Biomedical Signal Processing and Control 91 (2024) 105873

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

MCICT: Graph convolutional network-based end-to-end model for multi-label classification of imbalanced clinical text

Yao He [a,b], Qingyu Xiong [a,b,*], Cai Ke [b], Yaqiang Wang [c], Zhengyi Yang [b], Hualing Yi [b], Qilin Fan [b]

[a] Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, Chongqing, 401331, China
[b] School of Big Data and Software Engineering, Chongqing University, Chongqing, 401331, China
[c] School of Software Engineering, Chengdu University of Information Technology, Chengdu, 610225, China

ARTICLE INFO

Keywords:
Multi-label text classification
Clinical text classification
Graph convolutional neural network
Label co-occurrence
Additional information

ABSTRACT

The rapid growth of clinical text data requires accurate and powerful automated classification methods to support medical decision making and personalized healthcare. The multi-label classification task for clinical texts is designed to assign the most relevant set of labels to each clinical text. However, this task presents two significant challenges: (1) how to accurately extract fine-grained semantic features from complex clinical texts, and (2) how to effectively mitigate the issue of label class imbalance. To overcome these problems, we innovatively propose a novel Multi-label Classification of Imbalanced Clinical Text (MCICIT) model. In order to obtain fine-grained semantic features from clinical texts, we utilize the specialized pre-trained language model BioBERT, tailored for biomedical texts. To tackle the challenge of label class imbalance, we present a Co-occurrence Based and Embeddings with Additional Information Enhanced Graph Convolutional Network (CoEAI-GCN) module. On one hand, we enrich the label content by incorporating additional information to acquire more accurate word embeddings as the feature matrix. On the other hand, we combine the co-occurrence relationship of labels to construct a correlation matrix. Ultimately, label representations are learned through a graph convolutional network. By conducting multi-label classification experiments on two clinical text datasets extracted from real medical systems, our model achieves a 3.2% and 0.5% improvement in F1 scores, respectively, compared to state-of-the-art deep learning models. Additionally, we conduct ablation studies to explore the behaviors of the proposed model. These results together demonstrate that our proposed MCICT effectively enhances the classification performance of imbalanced clinical texts.

1. Introduction

With the continuous progress in the medical field and the rapid development of information technology, the quantity of clinical notes and electronic health records has been rapidly increasing [1]. These rich clinical medical data often contain essential information such as symptom descriptions, diagnostic results, treatment plans, etc. However, they usually exist in an unstructured form. Therefore, how to extract valuable knowledge from these massive unstructured data is of great significance.

Clinical medical text analysis is a very important research field [2]. Through systematic analysis and mining of these data using natural language processing (NLP) techniques [3], a profound understanding of diseases, treatment methods, and patient health can be obtained. Such data analysis can reveal hidden patterns, trends, and correlations, providing strong support for medical decision-making [4] and personalized healthcare [5]. Moreover, clinical text automatic classification can automatically classify text data into multiple related diagnostic result categories. This facilitates automated assistance for doctors in rapid diagnosis, treatment planning, and decision-making, thereby enhancing diagnostic efficiency and accuracy.

Clinical text automatic classification mainly includes two forms: multi-class classification and multi-label classification. Multi-class classification aims to divide text data into mutually exclusive categories, where each text can only belong to one category. On the other hand, multi-label classification allows a piece of text to be associated with multiple label categories, which better reflects the complexity and diversity of medical diagnoses in real-world scenarios.

Although previous research has made some progress in the field of multi-label clinical text classification, two major challenges remain.

Firstly, class imbalance is a common problem in this context [6,7], especially when confronted with a substantial number of labels. Taking datasets of Traditional Chinese Medicine (TCM) and Preoperative Physical Examination (PPE) as examples, they are two typical imbalanced

* Corresponding author.
E-mail addresses: yaohe@stu.cqu.edu.cn (Y. He), xiong03@cqu.edu.cn (Q. Xiong), caike@cqu.edu.cn (C. Ke), yaqwang@cuit.edu.cn (Y. Wang), zyyang@cqu.edu.cn (Z. Yang), yihualing@cqu.edu.cn (H. Yi), fanqilin@cqu.edu.cn (Q. Fan).

https://doi.org/10.1016/j.bspc.2023.105873
Received 20 October 2023; Received in revised form 2 December 2023; Accepted 21 December 2023
Available online 3 January 2024

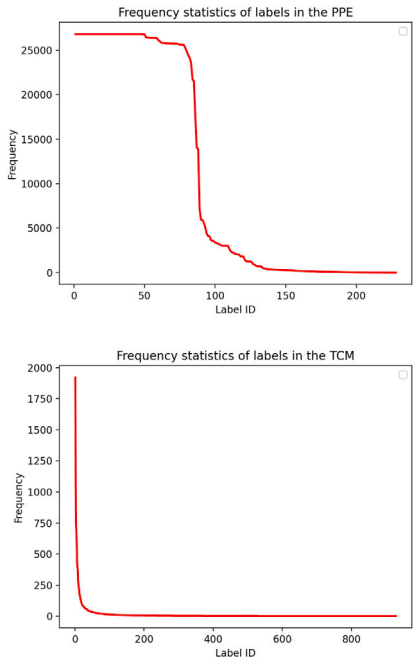

Fig. 1. Frequency statistics of labels in PPE and TCM.

clinical datasets collected from real medical systems. As shown in Fig. 1, we notice that a few of labels are linked to a large number of documents, while most labels are connected to a limited number of documents. This difference stems from the fact that certain disease or examination labels may be relatively common, while others may be rare. As a result, the model may exhibit an inappropriate bias for majority classes, resulting in poor classification performance for the minority classes and compromising the accuracy of the overall results.

Secondly, accurately extracting fine-grained semantic features from original clinical texts is also a major challenge [8,9]. Clinical medical texts typically encompass a multitude of specialized terms and intricate semantic structures. Enhancing the classification performance of clinical texts necessitates a profound comprehension of these terms and the precise capture of their meanings within the text. Therefore, the conversion of the original text into a more precise feature representation becomes imperative to achieve this goal.

To tackle the first challenge posed by label class imbalance, we propose a Co-occurrence Based and Embeddings with Additional Information Enhanced Graph Convolutional Network (CoEAI-GCN) module. Firstly, we present an Embeddings with Additional Information (EAI) method, which involves incorporating supplementary information to obtain label embeddings, so as to get richer and more accurate label representations. Secondly, through statistical analysis and visualization, we found that there are often co-occurrence relationships between labels, as shown in Fig. 2. Therefore, we consider taking advantages of these dependencies between labels to mitigate class imbalance and thus improve classification performance. Intuitively, if minority class labels appear infrequently in the data set, but often co-occur with other common labels, these co-occurrence relationships can help the model infer these minority class labels. Consequently, we utilize label co-occurrence to construct the correlation matrix to guide the propagation of nodes in a graph convolutional network (GCN).

