

Deep Learning for Multi-Label Learning

A Comprehensive Survey

Adane Nega Tarekegn, Mohib Ullah, *Member, IEEE*, Faouzi Alaya Cheikh, *Senior Member, IEEE*

Abstract—Multi-label learning is a rapidly growing research area that aims to predict multiple labels from a single input data point. In the era of big data, tasks involving multi-label classification (MLC) or ranking present significant and intricate challenges, capturing considerable attention in diverse domains. Inherent difficulties in MLC include dealing with high-dimensional data, addressing label correlations, and handling partial labels, for which conventional methods prove ineffective. Recent years have witnessed a notable increase in adopting deep learning (DL) techniques to address these challenges more effectively in MLC. Notably, there is a burgeoning effort to harness the robust learning capabilities of DL for improved modelling of label dependencies and other challenges in MLC. However, it is noteworthy that comprehensive studies specifically dedicated to DL for multi-label learning are limited. Thus, this survey aims to thoroughly review recent progress in DL for multi-label learning, along with a summary of open research problems in MLC. The review consolidates existing research efforts in DL for MLC, including deep neural networks, transformers, autoencoders, and convolutional and recurrent architectures. Finally, the study presents a comparative analysis of the existing methods to provide insightful observations and stimulate future research directions in this domain.

Index Terms— Multi-label learning, deep learning for MLC, multi-label classification, deep learning for MLL, transformers for multi-label classification, multi-label challenges.

I. INTRODUCTION

IN many real-world applications, an object may be associated with multiple labels concurrently, and such problems are recognized as multi-label learning (MLL) [1]. MLL is an extension of the standard single-label learning paradigm, where there is typically a finite set of potential labels that can be applied to the instances of multi-label data (MLD). The basic goal is to simultaneously predict a vector of outputs for a given single input, which means that it is possible to solve

more complex decision-making problems. This is opposed to the single-label classification, where each instance is associated with only one label. In the context of multi-label tasks, an instance is typically associated with a set of labels, constituting distinct combinations known as relevant labels (active labels), while labels not linked to the instance are termed irrelevant labels. Both the relevant and irrelevant labels are represented as a binary vector with its size aligning with the total number of labels in the MLD. Depending on the goal, there exist two major tasks in MLL: multi-label classification (MLC) and multi-label ranking (MLR) [2]. MLC is the main learning task concerned with learning a model that outputs a bipartition of the set of labels into relevant and irrelevant with respect to a query instance. MLR, on the other hand, is concerned with learning a model that outputs an ordering of the class labels according to their relevance to a query instance.

Although MLC applications traditionally concentrate on text analysis, multimedia, and biology, their significance is progressively growing across diverse domains, such as document classification [3][4][5], healthcare [6][7][8], environmental modelling [9][10], emotion recognition [11][12], commerce [13][14], social media [15][16][17], and more. Many other demanding applications, such as video annotation, web page categorization, and language modelling, can also derive benefits from being formulated as MLC tasks involving hundreds, thousands, or even millions of labels. Such extensive label space presents research challenges, such as issues related to data sparsity and scalability. MLC has additional complexities, including modelling label correlations [18][19], imbalanced labels [20] and nosily labels [21]. Traditional MLC methods, such as problem transformation and algorithm adaptation [22][23], demonstrate suboptimal performance in addressing these challenges.

Apart from the traditional approaches, deep learning (DL) techniques have gained increasing popularity in addressing the challenges of MLC. The formidable learning capacities of deep learning are particularly adaptable for addressing MLC challenges, as demonstrated by their notable success in addressing single-label classification tasks. Currently, a predominant trend in MLC involves extensively incorporating DL techniques even for more challenging problems, such as Extreme MLC [24][25][26], imbalanced MLC [27][28], weakly supervised MLC [29][30][31], and MLC with missing labels [32][33]. Effectively harnessing the strong learning capabilities of DL is crucial to better understand and model the label

The manuscript was submitted on Jan 24, 2024; revised dd, mm, yyyy. This work is supported by the European Union through the Horizon 2020 Research and Innovation Programme, in the context of the ALAMEDA project under grant agreement No GA 101017558 and the ADRIATIC project co-funded by the MarTERA partners Romanian Executive Unit for Financing Higher Education, Research, Development, and Innovation, the Scientific and Technological Research Council of Turkey and the Research Council of Norway. (Corresponding author: Adane Nega Tarekegn).

Adane Nega Tarekegn is affiliated with the MediaFutures, University of Bergen, Norway (email: adane.n.tarekegn@ntnu.no). Mohib Ullah, and Faouzi Alaya Cheikh are with the Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway (e-mails: mohib.ullah@ntnu.no, and faouzi.cheikh@ntnu.no).

correlations, thereby enabling DL to tackle MLC problems effectively. Several studies have shown that MLC methods explicitly designed to capture label dependencies typically demonstrate superior predictive performance [34][19]. This paper conducted a concise review of the existing literature to identify a broad range of DL-based techniques for MLC problems to inspire further exploration of innovative DL-based approaches for MLC. Few surveys are already available on traditional approaches for MLC, such as those referenced in [35][23] [36]. Additionally, there are surveys that contain both traditional and DL methods [37][38], but these have limited coverage of DL methods for MLC and are focused on specific domains. However, this paper uniquely concentrates on a range of DL architectures, including recurrent and convolutional networks, transformers, autoencoders, and hybrid models, for addressing MLC challenges across diverse domains. In Fig. 1, we present a taxonomy of multi-label learning methods comprising both traditional approaches and DL methods.

The main contributions can be outlined as follows:

1. To the best of the authors' knowledge, this survey is the first to thoroughly cover DL methods for solving MLC tasks, covering diverse domains and data modalities, including texts, music, images, and videos.
2. A comprehensive summary of the most recent DL methods for MLC across several publicly available datasets is provided (Tables I, II, and III), with a brief overview of each DL method and insightful discussions. Therefore, the survey provides the readers with state-of-the-art methods.
3. We have provided a brief description of the current challenges facing the realm of MLC. Additionally, we have included a summary of the multi-label datasets utilized in MLC, along with the definition of attributes used to evaluate the characteristics of these datasets.
4. Finally, this paper provides a comparative study of existing approaches involving various DL techniques and investigates the pros and cons of each approach (Table V). It offers insights that can guide the selection of suitable techniques and develop better DL approaches for MLC in future studies.

The subsequent sections of the paper are organized as follows. Section II presents the fundamental concepts of multi-label learning. Section III presents the research methodology, which focuses on the data source and search strategy, selection criteria, and statistical trends of publications. Section IV is the main section of this survey, which discusses various DL approaches for addressing the MLC challenges. Section V focuses on open challenges in MLC and datasets. Section VI provides a comparative analysis of solutions with advantages and limitations. Finally, section VII provides the conclusions of the paper.

II. FUNDAMENTAL CONCEPTS OF MLL

MLL is founded on a dataset where instances are associated with several target variables or labels simultaneously. The main goal when working with such data is MLC, which aims to categorize the target variables into relevant and irrelevant

groups for a specific instance. Additional tasks may include ranking the labels according to their relevance or creating a comprehensive joint distribution encompassing all possible assignments of values to the labels. The formal definition of MLL can be presented as follows [39]: Let X be a d -dimensional input space of categorical or numerical features and output space of q labels $L = \{\lambda_1, \lambda_2, \dots, \lambda_q\}$, $q > 1$. A multi-label example can be defined as a pair (\mathbf{x}, Y) where $\mathbf{x} = (x_1, x_2, \dots, x_q) \in X$ and $Y \subseteq L$ is called a label-set. $D = \{(\mathbf{x}_i, Y_i) \mid 1 \leq i \leq m\}$ is an MLD composed of a set of m instances. Let Q be a quality criterion that rewards models with high predictive performance and low complexity. If the task is an MLC, then the aim is to compute a function $h: X \rightarrow 2^L$ such that h maximizes Q . If the task is an MLR, then the goal is to find a function $f: X \times L \rightarrow \mathbb{R}$ such that f maximizes Q , where \mathbb{R} is the ranking of labels for a given sample.

MLC is currently receiving significant attention and is applicable to a variety of research domains, including bioinformatics [40][41], text classification [42][43], music categorization [44][45], medical diagnosis [46][47], image classification [48], and video annotation [49]. For example, in medical diagnosis, a patient can have multiple side effects for a disease, or a medical diagnosis might find a patient suffering from more than one disease at the same time. A real-world image can be annotated with several labels because an image is often associated with rich semantic information, such as objects, parts, scenes, actions, and their interactions or attributes. Modelling the rich semantic information and its dependencies is essential for image understanding. In-text categorization, a news article can cover multiple aspects of an event, thus being assigned to a set of multiple topics. In all these cases, the task is to assign a label set for each unseen instance [50]. In MLR, the goal is to not only predict a vector of outputs from a finite set of predefined labels but also to rank them based on their relevance to the provided input. In a multi-label learning scenario, the task extends beyond predicting relevant and irrelevant labels; it often involves generating a well-ordered ranking of relevant labels (i.e., a list of preferences) from the list of possible labels for each unseen example. MLR is an interesting problem as it subsumes many supervised learning tasks such as multi-label, multi-class, and hierarchical classifications [51]. Document classification is a prominent use case for MLR, involving the categorization of topics (such as technology, politics, and sports) within a collection of documents, such as news articles. It's common for a document to be associated with multiple topics, and the aim of the learning algorithm is to rank (order) the relevant topics higher than non-relevant ones for a given document query.

Two traditional approaches exist for solving the MLL task: problem transformation and algorithm adaptation approaches. The former transforms the MLC task into one or more single-label classification [52], or label ranking (LR) [2] tasks, while the latter aims to adapt or extend the traditional machine learning algorithms to handle an MLD directly [53]. The three most known problem transformation methods are binary relevance (BR) [1], label power-set (LP) [54], and classifier

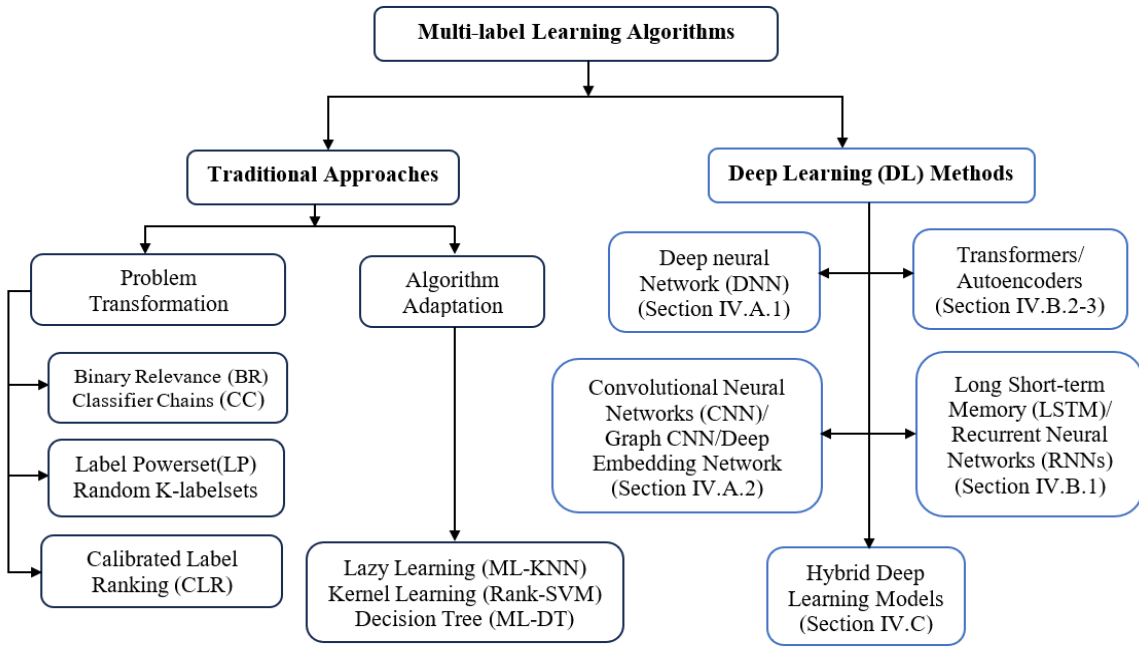


Fig. 1. A taxonomy of multi-label learning methods encompassing both traditional and deep learning methods.

chains (CC) [55]. The BR approach transforms the multi-label problem into a set of independent binary problems. Then, each binary problem is processed by using a traditional classifier. Finally, the individual predictions are combined to get the subset of labels relevant to each test instance. Although BR is relatively simple to implement, it is realized that BR ignores the possible relationship between labels (such as label dependency, cooccurrence, and correlation). To deal with the limitation of the BR method, the classifier chain (CC) [55] was introduced. This method interconnects binary classifiers in a sequential chain, where the predictions of preceding classifiers serve as features for subsequent classifiers. This allows the latter classifiers to leverage the correlation with earlier predictions to enhance the quality of their predictions. LP considers each unique set of labels as a class identifier, transforming the original MLD into a multi-class dataset. After using it to train a regular classifier, the predicted classes are back-transformed into the subsets of labels. Both CC and LP are traditional approaches to learning the interdependency between labels for MLL. However, in addition to high computation cost when dealing with a larger number of labels, CC and LP have limited ability to capture the high-order correlations between labels.

Regardless of the methods to solve the multi-label problem and the approaches to the label correlations, MLC has additional complexities, such as grappling with higher-order dependencies among labels, contending with an extensive array of labels requiring substantial computational resources, and handling scenarios involving partially or weakly supervised MLC as well as imbalanced MLC [20]. Moreover, the classical approaches mentioned earlier prove ineffective in addressing these challenges. Recently, deep learning (DL) techniques have gained increased popularity across diverse disciplines, and MLC has been no exception to benefiting from the latest developments in DL. Thus, this survey aims to provide a

thorough review of the recent progress in DL for MLC that aims to address these challenges and promote the application of DL-based MLC in various domains.

III. METHODOLOGY

This section reveals the search strategy, study selection criteria, and trends in publication to ensure a thorough and objective selection of literary sources.

A. Search Strategy

For this comprehensive review, an exploration of research articles pertaining to DL approaches for MLC, including publications from 2006 to 2023, was conducted. Initially, prominent library databases spanning various research domains were used as primary sources: Springer, IEEEExplore, DBLP, ACM Digital Library, Elsevier, Science Direct, and Google Scholar, among others. Boolean operators were employed to refine searches, combining terms with similar meanings, and delimiting the scope of investigation. Predetermined search keywords, incorporating phrases, such as ‘deep learning for multi-label classification,’ ‘multi-label classification using deep convolutional neural networks (CNN),’ ‘multi-label prediction using recurrent neural network (RNN), transformers, autoencoders’, or ‘hybrid deep learning for multi-label classification,’ were applied. Additionally, efforts were made to identify relevant articles from alternative sources, including peer-reviewed journals and conferences.

B. Selection Criteria

This paper primarily focuses on examining DL techniques for addressing MLC. The following eligibility criteria, which had to be jointly satisfied, were used to select the relevant publications: (1) the publication on MLL is based on MLD using DL; (2) the work adopts or proposes DL methods for solving MLC; (3) experimental results evaluate DL algorithms

using multi-label measures; (4) the publication is a full-text article written in English. Articles that present a proposal of DL approaches for solving MLC are selected for review without restrictions on the dates of publication. A total of 382 publications were initially gathered and identified during the search process. Out of these, 64 duplicates were identified, and 106 were excluded after screening titles and abstracts. Subsequently, a thorough examination of the full text of each paper was conducted, resulting in the identification of 212 relevant papers for inclusion in this study. Additionally, any publications with duplicated titles, abstracts, or content were carefully removed, ensuring the retention of only one copy of each publication in the final selection.

C. Publication Trends

Approximately 15 years ago, the landscape of MLC started to attract researchers, marking the emergence of this dynamic field as a compelling research topic. Fig. 2 shows an increasing trend of publications related to DL-based MLC, spanning from 2006 to 2023. Notably, the number of publications demonstrated consistent growth from 2012 to 2023. In 2019, there was a slight dip in publications compared to 2018; however, subsequent years displayed an upward trajectory. Particularly in recent times, the volume of published works on MLC employing DL techniques has significantly surpassed that of previous years. This observation underscores the continued exploration of innovative DL techniques to tackle MLC tasks as a noteworthy and actively researched area, attracting substantial attention and interest from the research community.

