



Research article

Double decoupled network for imbalanced obstetric intelligent diagnosis

Kunli Zhang^{1,2}, Shuai Zhang¹, Yu Song^{1,*}, Linkun Cai^{3,*} and Bin Hu¹

¹ College of Computer Intelligence, Zhengzhou University, Zhengzhou, China

² Pengcheng laboratory, Shenzhen, Guangdong, China

³ College of Biological Science and Medical Engineering, Beihang University, Beijing, China

* **Correspondence:** Email: ieysong@zzu.edu.cn, cailinkun@buaa.edu.cn.

Abstract: Electronic Medical Record (EMR) is the data basis of intelligent diagnosis. The diagnosis results of an EMR are multi-disease, including normal diagnosis, pathological diagnosis and complications, so intelligent diagnosis can be treated as multi-label classification problem. The distribution of diagnostic results in EMRs is imbalanced. And the diagnostic results in one EMR have a high coupling degree. The traditional rebalancing methods does not function effectively on highly coupled imbalanced datasets. This paper proposes Double Decoupled Network (DDN) based intelligent diagnosis model, which decouples representation learning and classifier learning. In the representation learning stage, Convolutional Neural Networks (CNN) is used to learn the original features of the data. In the classifier learning stage, a Decoupled and Rebalancing highly Imbalanced Labels (DRIL) algorithm is proposed to decouple the highly coupled diagnostic results and rebalance the datasets, and then the balanced datasets is used to train the classifier. This paper evaluates the proposed DDN using Chinese Obstetric EMR (COEMR) datasets, and verifies the effectiveness and universality of the model on two benchmark multi-label text classification datasets: Arxiv Academic Papers Datasets (AAPD) and Reuters Corpus1 (RCV1). Demonstrating the effectiveness of the proposed methods is an imbalanced obstetric EMRs. The accuracy of DDN model on COEMR, AAPD and RCV1 datasets is 84.17, 86.35 and 93.87% respectively, which is higher than the current optimal experimental results.

Keywords: imbalanced classification; decoupled; rebalancing methods; intelligent diagnosis

1. Introduction

With the implementation of China's three-child policy, obstetric clinical research is faced with unprecedented challenges, and the number of elderly parturient women is increasing rapidly. The incidence rate of obstetric and gynecological diseases in elderly maternal is significantly higher than that of pregnant women at the right age [1]. Electronic Medical Records (EMRs) are the most detailed and direct record of clinical medical activities. The clinical diagnosis process of doctors can be regarded as judging the probability of suffering from a certain disease according to the clinical manifestations and examination results of patients. An obstetric EMR usually contains multiple diagnostic results, that is, patients may be diagnosed with both "gestational diabetes mellitus" and "gestational hypertension", and the diagnosis results have strong coupling. If an EMR is regarded as one sample, each sample can be grouped into multiple categories. Therefore, the intelligent diagnosis problem can be regarded as a multi-label classification problem in machine learning, and multiple diagnostic results in an EMR have different labels [2].

However, the data distribution of EMRs is often imbalanced, and the sample number of rare diseases is far less than that of common diseases [3]. The imbalanced distribution of datasets will lead to the performance degradation of traditional classification algorithms [4]. Traditional algorithms tend to treat a few classes as noise or outliers, and ignore them in classification [5]. For imbalanced EMRs, the cost of false negative is much higher than that of false positive. For example, in 100 EMRs, 99 results were normal and one detected cancer. If the traditional classification algorithm is directly applied to this kind of data, the diagnostic results of EMR will be predicted as normal. Although the precision rate is as high as 99%, the most critical cancer information is ignored.

In neural networks, the features of input samples have a great influence on classifiers. The reason that rebalancing method works is that it can update the weights of classifiers and significantly improve the learning ability of deep network classifiers, but it damages the feature learning ability of deep network [6,7]. In addition, the coupling of high-frequency diagnostic results and low-frequency diagnostic results in an EMR needs to be considered in the multi-label rebalancing method: removing the EMR containing high-frequency diagnostic results also means losing low-frequency diagnostic results, and cloning the EMR containing low-frequency diagnostic results to add new instances will also increase the frequency of existing high-frequency diagnostic results.

In recent years, intelligent diagnosis has become a research focus. Yin et al. [8] proposed a method to extract signal features from heart rate variability signals and classify patients' states using the long short-term memory network. Yan et al. [9] collected 1880 endoscopic images, and developed a Gastric Intestinal Metaplasia (GIM) system with these images using a modified convolutional neural network algorithm. Wang et al. [10] proposed Patch Shuffle stochastic pooling neural network, which improved recognition performances of Corona Virus Disease 2019 (COVID-19) infection from chest CT (CCT) images, and will help assist radiologists to make diagnosis more quickly and accurately on COVID-19 cases.

In addition to using image data for intelligent diagnosis, text-based intelligent diagnosis is also a research focus. Rajkomar et al. [11] proved that in all the methods using EMRs, deep learning methods are superior to the most advanced statistical prediction model. Maxwell et al. [12] used physical examination data to predict the possible chronic diseases such as diabetes, hypertension and fatty liver. Yang et al. [13] proposed a Convolutional Neural Network (CNN) based auxiliary diagnosis method, which is based on self-learning to learn high-level semantic understanding from EMRs, and output the

prediction probability of common diseases including hypertension and diabetes. Liang et al. [14] proposed a system framework based on pediatric EMRs, which integrates medical knowledge in pediatric EMRs for intelligent diagnosis.

In the current researches, the imbalanced distribution of datasets is an important factor limiting the performance of intelligent diagnosis. The precision of low-frequency disease diagnosis is too low, which leads to the decline of the practicability of the diagnosis model. The existing researches based on imbalanced data can be divided into data-level methods and algorithm-level methods.