To address the second challenge of extracting semantic features from complex clinical texts and effectively improving classification performance, we propose a novel Multi-label Classification of Imbalanced Clinical Text (MCICT) model. For the representation learning of clinical texts, we utilize the BioBERT [10] and fine-tune it to fit specific medical domain tasks. This process allows us to extract fine-grained semantic

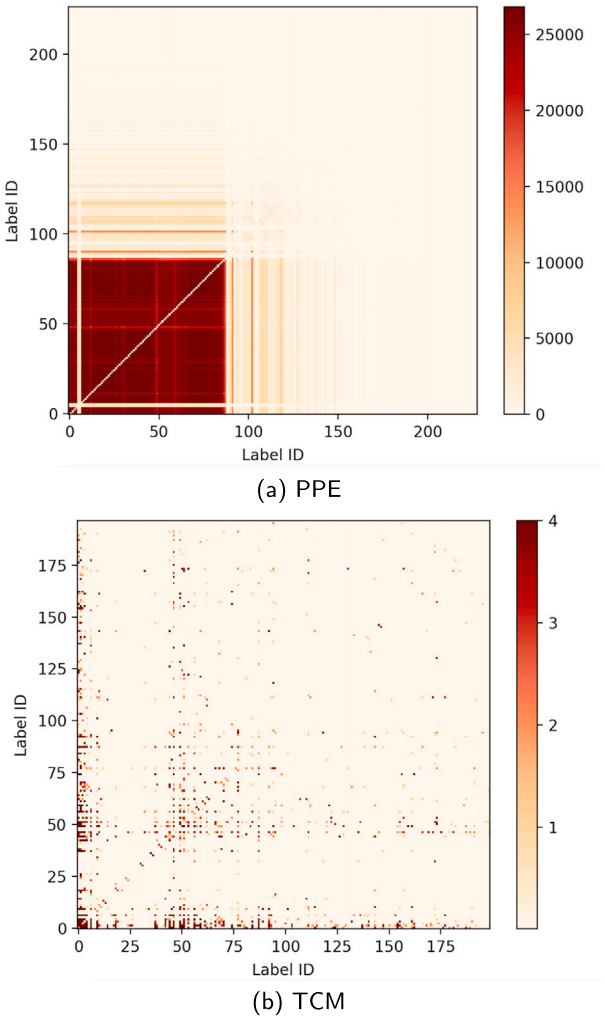

Fig. 2. Co-occurrence statistics of labels in the TCM and PPE. (Due to the large number of labels in the TCM dataset, we selected only the top two hundred most frequent labels for visualizing their co-occurrence.)

information from the texts. On the other hand, we employ the CoEAI-GCN to learn the feature representation of labels. Then, through a sequence of mapping and processing layers, the label and document representations generate prediction scores for each label and ultimately determine the most relevant set of labels for a given text.

On the whole, our research aims to extract fine-grained semantic features from clinical texts and alleviate class imbalance issues, so as to improve the multi-label classification performance of clinical medical text data. Through experiments and performance evaluations, we have demonstrated the superior performance of our proposed model on two authentic clinical text datasets.

The main contributions of this paper are as follows:

- In order to alleviate the issue of class imbalance, we present the CoEAI-GCN module based on GCN, which integrates additional information and leverages label co-occurrence;
- A novel end-to-end model, MCICT, is proposed for improving the performance of multi-label clinical text classification;
- Through a series of experiments on clinical datasets extracted from two real medical systems, our model outperforms state-of-the-art models and achieves remarkable classification performance.

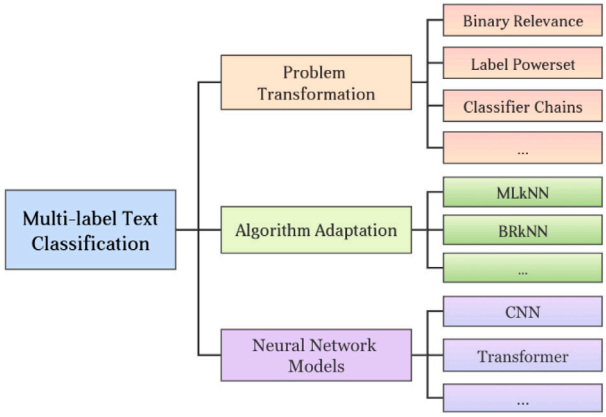

Fig. 3. Methods for multi-label text classification.

The rest of this paper is organized as follows: Section 2 introduces related research and existing methods; Section 3 provides a detailed description of our model and methods; Section 4 presents experimental results and analysis; finally, Section 5 summarizes and discusses future research directions.

2. Related work

2.1. Multi-label text classification

Text classification is one of the fundamental tasks in the field of NLP, aiming to accurately categorize and manage a large amount of text from different sources. In traditional single-label text classification tasks, each text corresponds to only one category label, and the labels are independent from each other, and the classification granularity is relatively coarse. In contrast, the task of multi-label text classification (MLTC) is more challenging because it requires assigning two or more class labels to each text, which is closer to real-world scenarios. It has been widely applied in fields such as web page tagging [11,12], question-answering systems [13], sentiment analysis [14,15], and biomedical text classification [16,17], etc. Due to the diversity of labels, complex correlations, and imbalanced sample distributions, constructing a simple and effective multi-label text classifier presents a major challenge.

At present, the mainstream approaches for MLTC can be divided into two main groups: traditional machine learning methods and deep learning methods. In the traditional field of machine learning, algorithms for MLTC can be further classified into two types: problem transformation and algorithm adaptation approaches. The methods of MLTC are shown in Fig. 3.

2.1.1. Problem transformation approaches

The aim of problem transformation methods is to develop data transformation approaches for converting multi-label classification problems into binary or multi-class classification tasks. There are several methods that fall into this category, including Binary Relevance (BR) [18], Label Powerset (LP) [19], Classifier Chains (CC) [20], Calibrated Ranking by Pairwise Comparison [21] and so on. BR is a simple and direct method that transforms multi-label classification problems into multiple independent binary classification subproblems. In other words, each label has an independent classifier, but this approach completely ignores the correlation between labels. In contrast, the LP method treats each possible combination of labels as a unique class, transforming the multi-label classification problem into a multi-class classification task. However, as the number of labels increases, this method may face the problem of label combination explosion. On the other hand, CC convert the multi-label classification problem into a Bayesian conditional chain of binary classification tasks. Overall, the problem transformation methods may require quite a bit of time and space when dealing with large-scale datasets and labels.

2.1.2. Algorithm adaptation approaches

Different from the problem transformation approaches, the algorithm adaptation approaches address multi-label classification problems by adapting or extending traditional single-label classification algorithms without data transformation. Within the framework of algorithmic adaptation, researchers have proposed several methods, such as MLkNN [22] and BRkNN [23], which are extensions of the kNN classifier, as well as IBLRML [24], which combines logistic regression with instance-based learning. Although algorithm adaptation methods can preserve correlations between labels to some extent, they are limited to utilizing first or second-order label correlations.

2.1.3. Neural network-based approaches

In recent years, with the rapid development of deep learning, MLTC algorithms based on deep neural networks have been widely concerned. Existing research on deep learning-based MLTC mainly focuses on learning enhanced document representations and modeling label dependencies to improve classification performance.

In terms of document representation, traditional feature engineering-based methods often fail to capture the complex semantics and contextual dependencies in the text. Therefore, in recent years, deep learning-based approaches have been widely applied to MLTC to learn more comprehensive and expressive document representations. These methods leverage technologies such as Convolutional Neural Networks (CNN) [9,25], Recurrent Neural Networks (RNN) [26,27], and Transformer [28] to explore the internal relationships within the text and obtain fine-grained text representations, thereby improving the effectiveness and generalization ability of MLTC. In addition, pre-trained language models such as ELMO [29], GPT [30] and BERT [31] have been introduced for MLTC tasks, as they can learn text representations with more semantic information. Furthermore, in specific domains, specialized pre-trained language models are available. For example, in the field of biomedicine, there exist pre-trained language models customized for biomedical texts, like BioBERT, which are designed to effectively handle the specialized terminology and knowledge present in medical texts.

On the other hand, the researchers also focused on modeling the dependencies between labels. In MLTC tasks, consideration of correlations between labels plays a crucial role in accurate classification. For instance, in clinical diagnosis tasks, the symptom label “fever” naturally correlates with “headache”. To capture such dependencies between labels, some studies have employed attention mechanisms [32–34], which aims to explore the semantic connections between labels and input documents, thereby learning label-specific document representations for classification. While these methods have shown promising results in MLTC, they encountered challenges in effectively distinguishing similar labels. The reason is that labels that belong to similar categories often appear together, which means that they correspond to the same documents, resulting in similar label features being extracted from these documents. As a result, these models face difficulty in accurately predicting minority class labels.