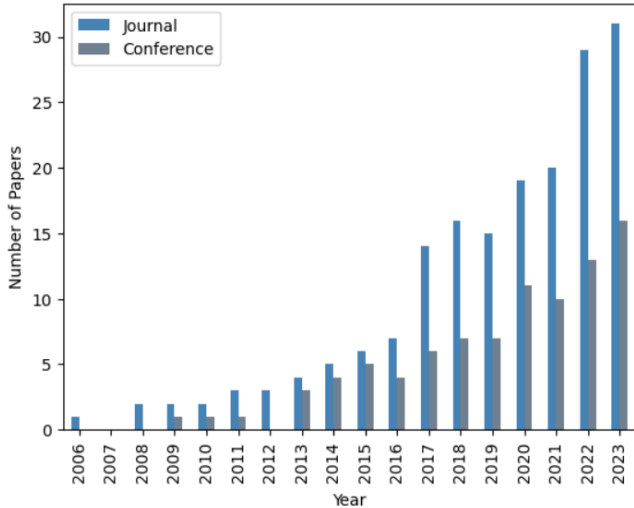


Fig. 2. A summary of the number of papers related to the topic of MLC utilizing DL methods according to Google Scholar and other sources from 2006-2023.

IV. DEEP LEARNING FOR MLC

Recent advances in DL have significantly enriched the landscape of MLC. DL architectures play a pivotal role in generating embedding representations for both input features and output space. The formidable learning capabilities of DL find widespread application in MLC tasks in various domains, such as images, textual, music, and videos. The most frequently used DL methods for MLC include deep neural networks, convolutional, recurrent, autoencoder and transformer architectures, and hybrid

models. Effectively harnessing the advantages of these DL approaches is crucial in addressing label dependencies and other challenges in MLC. This section provides an overview of these prominent DL methods for MLC with an overview and detailed examination of each technique specifically tailored for MLC.

A. Neural Networks for MLC

This section provides an in-depth exploration of deep neural networks (DNN) and convolutional neural networks (CNN) for MLC, along with a summary of the most recent DL methods, applications, and datasets in MLC.

1) Deep Neural Networks for MLC

Deep neural networks (DNNs) have been employed to address MLC problems, and the simplest approach is to decompose the MLC problem into several sets of binary classification problems, one for each label. However, when dealing with a large number of labels, this solution lacks scalability. Additionally, it considers missing labels as negatives, resulting in a performance decline, and ignores dependencies among labels, which is an important aspect of effective recognition. Therefore, a different approach that concentrates on the question of label dependencies needs to be exploited. One such approach is BP-MLL (Backpropagation for Multi-label Learning) [56], which formulates MLC problems as a neural network with multiple output nodes, one for each label. BP-MLL is applied to an MLC with sigmoidal neurons with one hidden layer and additional biases from the input and hidden layer. The input layer size corresponds to the total number of available features (plus a bias neuron). Let $X = \mathbb{R}^d$ denote a d -dimensional input space and $L = \{\lambda_1, \lambda_2, \dots, \lambda_q\}$, $q > 1$ be an output space of q labels, $D = \{(\mathbf{x}_i, Y_i) \mid 1 \leq i \leq m\}$ is an MLD comprising a set of m instances, where each instance $\mathbf{x}_i \in X$ is represented as a d -dimensional feature vector and a set of q labels associated with the feature vector and $Y_i \subseteq L$ is a label-set. A multi-layer perceptron (MLP) NN, in Fig. 3, can be built up to learn a model that has d input neurons that correspond to a d -dimension feature vector, while the last Q neurons represent a combination of output labels. The output layer size is equivalent to the number of labels. There are $d \times n$ weights (W_{ih} , $1 \leq i \leq d, 1 \leq h \leq n$) between the first two layers and $n \times q$ weights (W_{ho} , $1 \leq h \leq n, 1 \leq o \leq q$) between the latter two layers. The bias parameters are represented as I and H . Because the task of MLL is to predict labels of test samples, it needs to evaluate the global error of the model as:

$$E = \sum_{i=1}^m E_i \quad (1)$$

E_i is the error on the sample \mathbf{x}_i , which can be defined as:

$$E = \sum_{j=1}^q (c_j^i - d_j^i)^2 \quad (2)$$

where $c_j^i = c_j(\mathbf{x}_i)$ is the predicted j^{th} label on sample \mathbf{x}_i , and d_j^i is the actual j^{th} label of sample \mathbf{x}_i . The actual label has a value of either $+1 (j \in Y)$ or $-1 (j \notin Y)$.

Different learning algorithms are applicable for acquiring a model from training data. The backpropagation algorithm, for instance, is employed to learn from errors. However, the algorithm could be improper for MLL because the error

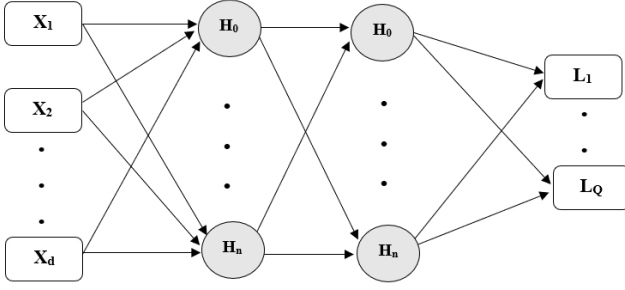


Fig. 3. Structure of the neural network for MLC

function (2) neglects the correlations among labels of a sample. In the original BP algorithm, the error function (2) limits individual label discrimination, whether a specific label $j \in L$ belongs to the sample x_i or not.

It should be taken into consideration that labels in Y_i are more important than those outside of Y_i . BP-MLL views each output node as a binary classification problem, and training is based on the classical BP algorithm, but in order to address the dependencies across labels, the new global error function is proposed that relies on the output layer to minimize pairwise ranking errors.

$$E = \sum_{i=1}^m E_i = \sum_{i=1}^m \frac{1}{|Y_i| |\hat{Y}_i|} \sum_{(k,l) \in Y_i \times \hat{Y}_i} \exp(-(C_k^i - C_l^i)) \quad (3)$$

In the error function (3), the i^{th} errors on the i^{th} sample (x_i, Y_i) are accumulated. \hat{Y}_i is a complementary set of Y_i in L and $|\cdot|$ computes the cardinality of a set. The item, $C_k^i - C_l^i$, represents the difference between the predicted labels and the actual labels. The sum of the i^{th} error term accumulates the difference between outputs of a pair of labels, which are the ones belonging to the sample and the other ones not belonging to it. The sum is normalized by the numbers of all pairs, $|Y_i| |\hat{Y}_i|$. Then, the correlations between the pair of labels are computed. In the previous statements, the error function calculates the difference between output labels. The task of learning is minimizing the error function via enlarging the output values of labels belonging to the training samples and diminishing the output values of the labels not belonging to it. If the training data can cover the distribution of the whole sample space, the model can learn it by minimizing the error function by feeding training samples. The next step to achieve the MLC classifier is to determine the set of labels belonging to the input instance. This information can be retrieved from the neural network output values through the threshold function, which depends on the input vector. If the output neuron value is higher than the threshold value, then the corresponding label belongs to the input instance. Otherwise, the label does not belong to the instance.

BP-MLL was among the first to utilize neural network architectures to solve MLC, which is supposed to perform better in multi-label problems since it takes label correlations into consideration compared to the standard form of neural network, which does not. However, it is found that BP-MLL requires more computational complexity and convergence speed as the number of labels grows. As a result, it was later

extended by state-of-the-art learning techniques. The authors in [57] proposed an improvement to the BP-MLL method by modifying the global error function. This modified error function allows the threshold value to be determined automatically by adaptation during neural network learning instead of using an additional step in BP-MLL to define the threshold function. Furthermore, [58] found the suboptimal performance of BP-MLL on textual datasets. In response to this limitation, [58] explored the constraints of BP-MLL by substituting ranking loss minimization with the more commonly employed cross-entropy error function. The authors show how a single hidden layer neural network can achieve state-of-the-art performance in large-scale multi-label text classification tasks through the use of recent developments in the area of DL, such as rectified linear units (ReLU), Dropout, and AdaGrad.

In a different study documented in [59], a label-decision module was integrated into DNNs, resulting in the attainment of state-of-the-art performance in multi-label image classification tasks. Building upon this framework, Du et al. [60] introduced ML-Net, a DNN designed for the MLC of biomedical texts. ML-Net adopts the label-decision module from [59], but it converts the framework from image processing to text classification. The ML-Net model integrates label prediction and decision-making within the same network, enabling the determination of output labels based on both label confidence scores and document context. Its objective is to minimize pairwise ranking errors of labels, allowing for end-to-end training and prediction of the label set without requiring an additional step for determining output labels.

Recently, [61] proposed a new loss for MLC, named ZLPR (Zero-bounded Log-sum-exp & Pairwise Rank-based) loss, to extend the application of DL in MLC. The authors extended the cross-entropy loss from the single-label classification, which is expressed in Eq. (4).

$$Loss_{zlpr} = \log \left(1 + \sum_{i \in \Omega_{pos}} e^{-s_i} \right) + \log \left(1 + \sum_{i \in \Omega_{neg}} e^{s_j} \right) \quad (4)$$

where Ω_{pos} is the label set and $\Omega_{neg} = L/\Omega_{pos}$, s_i is the model output score of the i_{th} category (λ_i). In contrast to earlier ranking-based losses, ZLPR exhibits the capability to dynamically determine the number of target categories while enhancing a model's label-ranking proficiency. In comparison to certain binary losses, the ZLPR loss excels in capturing more robust label dependencies and elucidating the ranking relationship between positive and negative categories.

Final Notes: DNNs have been one of the most widely used DL methods for MLC. To support the application of DNN for MLC tasks, various loss functions have been proposed to determine the range of tasks. BP-MLL loss, identified as one of the early studies, is acknowledged as one of the first DNNs for MLC, undergoing subsequent enhancements by a variety of researchers. The application of MLC using DNN architecture has been applied in various domains. For example, in healthcare domains, it has been used for tasks such as intelligent health risk prediction [62], protein function prediction [63], encoding

electronic medical records [64], and multi-label chronic disease prediction [65]. Other related tasks using DNN for MLC include SLA violation prediction [66], hierarchical DNN for peptide bioactivities [67], and Robust DNN for multi-label image classification [68].

2) Deep CNN for MLC

Deep CNN has shown promising performance in a single-label image learning problem. Multi-label image classification is, however, a more general and practical problem since most real-world images comprise objects from multiple different categories. The success demonstrated by deep CNN-based models in single-label image classification can be expanded and applied to tackle multi-label challenges. The multi-label task is addressed by improving the DL models from an architectural viewpoint, in particular, the loss layer. To establish multi-label losses, the research endeavour is primarily devoted to improving binary cross-entropy (BCE). A study by [69] delved into various multi-label losses, including SoftMax, Pairwise Ranking, and weighted approximated-ranking loss (WARP), when training CNN. The findings indicated that the WARP loss (Eq. (5)) works well in addressing multi-label annotation challenges.

$$L_{warp} = \sum_{u \notin Y_i} \sum_{u \in Y_i} w(r_i^u) \max(0, \alpha + f_u(x_i) - f_u(x_i)), \quad (5)$$

where each pairwise violation is weighted by a monotonically increasing function $w(\cdot)$, and r_i^u is the predicted rank of positive label u , and $f_u(x_i)$ is the u -th element of $f(x_i)$. The idea behind this is that when the positive label is assigned a lower rank, a greater penalty should be imposed for the violation. However, subsequent research by [59] pointed out that the WARP loss is non-smooth, making it challenging for optimization. Furthermore, it highlighted that the ranking objective falls short of fully optimizing the multi-label objective. To address this issue, [59] proposes an innovative loss function for pairwise ranking based on a log-sum-exp function. Their loss function serves as a smooth approximation to the traditional hinge loss, ensuring smoothness, and is easier to optimize. Other loss functions commonly used in DL for MLC tasks include the triplet loss function [70], which is used to draw images with similar label sets, and a resilient logistic loss function [71], which is employed to train CNNs from user-supplied tags.

In 2014, Kim et al. [72] introduced the TextCNN model that employs a CNN architecture for text classification and subsequently employs another CNN for sentence-level classification. However, a limitation of this model lies in its inability to overcome the drawback associated with fixed windows in CNNs, thereby hindering its capacity to model long sequence information effectively. Later [73] proposed an XML-CNN model that improved the TextCNN model [72] by incorporating dynamic pooling, refining the loss function with the binary-cross-entropy, and inserting a hidden layer between the pooling layer and output layer. This additional layer serves to map high-dimensional labels to a lower-dimensional space, thereby mitigating the computational burden.

Wei et al. [74] proposed the Hypothesis of CNN Pooling (HCP) as a novel approach, wherein a pre-defined number of object segment hypotheses serve as inputs. Subsequently, a shared CNN is integrated with each hypothesis, and the resulting CNN outputs from various hypotheses are amalgamated using max pooling to produce multi-label predictions. The architecture of HCP is shown in Fig. 4.

Weiwei et al. [75] proposed deep CNN for multi-label images, incorporating a novel objective function comprising three components: a max-margin objective, a max-correlation objective, and cross-entropy loss. Their proposed framework

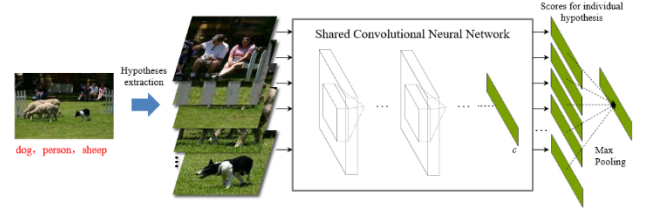


Fig. 4. Hypotheses-CNN-Pooling architecture [74]

aims to optimize the utilization of correlation information among labels. This is achieved by maximizing the score of labels present in the image over those that are absent, using a pre-defined margin. The correlation between extracted features and their corresponding labels is enhanced through learning in a semantic space. Zhu et al. [76] presented a Spatial Regularization Network (SRN) that leverages attention maps to capture both semantic and spatial relationships among multiple labels in image data. The SRN generates attention maps for all labels and captures the underlying relations through learnable convolutions. In a different approach, Kurata et al. [77] proposed a method for initializing neural networks to exploit label cooccurrence information in the context of multi-label text classification. This method applies CNN-based word embeddings to capture label correlations.

In [78], an ensemble of deep CNNs was suggested for multi-label image classification, incorporating well-known architectures such as VGG16 [79] and Resnet-101 [80]. This study delves into the effects of diverse image sizes and employs various data augmentation techniques in conjunction with a cross-entropy loss function for training and evaluating the model. More recently, Park et al. [81] introduced MarsNet, a CNN-based end-to-end network designed for MLC with the flexibility to accommodate inputs of varying sizes. To handle images of different dimensions, the authors adapted the dilated residual network (DRN) to generate higher-resolution feature maps. Additionally, they introduced horizontal-vertical pooling (HVP) to aggregate positional information from these feature maps efficiently. The approach further incorporated a multi-label scoring module and a threshold estimation module for MLC, and its effectiveness was validated through a series of diverse experiments. Table I summarizes the most recent CNN/DNN-based approaches for MLC.

MLC can also be executed through the use of a joint label embedding. An example of this is multi-view canonical correlation analysis [82], a three-way canonical analysis that aligns the image, label, and semantics within a shared latent space. Techniques such as WASABI [83] and DEVISE [84] employ a learning-to-rank framework with WARP loss to

TABLE I
DNN/CNN-based MLC methods with their applications and datasets.