The data-level methods are to transform the original datasets into a relatively balanced datasets from the data preparation stage. Liu et al. [15] proposed an algorithm based on information granulation. The algorithm assembles the data in most classes into particles to balance the proportion of classes in the data, and uses prostate cancer data to predict the survival rate of patients. Huang et al. [16] proposed random balanced sampling algorithm based on association rule selection, and verified the performance of the algorithm in a private diabetes EMRs.

Unlike the data-level methods, the algorithm-level methods do not change the distribution of training data; instead, they increase the importance of minority classes in the learning and decision-making process [17]. Li et al. [18] proposed dice loss to improve the weight of difficult samples and reduce the weight of simple negative samples. For high-dimensional imbalanced text data, researchers found that selecting features that are conducive to identifying minority classes can effectively deal with imbalanced data. Yang et al. [19] proposed a text feature selection method based on relation score. By calculating the relationship score of each feature and category, the relationship score of minority features is increased and the imbalance degree of data concentration categories.

At present, intelligent diagnosis researches based on EMRs are mostly for a single disease, and does not consider the multiple complications and other diagnostic results. In addition, the existence of rare diseases causes the imbalance of EMRs, and the diagnostic performance of the model needs to be further improved.

This paper proposes Double Decoupled Network (DDN) to improve the performance of intelligent diagnosis based imbalanced datasets. Our main contributions are summarized as follows:

- 1) DDN is proposed to decouple representation learning from classifier learning and high coupling diagnostic results.
- 2) In the classifier learning stage, a Decoupled and Rebalancing highly Imbalanced Labels (DRIL) algorithm is proposed to decouple the highly coupled diagnostic results and rebalance the datasets.
- 3) Experiments on a real Chinese Obstetric EMR (COEMR) datasets and two public datasets show that DDN method has better performance than comparison methods.

2. Materials and methods

Inspired by paper [6], we firstly analyze the performance of rebalancing strategies in neural network. Secondly, we discuss the rebalancing strategies in high coupling diagnostic results of COEMR datasets. This paper proposes DDN to decouple representation learning from classifier learning, and high coupling diagnostic results.

2.1. Overall framework

In order to solve the problem of high coupling of diagnosis results and obtain better features of

input samples, an intelligent diagnosis model based on double decoupling network is proposed in this section. The overall architecture of DDN is shown in Figure 1. First, DDN decouples representation learning from classifier learning. In the presentation learning stage, DDN uses CNN to learn the original features of COEMR datasets $D = [d_1, d_2, \dots, d_n]$, and fixed parameters of presentation learning, where d_i represents a sample in the datasets. Map the input text to a sequence of embedding vector. Words are embedded in a vector sequence through a convolutional layer and linearly transformed using a non-linear activation function to capture indicative information. Different types of useful information for prediction are selected in the pooling layer, and the maximum value is chosen from the feature mapping for each type. Finally, fully connected layers are used to integrate information with disease differentiation in the convolutional or pooling layers. In the classifier learning stage, DRIL algorithm is proposed to decouple the high coupling diagnostic results and rebalance the datasets $D' = [d'_1, d'_2, \dots, d'_n]$, using the datasets D' to train the classifier. The classifier consists a full connection layer and Softmax function. Both stages use the same CNN network structure and share all weights except for the last full connection layer.

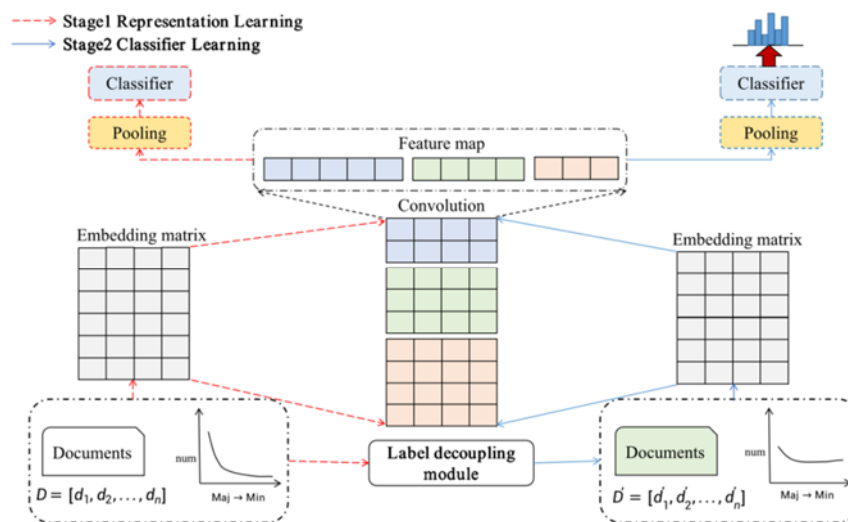


Figure 1. The architecture of DDN model.

2.2. Representation and classifier learning decoupling module

Two methods are adopted to deal with unbalanced data: resampling samples and reweighting sample loss in small batches. In order to explore the working mechanism of rebalancing method, we divide the training process of neural network into two stages, namely representation learning and classifier learning. Specifically, in the first stage, we use the common method (Cross Entropy, CE) or rebalancing methods (Re-sampling/R-weighting, RS/RW) to train the neural network to get the corresponding feature extractor. Then we fix the parameters of the feature extractor. In the second stage, the classifier is retrained by common method or rebalancing methods. In this section, representation learning and classifier learning are used to compare the effects of different training methods. Figure 2 shows the precision of different methods on COEMR and Arxiv Academic Papers Datasets (AAPD) datasets.