2.2. Biomedical multi-label text classification

Accurate classification of diverse biomedical texts holds crucial applications in biomedical and bioinformatics fields. These applications include classification of clinical notes [16], intent classification of medical texts [35], diagnosis of drug reaction [36], and more.

As biomedical literature rapidly expands, the demand for precise and robust automatic methods grows. These methods are crucial in effectively choosing relevant labels from a candidate set for specific biomedical documents, thereby aiding the process of knowledge

discovery. Deep learning methods have made significant advancements in biomedical MLTC tasks. Researchers have proposed various biomedical MLTC models based on deep learning. For example, MeSHProbeNet [37] was a self-attention deep learning neural network capable of predicting multiple relevant labels for biomedical articles using text information and journal titles. On the other hand, BERTMeSH [38] adopted a transfer learning strategy and utilizes BioBERT to extract information from full-text articles in PMC and titles or abstracts from MEDLINE to build the classifier. However, these methods encoded labels as one-hot vectors, resulting in increased storage and computational costs, and more significantly, overlooking the semantic correlations among labels. GHS-NET [17] was a generic hybridized shallow neural network used for biomedical MLTC. It employed CNN to extract the most discriminative features and Bi-LSTM layers to accurately capture the local features of biomedical text. However, this model did not take into account the use of the interconnected relationships between labels to enhance the classification process.

Some researchers have noticed the importance of capturing the interdependence between labels in the corpus. In 2017, Baker et al. [39] proposed a novel hierarchical biomedical MLTC method that utilized the co-occurrence relationship of corpus labels, including hypernyms, to initialize the final output layer. In 2018, Li et al. [40] introduced DeepLabeler, a neural network-based framework designed for automatic disease classification. DeepLabeler employed document vector feature representation and CNN to capture both local and global salient features. Furthermore, addressing the deficiency of label decision modules in MLTC methods, Du et al. [41] introduced a straightforward deep learning framework named ML-Net. This framework merged label prediction mechanisms with label counting prediction networks, thereby optimizing the collection of corpus labels. In 2022, Chen et al. [42] introduced the LIAR, a novel model to biomedical MLTC. This model adeptly incorporated both label independence and correlation, effectively harmonizing the two. This integration helps mitigate the challenge of imbalanced label distribution to a certain extent. Therefore, exploring and utilizing the dependencies between labels will be beneficial for biomedical MLTC tasks.

In summary, despite some progress in the field of biomedical text classification, there are still challenges in the multi-label classification of clinical medical texts. Existing methods struggle to achieve exceptional performance when dealing with slightly different types of texts [41]. For example, a method that performs well in the classification of biomedical literature may not achieve similar results in clinical record classification. Therefore, the aim of our paper is to propose a more general end-to-end model for multi-label classification of clinical texts.

3. Proposed model

As illustrated in Fig. 4, our MCICT model consists of two key modules: (1) Learning fine-grained representations of clinical document using BioBERT (2) Learning label representations using the proposed CoEAI-GCN. Ultimately, multi-label classification is achieved by generating prediction scores for multiple labels based on the learned document and label representations.

3.1. Problem formulation

Let $D = \{x_i, y_i\}^N$ be the set of clinical record documents, which consists of N document x_i and its corresponding category label $y_i \in \{0,1\}^{|C|}$, where $|C|$ denotes the total number of labels. Each document x_i contains J words $x_i = w_{i1}, w_{i2}, \ldots, w_{ij}$. The target of clinical multi-label text classification is to learn the mapping from input clinical text sequence to the most relevant labels.

3.2. Document representation learning

In this paper, we aim to improve the performance of clinical text classification tasks, and a crucial step in achieving this is to learn effective document representations. Representing clinical document involves transforming the original unstructured clinical medical document into machine-understandable and processable vector representations.

To accomplish this goal, we adopt the pre-trained language model BioBERT to extract fine-grained information from clinical document.

$$T' = f_{BioBERT}(x_i, \Theta), \tag{1}$$

x_i denotes the ith training sample, and Θ denotes pretrained model BioBERT parameters, resulting in the feature vector $T' \in R^{d_c}$.

3.3. Label representation learning

To mitigate the challenge posed by class imbalance, we adopt our novel CoEAI-GCN module to learn label representations. The module primarily comprises two components: (1) utilizing the EAI method to generate word embeddings for labels, and (2) incorporating label co-occurrence relationships into GCN learning to obtain the ultimate representation of labels.

3.3.1. Label embeddings with the EAI method

For the word embeddings of the labels, we adopt the EAI method to enrich the representation of clinical labels. First, we utilize the Wikipedia API[1] to extract the first two sentences of Wikipedia abstracts related to the input labels. These sentences serve as additional information, providing medical background and conceptual explanations for the labels, thereby enhancing the medical semantic understanding and context of the labels.

Next, we take the extracted sentences as input and use the Sentence-Transformer[2] to generate label embedding vectors, denoted as $V_{emb} \in R^{|C|*d}$. The Sentence-Transformer is based on a pre-trained language model, trained on a large-scale corpus to possess rich semantic representation capabilities.

Due to the high dimensionality of the obtained label embedding vectors at this stage, we apply Principal Component Analysis (PCA) for dimensionality reduction. PCA identifies the most significant feature directions (referred to as principal components) and maps the original data to a new low-dimensional space. The advantages of using PCA include reducing computational and storage costs, eliminating redundant information in the data, and retaining the most representative features. After this operation, we can obtain the final label embedding feature $V_{emb} \in R^{|C|*d_l}$.

Finally, we use the generated label embedding vectors, $V_{emb} \in R^{|C|*d_l}$, as the initialization node feature matrix $X \in R^{|C|*d_l}$ for the GCN.

3.3.2. GCN learning with label co-occurrence

a) Correlation matrix construction. Constructing an effective correlation matrix is crucial for the GCN model as it regulates the propagation of label information among nodes. To build the correlation matrix $A \in R^{|C|*|C|}$, we adopt a data-driven approach, approximating probabilities p(i) (representing the occurrence frequency of label i) and p(i, j) (representing the co-occurrence frequency of label i and label j) through statistical analysis of the labels' occurrence and co-occurrence counts. Specifically, the definition of each element $a_{ij} \in A$ in the correlation matrix is as follows:

$$a_{ij} = \frac{p(i,j)}{p(i)p(j)} = \frac{D_{(i,j)}D_T}{D_{(i)}D_{(j)}}, \tag{2}$$