Ref	Year	Technique	Application	Datasets	Evaluation
[85]	2023	Transfer Learning (VGG16, ResNet)	Movie Genre Prediction	IMDB dataset	Accuracy, AUC, F1-score, Hamming loss
[86]	2023	CNN based model	Emotion classification	COVID-19 Tweets	F1-score, Accuracy
[87]	2023	DNN based model	Music classification	OpenMIC Dataset	ROC-AUC, PR-AUC, F1-score
[88]	2023	Explainable CNN	ECG signals classification	Collection of ECG records	Hamming loss, accuracy,
[89]	2023	YOLOv5 model	Image Classification	MS-COCO Dataset	Mean average Precision
[90]	2022	Graph CNN	Image classification	MS-COCO and VOC2007	Mean average Precision, F1-score
[91]	2022	YOLOv4 Model	Waste detection	Waste image dataset	Mean average Precision
[7]	2022	DNN based a	Clinical profile identification	DEFT 2021 dataset	micro-F1 score
[92]	2022	SHO-CNN	News Classification	RCV1-v2, Reuters-21578,	Hamming loss, F1-score
[93]	2022	Multi-branch neural network	Image classification	Amazon forest, NusWide, Pascal VOC	F1-score, precision-Recall
[94]	2020	Encoder-decoder based model	Text classification	RCV1-v2, AAPD, and Ren-CECps datasets	Hamming loss, Micro-F1 score
[94]	2020	Seq2Seq-based CNN model	Text classification	RCV1-v2, AAPD and Ren-CECps	hamming loss and micro-F1 score
[95]	2017	DNN with AutoEncoder	Image and text classification	ESPGame, mirflickr, tmc2007, NUS-WIDE	Micro-F1 and Macro-F1

develop a joint embedding. Metric learning [96] focuses on acquiring a discriminative metric to gauge the similarity between images and labels. Additionally, label encodings can be achieved through methods like matrix completion [97] and bloom filter [98]. While these strategies effectively utilize label semantic redundancy, they often fall short of capturing the label cooccurrence dependency. Recognizing this limitation, the graph convolution network (GCN) [99] has been found effective in modelling label correlation in the MLC problem. Graph-based deep networks, like graph convolutional neural networks (GCN), offer an effective modelling approach for label dependencies. In this framework, each label is depicted as a node within the graph. Chen et al. [100] propose a directed graph for object labels, employing Graph Convolutional Networks (GCN) to model correlations among labels. This method maps label representations to inter-dependent object classifiers, enhancing the overall understanding of relationships between labels. Similarly, Semantic-Specific Graph Representation Learning (SSGRL) [101] incorporates semantic decoupling and interaction modules to learn and correlate semantic-specific representations. The correlation is established through GCN on a graph built from label cooccurrence data. In a related vein, a subsequent work [102] enhances label awareness by introducing lateral connections between GCN and CNN layers at different depths. This integration ensures improved injection of label information into the backbone CNN. Addressing the multi-label patent classification challenge as a text classification problem, a deep learning model proposed by [103] utilizes GCN to capture rich semantic information. The model incorporates an adaptive non-local second-order attention layer designed to model long-range semantic dependencies in text content, serving as label attention for patent categories.

The label dependency and cooccurrence can also be addressed by graph-based methods, such as conditional random field [104], dependency network [105], and cooccurrence matrix [106]. The label model [107] augments the label set with common label combinations. Most of these models only capture pairwise label correlations and have high computation costs when the number of labels is large. The low-dimensional

recurrent neurons in the RNN-based model are more computationally efficient representations for high-order label correlation, as discussed in the next section. More recent related studies include deep CNN for multi-label image classification [93], CNN-based cross-modal hashing methods [108], Improved sequence generation model via CNN [94], one-dimensional CNN (1D CNN) residual and attention mechanism for multi-label ECG recordings, graphical CNN to capture the label dependencies using the correlation between labels [90].

Final Notes: Several methods have proposed CNN-based techniques for MLC across diverse data modalities. However, deep CNN is particularly renowned for its effectiveness in multi-label image classification, and it has been applied through two main strategies. The first approach involves training the CNN individually for each label in the image, treating the multi-label problem as a series of single-label tasks [74][109][110]. This method often employs multiple local bounding boxes and instances of learning techniques, resulting in improved performance. However, it tends to overlook potential relationships among labels and struggles to assign labels describing the entire image accurately, as it processes only partial information at a time.

The second strategy adopts a holistic approach by extracting global features from the raw image and employing global loss functions that consider multiple labels simultaneously [111]. This method integrates the entire image into the classification task, enhancing the model's ability to assign labels to describe the overall content. For instance, in [69], a deep CNN model utilizing a multi-label loss function has been proposed for top-k ranking. This second-order strategy may compute label correlations, including rankings between relevant and irrelevant labels, resulting in good generalization. While the second-order strategy leverages label correlations to a certain extent, there are real-world applications where label connections extend beyond second-order relationships. This can be addressed by a high-order strategy, where MLL considers relations among labels beyond pairwise correlations. This can involve addressing connections among random subsets of labels to capture more complex relationships [55].

B. LSTMs and Transformers for MLC

A recurrent neural network (RNN) extends a regular feedforward neural network, enabling it to manage variable-length sequential data and undertake time-series prediction. RNN can also be considered as an extension of the hidden Markov model that employs a nonlinear transition function and is capable of modelling long-term temporal dependencies. Long-short Term Memory (LSTM) is an extension of RNN designed to tackle the vanishing gradient problem. It has significantly elevated the field of machine translation, speech recognition, and various other tasks. LSTMs have proven valuable for tasks that were traditionally non-sequential, such as MLC [112][113].

1) LSTM based MLC

To our knowledge, the work by Nam et al. [114] marked the first use of RNN to replace the classifier chains for sequence-to-sequence (seq2seq) text classification, thereby effectively capturing label correlations in MLC. After that, various seq2seq models were proposed to deal with MLC, such as Attentive RNN [115], Orderless RNN [113], and LSTM [116].

LSTM has been applied to solve MLC tasks. Yan et al. [116] proposed an LSTM-based multi-label ranking model for document classification, where the order of labels in the documents is rearranged in accordance with a semantic tree. In a similar domain, Yang et al. [117] consider an MLC problem as a sequence generation task and use a sequence generation model with a decoder structure to solve the MLC (Fig. 5). In their sequence generation model, the MLC task is modelled as searching for an ideal label sequence y^* that maximizes the conditional probability of $p(y|x)$. It is determined in Eq. (6) for a given text sequence x , and a subset y containing n labels.

$$p(y|x) = \prod_{i=1}^n p(y_i|y_1, y_2, \dots, y_{i-1}, x) \quad (6)$$

The decoder applies an LSTM to generate labels sequentially and predicts the next label based on the labels predicted previously. Thus, the model can consider the relationship between labels by processing label sequence dependencies through the LSTM structure. The sequence generation model with a decoder structure captures not only the correlations between labels but also selects the most informative words automatically when predicting different labels.

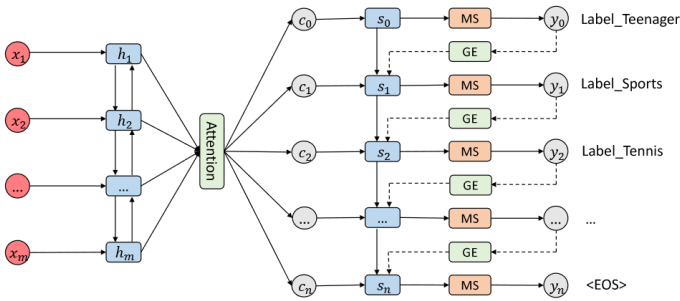


Fig. 5. Structure of the sequence generation model. MS denotes the masked Softmax layer. GE denotes the global embedding [117].

Yang et al. [118] proposed a DL model called RethinkNet to extend the limitation of the CC approach. It tackles the label ordering issue by using a global memory to store the

information about label correlation, and the global memory allows all learning models to share the same information without suffering from the label ordering issue. Other authors in [119] proposed a deep RNN for multi-label fault prediction in high-dimensional time series data with a loss function that is designed considering the class imbalance. The RNN used by authors consists of two connected LSTM networks, the encoder, and decoder, where each network is turned to capture the time series dynamics in either the historical data or the future segment data. More recently, Loris Nanni et al. [120] proposed an ensemble method that combines LSTM, GRU, and temporal CNN (TCN) for MLC tasks. Their proposed model undergoes training using various adaptations of Adam optimization, incorporating the concept of Multiple Clustering Centers (IMCC) to enhance the effectiveness of the multi-label classification system. The model employs the binary cross-entropy loss function, expressed by Eq. (4):

$$Loss = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^l y_i(j) \cdot \log(h_i(j)) + (1 - y_i(j)) \cdot \log(1 - h_i(j)) \quad (4)$$

where $y_i \in \{0,1\}^l$ and $h_i \in \{0,1\}^l$ represent the actual and predicted label vectors of each sample ($i \in 1, \dots, m$), respectively.

Zachary et al. [121] formulated clinical multivariate time series problems into MLC tasks based on LSTM, which marks the first use of LSTM for MLC within the medical domain. The authors employed target replication with LSTM (Fig. 6) to effectively classify diagnoses of critical care patients on time

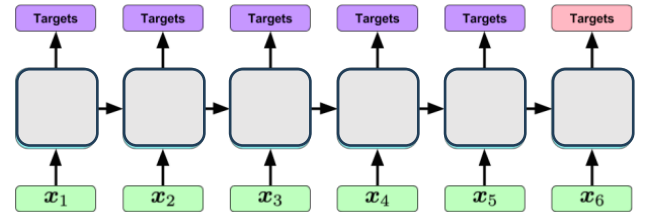


Fig. 6. LSTM model with target replication. The main target (red) is used during prediction, while during training, the model back-propagates errors from the intermediate targets (gray) at each sequence step.

series data. The model employing target replication produces an output $\hat{y}^{(t)}$ at each sequence step. The loss function is a convex combination of the ultimate loss and the average of losses across all steps (Eq. 7).

$$\alpha \cdot \frac{1}{T} \sum_{t=1}^T \text{loss}(\hat{y}^{(t)}, y^{(t)}) + (1 - \alpha) \cdot \text{loss}(\hat{y}^{(T)}, y^{(T)}) \quad (7)$$

where T is the total number of sequence steps, and $\alpha \in [0,1]$ is a hyper-parameter that determines the relative significance of achieving these intermediate targets.

The authors in [122] applied LSTM and Bayesian decision theory for multi-label lncRNA function prediction. They used LSTM for capturing the hierarchical relationships and Bayesian to change the hierarchical multi-label classification problem to

the conditional risk minimization problem to obtain final prediction results. Sagar et al. [123] proposed LSTM autoencoder-based multi-label classification for non-intrusive appliance load monitoring. Their proposed method takes electricity consumption as input from the smart meter and reconstructs a time-flipped version of the input using an encoder-decoder paradigm. In the context of multi-label emotion classification, a proposal by [124] presents latent emotion memory (LEM) to acquire latent emotion distribution without relying on external knowledge. LEM comprises latent emotion and memory modules designed to grasp emotion distribution and emotional features, respectively. The combination of these two components is then input into a Bi-directional Gated Recurrent Unit (BiGRU) for the purpose of making predictions.

Other RNN-based studies on MLC include extreme MLC [125], which uses stacked BiGRU for text encoding and incorporates cluster-sensitive attention to leverage correlations among the large label space. Li et al. [115] propose an end-to-end RNN for weakly-supervised MLC, and [122] designed hierarchical MLC based on LSTM and Bayesian decision theory for LncRNA function prediction.

2) AutoEncoder-based MLC

The autoencoders are unsupervised feature representation learning techniques [126] that aim to approximate the representation of the input through the collaboration of encoder and decoder layers. AutoEncoder techniques have found extensive application in MLC tasks. Notably, the Canonical Correlated AutoEncoder (C2AE) [95] stands out as the first DL-based label embedding method for MLC. Its structure is shown in Fig. 7. The fundamental concept behind C2AE involves

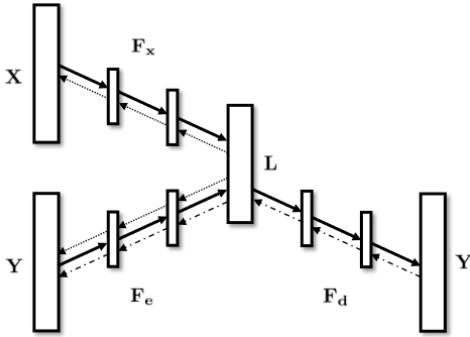


Fig. 7. Architecture of C2AE, which acquires a latent space (L) through nonlinear mappings of F_x , F_e , and F_d . X is the input, and Y is the label data.

exploring a profound latent space to concurrently incorporate instances and labels. C2AE conducts feature-conscious label encoding and label-correlation-aware prediction. The former is accomplished through the simultaneous learning of deep canonical correlation analysis (DCCA) and the encoding stage of an autoencoder, while the latter is attained through the introduction of a loss function designed for the decoding outputs. C2AE comprises two modules, namely DCCA, and autoencoder. It aims to discover three mapping functions: feature mapping F_x , encoding function F_e , and decoding function F_d . During the training process, C2AE takes input instances X and corresponding labels Y, associates them in the

latent space L, and ensures the recovery of Y through the autoencoder.

The objective function of C2AE is defined as follows:

$$\theta = \min_{F_x, F_e, F_d} \Phi(F_x, F_e) + \alpha \Gamma(F_e, F_d) \quad (8)$$

Where $\Phi(F_x, F_e)$ and $\Gamma(F_e, F_d)$ denote the losses in the latent and output spaces respectively, α is used to balance the two terms. By adapting the idea of CCA, C2AE learns the deep latent space by optimizing the correlation between instances and labels. Consequently, $\Phi(F_x, F_e)$ can be defined as:

$$\min_{F_x, F_e} \|F_x(X) - F_e(Y)\|_F^2$$

$$s.t \quad F_x(X)F_x(X)^T = F_e(Y)F_e(Y)^T = I \quad (9)$$

where $F_x(X)$ and $F_e(Y)$ denote the transformed feature and label data in the derived latent space L, respectively.

Later, Bai et al. [127] discovered that the learned deterministic latent space in C2AE lacks smoothness and structure. Minor perturbations within this latent space can result in vastly different decoding outcomes. Despite the proximity of the corresponding feature and label codes, there is no assurance that the decoded targets will exhibit similarity. In order to tackle this issue, [127] proposes an innovative framework, the Multivariate Probit Variational AutoEncoder (MPVAE), designed to acquire latent embedding spaces efficiently and capture label correlations in MLC. MPVAE adeptly learns and aligns two probabilistic embedding spaces—one for labels and another for features. The decoder in the MPVAE framework processes samples from these embedding spaces, effectively modelling the joint distribution of output targets using a multivariate probit model, which is achieved through the learning of a shared covariance matrix. Similar concepts are present in [128], which presents two-stage label embedding (TSLE) via a neural factorization machine for MLC, as depicted in Fig.8. Within this framework, the encoder is a Twin Encoding Network (TEN) composed of a singular feature network and a singular label network. The decoder's objective is to reconstruct the label based on the feature embedding. Both the feature network and the label network employ a factorization layer, enabling the computation of pairwise correlations between features and labels.

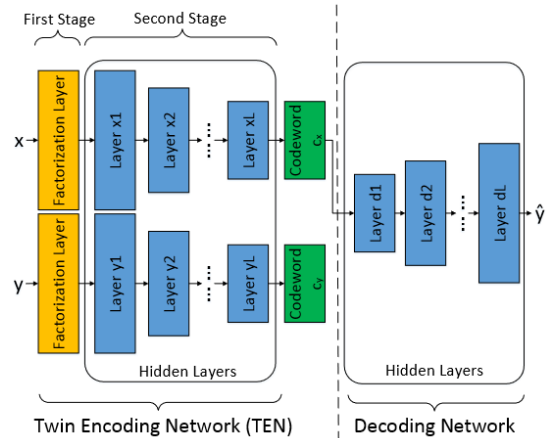


Fig. 8. Illustrates the structure of the TSLE paradigm [128].

Recently, Bai et al. [129] proposed an extension of their previous MPVAE [127] model by proposing contrastive learning boosted Gaussian mixture variational autoencoder (C-GMVAE). The loss function employed in their model is a combination of several components, including the KL loss of feature and label embeddings, the VAE reconstruction loss, a supervised contrastive loss of feature and label embeddings, and, ultimately, a cross-entropy loss for classification. Huang et al. [130] propose a dual-encoding layer autoencoder aimed at exchanging knowledge through the second encoding weight matrix. The autoencoder model is designed to jointly optimize representation learning and multi-label learning to enhance MLC performance. Nevertheless, existing methods based on autoencoders typically rely on a single autoencoder model, presenting challenges in the learning of multi-label feature representations and lacking the capability to assess similarities between data spaces.