1) CE: the traditional cross entropy loss is used to train the network on the original imbalanced data.

2) RS: in this section, class balanced resampling method is used to ensure that the probability of each class in each batch is the same. The sampling probability is p_j calculated as Eq (1).

$$p_j = \frac{1}{C} \quad (1)$$

where C is the number of all labels in the training set.

3) RW: reweighting all classes according to the reciprocal of their sample size.

For representation learning, when using the same classifier learning method (comparing the precision of three blocks in the horizontal direction), the precision of CE block is always higher than that of RW/RS block. We can find that CE can get better classification results since it obtains better features. The worse results of RW/RS block show that RW/RS method has a poor ability to identify depth features, which will damage the ability of representation learning.

For classifier learning, when using the same representation learning method (comparing the precision of three blocks in the vertical direction), it can be found that RW/RS methods can achieve higher precision than CE. The results show that the main reason why rebalancing methods can achieve a balanced performance on imbalanced data is that these methods directly affect the update of the weights of deep network classifiers, that is, promote the learning of classifiers.

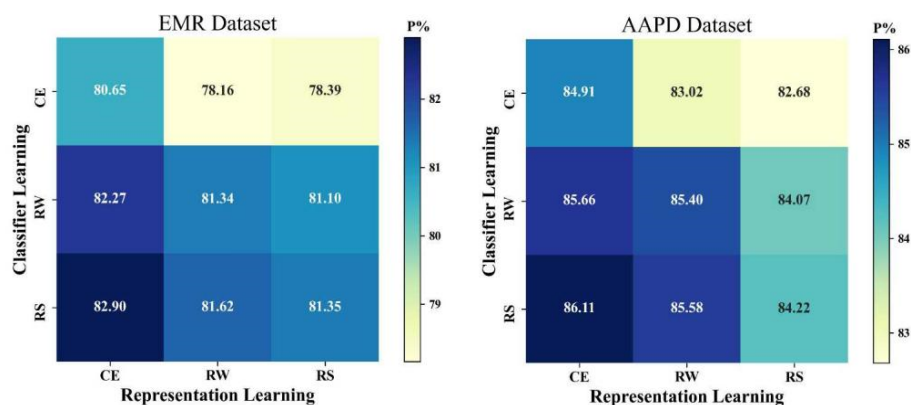


Figure 2. Precision of different methods on COEMR and AAPD datasets.

This section discusses the influence of rebalancing methods on representation learning and classifier learning in neural network. We can find that rebalancing methods can significantly promote classifier learning, but it also damages the ability of learning features to a certain extent. To solve these problems, we propose decoupled representation learning from classifier learning. In the representation learning stage, the original features of the datasets are learned. In the classifier learning stage, the rebalancing datasets are used for training, so as to give consideration to both representation learning and classifier learning, and improve the generalization ability of low-frequency data, and improve the classification performance of imbalanced data.

2.3. Label decoupling module

The rebalancing strategy is independent of classifier, so it is applicable to a wider range of

scenarios than does the adaptive classifier. It is difficult for traditional rebalancing methods to achieve good performance for multi-label data. The main problems include the huge difference of imbalanced degree between labels in multi-label datasets, and the highly coupling of low-frequency labels and high-frequency labels in the same sample.

The imbalance ratio and average imbalance ratio between labels in a multi-label dataset can be determined according to the method proposed in reference [20]. The first measure is the Imbalance Ratio per label (IR), as shown in Eq (2), which evaluates the imbalance ratio of a single label. Define a multi-label dataset $D = \{(X_i, Y_i) | 0 \leq i \leq n, Y_i \in L\}$, where X_i is the i^{th} sample in the datasets, Y_i is the label set of the datasets X_i , and L is the label set of the datasets.

$$IR_{(l)} = \frac{\sum_{i=1}^{|D|} h(l', Y_i)}{\sum_{i=1}^{|D|} h(l, Y_i)}, h(l, Y_i) = \begin{cases} 1, l \in Y_i \\ 0, l \notin Y_i \end{cases} \quad (2)$$

The second measure is Mean Imbalance Ratio (MeanIR), which is the overall estimation of the imbalance degree of multi-label datasets, that is, the average IR value of all labels, See Eq (3), where $|L|$ is the label set of the datasets.

$$MeanIR = \frac{1}{|L|} \sum_{l=L_1}^{L_{|L|}} (IR(l)) \quad (3)$$

According to MeanIR and IR value, this section defines high-frequency label and low-frequency label. When IR value is higher than mean IR value, it is a low-frequency label; otherwise, it is high-frequency label. For label y , if $IR(y) > MeanIR$, it belongs to minBags, otherwise it belongs to majBags.

In order to understand the coupling degree of low-frequency and high-frequency labels in the same sample in multi-label datasets, we can evaluate it by SCUMBLE measure [21]. As can be seen from Eq (4), SCUMBLE relies on the aforementioned IR metric. The coupling degree for each sample is first obtained $SCUMBLE_{ins}(i)$, and then, the average coupling degree $SCUMBLE(D)$ is calculated for the entire multi-label datasets, as shown in Eq (5). The values of SCUMBLE are normalized in the $[0, 1]$ range, and the larger the values, the higher the coupling between the unbalanced labels.

$$SCUMBLE_{ins}(i) = \frac{1}{IR_i} \left(\prod_{l=1}^{|L|} IR_{il} \right)^{(1/|L|)} \quad (4)$$

$$SCUMBLE(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} SCUMBLE_{ins}(i) \quad (5)$$

We calculate the average coupling degree $SCUMBLE(D) = 0.3$ of the COEMR datasets, and visualizes the label coupling of the COEMR datasets by using chord diagram, as shown in Figure 3. It

can be seen that there is a high coupling degree in the COEMR datasets, and the low-frequency labels are completely associated with some high-frequency labels.