[1] https://pypi.org/project/Wikipedia-API/
[2] https://www.sbert.net/

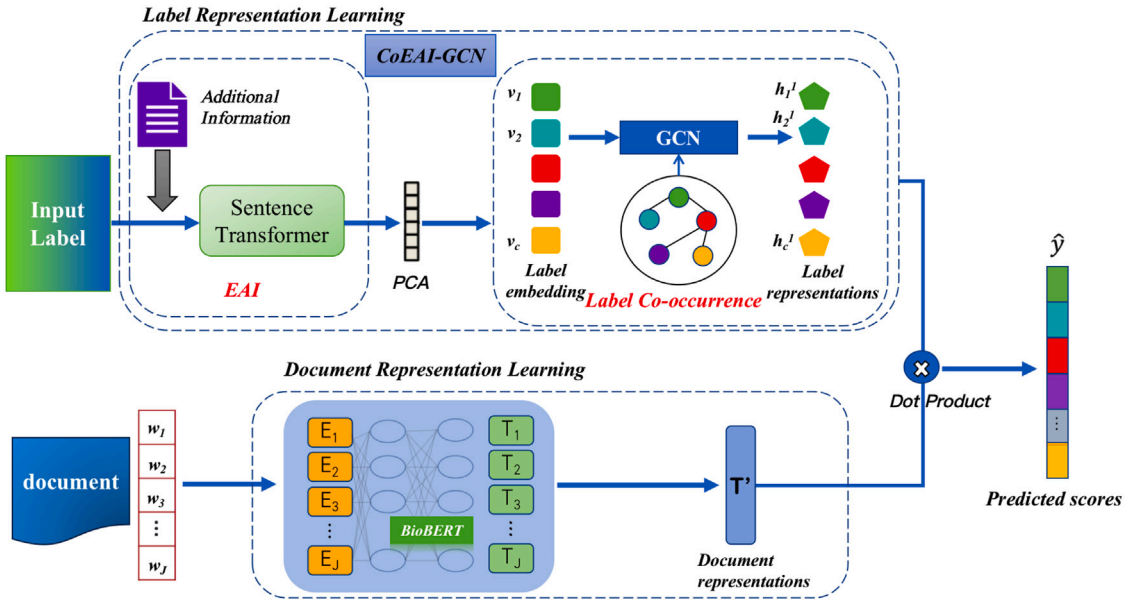

Fig. 4. The overall architecture of our MCICT model for clinical multi-label text classification.

$D_{(i,j)}$ represents the number of documents that have both labels i and j, $D_{(i)}$ represents the number of documents that have label i, and D_T is the total number of documents in the dataset. Intuitively, when the value of a_{ij} for an edge is higher, it means that the corresponding vertices connected by that edge have a higher correlation.

Constructing a correlation matrix based on the occurrences and co-occurrences of labels can help mitigate the issue of class imbalance to a certain extent. Specifically, this construction method compares the co-occurrence frequencies of labels with their respective occurrence probabilities to quantify the interrelationships between labels. As a result, for those labels that appear less frequently in the dataset but co-occur with other labels more frequently, the elements of the correlation matrix will be relatively larger, which will help predict labels with lower frequency.

b) Node updating mechanism in GCN. We utilize a GCN [43] to learn the deep relationships among label-specific semantic components guided by statistically derived label correlations. GCN is a neural network that operates on graphs, enhancing node representations by propagating messages between neighboring nodes.

Through the comparative experiments in Section 4.6, it is shown that the proposed model performs best when using a single layer of GCN. We take the component representations of the first layer H^0 (i.e., the initialized node feature matrix $X \in R^{|C|*d_l}$ obtained in Section 3.3.1) as input and output the enhanced component representations $H^1 \in R^{|C|*d_l'}$, where d_l' represents the dimensionality of the final node representations. The propagation rule between layers is as follows:

$$H^1 = \sigma(\hat{A}H^0W^0), \tag{3}$$

where $\sigma(.)$ represents the LeakyReLU [44] activation function. $W^0 \in R^{d_l*d_l'}$ is the transformation matrix to be learned. $\hat{A}$ represents the normalized adjacency matrix, and the normalization method is as follows:

$$\hat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}, \tag{4}$$

where D is a diagonal degree matrix with entries $D_{ij} = \sum_j A_{ij}$

3.4. Multi-label clinical text classification

After learning the label and document representations, the model will generate prediction scores $\hat{y}$, each comprising $|C|$ elements, with each element being a real number between 0 and 1. This is achieved by mapping the label correlations to the document representations, as shown below:

$$\hat{y} = HT'. \tag{5}$$

We use y to represent the true labels of the documents, where $y_i \in \{0, 1\}$ indicates whether label i appears in the document. The proposed model is trained using the multi-label cross-entropy loss:

$$L = \sum_{c=1}^{|C|} y^c log(\hat{y}^c) + (1 - y^c)log(1 - \hat{y}^c). \tag{6}$$

4. Experiments and results

In this section, we first describe datasets and present the baselines used for our comparison. Next, we discuss the settings and evaluation metrics employed in our study. Finally, we present a detailed comparison of the experimental results and conduct ablation studies to explore the behaviors of our model.

4.1. Datasets

To validate the effectiveness of our proposed MCICT model, we conduct experiments using two clinically relevant datasets, TCM and PPE. The detailed information of these datasets is shown in Table 1.

Traditional Chinese Medicine (TCM) dataset [45] is a collection of inquiry records extracted from real medical information systems, accumulated over years of clinical practice by traditional Chinese medicine practitioners. Each record consists of chief complaints and syndromes, as shown in Table 2. These records mostly consist of short text data, characterized by concise yet semantically rich content, with descriptive symptoms as labels.

Preoperative Physical Examination (PPE) dataset [46] comprises preoperative examination records from a medical institution in the year 2020. Each record includes preoperative diagnosis descriptions and a series of required physical examination items before surgery, as presented in Table 3. When processing this dataset, we utilize the pkuseg [47] segmentation tool specialized in the medical domain, which incorporates a dictionary containing more medical terminologies, ensuring more accurate segmentation results.

Table 1
Detailed statistics information of the datasets TCM and PPE.

Dataset	Number of instances	Number of labels	Number of labels in one instance			Number of words in one instance		
			Min	Max	Avg	Min	Max	Avg
TCM	10 000	929	1	5	1.85	1	35	8.84
PPE	34 679	228	50	120	89.94	1	591	21.64

Table 2
Sample examples of TCM.

Chief complaints	Syndrome labels
心悸，胸闷，气短，口干，不渴，左胁略胀，饮食正常，二便正常，舌暗红，形体胖，苔薄，脉时快时慢，节律不齐 (Palpitation, chest distress, breathe hard, dry mouth, hydroadipsia, left rib-side distention, normal diet, normal bowel function, dark red and swollen tongue, thin tongue fur, pulse waxes and wanes, rhythm not neat)	痰热内扰、心气不足 (Phlegm hot inside, heart-qi deficiency)
全身疼痛，气候变化加重，喜暖，恶风，畏寒，疲倦，头昏，下肢软，下腹痛，按压加重，腹胀，夜口干，舌红，苔薄黄腻少，脉软 (Generalized body pain, worsened by weather changes. Prefers warmth, aversion to wind, cold intolerance, fatigue, dizziness, weakness in lower limbs, lower abdominal pain aggravated by pressure, abdominal bloating, nocturnal dry mouth, red tongue, thin yellowish coating with scanty moisture, and a soft pulse)	脾肾阳虚、夹湿热、血瘀 (Asdthenic splenonephro-yang, dampness heat, blood stasis)

Table 3
Sample examples of PPE.

Preoperative diagnosis description	Physical examination items
风湿性心脏病 心脏瓣膜病 二尖瓣狭窄 重度 主动脉瓣狭窄 重度 反流 中度 窦性心律 心功能 Ⅲ级 (Rheumatic heart disease, heart valve disease, mitral stenosis, severe aortic stenosis, severe regurgitation, moderate sinus rhythm, heart function level III)	心脏瓣膜病、扩张性心肌病、充血性心力衰竭病史…… (Heart valve disease, dilated cardiomyopathy, history of congestive heart failure……)
胃 恶性肿瘤 慢性 支气管炎 肺气肿 肝 囊肿 胃窦 癌 (Gastric malignant tumor, chronic bronchitis, emphysema, liver cyst, gastric antrum cancer)	近2周呼吸系统感染病史、近2周内肺炎、肺不张、气管食管瘘…… (History of respiratory system infections in the past 2 weeks, pneumonia in the past 2 weeks, atelectasis, tracheoesophageal fistula……)

4.2. Baselines

Considering the scarcity of multi-label classification models for clinical texts, we select machine learning-based methods and recent deep learning-based universal models from the field of MLTC. Additionally, we choose several recently proposed multi-label classification models for biomedical texts as comparison benchmarks. This comprehensive evaluation allows us to assess the performance of our proposed MCICT model and highlight its practical application value in specific domains. Below is a brief introduction of the baseline models for comparison:

- Binary Relevance (BR) [18]: A fundamental multi-label classification method that treats each label as an independent binary classification task. For each label, a separate classifier is trained, and their predictions are combined for final prediction.
- Classifier Chains (CC) [20]: A multi-label classification method based on a chain-like structure. It considers the dependency between labels by using the predictions of preceding labels as inputs for subsequent label classification.
- Label Powerset (LP) [19]: A common approach in multi-label classification, where each possible combination of labels is treated as a separate category. By transforming the multi-label classification problem into a multi-class classification problem, the LP model can capture the correlations among labels.
- Transformer [28]: A neural network model based on self-attention mechanism is widely used in NLP. It can capture long-range dependency relationships and is suitable for handling long-text data in MLTC.
- Sequence Generation Model (SGM) [48]: Based on sequence generation models, multiple labels are generated one by one by predicting the probability distribution for each label.
- Label Embedding Enhanced SGM (LEE-SGM) [45]: The model alleviates the exposure bias problem based on the SGM and proposes a learning algorithm based on predetermined sampling, which effectively incorporates label embeddings into the label generation process.
- LSAN [33]: A generalized model for MLTC that uses both self-attention and label-attention mechanisms.
- LDGN [8]: A label-specific dual graph neural network for MLTC, modeling complete adaptive interactions based on statistical label co-occurrence and dynamic reconstruction of graph components.
- GHS-NET [17]: A generic hybridized shallow neural network used for biomedical MLTC. It employs CNN to extract the most discriminative features and Bi-LSTM layers to accurately capture the local features of biomedical text.
- LIAR [42]: It is employed for biomedical MLTC, effectively integrating label independence and correlation, and constructing a new loss function called AWLoss to alleviate the long-tail distribution.

4.3. Settings and evaluation metrics

We divide the TCM and PPE datasets into training, validation, and testing sets in a random manner, with a ratio of 7:1:2. As for network optimization, we use AdamW as the optimizer, with learning rates set to 1e-4, and a weight decay of 0.01 (excluding biases and LayerNorm). In the experiments, we utilize a single RTX3090 for training TCM and PPE separately, with 35 and 30 epochs of training, and a batch size of 16, respectively. For our model configuration, we set the number of GCN layers to 1, and the embedding size of the GCN layer to 768.

In addition, we perform a detailed statistical analysis of the parameter sizes within the three primary modules of our MCICT model. These modules encompass the BioBERT module for extracting document representations, the EAI module for obtaining label word embeddings, and the Co-GCN module (GCN learning with label co-occurrence) designed for label feature learning. The parameter sizes contained in each module are shown in the Table 4.

The multi-label classification task involves two evaluation metrics, namely sample-based metrics and label-based metrics. In this paper, we choose label-based metrics as the evaluation method, which includes $Precision_{micro}$ (P_{micro}), $Recall_{micro}$ (R_{micro}), and $F1_{micro}$, to assess the

Table 4
Parameter sizes of three primary modules.

Module	Composition	Parameter sizes
BioBert	Embedded Layer	86.46MB
	Hidden Layer (12 layers)	27.01MB x 12
	Output Layer	2.25MB
EAI	Sentence-Transformer	313.26MB
Co-GCN	GCN Layer (1 layer)	1.77MB

Table 5
Comparison of different results of various methods. For each dataset, boldface indicates the best results. Results of some baselines on the TCM dataset (marked with ∗) are directly cited from [45].

Model	TCM			PPE		
	P_{micro}	R_{micro}	$F1_{micro}$	P_{micro}	R_{micro}	$F1_{micro}$
BR∗	**0.843**	0.402	0.544	0.933	0.896	0.914
CC∗	0.764	0.460	0.574	0.922	0.894	0.908
LP∗	0.606	0.609	0.608	0.949	0.945	0.947
Transformer∗	0.713	0.484	0.576	0.972	0.947	0.959
SGM∗	0.559	0.566	0.552	0.966	0.944	0.955
LEE-SGM∗	0.620	**0.611**	0.615	0.964	0.951	0.957
LSAN	0.714	0.537	0.593	0.976	0.942	0.959
LDGN	0.772	0.531	0.629	0.970	0.954	0.960
GHS-NET	0.692	0.498	0.579	0.971	0.943	0.957
LIAR	0.762	0.542	0.633	0.973	0.950	0.961
MCICT(ours)	0.776	0.582	**0.665**	**0.978**	**0.955**	**0.966**

performance of different methods. These metrics focus on the overall classification accuracy and are suitable for cases with imbalanced class sample quantities. We calculate these metrics based on the values of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) in the confusion matrix. Specifically, the definitions and calculation methods of P_{micro}, R_{micro}, and $F1_{micro}$ are as follows:

$$P_{micro} = \frac{TP}{TP + FP}, \tag{7}$$

$$R_{micro} = \frac{TP}{TP + FN}, \tag{8}$$

$$F1_{micro} = \frac{2 \times P_{micro} \times R_{micro}}{P_{micro} + R_{micro}} \tag{9}$$

4.4. Experimental results and analysis

Table 5 presents the best $F1_{micro}$ results obtained by different methods under various settings. From the Table 5, it is evident that the proposed MCICT outperforms all other baselines in terms of F1 scores on both datasets.

It can be observed that, due to the impact of data class imbalance, BR, CC, and the neural network model Transformer exhibit a phenomenon of extremely high precision and low recall in the TCM dataset. Similar phenomena also appear in the PPE dataset. Although the BR method achieves the highest precision, its recall is also the lowest. Therefore, we are more concerned about the composite metric F1 score. In addition, it is worth noting that the evaluation results of the multi-class classification method LP on the TCM and PPE datasets indicate that several labels often co-occur, which confirms that the labels in these two datasets are somewhat correlated. Under the premise of label correlation, if the labels are sparse, for example, the TCM dataset has 929 syndrome labels, but the average number of syndromes corresponding to each sample is 1.85, then LP performs well on this dataset.

In comparison, the sequence-generating model SGM for labeling correlation modeling and its improved version LEE-SGM achieve more balanced precision and recall rates on both TCM and PPE datasets. However, their F1 scores are comparatively low.

Table 6
The results for ablation study. "BM" is short for base model, representing the foundational model. "EAI" represents the method of introducing additional information into label embedding. "Co" signifies the utilization of label co-occurrence relationships to construct correlation matrix.

Model	TCM			PPE		
	P_{micro}	R_{micro}	$F1_{micro}$	P_{micro}	R_{micro}	$F1_{micro}$
BM	0.761	0.550	0.639	0.967	0.942	0.954
BM + EAI	0.765	0.569	0.653	0.970	0.946	0.958
BM + EAI + Co	**0.776**	**0.582**	**0.665**	**0.978**	**0.955**	**0.966**

Additionally, the LDGN model also obtains good results, demonstrating the advantage of using GCN to learn label relationships. However, LDGN does not directly utilize the label correlations learned from GCN to represent label-specific texts. In contrast, our model not only incorporates label semantics but also capitalizes on label correlations through the construction of the proposed label correlation matrix. The matrix guides the information propagation between nodes in the GCN. Then, these label representations are directly mapped to document representations.

For the GHS-NET model used in biomedical MLTC, we observe that it achieves satisfactory performance on the PPE dataset, but its performance on the TCM dataset is relatively poor. Due to the severe class imbalance distribution in the TCM dataset, the GHS-NET model, however, does not consider utilizing the correlated information among labels to mitigate this issue.

The LIAR model achieves the second-best F1 score on both datasets, which is a Transformer-based MLTC method for biomedical literature, also capturing biomedical text features through BioBERT. Compared to existing MLTC methods in biomedical literature, the LIAR model captures not only the label-specific features but also the label correlation. However, one possible reason for its inferior performance compared to our model is that it initially adopts Word2vec to initialize word embeddings and then learns label representations solely by calculating the cosine similarity between text representations and label embeddings. In contrast, our model employs the CoEAI-GCN module, resulting in more accurate learned label representations and, consequently, providing a more beneficial enhancement to classification performance.