To tackle the single model limitation of [130], Zhu et al. [131] propose a new approach called the Representation Learning method with Dual Autoencoder RLDA. This method effectively captures diverse characteristics and abstract features from data through the sequential integration of two distinct autoencoder types. First, the algorithm utilizes Reconstruction Independent Component Analysis (RICA) within a sparse autoencoder framework, training on patches from both the training and test datasets to learn global features robustly. Subsequently, leveraging the output from RICA, a stacked autoencoder with manifold regularization is employed to enhance the quality of multi-label feature representations. Ultimately, the sequential connection of these two autoencoder types yields novel feature representations for multi-label classification.

3) Transformer based MLC

Originally introduced for capturing long-term dependencies in sequence learning challenges, transformers [132] have found extensive application in various natural language processing tasks. More recently, models based on transformers have been developed for numerous vision-related tasks, demonstrating significant promise in the field. The application of transformers in addressing MLC arises from the necessity to dynamically extract local discriminative features tailored to different labels. This adaptive feature extraction is a highly desirable property, particularly in scenarios involving multiple objects within a single image.

Ramil and Pavel [133] were the first to apply the BERT model to the multi-label problem and investigate its efficacy in hierarchical text classification challenges. They proposed a sequence-generating BERT model within the realm of multi-label text classification. Later, Gong et al. [134] proposed an HG-transformer, a deep learning model that initially represents input text as a graph structure. Subsequently, it employs a multi-layer transformer structure featuring a multi-attention mechanism at word, sentence, and graph levels to comprehensively capture text characteristics. The model then leverages hierarchical relationships among labels to generate label representations, incorporating a weighted loss function designed based on semantic distances among labels. While the effectiveness of the transformer-based MLC model surpasses that of CNN and RNN structures, it is noteworthy that

transformer models often involve a substantial number of parameters and a complex network structure, leading to practical limitations. In the pursuit of enhancing the applicability of transformers in MLC, [135] proposes a hybrid model named tALBERT that combines LDA and ALBERT to derive diverse multi-level document representations. Extensive experiments conducted on three datasets substantiate the superior performance of their hybrid model compared to the current state-of-the-art methods in the realm of multi-label text classification.

The framework Query2Label, introduced in the study [136], presents a novel approach to MLC by employing a transformer decoder. To our knowledge, this marks the first application of such a framework to address MLC challenges. Query2Label operates in two stages, utilizing Transformer decoders to extract features. The multi-head attention employed in this process focuses on distinct facets or perspectives of an object category. Furthermore, the framework autonomously learns label embeddings from the provided data. To better handle the imbalance problem, the framework adopts a simplified asymmetric focal loss for computing the loss of each training sample, as shown in Eq. (10).

$$L = \frac{1}{K} \sum_{k=1}^K \begin{cases} (1 - p_k)^r + \log(p_k), & y_k = 1 \\ (p_k)^r - \log(1 - p_k), & y_k = 0 \end{cases} \quad (10)$$

where y_k represents a binary label that indicates whether the image x is associated with label k , the overall loss is calculated by averaging this specific loss across all samples within the training dataset. r denotes the focal parameter, and stochastic gradient descent is employed for optimization. The default values for $r +$ and $r -$ are set to 0 and 1, respectively.

Ridnik et al. [137] propose an ML-Decoder model, which can provide a unified solution for multi-label classification. The ML-Decoder can be seamlessly applied by excluding any pre-trained, fully connected layers, demonstrating a consistently improved balance between speed and accuracy in tests conducted on the MS-COCO MLC task. [138] A triplet transformer architecture designed for multi-label document classification was presented, as shown in Fig. 9.

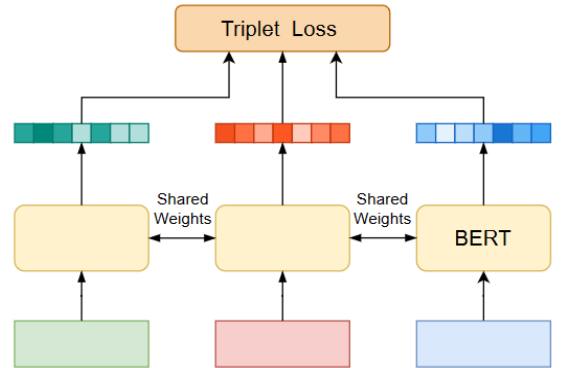


Fig. 9. Triplet transformer network for MLC [138].

This model is adept at embedding both labels and documents into a unified vector space. The architecture comprises three

BERT Networks with shared weights, enabling the classification of documents based on identifying the closest and, consequently, the most similar labels. More recently, [139] proposed a graph attention transformer network (GATN) for the MLC problem. This network is specifically designed to efficiently discover intricate relationships among labels. GATN employs a two-step process to enhance the expressive capacity of label relations. Initially, cosine similarity is applied using label word embeddings to create an initial correlation matrix, capturing extensive semantic information. Subsequently, a graph attention transformer layer is crafted to adapt this adjacency matrix to the present domain. They adopted node embeddings to construct their final correlation matrix, as shown in Eq. (12). The generation of the correlation matrix is performed as follows. For an MLC having n categories, let Z_i represent the embedding vector of the i -th label of length d and $R_{i,j}$ denote the correlation value between the i -th label and the j -th label. $R_{i,j}$ is calculated based on the cosine similarity of the embedding vectors (Eq. 11).

$$R_{i,j} = \frac{\sum_{k=1}^d Z_{i,k} \times Z_{j,k}}{\sqrt{\sum_{k=1}^d (Z_{i,k})^2} \times \sqrt{\sum_{k=1}^d (Z_{j,k})^2}} \quad (11)$$

The resulting matrix R is symmetric, where $R_{i,j}$ is always equal to $R_{j,i}$, indicating a consistent relationship between any two categories. By using a threshold τ and weight parameter p , the ultimate correlation matrix A can be derived.

$$R'_{i,j} = \begin{cases} 0, & \text{if } R_{i,j} < \tau \\ 1, & \text{if } R_{i,j} \geq \tau \end{cases}$$

$$A_{i,j} = \begin{cases} p / \sum_{j=1}^n R'_{i,j}, & \text{if } i \neq j \\ 1 - p, & \text{if } i = j \end{cases} \quad (12)$$

In a different study, Chen et al. [148] developed multi-label image recognition based on spatial and semantic transformers (SST). SST functions as a modular, plug-and-play system that concurrently extracts spatial and semantic correlations in multi-label images. It consists of two independent transformers with distinct objectives: the spatial transformer focuses on modelling correlations among features from various spatial positions, whereas the semantic transformer is employed to apprehend the

coexistence of labels without the need for manually defined rules. More recently, large language models pre-trained on large datasets using transformer architectures have been suggested for multi-label text classification, such as LP-MTC [42], prompt tuning [149], and SciBERT [150]. Table II summarizes state-of-the-art transformers and LSTMs adopted to address MLC challenges.

Final Notes: Transformers and autoencoders have become some of the most successful DL approaches actively adopted for MLC in recent years. They have been applied in many real-world applications, such as multi-label emotion classification [4], multi-label video classification for underwater ship inspection [151], multi-label text classification [135], and multi-label disease classification [152]. MLC models built on transformer structures often outperform those based on RNN and LSTM. However, it's worth noting that transformer models typically entail a substantial number of parameters and a complex network structure, introducing certain limitations in practical applications. Moreover, addressing label correlations in MLC is crucial for certain objectives, presenting a challenge due to the nature of the label space in the data. In future studies, exploring effective methods for capturing label correlations and other related challenges will be the main research focus of autoencoders and transformers for MLC.

C. Hybrid DL for MLC Problem

One common approach to extending CNNs to MLC involves transforming the problem into multiple single-label classification tasks, which use either ranking loss [69] or cross-entropy loss [153]. However, these methods fall short of capturing the dependencies between multiple labels when treating them independently. Several studies have demonstrated the significant label cooccurrence dependencies in MLC problems. To model label dependency, existing works have used a variety of techniques, such as nearest-neighbors-based models [154][155], ranking-based methods [156][157], structured inference models [158][159], and graphical models [106][160][107]. A prevalent strategy involves representing dependencies and cooccurrence relationships through pairwise compatibility probabilities or cooccurrence probabilities. Subsequently, Markov random fields [105] are often utilized to deduce the ultimate joint label probability. However, when dealing with a large number of labels, the parameters associated with these pairwise probabilities can become excessively large,

TABLE II
Review of state-of-the-art LSTM/transformer-based approaches to MLC applications.

Ref	Year	Technique	Application	Datasets	Evaluation
[140]	2023	BERT based models	Toxicity content identification	Online Corpus	Precision, recall, and F1-score
[3]	2023	Transformers	Document classification	AAPD dataset, Reuters-21578	F1-score, Precision, recall
[141]	2023	Transformers	Chest X-ray diagnosis	PadChest dataset	AUC, and mean AUC
[142]	2023	Transformer network	Retinal Disease Classification	MuReD, ARIA, STARE, etc.	Precision, recall, F1, AUC
[143]	2022	LSTM with GloVe method	Cardiovascular disease classification	Cardiovascular text dataset	Accuracy
[144]	2022	Transformer model	Image Classification	MS-COCO, Pascal-VOC, NUSWIDE	Precision, recall, F1-score
[122]	2022	LSTM Network	lncRNA function prediction	GOA-lncRNA dataset	Micro F1, Macro F1
[145]	2022	Transformer-CNN	Text classification	RCV1 and AAPD	Micro and macro F1
[146]	2022	encoder-decoder	Text classification	AAPD and RCV1-V2	Hamming loss, Micro-F1
[147]	2021	Deep graph LSTM	Legal Text classification	Legal case of Indian judiciary	Accuracy

and many of these parameters may be redundant if labels have highly overlapping meanings. Furthermore, most of these methods either cannot model higher-order correlations [106], or trade computational complexity to capture more intricate label relationships [107]. In the last few years, there has been a shift towards RNN, particularly LSTM [161] following CNN, as an effective method to exploit high-order label correlations. LSTMs demonstrate the capability to capture higher-order inter-label relationships while maintaining tractable computational complexity.

The concept of exploiting RNN models to capture label correlations in MLC was initially introduced in [162] and [163], where the fusion of CNN with RNN architecture was proposed. Because classifier chains (CC) are considered a memory mechanism that stores the label predictions of the earlier classifiers, CNN-RNN-based algorithms can extend CC by replacing the mechanism with a more sophisticated memory-based model. Wang et al.[162] put forward a unified CNN-RNN framework that learns a joint image-label embedding to characterize semantic label dependencies. The CNN-RNN structure comprises a CNN feature mapping layer (encoder) for extracting semantic representations from images and an RNN inference layer (decoder) that utilizes the encoding to generate a label sequence, modelling image/label relationships and dependencies. RNNs adopt a frequent-first ordering approach for sequential outputs, and multiple label outputs are generated at the prediction layer through the nearest neighbour search. In [163], it was demonstrated that the order of labels in the training phase had an impact on the annotation performance, with rare-to-frequent order yielding the best results, which was further validated in subsequent works such as [162][164]. Jin et al. [107] utilized CNN to represent images, feeding them into an RNN for predictions. They experimented with frequent-first, dictionary-order, rare-first, and random label ordering, comparing the outcomes of each method. Liu et al.[164] employ a comparable framework, wherein they explicitly assign the tasks of label prediction and label correlation to the CNN and RNN models, respectively. Rather than employing a fully connected layer linking the CNN and RNN models, the researchers feed the RNN with class probabilities predicted by the CNN model, thereby supervising both models throughout training. They adopt a rare-first ordering in their model to accord greater significance to less common labels. The authors explore various visual representations to input into the RNN. In [162], images and labels are projected to the same low-dimensional space to model the image-text relationship, while [164] uses the predicted class probabilities and [163] experiments with different internal layers of the CNN. Other works based on CNN-RNN architecture include an ensemble of CNN-RNN for text categorization [165], CNN-RNN for satellite images of the Amazon rainforest [166], hybrid CNN, and bidirectional LSTM network for multi-label aerial image classification [167] and CNN-ConvLSTM for pedestrian attribute recognition [168], as shown in Fig. 10.

Despite the encouraging performance exhibited by the CNN-RNN architecture, its reliance on a predetermined label order for learning poses a significant challenge. Since RNN-based models produce sequential outputs, a pre-defined label order is required during training for the MLC task. For example, Wang et al.[162] determine the label order based on label frequencies

observed in the training data. However, employing such predetermined label orders may not accurately capture natural label dependencies, introducing a rigid constraint on the RNN model. Imposing a frequent-to-rare label order biases the model towards frequent labels, requiring it to make numerous correct predictions before addressing the rare label(s). Conversely, a

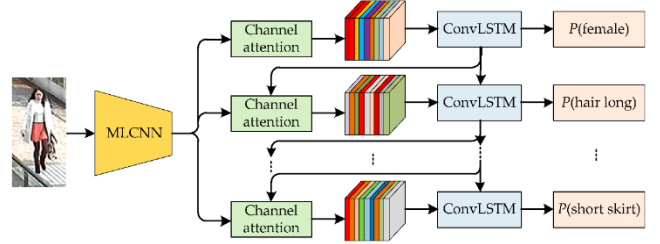


Fig.10. CNN- ConvLSTM for pedestrian attribute recognition [168].

rare-to-frequent label order compels the model to prioritize learning rare labels, a challenging task given the scarcity of training examples. In general, any frequency-based pre-defined label order fails to capture true label dependencies, as each label in a multi-labelled image is intricately connected to many other labels within the global context, even though a label may exhibit a stronger connection to only a few of them. Moreover, defining such orders introduces a bias towards dataset-specific statistics, undermining the generalizability of the model. The lack of robustness in learning optimal label orders, as verified in [162], is exacerbated by the difficulty in predicting labels for smaller-sized objects when visual attention information is underutilized. Consequently, addressing how to introduce flexibility in learning optimal label orders while concurrently exploiting associated visual information becomes an important focus of research.

In resolving these constraints related to the order of labels, some studies [112][113] have proposed techniques that do not require feeding the ground-truth labels to the RNN in any particular sequence. Chen et al. [112] introduce an order-free RNN for MLC, incorporating visual attention and LSTM models. They utilize binary cross-entropy loss at each time step to predict labels without considering their order. The simultaneous learning of attention and LSTM models enables the identification of regions of interest associated with each label, automatically capturing label order without pre-defined

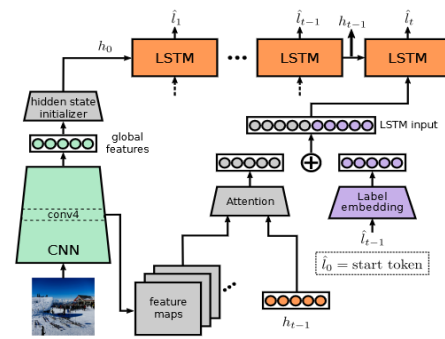


Fig.11. Illustration of CNN-LSTM architecture where the orderless loss function substitutes the loss function [113].

constraints. The integration of attention with LSTM results in slight enhancements over CNN-RNN models. Additional work

presented in [113] proposes an order-less recurrent model for MLC, as depicted in Fig. 11. This approach explores the dynamic ordering of ground truth labels with predicted label sequences, facilitating faster training of more optimal LSTM models. Notably, this method avoids duplicate generation, contrasting with Chen et al.'s [112] explicit duplication removal module. Subsequent studies, including [169] and [115], have also embraced order-free strategies in MLC. In [169], the authors propose an end-to-end trainable framework for multi-label image recognition. This framework comprises CNN for extracting deep feature representations and an attention-aware module based on LSTM for iteratively identifying class-related regions and predicting label scores for these identified regions. The end-to-end training of the framework relies solely on image-level labels facilitated by reinforcement learning techniques. The process is initiated by inputting the image into the CNN, yielding feature maps. Subsequently, an LSTM unit processes these features, incorporating the hidden state from the previous iteration to forecast scores for each region. These scores play a pivotal role in determining the location for the subsequent iteration. To arrive at the ultimate label distribution, the predicted scores undergo consolidation through category-wise max-pooling. However, these methods tend to internally select a specific label order at the initial time step and subsequently iterate over the same sequence in the subsequent time steps. In essence, these strategies enable the RNN to implicitly favour one among numerous sequences, thereby introducing inherent bias.