The high coupling of imbalanced labels can be alleviated with label decoupling strategy. Chart et al. [22] proposed remedial algorithm, which is independent of resampling algorithm and uses $SCUMBLE_{ins}(i) > SCUMBLE(D)$ as the judgment condition of whether to decouple labels, the coupling degree between high-frequency labels and low-frequency labels is reduced by decoupling them. Based on this, this paper proposes a DRIL algorithm, which decouples and clones the samples with high scumble value, and obtains two examples, one of which is associated with high-frequency labels, and the other is associated with low-frequency labels, so as to reduce the coupling degree, and then uses the method of combining oversampling and under-sampling to rebalance the datasets. It can reduce the loss of high-frequency disease sample information and the over fitting of low-frequency disease.

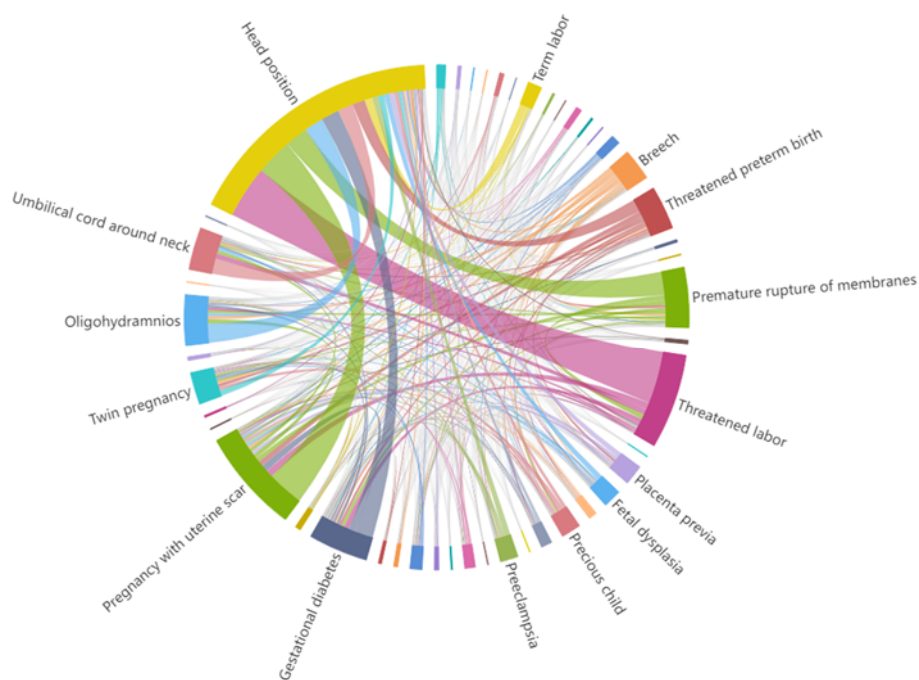


Figure 3. Visualization of label coupling in COEMR datasets.

Table 1 shows the pseudo code of DRIL. Specifically, DRIL first calculates the IR value and MeanIR value of each label to determine which category the label belongs to. Resampling rate P indicates that the proportion of samples needs to be adjusted. Then, DRIL calculates the SCUMBLE of each sample and whole dataset $SCUMBLE(D)$. The NumOfMinBag is calculated:

$$\text{NumOfMinBag} = \sum_{l=1}^{Y_i} h(l, \text{minBags}) \quad (6)$$

When $SCUMBLE_{ins}(i) > SCUMBLE(D)$ and $\text{NumOfMinBag} < |Y_i|$, the label will be decoupled. Find sample D_i with high coupling label according to step 5. From step 5 to step 7, one sample is decoupled into two samples, namely clone sample D_i is D'_i , L_i is the label set of

D_i , L'_i is the label set of D'_i , $L_i = L_{i[IR(y \leq MeanIR)]}$, $L'_i = L_{i[IR(y \geq MeanIR)]}$. By decoupling, the low-frequency and high-frequency labels in the samples can be separated, and the decoupled datasets D'_i is obtained. MeanSamples is the number of samples required for all labels to reach the mean state of MeanIR. Its calculation method is to divide the number of samples of the most frequent labels by the value of MeanIR:

$$MeanSamples = \frac{L_{|L|} \operatorname{argmax}(\sum_{i=1}^{|D|} h(l', Y_i))}{MeanIR} \quad (7)$$

From step 9 to step 21, the datasets are balanced by the method of combining oversampling and under-sampling. Firstly, a random label y is generated and X samples are randomly selected from y . If y belongs to minBags , $x = \text{Random}(0, \text{MeanSamples} - |y|)$, then the selected samples are added to the datasets D' . If y belongs to majBags , $x = \text{Random}(0, |y| - \text{MeanSamples})$, delete the X samples from the datasets D' . DRIL uses Mean samples to limit the number of samples x to balance the distribution among samples, and the number of samples needed to achieve the balance is not more than or less than the number of samples needed to achieve the balance. At the end of each rebalancing, MeanIR and IR are recalculated. Mean samples always use the initial value, and the original distribution of the datasets will not be greatly affected. Finally, we get the datasets D' which is decoupled and balanced by the DRIL algorithm.

Table 1. DRIL algorithm.