In general, the experimental results indicate that the proposed MCICT achieves the highest F1 scores on both the TCM and PPE datasets, surpassing the compared state-of-the-art models by 3.2% and 0.5%, respectively. This demonstrates the effectiveness of fine-tuning using domain specific pre-trained models and the importance of utilizing label correlation.

4.5. Further discussion

4.5.1. Ablation studies

As shown in Table 6, we conduct ablation studies to demonstrate the effectiveness of the proposed modules. The data from the table reveals that the model incorporating the EAI method achieved better results compared to the base model, thus confirming the efficacy of integrating additional information into label embedding. Furthermore, by incorporating the strategy for constructing a correlation matrix based on label co-occurrence relationships into the model, the results of both tasks are improved to varying degrees. This indicates that leveraging the correlations between labels helps the model better predict less common labels, thereby improving overall performance.

4.5.2. Comparison of deep learning models for document representation learning

We validate the effectiveness of using the BioBERT pre-trained model to learn clinical document representation and apply it to downstream MLTC tasks. For this purpose, we compare it with several common document representation learning models, including simple

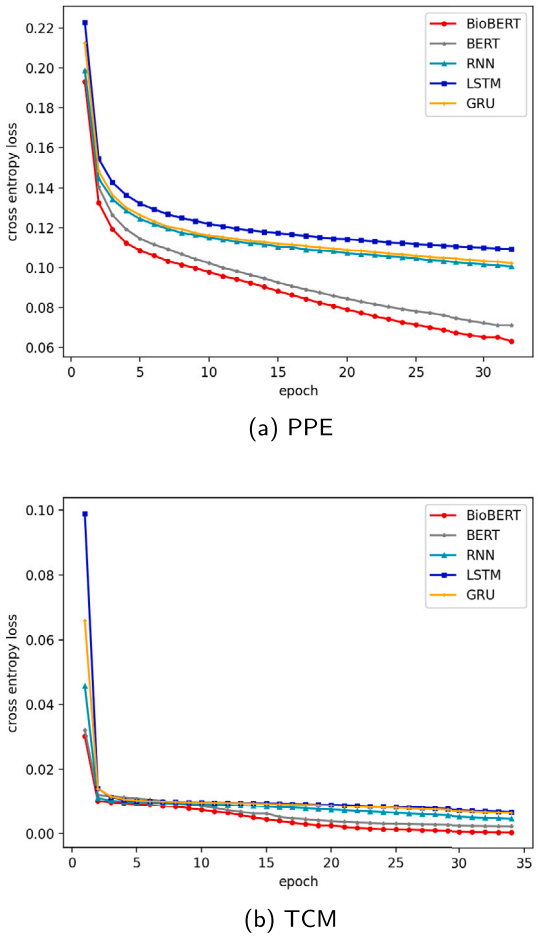

Fig. 5. Comparison of training loss for different models on text features.

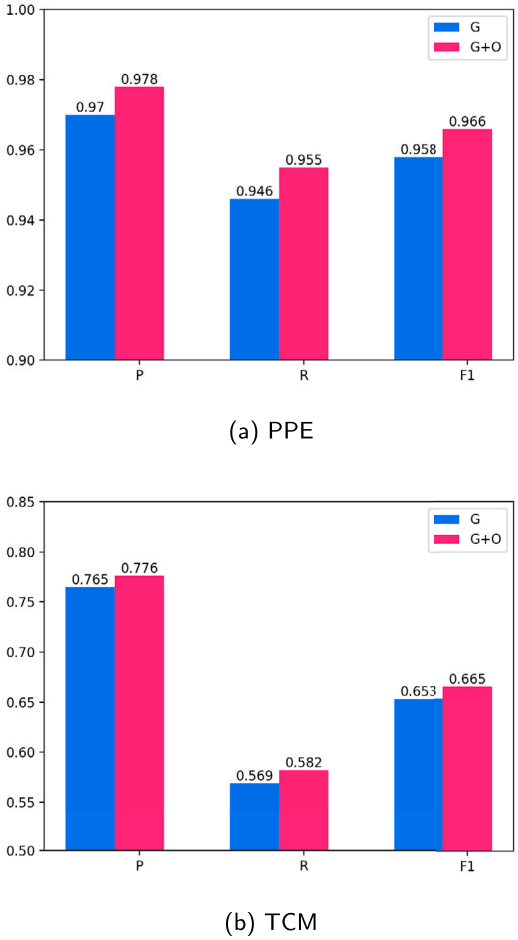

Fig. 6. Effects of introducing label co-occurrence. ('G' represents using the unit matrix as the correlation matrix, while 'G+O' represents using the correlation matrix constructed in our paper).

RNN, LSTM, GRU and BERT. In the experiments, we examine the loss convergence results during the training process, as shown in Fig. 5.

The experimental results show that BERT and BioBERT pre-trained models have the best performance. Notably, BioBERT stands out by demonstrating the fastest and most significant reduction in loss during the training process.

Furthermore, the performance of document feature learning using simple RNN is slightly better than LSTM and GRU, which may be due to the advantages of LSTM and GRU in processing long sequence data. However, since our samples of clinical text are not particularly long, the simpler RNN performs better at this task.

4.5.3. Effects of introducing label co-occurrence

We conduct a series of experiments to demonstrate the advantages of combining co-occurrence information between labels to construct GCN correlation matrix. The experimental results are shown in Fig. 6.

Specifically, we compare a correlation matrix constructed using the unit matrix as the initial correlation matrix with one that considers label co-occurrence. Firstly, when using the unit matrix as the initial correlation matrix, it means that the correlations between labels are not modeled and all labels are assumed to have the same level of correlation. In our experiments, we find that the effect of using the unit matrix was relatively weak.

Then, the correlation matrix is constructed using label co-occurrence, and the matrix elements are calculated according to label occurrence frequency and co-occurrence frequency. We observe that this method of constructing the correlation matrix yields better results than using the unit matrix. The precision, recall, and F1 scores all improved, with F1 scores increasing by 0.8% and 1.2%, respectively, on both tasks. The reason is that label co-occurrence can more accurately reflect the association between labels, so that the model can better capture the commonalities and characteristics between labels. Therefore, it can alleviate class imbalance to a certain extent and improve classification accuracy.

4.5.4. Effects of additional information on label embedding

We test the effect of different label embedding methods on the performance of our proposed MCICT model, also compare well-known embedding methods, such as OneHot, Word2vec [49], FastText [50], and ELMo [29], with our proposed embedding method EAI applied to the TCM and PPE datasets. As shown in Fig. 7, experimental results demonstrate that our method is superior to other methods in the clinical MLTC task.

Due to the specialized and complex nature of clinical texts, traditional word embedding methods may not effectively capture their characteristics. However, our approach utilizes a specific source of clinical text — Wikipedia, to enrich the semantic representation of labels, so that they can better adapt to the unique features of the medical domain.

The above shows that the introduction of additional information into the label embedding methods can effectively improve the classification performance of the model, especially in specialized text tasks such as the medical domain.