Later, Ayushi et al. [175] proposed multi-order RNN to address the limitation of these order-free approaches (Fig. 12). Their approach provides RNNs with the flexibility to explore and learn multiple relevant inter-label dependencies in the form of multiple label orders instead of a fixed and pre-defined one. The architecture of multi-order RNN comprises a deep CNN fine-tuned with ground-truth data from a specified dataset and an LSTM model utilizing the soft confidence vector derived from the CNN as its initial state. For each sample, at each time step, a cross-entropy loss is calculated, considering all true labels except the one from the preceding time step as potential candidates for prediction at that specific time step. The ultimate predictions are acquired by max-pooling individual label scores across all time steps. They showed that multi-order RNN consistently outperforms the existing CNN-RNN-based approaches and provides an intuitive way of adapting a sequence prediction framework for the image annotation task.

More recently, Wang et al. [176] proposed cross-modal fusion with an attention mechanism for multi-label image

classification that combines the attention mechanism and graph convolution network to capture the local and global label dependencies simultaneously in an end-to-end manner. Their method involves a feature extraction module with an attention mechanism, a label cooccurrence embedding learning module and a cross-modal fusion module with Multi-modal Factorized Bilinear pooling which efficiently fuses the above image features and label cooccurrence embeddings. Their method was

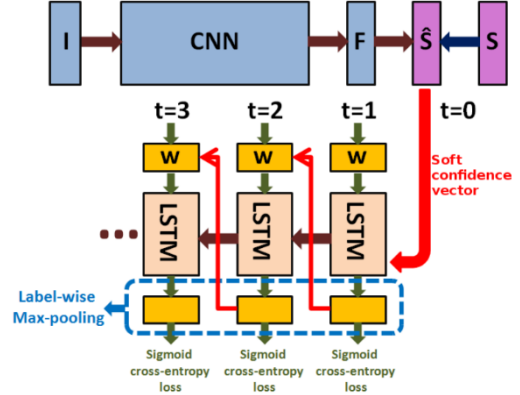


Fig. 12. Multi-order RNN for Image Annotation

tested on COCO and VOC2007 datasets and showed better classification results compared to other similar approaches. In a different study, [177] introduced two hybrid DL models for taxonomic categorization of DNA sequences. The first model combines a stacked convolutional autoencoder (SCAE) with a Multi-label Extreme Learning Machine (MLELM), while the second model incorporates a variational convolutional autoencoder (VCAE) into the MLELM framework. These models efficiently generate precise feature maps capturing individual and inter-label interactions in DNA sequences, encompassing both spatial and temporal properties. Extracted features are input into MLELM networks, producing soft classification scores and hard labels. The VCAE-MLELM model consistently outshone the SCAE-MLELM model, while the latter demonstrated superior performance in soft categorization, surpassing existing methods, such as CNN-BiLSTM and DeepMicrobe [178]. The convolutional autoencoder aids in extracting spatial organization, enhancing computational efficiency by identifying latent associations. The extracted features are then fed into two MLELMs, with the initial model generating probabilistic labels and the subsequent model establishing relationships between deterministic and probabilistic labels. Table III summarizes some of the state-of-

TABLE III
Review of state-of-the-art hybrid-based approaches to MLC applications.

Ref	Year	Technique	Application	Datasets	Evaluation
[170]	2023	Transformer + CNN	Medical image classification	ChestX-ray11, NIH ChestX-ray14	Mean avg. Precision, recall, F1
[171]	2023	CNN-BiLSTM	Short texts Classification	Online large corpuses	Micro-f1, Macro F1
[146]	2023	Hybrid transformer	Text Classification	AAPD and RCV1-V2	Hamming loss, F1
[172]	2022	AutoEncoder+CNN	Movie and image classification	IMDb Movie, MS-COCO	precision, Recall, Hamming loss
[120]	2022	LSTM+GRU+TCN	Music, image, drug	Cal500, Image, Scene, Yeast, etc.	Hamming loss, Ranking loss,
[173]	2021	Graph CNN + CNN	Breast cancer	Mammograms 2018	Accuracy, TPR
[174]	2019	Ensemble of CNN	Image classification	NUS-WIDE, MS-COCO, PASCAL	Mean average Precision, F1 score
[162]	2017	CNN +RNN	Image classification	NUS-WIDE, MS-COCO, VOC PASCAL 2007	per-class and overall (recall, Precision, etc.)

the-art Hybrid DL techniques in MLC, along with their domain of application, datasets, and evaluation metrics.

V. OPEN CHALLENGES AND DATASETS

The sustained appeal of MLC can be attributed to the widespread prevalence of multi-label data, which is pervasive across various domains, such as biology, environment, healthcare, commerce, recommender systems, social media, retail, sentiment analysis, energy, transportation, and robotics. Furthermore, the internet consistently generates quintillions of bytes of streaming data daily, posing significant challenges for MLC tasks. In real-world applications, MLC remains challenging due to the complex presence of labels. As an example, there are scenarios where the number of labels is very large, the labels are only partially or weakly provided, and they may emerge continuously or be entirely unseen beforehand. This section reviews some of the open challenges in MLC and associated datasets.

- **Label Dependencies:** The presence of multiple data labels can suggest associations between different categories. For instance, in an image object detection task, cats and dogs may frequently coexist, while cats and sharks typically do not share the same space. Thus, modelling and learning correlations among categories have consistently remained a fundamental focus in MLC [190]. However, effectively leveraging label dependencies continues to be a persistent challenge in MLC. Regarding the modelling of these dependencies, learning approaches can be classified as first-order (addressing each distinct label independently), second-order (modelling pairs of labels), and higher-order (simultaneously addressing more than two labels). The powerful learning capabilities of DL methods are commonly harnessed to tackle second-order label dependencies in various ways, including through graph CNN [90][191], autoencoder-based [127][129], transformers [3][192], and hybrid DL models [170][171][162]. However, the challenge of higher-order label dependence in MLC continues to be a central focus for researchers, including both practical and theoretical considerations, persisting into the present day.

- **Extreme MLC:** Another important challenge in MLC is the presence of a very large number of labels, also known as extreme MLC. It is an active area of research wherein the number of labels can be exceptionally high, reaching into the millions in certain cases. Traditional classifiers like one-vs-all, SVM, and neural networks face two primary impediments when applied directly in the context of extreme MLC [193]. Firstly, the substantial number of labels poses a significant bottleneck, as it is impractical to implement a straightforward classifier for each label due to memory limitations. Secondly, the existence of labels with very few samples in their support adds complexity to the learning process for these labels. Several efforts have been made, such as Ranking-based Auto-Encoder (Rank-AE)[194], DeepXML framework [195], AttentionXML [24], and two-stage XMTC framework (XRR)[26] to address the challenges posed by extreme MLC.
- **Weakly-Supervised MLC:** Weakly supervised learning focuses on the more demanding aspect of MLC, wherein certain labels in the training set are missing. Given the extensive data volumes and diverse domains involved in such tasks, fully supervised learning demands manually annotated datasets, incurring significant costs and time. The weakly supervised MLC task, involving the training of MLC models with partially observed labels per sample, is gaining importance due to its potential for substantial savings in annotation costs [29]. In addressing MLC with missing or partial labels, several notable approaches have been suggested, including Graph Neural Networks (GNNs) [33], deep generative models[30], and hierarchical MLC [196]. Moreover, learning paradigms, such as zero-shot learning [197], few-shot learning [198], and self-supervised learning [199] are emerging research directions for partial or weakly supervised MLC.
- **Imbalanced MLC:** Imbalanced learning is a widely recognized and intrinsic characteristic observed in multi-label datasets, influencing the learning dynamics of various MLC algorithms. The issue of imbalance in multi-label data can be analyzed from three perspectives [20]: intra-label imbalance, inter-label imbalance, and imbalance among label

TABLE IV
Multi-label datasets and their description

Datasets	Domain	#Labels	#Features	#Instances	Cardinality	Density	Diversity
Bibtex [179]	Text	159	1836	7,395	2.402	0.015	0.386
EUR-Lex [180]	Text	3993	5000	19,348	1.292	0.003	0.083
RCV1[181]	Text	103	100000	-	-	-	-
Corel5k [182]	Images	374	499	5000	3.522	0.009	0.635
NUS-WIDE [183]	Images	81	500	269,648	1.869	0.023	-
AmazonCat-13K	Text	13330	-	1.5 M	-	-	-
CNIPA-data [61]	Text (patent)	618	17/242	212,095	1.330	0.002	-
USPTO-data [61]	Text (patent)	632	8/111	353,701	2.174	0.003	-
Birds [184]	audio	19	260	645	1.014	0.053	-
CAL500 [185]	Music	174	68	502	26.044	0.150	1.000
Emotions [186]	Music	6	72	593	1.869	0.311	0.422
CMU-Movie [61]	Movie	372	3/312	42,204	3.691	0.010	-
Liu [187]	Drugs	1385	2892	832	-	-	-
Yeast [111]	Biology	14	103	2,417	.237	0.303	0.082
Genbase [188]	Biology	27	1186	662	1.252	0.046	0.048
Water quality [189]	Chemistry	14	16	1,060	5.073	0.362	-
GoEmotions [61]	Emotion	28	13	211,225	1.181	0.042	-
Toxic [61]	Comments	6	70	159,571	0.220	0.037	-

sets. These factors may also co-occur, further intensifying the complexity of the MLC task. Although traditional independent approaches are commonly used for addressing imbalances in MLC [200], DL model adaptation methods are still underdeveloped [201][202].

- **High Data Dimensionality:** Much like numerous learning tasks, MLC faces the challenge of dimensionality. The rapid expansion of data scaling in multi-label datasets often results in high-dimensional features [203], contributing to prolonged processing times and diminished classifier performance. The occurrence of this problem stems from an abundance of redundant, noisy, and irrelevant features, giving rise to overfitting problems. To mitigate these challenges, it becomes imperative to reduce feature dimensionality with two primary approaches: feature extraction and feature selection. The former involves mapping high-dimensional features into a lower-dimensional space [204], while the latter

entails choosing a smaller subset of features to replace the entire original set [205]. Feature extraction produces new features that lack physical meaning, whereas feature selection preserves physical meaning and enhances explanatory power. Table IV provides a summary of some multi-label datasets spanning diverse domains. The table includes information such as the number of instances, features, labels, cardinality, density, and diversity. The names listed in Table IV may not necessarily match those mentioned in the original papers; instead, they are the ones commonly used in the literature. To describe the characteristics of a multi-label dataset, three metrics can be employed for measurement, namely, label cardinality, label density, and label diversity [206]. Given a multi-label dataset $M = \{(X_i, Y_i) | 1 \leq i \leq N\}$, the three-attribute metrics can be defined as follows, where $m = |M|$ represents the size of the dataset, $|Y_i|$ represents the number of labels for i^{th} instance, and $|L|$ the number of labels in M .

TABLE V
Comparison of deep learning techniques for solving multi-label classification

Ref	Neural Network Type used	Label Correlation Strategy used	Data Modality	Advantages	Limitations
[3]	Transformers	A dependency regularization approach is added to the loss function	Textual/ Document	Effective use of transformers to learn pairwise label co-occurrence and dependencies	Limited to handle effectively higher order label correlation.
[90]	Graph CNN	Semantic label embedding	Images	The method can resolve the scalability and label cooccurrence issues in MLC	Limited to handling interdependent relationships among the label
[56]	Multi-layer Feedforward DNN	Employing a new error function with pairwise ranking loss	Genomics and Textual	Perform better in MLC problems since it takes label correlations into consideration	Computational complexity with label size. Does not perform well on textual data
[95]	DNN with AutoEncoder	Canonical Correlated AutoEncoder (C2AE)	Texts and Images	Allows better exploitation of cross-label dependency during prediction processes	Different objectives are mixed and difficult to understand
[163]	CNN-RNN	Based on pre-defined label ordering using frequent-to-rare or rare-to-frequent	Image Datasets	Captures the characteristics of MLC and performs better for sequence generation problem	Pre-defined label order does not reflect the true label dependencies
[207]	SHO-LSTM	Word embedding	Textual	good convergence capability and optimal solution	Limited to handle label dependency
[112]	CNN-RNN	Order-free RNN with Visual Attention	Image Datasets	Doesn't require pre-defined label order for label correlation and prediction	Label label-ordering approach doesn't work for some problems
[208]	Transformers	Linguistic and Semantic Cross-Attention	Multi-modal data	Handles high-level semantics and linguistic embeddings	Does not consider label cooccurrences
[175]	CNN-RNN	Multi-order RNN		Multi-order RNN	Computational complexity
[120]	Ensemble Model	Ensemble of GRU, LSTM, and TCN	Music, Biological	Better performance and stability in prediction	Computational complexity
[58]	Deep NN	Cross entropy loss function	Textual/ Document	Addresses limitation of ranking loss minimization	Doesn't contain a deep-layer network for large-scale image data
[77]	Deep NN	Word embeddings based on CNN with binary cross-entropy	Textual datasets	Less computational overhead during training and evaluation	weight initialization requires a sophisticated approach
[174]	Deep CNN	Sigmoid cross-entropy loss	Images Dataset	Ability to handle various size inputs with a strong baseline	Doesn't consider the correlation between labels
[116]	LSTM2 and rankLSTM	Word embedding based on a hybrid loss function	MEDLINE Dataset	Uses a semantic tree for label ranking and addresses the error propagation for a variable number of labels	Not efficient for full-text document classification.
[118]	RethinkNet	Adopt a global memory approach	Emotions, medical,	Tackles both label correlation and class imbalance	Running time complexity
[162]	CNN-RNN	Joint image-label embedding for dependency	Imaging dataset	Hybrid use of CNN and RNN makes the task tractable	frequency-based label order does not reflect the true label correlation
[209]	Hyperspectral CNN	Twin network structure and uses a hybrid mechanism	Textual data	Reduced imbalanced problem in MLC	Does not capture the relationship between labels
[135]	tALBERT (LDA+ ALBERT)	Deep semantic information	Documents	used to extract the depth features of documents	The model is only effective on fewer data instances

- **Label cardinality (Card):** it is expressed as the average number of labels per sample.

$$\text{Card}(M) = \frac{1}{m} \sum_{i=1}^m |Y_i| \quad (11)$$

- **Label density (Dense):** is the average number of labels of samples in dataset M divided by the number of labels.

$$\text{Dens}(M) = \frac{1}{m} \sum_{i=1}^m \frac{|Y_i|}{|L|} \quad (12)$$

- **Label Diversity (Diver):** denotes the number of different label combinations in the sample set.

$$\text{Diver}(M) = |\{Y_x \ni x: (X, Y_x) \in M\}| \quad (13)$$

MLC problem is challenging due to the high feature/label dimensionality. Several suggestions have been made regarding dimensionality reduction [205], and feature selection [210] in MLC. A recent contribution by Zan et al. [211] introduces a novel algorithm known as Global and Local Feature Selection (GLFS), demonstrating its superiority over existing state-of-the-art multi-label feature selection methods. However, real-world MLC scenarios present additional complexities in the feature space, such as the potential disappearance or augmentation of certain features and alterations in distribution. Effectively addressing these challenges concurrently in both label and feature spaces represents a more formidable task and constitutes a prospective avenue for future research in addressing the intricate nature of the MLC problem.

VI. COMPARATIVE ANALYSIS

This section presents the comparative study of the various DL methods proposed in the literature for addressing MLL. The comparison parameters used include the network architecture, the label correlation approach, the evaluation metric, key findings/advantages, and limitations of the proposed approach. Table V depicts detailed information on the various methods proposed so far for solving the MLC using DL approaches.

VII. CONCLUSIONS

Due to its robust learning capability, deep learning (DL) has demonstrated state-of-the-art performance in numerous real-world multi-label applications, including tasks like multi-label image and text classification. In solving multi-label learning problems, the main challenge lies in effectively leveraging DL to capture label dependencies more adeptly. This paper presents a comprehensive review of DL for multi-label learning problems, with a primary emphasis on DL for multi-label classification (MLC) involving label correlations. We have compiled and scrutinized numerous articles pertaining to DL techniques for MLC published from 2006 to 2023. The survey details recent methods related to various DL approaches, including DNN, CNN, LSTM, autoencoders, transformers, and hybrid models to tackle MLC challenges. The study provides an overview of the representative works cited, delving into the latest DL techniques applied in MLC and scrutinizing their limitations. It also covers a concise depiction of the existing challenges in MLC and a brief description of the publicly

available multi-label datasets. Furthermore, we conducted a comparative analysis of DL approaches for MLC, highlighting the strengths and weaknesses of existing methods and offering insights into promising avenues for future research.