Input: multi-label datasets D , resampling rate P	
Output: decoupled datasets D'	
1	Calculate $\text{samplesToResampling} = D * P$, IR, MeanIR & MeanSamples
2	Calculate the $SCUMBLE_{ins}$ of each sample D_i in D , calculate the $SCUMBLE(D)$ of the datasets
3	For each instance D_i in D do
4	Calculate number of low-frequency labels of sample D_i : NumOfMinBag
5	If $SCUMBLE_{ins}(i) > SCUMBLE$ and $\text{NumOfMinBag} < Y_i $ then
6	clone $D'_i \leftarrow D_i$, $L'_i = L_{i[IR(y \geq MeanIR)]}$, $L_i = L_{i[IR(y \leq MeanIR)]}$
7	$D_d = D_d + D_i + D'_i$, $D = D - D_i$ (D_d is a decoupled dataset)
8	$D = D + D_d$, $D' = D$
9	While $SCUMBLE(D') > 0.1$ or $\text{samplesToResampling} > 0$
10	Randomly select label y
11	If $ y < \text{MeanSamples}$ and $y \in \text{minBags}$ then
13	$x = \text{Random}(0, \text{MeanSamples} - y)$ Select x samples from the samples of label y
14	Add x to D' , $D' += x$,
16	If $ y > \text{MeanSamples}$ and $y \in \text{majBags}$ then
18	$x = \text{Random}(0, y - \text{MeanSamples})$ Select x samples from the samples of label y
19	Remove x from D' , $D' -= x$,
20	$\text{samplesToResampling} -= x$, Recalculate MeanIR and IR
21	Return D'

3. Experiments

We evaluate the proposed intelligent diagnosis model using COEMR datasets, and verifies the effectiveness and universality of the model in two benchmark multi-label text classification datasets: AAPD and RCV1. Table 2 shows the descriptive statistics of the datasets used in the experiment. In the presentation learning stage, the data filter widths of CNN are set to (2, 3, 4) and the number of each filter is 25. When training the classifier, the Xavier method [23] is used to randomly initialize the classifier parameters. The resampling rate P of DRIL is set to 0.1, which is the best resampling rate for multi-label data [24]. Adam [25] is employed as the optimizer and the learning rate set to 0.001, batch size set to 32 and dropout set to 0.3.

Table 2. Statistical information of COEMR, AAPD and RCV1 datasets.

Dataset	Total	Train	Test	Label	MeanIR	SCUMBLE
COEMR	24,339	21,905	2434	73	246.5693	0.3028
AAPD	55,840	54,840	1000	54	16.9971	0.1158
RCV1	804,414	23,149	781,265	103	279.6319	0.3497

Table 3. Sample of an obstetric COEMR datasets.

Title	Content
Sex	Female
Age	Thirty-six years old
Chief complaint	Taking "rest of June, vaginal bleeding for 4 hours" as the chief complaint, the pregnant woman had regular menses with normal menses, and the urine HCG was positive by self-measurement for 30 days after menorrhea. After rest of January, the patient was diagnosed with ectopic pregnancy by B-ultrasound examination, and 40 days of menorrhea showed nausea and vomiting and other early pregnancy reactions...
Admission physical examination	T: 36.6 °C. P: 80/min R: 20/min BP:120/80 mm Hg. Normal development, nutrition medium, Clear headed, spirit can, walk into the ward, autonomous body position, check body cooperation. the whole-body skin mucous membrane is rudimentary and without yellow stain, rash, bleeding points, not touch the enlarged superficial lymph nodes...
Obstetric examination	Extrapelvic measurements is: 24.0 cm IC: 27.0 cm ec19.0 cm to: 9.0 CM. Uterine height: 29.0 cm abdominal circumference: 93.0 cm fetal heart 144 beats/min fetal estimated weight 2600 g, no contractions
Auxiliary examination	Fetal color ultrasound: BPD: 74.0 mm, FL: 53.0 mm, AFI: 165.0 mm, fetal orientation: breech s/D 2.2 placenta grade I
Admission diagnosis	Threatened preterm birth Placenta previa (marginal) Intrauterine pregnancy 28 ⁺² weeks G3P1 Breech presentation One week around umbilicus
Diagnostic basis	Pregnancy greater than or equal to 28 weeks and less than 37 weeks Presence of irregular or regular contractions with or without distension of the endocervical OS Minimal vaginal bleeding

COEMR: In this dataset, 24,339 EMRs were randomly selected from the inpatient departments of several hospitals. EMRs are mainly composed of structured and unstructured text data. Structured data includes basic patient information, such as age, ethnicity, and laboratory test data. Unstructured data mainly refers to patients' statement, condition of hospitalization and objective examination, etc. In order to protect the privacy of patients, the patient's name, ID number and other privacy information were removed. Table 3 shows the detailed description of the obstetric COEMR.

The distribution of COEMR datasets diagnostic results is visualized and shown in Table 4. In the COEMR datasets, more than 90% of the diagnostic results have "head position", while "gestational hypertension" accounts for less than 10%. According to the diagnostic results of 73 diseases with high coupling degree, all 24,339 samples were divided into training set (21,905) and test set (2434) according to 9 : 1.

Arxiv academic papers datasets (AAPD): AAPD datasets is a large dataset of MLTC constructed by Yang et al. [26]. It includes 55,840 abstracts from Arxiv¹ on computer science.

Reuters Corpus1 (RCV1): RCV1 is provided by Gleviset et al. [27] and consists of artificially annotated Reuters news from 1996 to 1997. Each piece of news can be assigned multiple topics, a total of 103 topics.

Table 4. Distribution diagnostic results in COEMR datasets.