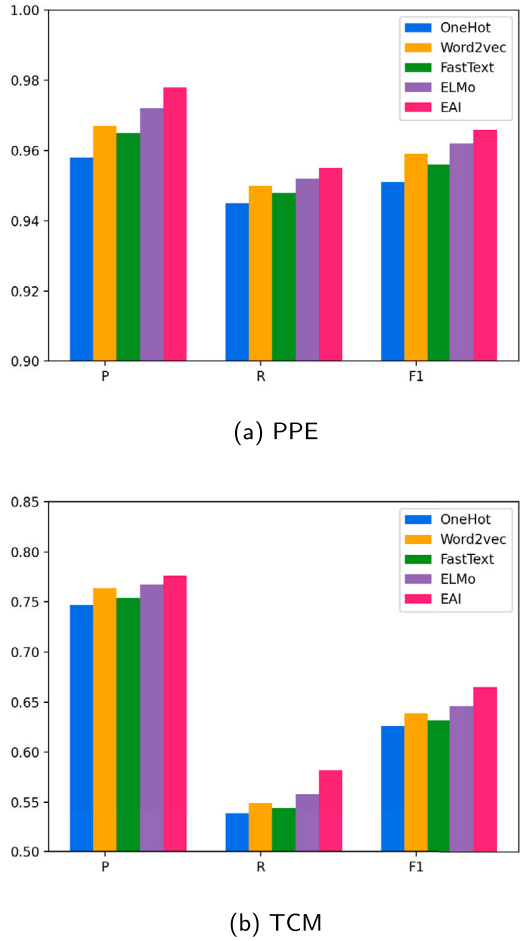

Fig. 7. Comparison of different label embedding methods.

4.6. Parameter sensitivity of CoEAI-GCN

In this section, we first test the classification performance on two datasets using GCNs of different layers.

As shown in the left figure of Fig. 8, the model achieve good results with only one layer. We also observe that as the number of GCN layers increased, the classification performance decreased. One possible reason is that as the number of layers of the network increases, so does the complexity of the model, which may lead to overfitting. Overfitting can cause the model to perform well on the training set, but poorly on data it has not seen before.

Next, we use only one layer on the different embedding dimensions (256, 512, 768, 1024, and 2048) to evaluate our model. As shown in the right figure in Fig. 8, we observe that the best performance is achieved with the embedding dimension is 768. This is because a lower embedding dimension may not be sufficient to represent label information, and a higher embedding dimension does not necessarily guarantee improved classification performance.

5. Conclusion

In this paper, we proposed an end-to-end model MCICT based on GCN to tackle the MLTC task in clinical text. The MCICT model consisted of two primary components. Firstly, BioBERT was used to extract intricate semantic features from clinical texts. Secondly, the CoEAI-GCN module was adopted to acquire label feature representations. The introduced CoEAI-GCN module employed additional information to enhance the precision of label embeddings and integrates label co-occurrence knowledge in GCN to learn label representations. This strategy effectively mitigated the problem of class imbalance. Through a series of experiments on two real clinical text datasets, we demonstrated that the MCICT model achieved good performance in multi-label classification of imbalanced clinical texts.

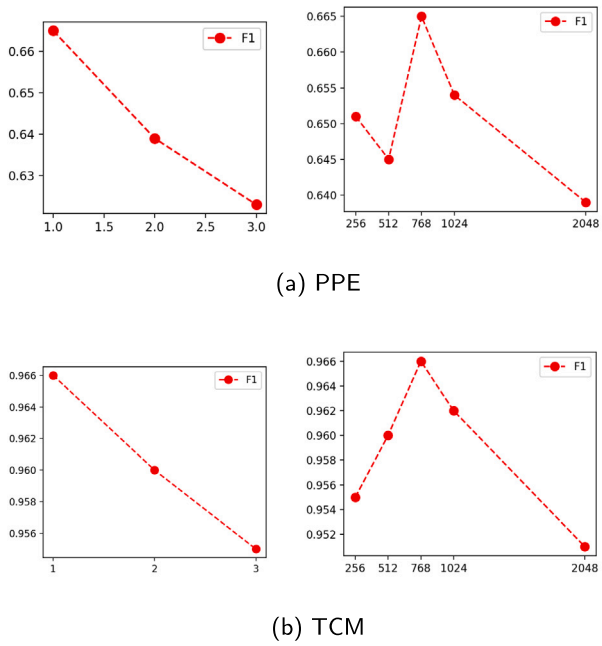

Fig. 8. Test performance with 1, 2 and 3 GCN layers. (left) Test performance under varying GCN embedding dimensions. (right).

In future work, we are considering incorporating a label attention mechanism. This mechanism will direct attention to the most relevant labels for a given text in the classification task. The exploration is expected to more accurately capture correlations between text and labels, leading to potential improvements in model performance.

CRediT authorship contribution statement

Yao He: Methodology, Validation, Writing – original draft. **Qingyu Xiong:** Writing – review & editing, Supervision. **Cai Ke:** Conceptualization, Visualization, Writing – review & editing. **Yaqiang Wang:** Data curation, Investigation. **Zhengyi Yang:** Resources, Supervision, Validation. **Hualing Yi:** Conceptualization, Formal analysis, Investigation. **Qilin Fan:** Data curation, Resources, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported by the National NSFC (Grant No. 62102053), the Natural Science Foundation of Chongqing, China (Grant No. CSTB2022NSCQ-MSX1104), and Graduate Scientific Research and Innovation Foundation of Chongqing, China (No. CYS22128). In addition, we would like to thank Professor Zhu Tao and Professor Hao Xuechao from the Department of Anesthesiology, West China Hospital, Sichuan University, for their data set supporting this work.

References

[1] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, Sandeep Kaushik, Big data in healthcare: Management, analysis and future prospects, J. Big Data 6 (1) (2019) 1–25.

[2] Kornelia Batko, Andrzej Ślęzak, The use of big data analytics in healthcare, J. Big Data 9 (1) (2022) 3.

[3] Andrew Wen, Sunyang Fu, Sungrim Moon, Mohamed El Wazir, Andrew Rosenbaum, Vinod C Kaggal, Sijia Liu, Sunghwan Sohn, Hongfang Liu, Jungwei Fan, Desiderata for delivering NLP to accelerate healthcare AI advancement and a mayo clinic NLP-as-a-service implementation, NPJ Digit. Med. 2 (1) (2019) 130.

[4] Miriam Reisman, EHRs: The challenge of making electronic data usable and interoperable, Pharm. Ther. 42 (9) (2017) 572.

[5] Davide Cirillo, Alfonso Valencia, Big data analytics for personalized medicine, Curr. Opin. Biotechnol. 58 (2019) 161–167.

[6] Kevin De Angeli, Shang Gao, Ioana Danciu, Eric B Durbin, Xiao-Cheng Wu, Antoinette Stroup, Jennifer Doherty, Stephen Schwartz, Charles Wiggins, Mark Damesyn, et al., Class imbalance in out-of-distribution datasets: Improving the robustness of the TextCNN for the classification of rare cancer types, J. Biomed. Inform. 125 (2022) 103957.

[7] Hongxia Lu, Louis Ehwerhemuepha, Cyril Rakovski, A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance, BMC Med. Res. Methodol. 22 (1) (2022) 181.

[8] Qianwen Ma, Chunyuan Yuan, Wei Zhou, Songlin Hu, Label-specific dual graph neural network for multi-label text classification, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 3855–3864.

[9] Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, Yiming Yang, Deep learning for extreme multi-label text classification, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 115–124.

[10] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang, BioBERT: A pre-trained biomedical language representation model for biomedical text mining, Bioinformatics 36 (4) (2020) 1234–1240.

[11] Himanshu Jain, Yashoteja Prabhu, Manik Varma, Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 935–944.

[12] Tirath Prasad Sahu, Reswanth Sai Thummalapudi, Naresh Kumar Nagwani, Automatic question tagging using multi-label classification in community question answering sites, in: 2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2019 5th IEEE International Conference on Edge Computing and Scalable Cloud, EdgeCom, IEEE, 2019, pp. 63–68.

[13] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, Richard Socher, Ask me anything: Dynamic memory networks for natural language processing, in: International Conference on Machine Learning, PMLR, 2016, pp. 1378–1387.

