002**	**0.748±0.001**	0.747±0.002

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

introduced methods that leverage constraints not only as requirements to be satisfied but also as an additional source of weak supervision. This perspective allows relationships among labels to compensate for the limited information carried by complementary labels, ultimately improving learning effectiveness.

Across multiple datasets and experimental settings, our methods consistently demonstrated that integrating logical constraints with complementary-label learning yields clear benefits. These results validate our claim that constraints can help close the supervision gap and enhance model performance, even in low-resource scenarios.

Future research may expand this line of work by integrating other complementary label approaches, such as CTL [8], or by incorporating hard constraints through methods like CCN+ [14]. It would also be valuable to assess their effectiveness on additional datasets and model families.

References

[1] M. L. Zhang and Z. H. Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014.
[2] M. L. Zhang and L. Wu. Lift: Multi-label learning with label-specific features. *IEEE transactions on pattern analysis and machine intelligence*, 37(1):107–120, 2014.
[3] G. Holste, Y. Zhou, S. Wang, A. Jaiswal, and Others. Towards long-tailed, multi-label disease classification from chest x-ray: Overview of the cxr-lt challenge. *Medical Image Analysis*, 97:103224, 2024.
[4] J. Xie, X. Li, Y. Yuan, Y. Guan, J. Jiang, X. Guo, and X. Peng. Knowledge-based dynamic prompt learning for multi-label disease diagnosis. *Knowledge-Based Systems*, 286:111395, 2024.
[5] Z. Guo, B. Dong, Z. Ji, J. Bai, Y. Guo, and W. Zuo. Texts as images in prompt tuning for multi-label image recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
[6] L. Ma, H. Xie, L. Wang, Y. Fu, D. Sun, and H. Zhao. Text-region matching for multi-label image recognition with missing labels. In *ACM International Conference on Multimedia*, 2024.
[7] S. S. Mohanrasu, L. T. Phan, R. Rajan, and Balachandran Manavalan. Cost-sensitive feature selection for multi-label classification: multi-criteria decision-making approach. *Applied Computing and Informatics*, 2025.
[8] Y. Gao, J. Y. Zhu, M. Xu, and M. L. Zhang. Multi-label learning with multiple complementary labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(9):8013–8024, 2025.
[9] Y. Gao, M. Xu, and M. L. Zhang. Complementary to multiple labels: A correlation-aware correction approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9179–9191, 2024.
[10] Y. Gao, M. Xu, and M.-L. Zhang. Unbiased risk estimator to multi-labeled complementary label learning. In *International Joint Conference on Artificial Intelligence*, 2023.
[11] T. Ishida, G. Niu, W-H. Hu, and Masashi Sugiyama. Learning from complementary labels. In *Neural Information Processing Systems 30*, 2017.
[12] Y. Gao and M.-L. Zhang. Discriminative complementary-label learning with weighted loss. In *International Conference on Machine Learning*, 2021.
[13] E. Giunchiglia and T. Lukasiewicz. Multi-label classification neural networks with hard logical constraints. *Journal of Artificial Intelligence Research*, 72:759–818, 2021.
[14] E. Giunchiglia, A. Tatomir, M. C. Stoian, and T. Lukasiewicz. Ccn+: A neuro-symbolic framework for deep learning with requirements. *International Journal of Approximate Reasoning*, 171:109124, 2024.
[15] N. Hoernle, R.-M Karampatsis, V. Belle, and K. Gal. Multiplexnet: Towards fully satisfied logical constraints in neural networks. In *AAAI Conference on Artificial Intelligence*, 2022.
[16] C. M. Bishop and H. Bishop. *Deep learning: Foundations and concepts*. Springer Nature, 2023.
[17] A. Paszke, S. Gross, F. Massa, A. Lerer, and Others. Pytorch: An imperative style, high-performance deep learning library. 2019.