Label	Number	Label	Number	Label	Number
Head position	18,139	Fetal dysplasia	1249	Induced labor	265
Threatened labor	6257	threatened abortion	1112	RH negative blood	259
Pregnancy with uterine scar	5757	Placenta previa	1033	Fetal distress	257
Premature rupture of membranes	3239	Preeclampsia	1029	Pregnancy induced hypertension	251
Oligohydramnios	2897	Precious child	819	Cervical insufficiency	217
Gestational diabetes mellitus	2661	Polyhydramnios	496	Pregnancy complicated with hysteromyoma	201
Threatened preterm birth	2130	Intrauterine fetal growth restriction	405	Diabetes complicated with pregnancy	189
Umbilical cord around neck	2054	Group B streptococcal infection	374	Pregnancy complicated with hyperthyroidism	182
Breech	1806	Pregnancy with hypothyroidism	335	Pregnancy complicated with anemia	178
Twin pregnancy	1329	Low placental	287	Inevitable abortion	177

¹ <https://arxiv.org/>

4. Results

This paper compares DDN model with common multi-label text classification methods: Binary Relevance (BR) [28], Label Powerset (LP) [29] and CNN. BP algorithm establishes a two classifier for each tag in the tag data set to predict, which has the advantages of simplicity and efficiency. LP algorithm regards the combination of each label as a new class, and transforms the multi label problem into a multi classification problem. Its advantage is that it pays attention to the semantic correlation between each label. Chen [30] used CNN for the first time in the text classification task. The key information in the sentence is extracted by multiple filters of different sizes in CNN, and the local correlation of the text can be better captured.

In addition, in order to verify the effectiveness of the proposed decoupled module, DDN is compared with BR + DRIL, LP + DRIL and RS + CNN + DRIL is to use the datasets balanced by the DRIL algorithm proposed in this section for classification. RS + CNN means that the data after class rebalancing is directly input into CNN for classification.

From the mean IR value in Table 2, it can be seen that the three datasets are imbalanced, so theoretically, the three datasets can benefit from the rebalancing method. It can be seen from the SCUMBEL value that the SCUMBEL value of COEMR datasets and RCV1 datasets is larger, which belongs to the difficult multi-label datasets, and has high coupling on the labels of different imbalance levels. Tables 5–7 show the experimental results on COEMR datasets, AAPD and RCV1 datasets respectively. In each column, the best results are expressed in bold.

Table 5. Experimental results of DDN in COEMR datasets.

Model	HL	P	R	F1
BR	0.0307	0.6114	0.5442	0.5758
LP	0.0305	0.6073	0.5030	0.5503
BR+DRIL	0.0281	0.6438	0.5581	0.5979
LP+DRIL	0.0293	0.6352	0.5176	0.5704
CNN	0.0266	0.8065	0.5427	0.6488
RS+CNN	0.0251	0.8134	0.5496	0.6560
DDN	0.0241	0.8417	0.5534	0.6678

Table 6. Experimental results of DDN in AAPD datasets.

Model	HL	P	R	F1
BR	0.0316	0.6642	0.6476	0.6558
LP	0.0312	0.6624	0.6082	0.6344
BR+DRIL	0.0433	0.6665	0.6491	0.6576
LP+DRIL	0.0320	0.6630	0.6113	0.6361
CNN	0.0256	0.8491	0.5456	0.6643
RS+CNN	0.0253	0.8540	0.5587	0.6755
DDN	0.0243	0.8635	0.5986	0.7071

Table 7. Experimental results of DDN in RCV1 datasets.

Model	HL	P	R	F1
BR	0.0086	0.9043	0.8160	0.8578
LP	0.0087	0.8956	0.8244	0.8585
BR+DRIL	0.0081	0.9192	0.8225	0.8682
LP+DRIL	0.0084	0.9034	0.8251	0.8625
CNN	0.0089	0.9223	0.7984	0.8559
RS+CNN	0.0086	0.9244	0.7986	0.8569
DDN	0.0076	0.9387	0.8231	0.8771

5. Discussion

It can be seen from the results in Tables 5 and 6 that after the application of the proposed DRIL on COEMR and RCV1 datasets, all measurements have been improved. Compared with the traditional BR algorithm, BR + DRIL algorithm improves the P, R and F1 by 5.30, 2.55 and 3.84% respectively. Compared with the traditional LP algorithm, LP + DRIL algorithm improves the P, R and F1 by 4.59, 2.90 and 3.65% respectively. For RCV1 datasets, due to the large number of labels and the complex hierarchical structure between labels, the improvement of each index is relatively small. Compared with the traditional BR algorithm, the F1 value of BR + DRIL algorithm is increased by 1.21%, and that of LP + DRIL algorithm is increased by 0.27%. This is because the DRIL algorithm decouples the high coupling labels, and combines the over-sampling and under-sampling methods to balance the distribution of the datasets, which makes it easier for the multi-label classification algorithm to process. In addition, the smaller SCUMBLE value of AAPD datasets indicates that there is almost no imbalanced label concurrency in the datasets, so the impact of using DRIL algorithm on the results is relatively small. DDN model is based on CNN. CNN is suitable for extracting local features. Using Bert and other models can have higher recall than DDN model, but F1 value may be reduced.

The neural network model can capture more abundant features and deeper semantic information, so CNN has a certain improvement in most of the evaluation indexes than the traditional classification methods BR and LP. Because CNN model is suitable for extracting local features, CNN model tends to select features favorable to high-frequency samples for unbalanced data, so the recall increment of CNN is relatively low. For the RS + CNN model, the traditional class resampling method can improve the frequency of low-frequency class samples, but without considering the high coupling between labels, the performance is only slightly improved. Compared with CNN model, the P value and F1 value of DDN are improved by 4.36 and 2.93%, respectively, reaching 84.17 and 66.78%, and the Hamming loss is reduced by 9.40%.

DDN model can solve the problem of high coupling of imbalanced labels by decoupling of representation learning from classifier learning and high coupling label so that the model can learn high-quality text feature representation, and performance has been further improved. DDN also performs well on AAPD and RCV1, which indicates that DDN model can also be applied to other multi-label text classification tasks.