Overall, despite the increasing demand for MLC across diverse domains, the research on developing an efficient and comprehensive DL framework for MLC, along with an effective model to address the associated challenges, such as label correlations remains still a challenge. Consequently, there is a need for further exploration to identify more effective solutions in the future.

REFERENCES

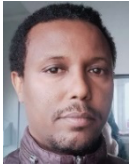
- [1] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-label Data," in *Data Mining and Knowledge Discovery Handbook*, Boston, MA: Springer US, 2009, pp. 667–685. doi: 10.1007/978-0-387-09823-4_34.
- [2] J. Fürnkranz, E. Hüllermeier, E. Loza Mencía, and K. Brinker, "Multilabel classification via calibrated label ranking," *Mach. Learn.*, vol. 73, no. 2, pp. 133–153, Nov. 2008, doi: 10.1007/s10994-008-5064-8.
- [3] H. Fallah, E. Bruno, P. Bellot, and E. Murisasco, "Exploiting Label Dependencies for Multi-Label Document Classification Using Transformers," in *Proceedings of the ACM Symposium on Document Engineering 2023*, Aug. 2023, pp. 1–4. doi: 10.1145/3573128.3609356.
- [4] I. Ameer, N. Bölücü, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Expert Syst. Appl.*, 2023, doi: 10.1016/j.eswa.2022.118534.
- [5] A. Haghighian Roudsari, J. Afshar, W. Lee, and S. Lee, "PatentNet: multi-label classification of patent documents using deep learning based language understanding," *Scientometrics*, 2022, doi: 10.1007/s11192-021-04179-4.
- [6] M. Priyadarshini, A. F. Banu, B. Sharma, S. Chowdhury, K. Rabie, and T. Shongwe, "Hybrid Multi-Label Classification Model for Medical Applications Based on Adaptive Synthetic Data and Ensemble Learning," *Sensors*, 2023, doi: 10.3390/s23156836.
- [7] C. Gérardin, P. Wajsbürt, P. Vaillant, A. Bellamine, F. Carrat, and X. Tannier, "Multilabel classification of medical concepts for patient clinical profile identification," *Artif. Intell. Med.*, 2022, doi: 10.1016/j.artmed.2022.102311.
- [8] Y. Ren et al., "Multi-label classification for multi-drug resistance prediction of *Escherichia coli*," *Comput. Struct. Biotechnol. J.*, 2022, doi: 10.1016/j.csbj.2022.03.007.
- [9] I. D. Lopez, A. Figueroa, and J. C. Corrales, "Multi-Label Data Fusion to Support Agricultural Vulnerability Assessments," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3089665.
- [10] S. Chen, T. Shu, H. Zhao, and Y. Y. Tang, "MASK-CNN-Transformer for real-time multi-label weather recognition," *Knowledge-Based Syst.*, 2023, doi: 10.1016/j.knosys.2023.110881.
- [11] X. Liu et al., "Emotion classification for short texts: an improved multi-label method," *Humanit. Soc. Sci. Commun.*, 2023, doi: 10.1057/s41599-023-01816-6.
- [12] I. Ameer, G. Sidorov, H. Gomez-Adorno, and R. M. A. Nawab, "Multi-Label Emotion Classification on Code-Mixed Text: Data and Methods," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3143819.
- [13] Z. Y. Poo, C. Y. Ting, Y. P. Loh, and K. I. Ghauth, "Multi-Label Classification with Deep Learning for Retail Recommendation," *J. Informatics Web Eng.*, 2023, doi: 10.33093/jiwe.2023.2.2.16.
- [14] E. Deniz, H. Erbay, and M. Coşar, "Multi-Label Classification of E-Commerce Customer Reviews via Machine Learning," *Axioms*, 2022, doi: 10.3390/axioms11090436.
- [15] A. Huang, R. Xu, Y. Chen, and M. Guo, "Research on multi-label user classification of social media based on ML-KNN algorithm," *Technol. Forecast. Soc. Change*, 2023, doi: 10.1016/j.techfore.2022.122271.
- [16] S. P. Singh, R. Tiwari, S. Sharma, and A. Kumar, "AI Based Multi Label Data Classification of Social Media," 2022. doi: 10.1007/978-981-19-0332-8_24.
- [17] S. A. Azzi and C. B. O. Zribi, "A new Classifier Chain method of BERT Models For Multi-label Classification of Arabic Abusive Language on Social Media," *Procedia Comput. Sci.*, vol. 225, pp. 476–485, 2023, doi: 10.1016/j.procs.2023.10.032.
- [18] S. Wang, G. Peng, and Z. Zheng, "Capturing Joint Label Distribution for

- Multi-Label Classification through Adversarial Learning,” *IEEE Trans. Knowl. Data Eng.*, 2020, doi: 10.1109/TKDE.2019.2922603.
- [19] Y. Zhu, J. T. Kwok, and Z. H. Zhou, “Multi-label learning with global and local label correlation,” *IEEE Trans. Knowl. Data Eng.*, 2018, doi: 10.1109/TKDE.2017.2785795.
 - [20] A. N. Tarekgn, M. Giacobini, and K. Michalak, “A review of methods for imbalanced multi-label classification,” *Pattern Recognition*. 2021. doi: 10.1016/j.patcog.2021.107965.
 - [21] J. Chen, R. Zhang, J. Xu, C. Hu, and Y. Mao, “A Neural Expectation-Maximization Framework for Noisy Multi-Label Text Classification,” *IEEE Trans. Knowl. Data Eng.*, 2023, doi: 10.1109/TKDE.2022.3223067.
 - [22] Y. Jiao, P. Zhao, J. Wu, X. Xian, H. Xu, and Z. Cui, “Active multi-label learning with optimal label subset selection,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2014, doi: 10.1007/978-3-319-14717-8_41.
 - [23] M. L. Zhang and Z. H. Zhou, “A review on multi-label learning algorithms,” *IEEE Transactions on Knowledge and Data Engineering*. 2014. doi: 10.1109/TKDE.2013.39.
 - [24] R. You, Z. Zhang, Z. Wang, S. Dai, H. Mamitsuka, and S. Zhu, “AttentionXML: Label tree-based attention-aware deep model for high-performance extreme multi-label text classification,” 2019.
 - [25] I. Mishra, A. Dasgupta, P. Jawanpuria, B. Mishra, and P. Kumar, “Light-weight Deep Extreme Multilabel Classification,” 2023. doi: 10.1109/IJCNN54540.2023.10191716.
 - [26] J. Xiong, L. Yu, X. Niu, and Y. Leng, “XRR: Extreme multi-label text classification with candidate retrieving and deep ranking,” *Inf. Sci. (Ny.)*, 2023, doi: 10.1016/j.ins.2022.11.158.
 - [27] H. Liz, J. Huertas-Tato, M. Sánchez-Montañés, J. Del Ser, and D. Camacho, “Deep learning for understanding multilabel imbalanced Chest X-ray datasets,” *Futur. Gener. Comput. Syst.*, 2023, doi: 10.1016/j.future.2023.03.005.
 - [28] L. Chen, Y. Wang, and H. Li, “Enhancement of DNN-based multilabel classification by grouping labels based on data imbalance and label correlation,” *Pattern Recognit.*, 2022, doi: 10.1016/j.patcog.2022.108964.
 - [29] Y. Zhang et al., “Weakly Supervised Multi-Label Classification of Full-Text Scientific Papers,” 2023. doi: 10.1145/3580305.3599544.
 - [30] H. M. Chu, C. K. Yeh, and Y. C. F. Wang, “Deep Generative Models for Weakly-Supervised Multi-Label Classification,” 2018. doi: 10.1007/978-3-030-01216-8_25.
 - [31] Y. Dai, W. Song, Z. Gao, and L. Fang, “Global-guided weakly-supervised learning for multi-label image classification,” *J. Vis. Commun. Image Represent.*, 2023, doi: 10.1016/j.jvcir.2023.103823.
 - [32] D. Huynh and E. Elhamifar, “Interactive multi-label CNN learning with partial labels,” 2020. doi: 10.1109/CVPR42600.2020.00944.
 - [33] T. Durand, N. Mehrasa, and G. Mori, “Learning a deep convnet for multi-label classification with partial labels,” 2019. doi: 10.1109/CVPR.2019.00074.
 - [34] C. Si, Y. Jia, R. Wang, M.-L. Zhang, Y. Feng, and Q. Chongxiao, “Multi-Label Classification With High-Rank and High-Order Label Correlations,” *IEEE Trans. Knowl. Data Eng.*, pp. 1–13, 2023, doi: 10.1109/TKDE.2023.3330449.
 - [35] J. Bogatinovski, L. Todorovski, S. Džeroski, and D. Kocev, “Comprehensive comparative study of multi-label classification methods,” *Expert Syst. Appl.*, vol. 203, pp. 1–23, 2022, doi: 10.1016/j.eswa.2022.117215.
 - [36] N. A. Sajid et al., “Single vs. Multi-Label: The Issues, Challenges and Insights of Contemporary Classification Schemes,” *Applied Sciences (Switzerland)*. 2023. doi: 10.3390/app13116804.
 - [37] X. Chen et al., “A Survey of Multi-label Text Classification Based on Deep Learning,” 2022. doi: 10.1007/978-3-031-06794-5_36.
 - [38] W. Liu, H. Wang, X. Shen, and I. W. Tsang, “The Emerging Trends of Multi-Label Learning,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, doi: 10.1109/TPAMI.2021.3119334.
 - [39] E. Gibaja and S. Ventura, “A Tutorial on Multilabel Learning,” *ACM Comput. Surv.*, vol. 47, no. 3, pp. 1–38, Apr. 2015, doi: 10.1145/2716262.
 - [40] X. Liu, H. Yang, C. Ai, Y. Ding, F. Guo, and J. Tang, “MVML-MPI: Multi-View Multi-Label Learning for Metabolic Pathway Inference,” *Brief. Bioinform.*, vol. 24, no. 6, Sep. 2023, doi: 10.1093/bib/bbad393.
 - [41] Y. Liu et al., “Predicting the multi-label protein subcellular localization through multi-information fusion and MLSI dimensionality reduction based on MLFE classifier,” *Bioinformatics*, 2022, doi: 10.1093/bioinformatics/btab811.
 - [42] R. Song et al., “Label prompt for multi-label text classification,” *Appl. Intell.*, vol. 53, no. 8, pp. 8761–8775, Apr. 2023, doi: 10.1007/s10489-022-03896-4.
 - [43] Z. Liu, C. Tang, S. E. Abhadiomhen, X. J. Shen, and Y. Li, “Robust Label and Feature Space Co-Learning for Multi-Label Classification,” *IEEE Trans. Knowl. Data Eng.*, 2023, doi: 10.1109/TKDE.2022.3232114.
 - [44] R. B. Mangolin et al., “A multimodal approach for multi-label movie genre classification,” *Multimed. Tools Appl.*, 2022, doi: 10.1007/s11042-020-10086-2.
 - [45] A. Noumida and R. Rajan, “Multi-label bird species classification from audio recordings using attention framework,” *Appl. Acoust.*, 2022, doi: 10.1016/j.apacoust.2022.108901.
 - [46] T. Pham, X. Tao, J. Zhang, J. Yong, Y. Li, and H. Xie, “Graph-based multi-label disease prediction model learning from medical data and domain knowledge,” *Knowledge-Based Syst.*, 2022, doi: 10.1016/j.knosys.2021.107662.
 - [47] P. Tang, X. Yan, Y. Nan, S. Xiang, S. Krammer, and T. Lasser, “FusionM4Net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification,” *Med. Image Anal.*, 2022, doi: 10.1016/j.media.2021.102307.
 - [48] Y. Chen, C. Zou, and J. Chen, “Label-aware graph representation learning for multi-label image classification,” *Neurocomputing*, 2022, doi: 10.1016/j.neucom.2022.04.004.
 - [49] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Džeroski, “An extensive experimental comparison of methods for multi-label learning,” *Pattern Recognit.*, vol. 45, no. 9, pp. 3084–3104, Sep. 2012, doi: 10.1016/j.patcog.2012.03.004.
 - [50] M. L. Zhang and Z. H. Zhou, “ML-KNN: A lazy learning approach to multi-label learning,” *Pattern Recognit.*, 2007, doi: 10.1016/j.patcog.2006.12.019.
 - [51] S. Vembu and T. Gärtner, “Label Ranking Algorithms: A Survey,” in *Preference Learning*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 45–64. doi: 10.1007/978-3-642-14125-6_3.
 - [52] G. Tsoumakas and I. Vlahavas, “Random k-Labelsets: An Ensemble Method for Multilabel Classification,” in *Machine Learning: ECML 2007*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 406–417. doi: 10.1007/978-3-540-74958-5_38.
 - [53] M.-L. Zhang and Z.-H. Zhou, “ML-KNN: A lazy learning approach to multi-label learning,” *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007, doi: 10.1016/j.patcog.2006.12.019.
 - [54] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, “Learning multi-label scene classification,” *Pattern Recognit.*, vol. 37, no. 9, pp. 1757–1771, Sep. 2004, doi: 10.1016/j.patcog.2004.03.009.
 - [55] J. Read, B. Pfahringer, G. Holmes, and E. Frank, “Classifier chains for multi-label classification,” *Mach. Learn.*, vol. 85, no. 3, pp. 333–359, Dec. 2011, doi: 10.1007/s10994-011-5256-5.
 - [56] M. L. Zhang and Z. H. Zhou, “Multilabel neural networks with applications to functional genomics and text categorization,” *IEEE Trans. Knowl. Data Eng.*, 2006, doi: 10.1109/TKDE.2006.162.
 - [57] R. Grodzicki, J. Mańdziuk, and L. Wang, “Improved Multilabel Classification with Neural Networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, pp. 409–416. doi: 10.1007/978-3-540-87700-4_41.
 - [58] J. Nam, J. Kim, E. Loza Mencía, I. Gurevych, and J. Fürnkranz, “Large-Scale Multi-label Text Classification — Revisiting Neural Networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, pp. 437–452. doi: 10.1007/978-3-662-44851-9_28.
 - [59] Y. Li, Y. Song, and J. Luo, “Improving Pairwise Ranking for Multi-label Image Classification,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 1837–1845. doi: 10.1109/CVPR.2017.199.
 - [60] J. Du, Q. Chen, Y. Peng, Y. Xiang, C. Tao, and Z. Lu, “ML-Net: multi-label classification of biomedical texts with deep neural networks,” *J. Am. Med. Informatics Assoc.*, vol. 26, no. 11, pp. 1279–1285, Nov. 2019, doi: 10.1093/jamia/ocz085.
 - [61] J. Su, M. Zhu, A. Murtadha, S. Pan, B. Wen, and Y. Liu, “ZLPR: A Novel Loss for Multi-label Classification,” 2022, [Online]. Available: <http://arxiv.org/abs/2208.02955>
 - [62] A. Maxwell et al., “Deep learning architectures for multi-label classification of intelligent health risk prediction,” *BMC Bioinformatics*, 2017, doi: 10.1186/s12859-017-1898-z.
 - [63] X. Yuan, W. Li, K. Lin, and J. Hu, “A deep neural network based