[14] Yaqi Wang, Shi Feng, Daling Wang, Ge Yu, Yifei Zhang, Multi-label Chinese microblog emotion classification via convolutional neural network, in: Web Technologies and Applications: 18th Asia-Pacific Web Conference, APWeb 2016, Suzhou, China, September 23-25, 2016. Proceedings, Part I, Springer, 2016, pp. 567–580.

[15] Mohammed Jabreel, Antonio Moreno, A deep learning-based approach for multi-label emotion classification in tweets, Appl. Sci. 9 (6) (2019) 1123.

[16] Ghulam Mujtaba, Liyana Shuib, Norisma Idris, Wai Lam Hoo, Ram Gopal Raj, Kamran Khowaja, Khairunisa Shaikh, Henry Friday Nweke, Clinical text classification research trends: Systematic literature review and open issues, Expert Syst. Appl. 116 (2019) 494–520.

[17] Muhammad Ali Ibrahim, Muhammad Usman Ghani Khan, Faiza Mehmood, Muhammad Nabeel Asim, Waqar Mahmood, GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification, J. Biomed. Inform. 116 (2021) 103699.

[18] Matthew R Boutell, Jiebo Luo, Xipeng Shen, Christopher M Brown, Learning multi-label scene classification, Pattern Recogn. 37 (9) (2004) 1757–1771.

[19] Grigorios Tsoumakas, Ioannis Vlahavas, Random k-labelsets: An ensemble method for multilabel classification, in: European Conference on Machine Learning, Springer, 2007, pp. 406–417.

[20] Jesse Read, Bernhard Pfahringer, Geoff Holmes, Eibe Frank, Classifier chains for multi-label classification, Mach. Learn. 85 (2011) 333–359.

[21] Johannes Fürnkranz, Eyke Hüllermeier, Eneldo Loza Mencía, Klaus Brinker, Multilabel classification via calibrated label ranking, Mach. Learn. 73 (2008) 133–153.

[22] Min-Ling Zhang, Zhi-Hua Zhou, ML-KNN: A lazy learning approach to multi-label learning, Pattern Recogn. 40 (7) (2007) 2038–2048.

[23] Eleftherios Spyromitros, Grigorios Tsoumakas, Ioannis Vlahavas, An empirical study of lazy multilabel classification algorithms, in: Artificial Intelligence: Theories, Models and Applications: 5th Hellenic Conference on AI, SETN 2008, Syros, Greece, October 2-4, 2008. Proceedings 5, Springer, 2008, pp. 401–406.

[24] Weiwei Cheng, Eyke Hüllermeier, Combining instance-based learning and logistic regression for multilabel classification, Mach. Learn. 76 (2009) 211–225.

[25] Francesco Gargiulo, Stefano Silvestri, Mario Ciampi, Deep convolution neural network for extreme multi-label text classification, in: Healthinf, 2018, pp. 641–650.

[26] Priyanka Nigam, Applying Deep Learning to ICD-9 Multi-Label Classification from Medical Records, Technical report, Technical report, Stanford University, 2016.

[27] Guibin Chen, Deheng Ye, Zhenchang Xing, Jieshan Chen, Erik Cambria, Ensemble application of convolutional and recurrent neural networks for multi-label text categorization, in: 2017 International Joint Conference on Neural Networks, IJCNN, IEEE, 2017, pp. 2377–2383.

[28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, vol.30, 2017.

[29] E. Matthew, Mark Neumann Peters, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer, Deep contextualized word representations, in: Proc. of NAACL. Vol. 5, 2018.

[30] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al., Improving language understanding by generative pre-training, 2018, pp. 1–12, Preprint.

[31] Jacob Devlin Ming-Wei Chang Kenton, Lee Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of NaacL-HLT. vol. 1, 2019, p. 2.

[32] Cunxiao Du, Zhaozheng Chen, Fuli Feng, Lei Zhu, Tian Gan, Liqiang Nie, Explicit interaction model towards text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, no. 01, 2019, pp. 6359–6366.

[33] Lin Xiao, Xin Huang, Boli Chen, Liping Jing, Label-specific document representation for multi-label text classification, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, 2019, pp. 466–475.

[34] Linkun Cai, Yu Song, Tao Liu, Kunli Zhang, A hybrid BERT model that incorporates label semantics via adjustive attention for multi-label text classification, Ieee Access 8 (2020) 152183–152192.

[35] Helong Yu, Chunliu Liu, Lina Zhang, Chengwen Wu, Guoxi Liang, José Escorcia-Gutierrez, Osama A Ghoneim, An intent classification method for questions in" treatise on febrile diseases" based on TinyBERT-CNN fusion model, Comput. Biol. Med. (2023) 107075.

[36] Trung Huynh, Yulan He, Alistair Willis, Stefan Rueger, Adverse drug reaction classification with deep neural networks, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 877–887.

[37] Guangxu Xun, Kishlay Jha, Ye Yuan, Yaqing Wang, Aidong Zhang, MeSHProbeNet: A self-attentive probe net for MeSH indexing, Bioinformatics 35 (19) (2019) 3794–3802.

[38] Ronghui You, Yuxuan Liu, Hiroshi Mamitsuka, Shanfeng Zhu, BERTMeSH: Deep contextual representation learning for large-scale high-performance MeSH indexing with full text, Bioinformatics 37 (5) (2021) 684–692.

[39] Simon Baker, Anna Korhonen, Initializing neural networks for hierarchical multi-label text classification, in: BioNLP 2017, 2017, pp. 307–315.

[40] Min Li, Zhihui Fei, Min Zeng, Fang-Xiang Wu, Yaohang Li, Yi Pan, Jianxin Wang, Automated ICD-9 coding via a deep learning approach, IEEE/ACM Trans. Comput. Biol. Bioinform. 16 (4) (2018) 1193–1202.

[41] Jingcheng Du, Qingyu Chen, Yifan Peng, Yang Xiang, Cui Tao, Zhiyong Lu, ML-net: Multi-label classification of biomedical texts with deep neural networks, J. Am. Med. Inform. Assoc. 26 (11) (2019) 1279–1285.

[42] Zihao Chen, Jing Peng, Learning label independence and relevance for multi-label biomedical text classification, in: 2022 IEEE International Conference on Systems, Man, and Cybernetics, SMC, IEEE, 2022, pp. 2776–2781.

[43] Thomas N. Kipf, Max Welling, Semi-supervised classification with graph convolutional networks, in: International Conference on Learning Representations, 2016.

[44] Andrew L. Maas, Awni Y. Hannun, Andrew Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. Icml. vol. 30, no. 1, Atlanta, GA, 2013, p. 3.

[45] Yaqiang Wang, Feifei Yan, Xiaofeng Wang, Wang Tang, Hongping Shu, Label embedding enhanced multi-label sequence generation model, in: Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II 9, Springer, 2020, pp. 219–230.

[46] Yaqiang Wang, Xiao Yang, Xuechao Hao, Hongping Shu, Guo Chen, Tao Zhu, An unstructured data representation enhanced model for postoperative risk prediction, in: Proceedings of the 21st Chinese National Conference on Computational Linguistics, 2022, pp. 580–590.

[47] Ruixuan Luo, Jingjing Xu, Yi Zhang, Zhiyuan Zhang, Xuancheng Ren, Xu Sun, Pkuseg: A toolkit for multi-domain chinese word segmentation, 2019, arXiv preprint arXiv:1906.11455.

[48] Pengcheng Yang, Xu Sun, Wei Li, Shuming Ma, Wei Wu, Houfeng Wang, SGM: Sequence generation model for multi-label classification, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 3915–3926.

[49] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, Efficient estimation of word representations in vector space, 2013, arXiv preprint arXiv:1301.3781.

[50] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, Tomas Mikolov, Fasttext. zip: Compressing text classification models, 2016, arXiv preprint arXiv:1612.03651.