This paper further evaluates the COEMR datasets to explore the effect of DDN model on different parts of the dataset distribution. As shown in the Figure 4, the horizontal axis is the diagnostic result sorted in descending order according to the number of corresponding samples, and the vertical axis is the accuracy increment of each category.

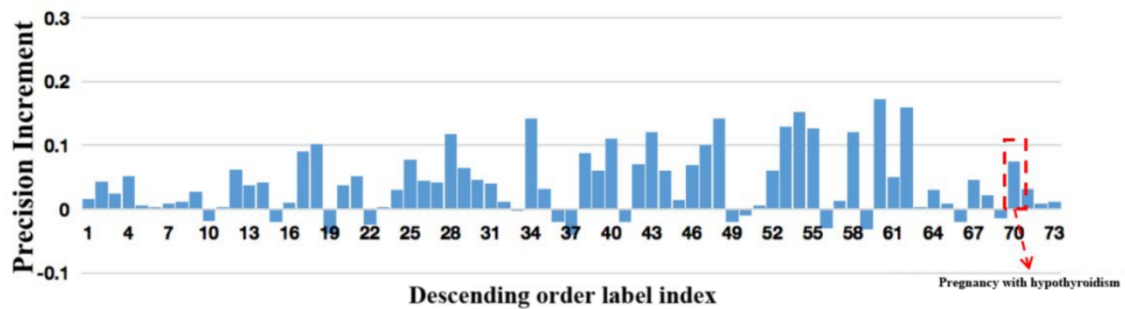


Figure 4. DDN increases the accuracy of diagnostic results on COEMR datasets.

It can be seen that DDN has a certain improvement in the diagnostic performance of low-frequency diseases. Taking the disease “Pregnancy with hypothyroidism” as an example, after label decoupling, the number of electronic medical records containing the disease will increase, and the model is more sensitive to the characteristics of this type of disease, such as “Excessive serum TSH”, “Edema” and other disease-related signs and symptoms. The decoupled low-frequency labels will not be restrained by the high-frequency labels. In addition, DDN does not improve the performance of low-frequency diseases by sacrificing the diagnostic accuracy of high-frequency diseases. After label decoupling, the performance of the model can be improved on most diseases. As mentioned above, the resampling method often results in over fitting of low-frequency data, and DDN decouples representation learning from classifier learning to learn good feature representations, and improve the generalization ability of low-frequency data. In conclusion, the DDN model including representation and classifier learning decoupling and label decoupling proposed in this paper has good performance for the diagnosis of low-frequency diseases.

6. Conclusions

This paper proposes a DDN model for intelligent diagnosis based on imbalanced EMRs. A two-stage training method is proposed to decouple the representation learning and classifier learning: in the representation learning stage, CNN model is used to learn the original features of data. In the classifier stage, considering the high coupling diagnostic results of EMRs, a DRIL algorithm is proposed to decouple the high coupling diagnostic results and balance the data distribution. The experimental results on COEMR datasets show that DDN can effectively improve the performance of intelligent diagnosis based on imbalanced EMRs, especially the precision of low-frequency disease diagnosis. In the future, we will try to use DDN for intelligent diagnosis in diseases with more complications such as diabetes.

Acknowledgments

We thank the anonymous reviewers for their constructive comments, and gratefully acknowledge the support of Major Science and Technology Project of Yunnan Province (202102AA100021),

Zhengzhou City Collaborative Innovation Major Projects (20XTZX11020), National Key Research and Development Program (2017YFB1002101), National Natural Science Foundation of China (62006211), Henan Science and Technology Research Project (192102210260), Henan Medicine Science and Technology Research Plan: Provincial and Ministry Co-construction Project (SB201901021), Henan Provincial Key Scientific Research Project of Colleges and Universities (19A520003, 20A520038), The MOE Layout Foundation of Humanities and Social Sciences (Grant No. 20YJA740033), Henan Social Science Planning Project (Grant No. 2019BYY016).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Han, M. Tong, L. Jin, W. Meng, A. Ren, Maternal age at pregnancy and risk for gestational diabetes mellitus among Chinese women with singleton pregnancies, *Int. J. Diabetes Dev. Countries*, **41** (2021), 114–120. <https://doi.org/10.1007/s13410-020-00859-8>
2. K. Zhang, H. Ma, Y. Zhao, H. Zan, L. Zhuang, The comparative experimental study of multilabel classification for diagnosis assistant based on Chinese obstetric EMRs, *J. Healthcare Eng.*, (2018), 1–9. <https://doi.org/10.1155/2018/7273451>
3. C. Xu, P. Liu, Y. Sun, Research on disease prediction model for unbalanced medical datasets, *Chin. J. Comput.*, **42** (2019), 596–609. <https://doi.org/10.11897/SP.J.1016.2019.00596>
4. Y. Liu, H. Loh, A. Sun, Imbalanced text classification: A term weighting approach, *Expert Syst. Appl.*, **36** (2009), 690–701. <https://doi.org/10.1016/j.eswa.2007.10.042>
5. J. Stefanowski, Dealing with data difficulty factors while learning from imbalanced data, in *Challenges in Computational Statistics and Data Mining*, Springer, Cham, (2016), 333–363. https://doi.org/10.1007/978-3-319-18781-5_17
6. B. Zhou, Q. Cui, X. Wei, Z. Chen, BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 9719–9728. <https://doi.org/10.1109/CVPR42600.2020.00974>
7. B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, et al., Decoupling representation and classifier for long-tailed recognition, preprint, arXiv:1910.09217.
8. Q. Yin, D. Shen, Y. Tang, Q. Ding, Intelligent monitoring of noxious stimulation during anaesthesia based on heart rate variability analysis, *Comput. Biol. Med.*, **145** (2022), 105408. <https://doi.org/10.1016/j.combiomed.2022.105408>
9. T. Yan, P. Wong, C. Choi, C. Vong, H. Yu, Intelligent diagnosis of gastric intestinal metaplasia based on convolutional neural network and limited number of endoscopic images, *Comput. Biol. Med.*, **126** (2020), 104026. <https://doi.org/10.1016/j.combiomed.2020.104026>
10. S. Wang, Y. Zhang, X. Cheng, X. Zhang, Y. Zhang, PSSPNN: PatchShuffle stochastic pooling neural network for an explainable diagnosis of COVID-19 with multiple-way data augmentation, *Comput. Math. Methods Med.*, (2021), 1–18. <https://doi.org/10.1155/2021/6633755>
11. A. Rajkomar, E. Oren, K. Chen, A. Dai, N. Hajaj, M. Hardt, et al., Scalable and accurate deep learning with electronic health records, *NPJ Digital Med.*, **1** (2018), 1–10. <https://doi.org/10.1038/s41746-018-0029-1>