- hierarchical multi-label classifier for protein function prediction,” 2019. doi: 10.1109/CITS.2019.8862034.
- [64] A. Rios, “Deep Neural Networks for Multi-Label Text Classification: Application to Coding Electronic Medical Records,” 2018. doi: <https://orcid.org/0000-0003-1781-3975>.
- [65] X. Zhang, H. Zhao, S. Zhang, and R. Li, “A novel deep neural network model for multi-label chronic disease prediction,” *Front. Genet.*, 2019, doi: 10.3389/fgene.2019.00351.
- [66] N. Jalodia, M. Taneja, and A. Davy, “A Deep Neural Network-Based Multi-Label Classifier for SLA Violation Prediction in a Latency Sensitive NFV Application,” *IEEE Open J. Commun. Soc.*, 2021, doi: 10.1109/OJCOMS.2021.3122844.
- [67] A. G. B. Gronning, T. Kacprowski, and C. Schéele, “MultiPep: A hierarchical deep learning approach for multi-label classification of peptide bioactivities,” *Biol. Methods Protoc.*, 2021, doi: 10.1093/biomethods/bpab021.
- [68] H. Cevikalp, B. Benligiray, and O. N. Gerek, “Semi-supervised robust deep neural networks for multi-label image classification,” *Pattern Recognit.*, 2020, doi: 10.1016/j.patcog.2019.107164.
- [69] Y. Gong, Y. Jia, T. K. Leung, A. Toshev, and S. Ioffe, “Deep convolutional ranking for multilabel image annotation,” 2014.
- [70] F. Wu et al., “Weakly Semi-Supervised Deep Learning for Multi-Label Image Annotation,” *IEEE Trans. Big Data*, 2015, doi: 10.1109/tbdata.2015.2497270.
- [71] H. Izadinia, B. C. Russell, A. Farhadi, M. D. Hoffman, and A. Hertzmann, “Deep classifiers from image tags in the wild,” 2015. doi: 10.1145/2814815.2814821.
- [72] Y. Kim, “Convolutional neural networks for sentence classification,” 2014. doi: 10.3115/v1/d14-1181.
- [73] J. Liu, W. C. Chang, Y. Wu, and Y. Yang, “Deep learning for extreme multi-label text classification,” 2017. doi: 10.1145/3077136.3080834.
- [74] Y. Wei et al., “HCP: A flexible CNN framework for multi-label image classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, doi: 10.1109/TPAMI.2015.2491929.
- [75] W. Shi, Y. Gong, X. Tao, and N. Zheng, “Training DCNN by combining max-margin, max-correlation objectives, and correntropy loss for multilabel image classification,” *IEEE Trans. Neural Networks Learn. Syst.*, 2018, doi: 10.1109/TNNLS.2017.2705222.
- [76] F. Zhu, H. Li, W. Ouyang, N. Yu, and X. Wang, “Learning spatial regularization with image-level supervisions for multi-label image classification,” 2017. doi: 10.1109/CVPR.2017.219.
- [77] G. Kurata, B. Xiang, and B. Zhou, “Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence,” 2016. doi: 10.18653/v1/n16-1063.
- [78] Q. Wang and T. Breckon, “A Baseline for Multi-Label Image Classification Using Ensemble Deep CNN,” *arXiv:1811.08412 [cs.CV]*. 2018.
- [79] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
- [80] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016. doi: 10.1109/CVPR.2016.90.
- [81] J.-Y. Park, Y. Hwang, D. Lee, and J.-H. Kim, “MarsNet: Multi-Label Classification Network for Images of Various Sizes,” *IEEE Access*, vol. 8, pp. 21832–21846, 2020, doi: 10.1109/ACCESS.2020.2969217.
- [82] Y. Gong, Q. Ke, M. Isard, and S. Lazebnik, “A multi-view embedding space for modeling internet images, tags, and their semantics,” *Int. J. Comput. Vis.*, 2014, doi: 10.1007/s11263-013-0658-4.
- [83] J. Weston, S. Bengio, and N. Usunier, “WSABIE: Scaling up to large vocabulary image annotation,” 2011. doi: 10.5591/978-1-57735-516-8/IJCAI11-460.
- [84] A. Frome et al., “DeViSE: A deep visual-semantic embedding model,” 2013.
- [85] F. Z. Unal, M. S. Guzel, E. Bostanci, K. Acici, and T. Asuroglu, “Multilabel Genre Prediction Using Deep-Learning Frameworks,” *Appl. Sci.*, 2023, doi: 10.3390/app13158665.
- [86] K. Anuratha and M. Parvathy, “Multi-label Emotion Classification of COVID-19 Tweets with Deep Learning and Topic Modelling,” *Comput. Syst. Sci. Eng.*, 2023, doi: 10.32604/csse.2023.031553.
- [87] Z. Zhong, M. Hirano, K. Shimada, K. Tateishi, S. Takahashi, and Y. Mitsufuji, “An Attention-Based Approach to Hierarchical Multi-Label Music Instrument Classification,” 2023. doi: 10.1109/icassp49357.2023.10095162.
- [88] M. Ganeshkumar, V. Ravi, V. Sowmya, E. A. Gopalakrishnan, and K. P. Soman, “Explainable Deep Learning-Based Approach for Multilabel Classification of Electrocardiogram,” *IEEE Trans. Eng. Manag.*, 2023, doi: 10.1109/TEM.2021.3104751.
- [89] C. Shen, M. Yi, and M. Fu, “An improved multi-label classification algorithm based on YOLOV5s,” 2023. doi: 10.1109/NNICE58320.2023.10105674.
- [90] V. Chauhan, A. Tiwari, B. Venkata, and V. Naik, “Tackling over-smoothing in multi-label image classification using graphical convolution neural network,” *Evol. Syst.*, no. 0123456789, 2022, doi: 10.1007/s12530-022-09463-z.
- [91] Q. Zhang et al., “A multi-label waste detection model based on transfer learning,” *Resour. Conserv. Recycl.*, 2022, doi: 10.1016/j.resconrec.2022.106235.
- [92] M. I. Nadeem et al., “SHO-CNN: A Metaheuristic Optimization of a Convolutional Neural Network for Multi-Label News Classification,” *Electron.*, 2023, doi: 10.3390/electronics12010113.
- [93] S. Coulbaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, “Deep Convolution Neural Network sharing for the multi-label images classification,” *Mach. Learn. with Appl.*, vol. 10, p. 100422, Dec. 2022, doi: 10.1016/j.mlwa.2022.100422.
- [94] W. Liao, Y. Wang, Y. Yin, X. Zhang, and P. Ma, “Improved sequence generation model for multi-label classification via CNN and initialized fully connection,” *Neurocomputing*, 2020, doi: 10.1016/j.neucom.2019.11.074.
- [95] C. K. Yeh, W. C. Wu, W. J. Ko, and Y. C. F. Wang, “Learning deep latent spaces for multi-label classification,” 2017. doi: 10.1609/aaai.v31i1.10769.
- [96] J. Li, X. Lin, X. Rui, Y. Rui, and D. Tao, “A distributed approach toward discriminative distance metric learning,” *IEEE Trans. Neural Networks Learn. Syst.*, 2015, doi: 10.1109/TNNLS.2014.2377211.
- [97] R. Cabral, F. De La Torre, J. P. Costeira, and A. Bernardino, “Matrix completion for weakly-supervised multi-label image classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, doi: 10.1109/TPAMI.2014.2343234.
- [98] M. Cissé, N. Usunier, T. Artieres, and P. Gallinari, “Robust Bloom filters for large multilabel classification tasks,” 2013.
- [99] J. Zhou et al., “Graph neural networks: A review of methods and applications,” *AI Open*. 2020. doi: 10.1016/j.aiopen.2021.01.001.
- [100] Z. M. Chen, X. S. Wei, P. Wang, and Y. Guo, “Multi-label image recognition with graph convolutional networks,” 2019. doi: 10.1109/CVPR.2019.00532.
- [101] T. Chen, M. Xu, X. Hui, H. Wu, and L. Lin, “Learning semantic-specific graph representation for multi-label image recognition,” 2019. doi: 10.1109/ICCV.2019.00061.
- [102] Y. Wang et al., “Multi-label classification with label graph superimposing,” 2020. doi: 10.1609/aaai.v34i07.6909.
- [103] P. Tang, M. Jiang, B. Xia, J. W. Pitera, J. Welsner, and N. V. Chawla, “Multi-label patent categorization with non-local attention-based graph convolutional network,” 2020. doi: 10.1609/aaai.v34i05.6435.
- [104] N. Ghamrawi and A. McCallum, “Collective multi-label classification,” 2005. doi: 10.1145/1099554.1099591.
- [105] Y. Guo and S. Gu, “Multi-label classification using conditional dependency networks,” 2011. doi: 10.5591/978-1-57735-516-8/IJCAI11-220.
- [106] X. Xue, W. Zhang, J. Zhang, B. Wu, J. Fan, and Y. Lu, “Correlative multi-label multi-instance image annotation,” 2011. doi: 10.1109/ICCV.2011.6126300.
- [107] X. Li, F. Zhao, and Y. Guo, “Multi-label image classification with a probabilistic label enhancement model,” 2014.
- [108] X. Zou, S. Wu, E. M. Bakker, and X. Wang, “Multi-label enhancement based self-supervised deep cross-modal hashing,” *Neurocomputing*, 2022, doi: 10.1016/j.neucom.2021.09.053.
- [109] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks.” In *Advances in neural information processing systems*, IEEE Trans. Pattern Anal. Mach. Intell., 2017, doi: 10.1109/TPAMI.2016.2577031.
- [110] H. Yang, J. T. Zhou, Y. Zhang, B. Bin Gao, J. Wu, and J. Cai, “Exploit bounding box annotations for multi-label object recognition,” 2016. doi: 10.1109/CVPR.2016.37.
- [111] A. Elisseeff and J. Weston, “A kernel method for multi-labelled classification,” 2002. doi: 10.7551/mitpress/1120.003.0092.
- [112] S. F. Chen, Y. C. Chen, C. K. Yeh, and Y. C. F. Wang, “Order-free RNN with visual attention for multi-label classification,” 2018.
- [113] V. Oguz Yazici, A. Gonzalez-Garcia, A. Ramisa, B. Twardowski, and J. van de Weijer, “Orderless Recurrent Models for Multi-Label Classification,” 2020. doi: 10.1109/cvpr42600.2020.01345.
- [114] J. Nam, E. L. Mencia, H. J. Kim, and J. Fürnkranz, “Maximizing subset

- accuracy with recurrent neural networks in multi-label classification,” 2017.
- [115] L. Li, S. Wang, S. Jiang, and Q. Huang, “Attentive recurrent neural network for weak-supervised multi-label image classification,” 2018. doi: 10.1145/3240508.3240649.
- [116] Y. Yan, Y. Wang, W. C. Gao, B. W. Zhang, C. Yang, and X. C. Yin, “LSTM 2: Multi-Label Ranking for Document Classification,” *Neural Process. Lett.*, 2018, doi: 10.1007/s11063-017-9636-0.
- [117] P. Yang, X. Sun, W. Li, S. Ma, W. Wu, and H. Wang, “SGM: Sequence Generation Model for Multi-label Classification,” Jun. 2018.
- [118] Y. Y. Yang, Y. A. Lin, H. M. Chu, and H. T. Lin, “Deep Learning with a Rethinking Structure for Multi-label Classification,” 2019.
- [119] W. Zhang, D. K. Jha, E. Laftchiev, and D. Nikovski, “Multi-label Prediction in Time Series Data using Deep Neural Networks,” Jan. 2020.
- [120] L. Nanni, L. Trambaiollo, S. Brahnam, X. Guo, and C. Woolsey, “Ensemble of Networks for Multilabel Classification,” *Signals*, 2022, doi: 10.3390/signals3040054.
- [121] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, “Learning to diagnose with LSTM recurrent neural networks,” 2016.
- [122] S. Feng, H. Li, and J. Qiao, “Hierarchical multi-label classification based on LSTM network and Bayesian decision theory for LncRNA function prediction,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–19, 2022, doi: 10.1038/s41598-022-09672-1.
- [123] S. Verma, S. Singh, and A. Majumdar, “Multi-label LSTM autoencoder for non-intrusive appliance load monitoring,” *Electr. Power Syst. Res.*, vol. 199, no. November 2020, p. 107414, 2021, doi: 10.1016/j.epsr.2021.107414.
- [124] X. Shen, W. Liu, Y. Luo, Y.-S. Ong, and I. W. Tsang, “Deep Discrete Prototype Multilabel Learning,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Jul. 2018, pp. 2675–2681. doi: 10.24963/ijcai.2018/371.
- [125] K. Halder, L. Poddar, and M. Y. Kan, “Cold Start Thread Recommendation as Extreme Multi-label Classification,” 2018. doi: 10.1145/3184558.3191659.
- [126] S. Chen and W. Guo, “Auto-Encoders in Deep Learning—A Review with New Perspectives,” *Mathematics*. 2023. doi: 10.3390/math11081777.
- [127] J. Bai, S. Kong, and C. Gomes, “Disentangled variational autoencoder based multi-label classification with covariance-aware multivariate probit model,” 2020. doi: 10.24963/ijcai.2020/595.
- [128] C. Chen, H. Wang, W. Liu, X. Zhao, T. Hu, and G. Chen, “Two-stage label embedding via neural factorization machine for multi-label classification,” 2019. doi: 10.1609/aaai.v33i01.33013304.
- [129] J. Bai, S. Kong, and C. Gomes, “Gaussian Mixture Variational Autoencoder with Contrastive Learning for Multi-Label Classification,” 2022.
- [130] M. Huang et al., “Supervised representation learning for multi-label classification,” *Mach. Learn.*, 2019, doi: 10.1007/s10994-019-05783-5.
- [131] Y. Zhu, Y. Yang, Y. Li, J. Qiang, Y. Yuan, and R. Zhang, “Representation Learning with Dual Autoencoder for Multi-Label Classification,” *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3096194.
- [132] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2019.
- [133] R. Yarullin and P. Serdyukov, “BERT for Sequence-to-Sequence Multi-label Text Classification,” 2021. doi: 10.1007/978-3-030-72610-2_14.
- [134] J. Gong et al., “Hierarchical Graph Transformer-Based Deep Learning Model for Large-Scale Multi-Label Text Classification,” *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2972751.
- [135] W. Liu, J. Pang, N. Li, X. Zhou, and F. Yue, “Research on Multi-label Text Classification Method Based on tALBERT-CNN,” *Int. J. Comput. Intell. Syst.*, 2021, doi: 10.1007/s44196-021-00055-4.
- [136] S. Liu, L. Zhang, X. Yang, H. Su, and J. Zhu, “Query2Label: A Simple Transformer Way to Multi-Label Classification,” 2021, [Online]. Available: <http://arxiv.org/abs/2107.10834>
- [137] T. Ridnik, G. Sharir, A. Ben-Cohen, E. Ben-Baruch, and A. Noy, “ML-Decoder: Scalable and Versatile Classification Head,” 2023. doi: 10.1109/WACV56688.2023.00012.
- [138] J. Melsbach, S. Stahlmann, S. Hirschmeier, and D. Schoder, “Triplet transformer network for multi-label document classification,” 2022. doi: 10.1145/3558100.3563843.
- [139] J. Yuan et al., “Graph Attention Transformer Network for Multi-label Image Classification,” *ACM Trans. Multimed. Comput. Commun. Appl.*, 2023, doi: 10.1145/3578518.
- [140] Y. Sagama and A. Alamsyah, “Multi-Label Classification of Indonesian Online Toxicity using BERT and RoBERTa,” 2023. doi: 10.1109/IAICT59002.2023.10205892.
- [141] Z. Sun, L. Qu, J. Luo, Z. Song, and M. Wang, “Label correlation transformer for automated chest X-ray diagnosis with reliable interpretability,” *Radiol. Medica*, 2023, doi: 10.1007/s11547-023-01647-0.
- [142] M. A. Rodriguez, H. Almarzouqi, and P. Liatsis, “Multi-Label Retinal Disease Classification Using Transformers,” *IEEE J. Biomed. Heal. Informatics*, 2023, doi: 10.1109/JBHI.2022.3214086.
- [143] R. Chaib, N. Azizi, N. E. Hammami, I. Gasmi, D. Schwab, and A. Chaib, “GL-LSTM Model For Multi-label Text Classification Of Cardiovascular Disease Reports,” 2022. doi: 10.1109/IRASET52964.2022.9738147.
- [144] X. Cheng et al., “MLTR: Multi-Label Classification with Transformer,” 2022. doi: 10.1109/ICME52920.2022.9860016.
- [145] J. Li et al., “Multi-label text classification via hierarchical Transformer-CNN,” 2022. doi: 10.1145/3529836.3529912.
- [146] L. Duan, Q. You, X. Wu, and J. Sun, “Multilabel Text Classification Algorithm Based on Fusion of Two-Stream Transformer,” *Electron.*, 2022, doi: 10.3390/electronics1142138.
- [147] V. Mittal, D. Gangodkar, and B. Pant, “Deep Graph-Long Short-Term Memory: A Deep Learning Based Approach for Text Classification,” *Wirel. Pers. Commun.*, 2021, doi: 10.1007/s11277-021-08331-4.
- [148] Z. M. Chen, Q. Cui, B. Zhao, R. Song, X. Zhang, and O. Yoshie, “SST: Spatial and Semantic Transformers for Multi-Label Image Recognition,” *IEEE Trans. Image Process.*, 2022, doi: 10.1109/TIP.2022.3148867.
- [149] L. Wei, Y. Li, Y. Zhu, B. Li, and L. Zhang, “Prompt Tuning for Multi-Label Text Classification: How to Link Exercises to Knowledge Concepts?,” *Appl. Sci.*, 2022, doi: 10.3390/app122010363.
- [150] P. D. Thushari, S. Niazi, and S. Meena, “Transfer Learning Approach to Multilabel Biomedical Literature Classification using Transformer Models,” 2023. doi: 10.1109/I2CT57861.2023.10126262.
- [151] M. A. Azad, A. Mohammed, M. Waszak, B. Elvesæter, and M. Ludvigsen, “Multi-label Video Classification for Underwater Ship Inspection,” *Ocean. 2023 - Limerick, Ocean. Limerick 2023*, 2023, doi: 10.1109/OCEANSLimerick52467.2023.10244578.
- [152] J. Xiao, Y. Bai, A. Yuille, and Z. Zhou, “Delving into Masked Autoencoders for Multi-Label Thorax Disease Classification,” 2023. doi: 10.1109/WACV56688.2023.00358.
- [153] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, “TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation,” 2009. doi: 10.1109/ICCV.2009.5459266.
- [154] A. Makadia, V. Pavlovic, and S. Kumar, “A new baseline for image annotation,” 2008. doi: 10.1007/978-3-540-88690-7-24.
- [155] Y. Verma and C. V. Jawahar, “Image Annotation by Propagating Labels from Semantic Neighbourhoods,” *Int. J. Comput. Vis.*, 2017, doi: 10.1007/s11263-016-0927-0.
- [156] B. Hariharan, L. Zelnik-Manor, S. V. N. Vishwanathan, and M. Varma, “Large scale max-margin multi-label classification with priors,” 2010.
- [157] S. S. Bucak, R. Jin, and A. K. Jain, “Multi-label learning with incomplete class assignments,” 2011. doi: 10.1109/CVPR.2011.5995734.
- [158] H. Hu, G. T. Zhou, Z. Deng, Z. Liao, and G. Mori, “Learning Structured Inference Neural Networks with Label Relations,” 2016. doi: 10.1109/CVPR.2016.323.
- [159] T. Lan and G. Mori, “A max-margin riffled independence model for image tag ranking,” 2013. doi: 10.1109/CVPR.2013.399.
- [160] Q. Li, M. Qiao, W. Bian, and D. Tao, “Conditional Graphical Lasso for Multi-label Image Classification,” 2016. doi: 10.1109/CVPR.2016.325.
- [161] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [162] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, “CNN-RNN: A Unified Framework for Multi-label Image Classification,” 2016. doi: 10.1109/CVPR.2016.251.
- [163] J. Jin and H. Nakayama, “Annotation order matters: Recurrent Image Annotator for arbitrary length image tagging,” 2016. doi: 10.1109/ICPR.2016.7900004.
- [164] F. Liu, T. Xiang, T. M. Hospedales, W. Yang, and C. Sun, “Semantic regularisation for recurrent image annotation,” 2017. doi: 10.1109/CVPR.2017.443.
- [165] G. Chen, D. Ye, Z. Xing, J. Chen, and E. Cambria, “Ensemble application of convolutional and recurrent neural networks for multi-label text categorization,” 2017. doi: 10.1109/IJCNN.2017.7966144.
- [166] C. Zanolini and J. Andress, “Exploring CNN-RNN Architectures for