12. A. Maxwell, R. Li, B. Yang, H. Weng, A. Ou, H. Hong, et al., Deep learning architectures for multi-label classification of intelligent health risk prediction, *BMC. Bioinf.*, **18** (2017), 523. <https://doi.org/10.1186/s12859-017-1898-z>
13. Z. Yang, Y. Huang, Y. Jiang, Y. Sun, Y. Zhang, P. Luo, et al., Clinical assistant diagnosis for electronic medical record based on convolutional neural network, *Sci. Rep.*, **8** (2018), 6329. <https://doi.org/10.1038/s41598-018-24389-w>
14. H. Liang, B. Tsui, H. Ni, C. Valentim, S. Baxter, G. Liu, et al., Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence, *Nat. Med.*, **25** (2019), 433. <https://doi.org/10.1038/s41591-018-0335-9>
15. N. Liu, E. Qi, M. Xu, B. Gao, G. Liu, A novel intelligent classification model for breast cancer diagnosis, *Inf. Process. Manage.*, **56** (2019), 609–623. <https://doi.org/10.1016/j.ipm.2018.10.014>
16. C. Huang, X. Huang, Y. Fang, J. Xu, Y. Qu, P. Zhai, et al., Sample imbalance disease classification model based on association rule feature selection, *Pattern Recognit. Lett.*, **133** (2020), 280–286. <https://doi.org/10.1016/j.patrec.2020.03.016>
17. B. Krawczyk, Learning from imbalanced data: Open challenges and future directions, *Prog. Artif. Intell.*, **5** (2016), 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
18. X. Liu, X. Sun, Y. Meng, J. Liang, F. Wu, J. Li, Dice loss for data-imbalanced NLP tasks, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), 465–476. <https://doi.org/10.48550/arXiv.1911.02855>
19. J. Yang, Z. Qu, Z. Liu, Improved feature-selection method considering the imbalance problem in text categorization, *Sci. World J.*, (2014), 625342. <https://doi.org/10.1155/2014/625342>
20. F. Charte, A. Rivera, M. Jesus, F. Herrera, A first approach to deal with imbalance in multi-label datasets, in *International Conference on Hybrid Artificial Intelligence Systems*, (2013), 150–160. https://doi.org/10.1007/978-3-642-40846-5_16
21. F. Charte, A. Rivera, M. Jesus, F. Herrera, Concurrence among imbalanced labels and its influence on multilabel resampling algorithms, in *International Conference on Hybrid Artificial Intelligence Systems*, (2014), 110–121. https://doi.org/10.1007/978-3-319-07617-1_10
22. F. Charte, A. Rivera, M. Jesus, F. Herrera, Resampling multilabel datasets by decoupling highly imbalanced labels, in *International Conference on Hybrid Artificial Intelligence Systems*, (2015), 489–501. https://doi.org/10.1007/978-3-319-19644-2_41
23. X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, (2010), 249–256. <http://proceedings.mlr.press/v9/glorot10a>
24. F. Charte, A. Rivera, M. Jesus, F. Herrera, Addressing imbalance in multilabel classification: Measures and random resampling algorithms, *Neurocomputing*, **163** (2015), 3–16. <https://doi.org/10.1016/j.neucom.2014.08.091>
25. D. Kingma, J. B. Adam, A method for stochastic optimization, preprint, arXiv:1412.6980.
26. P. Yang, X. Sun, W. Li, S. Ma, W. Wu, H. Wang, SGM: Sequence generation model for multi-label classification, in *Proceedings of the 27th COLING*, (2018), 3915–3926. <https://doi.org/10.48550/arXiv.1806.04822>
27. D. Lewis, Y. Yang, T. Rose, F. Li, Rcv1: A new benchmark collection for text categorization research, *Mach. Learn. Res.*, (2004), 361–397. <https://research.gold.ac.uk/id/eprint/29758>

28. M. Boutell, J. Luo, X. Shen, C. Brown, Learning multi-label scene classification, *Pattern Recognit.*, **37** (2004), 1757–1771. <https://doi.org/10.1016/j.patcog.2004.03.009>
29. G. Tsoumakas, I. Katakis, Multi-label classification: An overview, *Int. J. Data Warehous. Min.*, **3** (2007), 1–13. <https://doi.org/10.4018/jdwm.2007070101>
30. Y. Chen, Convolutional neural networks for sentence classification, in *Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2014), 1–62. <http://hdl.handle.net/10012/9592>



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)