- Multilabel Classification of the Amazon,” *Med. World News*, 1979.
- [167] Y. Hua, L. Mou, and X. X. Zhu, “Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification,” *ISPRS J. Photogramm. Remote Sens.*, 2019, doi: 10.1016/j.isprsjprs.2019.01.015.
- [168] Y. Li, H. Xu, M. Bian, and J. Xiao, “Attention based CNN-convlstm for pedestrian attribute recognition,” *Sensors (Switzerland)*, 2020, doi: 10.3390/s20030811.
- [169] T. Chen, Z. Wang, G. Li, and L. Lin, “Recurrent Attentional Reinforcement Learning for Multi-label Image Recognition,” 2018.
- [170] X. Wu et al., “CTransCNN: Combining transformer and CNN in multilabel medical image classification,” *Knowledge-Based Syst.*, vol. 281, p. 111030, Dec. 2023, doi: 10.1016/j.knsys.2023.111030.
- [171] G. Lu, Y. Liu, J. Wang, and H. Wu, “CNN-BiLSTM-Attention: A multi-label neural classifier for short texts with a small set of labels,” *Inf. Process. Manag.*, 2023, doi: 10.1016/j.ipm.2023.103320.
- [172] H. Dhole, V. Susheela Devi, R. Aparna, and R. Raj, “Multilabel Image classification using optimized ensemble deep learning,” 2022, doi: 10.1109/AIC55036.2022.9848954.
- [173] Y. D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S. H. Wang, “Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network,” *Inf. Process. Manag.*, 2021, doi: 10.1016/j.ipm.2020.102439.
- [174] Q. Wang, N. Jia, and T. P. Breckon, “A Baseline for Multi-Label Image Classification Using an Ensemble of Deep Convolutional Neural Networks,” 2019, doi: 10.1109/ICIP.2019.8803793.
- [175] A. Dutta, Y. Verma, and C. V. Jawahar, “Recurrent Image Annotation with Explicit Inter-label Dependencies,” 2020, pp. 191–207, doi: 10.1007/978-3-030-58526-6_12.
- [176] Y. Wang, “Cross-modal fusion for multi-label image classification with attention mechanism ☆,” vol. 101, no. July, pp. 1–7, 2022.
- [177] P. S. Hossain, K. Kim, J. Uddin, M. A. Samad, and K. Choi, “Enhancing Taxonomic Categorization of DNA Sequences with Deep Learning: A Multi-Label Approach,” *Bioengineering*, vol. 10, no. 11, p. 1293, Nov. 2023, doi: 10.3390/bioengineering10111293.
- [178] Q. Liang, P. W. Bible, Y. Liu, B. Zou, and L. Wei, “DeepMicrobes: taxonomic classification for metagenomics with deep learning,” *NAR Genomics Bioinforma.*, 2020, doi: 10.1093/nargab/lqaa009.
- [179] I. Katakis, G. Tsoumakas, and I. Vlahavas, “Multilabel Text Classification for Automated Tag Suggestion,” *Proc. ECMLPKDD 2008 Discov. Chall.*, 2008.
- [180] E. Loza Mencía and J. Fürnkranz, “Efficient pairwise multilabel classification for large-scale problems in the legal domain,” 2008, doi: 10.1007/978-3-540-87481-2_4.
- [181] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, “RCV1: A new benchmark collection for text categorization research,” *J. Mach. Learn. Res.*, 2004.
- [182] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth, “Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary,” 2002, doi: 10.1007/3-540-47979-1_7.
- [183] T. S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, “NUS-WIDE: A real-world web image database from National University of Singapore,” 2009, doi: 10.1145/1646396.1646452.
- [184] F. Briggs et al., “The 9th annual MLSP competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment,” 2013, doi: 10.1109/MLSP.2013.6661934.
- [185] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, “Semantic annotation and retrieval of music and sound effects,” *IEEE Trans. Audio, Speech Lang. Process.*, 2008, doi: 10.1109/TASL.2007.913750.
- [186] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, “Multi-label classification of music by emotion,” *EURASIP J. Audio, Speech, Music Process.*, vol. 2011, no. 1, p. 4, Dec. 2011, doi: 10.1186/1687-4722-2011-426793.
- [187] W. Zhang, F. Liu, L. Luo, and J. Zhang, “Predicting drug side effects by multi-label learning and ensemble learning,” *BMC Bioinformatics*, 2015, doi: 10.1186/s12859-015-0774-y.
- [188] S. Diplaris, G. Tsoumakas, P. A. Mitkas, and I. Vlahavas, “Protein classification with multiple algorithms,” 2005, doi: 10.1007/11573036_42.
- [189] H. Blockeel, S. Džeroski, and J. Grbović, “Simultaneous prediction of multiple chemical parameters of river water quality with TILDE,” 1999, doi: 10.1007/978-3-540-48247-5_4.
- [190] Y. Wang, Y. Xie, Y. Liu, K. Zhou, and X. Li, “Fast Graph Convolution Network Based Multi-label Image Recognition via Cross-modal Fusion,” 2020, doi: 10.1145/3340531.3411880.
- [191] B. Chen, Z. Zhang, Y. Lu, F. Chen, G. Lu, and D. Zhang, “Semantic-Interactive Graph Convolutional Network for Multilabel Image Recognition,” *IEEE Trans. Syst. Man, Cybern. Syst.*, 2022, doi: 10.1109/TSMC.2021.3103842.
- [192] J. Lanchantin, T. Wang, V. Ordóñez, and Y. Qi, “General Multi-label Image Classification with Transformers,” 2021, doi: 10.1109/CVPR46437.2021.01621.
- [193] P. Prajapati and A. Thakkar, “Performance improvement of extreme multi-label classification using K-way tree construction with parallel clustering algorithm,” *J. King Saud Univ. - Comput. Inf. Sci.*, 2022, doi: 10.1016/j.jksuci.2021.02.014.
- [194] B. Wang, L. Chen, W. Sun, K. Qin, K. Li, and H. Zhou, “Ranking-based autoencoder for extreme multi-label classification,” 2019, doi: 10.18653/v1/n19-1289.
- [195] K. Dahiya et al., “DeepXML: A Deep Extreme Multi-Label Learning Framework Applied to Short Text Documents,” 2021, doi: 10.1145/3437963.3441810.
- [196] S. Jo, D. Shin, B. Na, J. Jang, and I.-C. Moon, “Hierarchical Multi-Label Classification with Partial Labels and Unknown Hierarchy,” in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Oct. 2023, pp. 1025–1034, doi: 10.1145/3583780.3614912.
- [197] C. W. Lee, W. Fang, C. K. Yeh, and Y. C. F. Wang, “Multi-label Zero-Shot Learning with Structured Knowledge Graphs,” 2018, doi: 10.1109/CVPR.2018.00170.
- [198] R. Zhang, S. Luo, L. Pan, Y. Ma, and Z. Wu, “Strengthened multiple correlation for multi-label few-shot intent detection,” *Neurocomputing*, 2023, doi: 10.1016/j.neucom.2022.12.047.
- [199] K. Zhu, M. Fu, and J. Wu, “Multi-Label Self-Supervised Learning with Scene Images,” pp. 6694–6703, 2023, [Online]. Available: <http://arxiv.org/abs/2308.03286>
- [200] Tarekegn, Adane Nega, Krzysztof Michalak, and Mario Giacobini. “Cross-validation approach to evaluate clustering algorithms: An experimental study using multi-label datasets.” *SN Computer Science* 1 (2020): 1-9. doi: <https://doi.org/10.1007/s42979-020-00283-z>
- [201] G. Wu, Y. Tian, and D. Liu, “Cost-sensitive multi-label learning with positive and negative label pairwise correlations,” *Neural Networks*, 2018, doi: 10.1016/j.neunet.2018.09.003.
- [202] S. Pouyanfar, T. Wang, and S. C. Chen, “A Multi-label Multimodal Deep Learning Framework for Imbalanced Data Classification,” 2019, doi: 10.1109/MIPR.2019.00043.
- [203] Z. He, Y. Lin, C. Wang, L. Guo, and W. Ding, “Multi-Label Feature Selection Based on Correlation Label Enhancement,” *Inf. Sci. (N.Y.)*, 2023, doi: 10.1016/j.ins.2023.119526.
- [204] Y. Lin, H. Liu, H. Zhao, Q. Hu, X. Zhu, and X. Wu, “Hierarchical Feature Selection Based on Label Distribution Learning,” *IEEE Trans. Knowl. Data Eng.*, 2023, doi: 10.1109/TKDE.2022.3177246.
- [205] W. Siblini, P. Kuntz, and F. Meyer, “A Review on Dimensionality Reduction for Multi-Label Classification,” *IEEE Trans. Knowl. Data Eng.*, 2021, doi: 10.1109/TKDE.2019.2940014.
- [206] W. Qian, J. Huang, F. Xu, W. Shu, and W. Ding, “A survey on multi-label feature selection from perspectives of label fusion,” *Information Fusion*, 2023, doi: 10.1016/j.inffus.2023.101948.
- [207] H. K. Maragheh, F. S. Gharehchopogh, K. Majidzadeh, and A. B. Sangar, “A New Hybrid Based on Long Short-Term Memory Network with Spotted Hyena Optimization Algorithm for Multi-Label Text Classification,” *Mathematics*, 2022, doi: 10.3390/math10030488.
- [208] J. Zhao, Y. Zhao, and J. Li, “M3TR: Multi-modal Multi-label Recognition with Transformer,” 2021, doi: 10.1145/3474085.3475191.
- [209] W. Yang, J. Li, F. Fukumoto, and Y. Ye, “HSCNN: A hybrid-siamese convolutional neural network for extremely imbalanced multi-label text classification,” 2020, doi: 10.18653/v1/2020.emnlp-main.545.
- [210] X. Wang and Y. Zhou, “A Review of Multi-Label Feature Selection,” 2023, doi: 10.1088/1742-6596/2504/1/012007.
- [211] Z. Zhang, L. Liu, J. Li, and X. Wu, “Integrating Global and Local Feature Selection for Multi-Label Learning,” *ACM Trans. Knowl. Discov. Data*, 2023, doi: 10.1145/3532190.



Adane Nega Tarekegn received his Ph.D. in modeling and data science from the University of Turin, Italy in 2020 and his master's degree in information technology from the University of Gondar in 2015 and his bachelor's degree in computer science from Bahir Dar University, Ethiopia, 2008. He is currently a postdoc at MediaFutures, University of Bergen, Norway. He has been the head of the AI and multimedia chair and has worked as an assistant professor at Bahir Dar University. His research interests include machine learning, computer vision, deep learning, and pattern recognition. In addition, he has been a reviewer for international conferences and prestigious journals such as IEEE Transactions on Multimedia, Applied Soft Computing, and JMIR Internet Research.



Mohib Ullah (Member, IEEE) received his bachelor's degree in electronic and computer engineering from the Politecnico di Torino, Italy, in 2012, and his master's degree in telecommunication engineering from the University of Trento, Italy, in 2015. Subsequently, he earned his Ph.D. in computer science from NTNU, Norway, in 2019. He is currently a researcher at the NTNU and participates in several industrial projects, including machine learning, data analysis, and video surveillance. He has also taught machine learning and video processing courses. He has published over 75 peer-reviewed articles in reputable journals, conferences, and workshops. His research interests include machine learning, computer vision, and image analyses. In addition to his research activities, he has participated in various academic conferences, serving as a Program Committee Member for the International Workshop on Computer Vision in Sports (CVsports) and a Chair for the Technical Program at the European Workshop on Visual Information Processing. He is also a reviewer for prestigious journals and conferences, such as Neurocomputing (Elsevier), Neural Computing and Applications (Elsevier), Multimedia Tools and Applications (Spring), IEEE ACCESS, ACM Transactions of Computer Systems, IEEE CVPRW, IEEE ICIP, and IEEE AVSS.



Faouzi Alaya Cheikh (Senior Member, IEEE) received a Ph.D. degree in Information Technology from Tampere University of Technology, Tampere, Finland, in April 2004. Since 1994, he has been a Researcher with the Signal Processing Algorithm Group, at Tampere University of Technology. Since 2006, he has been with the Department of Computer Science, Gjøvik University College, Gjøvik, Norway, as an associate professor. Since January 2016, he has been with NTNU in Trondheim, Norway. He has taught courses in image and video processing and analysis as well as media security. He is currently a co-supervisor with nine PhD students. He has been involved in several European and national projects such as ESPRIT, NOBLESS, COST 211Quat, HyPerCept, IQ-Med, H2020 ITN HiPerNav, and H2020 RIA ALAMEDA. His research interests include e-learning, 3-D imaging, image/video processing and analysis, video-based navigation, biometrics, pattern recognition, embedded systems, and content-based image retrieval. He has published more than 150 journal and conference articles and has supervised four postdoctoral researchers, nine PhDs, and several M.Sc. thesis projects. Dr. Cheikh is an expert reviewer for several scientific journals and conferences. He is on the Editorial Board of the IET Image Processing and the International Journal of Advanced Robotics and Automation, and technical committees of several international conferences.