IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ON IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received July 11, 2020, accepted July 19, 2020, date of publication July 23, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011374

Prediction of Chronic Diseases With Multi-Label Neural Network

RUIQUAN GE[1], RENFENG ZHANG[2], AND PU WANG[3]

[1]Key Laboratory of Complex Systems Modeling and Simulation, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021, China
[3]Computer School, Hubei University of Arts and Science, Xiangyang 441053, China

Corresponding author: Pu Wang (wp3751@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702146, Grant 61841104, and Grant 81602490; in part by the Zhejiang Postdoctoral Foundation under Grant zj 20180025; in part by the Shandong Province Key Research and Development Projects under Grant 2017GSF218067; and in part by the China Scholarship Council under Grant 201808330081.

ABSTRACT Chronic diseases have seriously affected human activities, especially in many developing countries and underdeveloped countries. The long duration of chronic diseases and the high cost of medical care have placed a huge economic burden on society and families. Meanwhile, chronic patients tend to have a variety of complications over time. So, it is difficult for doctors to find effective diagnosis and appropriate treatment. Machine learning techniques can integrate their heterogeneous data of various body indicators. Meanwhile, for chronic patients, multi-label learning methods can be used to help doctors identify the types of the chronic diseases. This paper proposes a novel multi-label neural network method (ML-NN) to predict the chronic diseases combining neural network and multi-label learning technology based on cross entropy lost function and backward propagation algorithm. Compared with 14 traditional multi-label learning methods on 10 chronic diseases and 19733 patients, the proposed method achieved a consistently best in 5 performance measurements. The results demonstrate the proposed method can effectively predict chronic diseases and assist doctors to diagnose and treat patients.

INDEX TERMS Chronic diseases prediction, multi-label learning, neural network.

I. INTRODUCTION

Chronic diseases are also known as noncommunicable diseases (NCDs), which are distinguished by a long duration and slow development. Chronic diseases usually include cardiovascular diseases (such as hypertension, coronary heart disease, and stroke), chronic respiratory diseases, and diabetes [1]. The long duration of chronic diseases and the high cost of medical care have placed a huge economic burden on society and families [2]. At present, the situation of prevention and treatment of chronic diseases in all countries is still severe, especially in low- and middle-income countries [3]. The burden of diseases caused by chronic diseases accounts a large proportion of the total diseases, and the number of deaths caused by chronic diseases is gradually increasing [4].

Because chronic diseases with complex causes often develop into complications, it is difficult for doctors to find appropriate treatment. Although continuous monitoring of patients using modern advanced medical technologies and devices (e.g., wearable medicine, mobile health [5], etc.) can obtain a large amount of patient-related data, these data is often heterogeneous, such as laboratory test values, physical values, or electrocardiograms [6]. Doctors need to use data mining or machine learning tools to summarize and analyze these complex data to assist them to make optimal decisions [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Kin Fong Lei.

A. RELATED WORK

It is very common that the research object has multi-label feature in the field of biomedicine, such as protein subcellular localization [8]–[12], bioenzyme function [13], [14], membrane protein function [15], [16], protein function prediction [17], [18] etc. As the course of chronic diseases increases, other diseases are often caused. Therefore, chronic patients often have the characteristics of multiple diseases. So, multiple complications of chronic patients also can be modeled analytically using multi-label learning algorithms [19], [20].

To solve these problems, the researchers have proposed many multi-label learning algorithms [21], especially in the field of images and text [22]–[24]. Usually these methods can be divided into two categories: problem transformation methods and algorithm adaptation methods [25]. The problem transformation method commonly uses data to adapt to the algorithm, which directly calls or combines the existing classification algorithm without changing it. The typical representatives of problem transformation methods have Binary Relevance (BR) [26], Classifier Chains (CC) [27], Label Powerset (LP) [28], Hierarchy Of Multi-label Learners (HOMER) [29], and Random k-label sets (RAkEL) [30]. BR method uses one-to-all binary classification strategy to realize the multi-label classification. CC has an extension to amend the drawback of BR method which does not consider the correlation between the labels. LP methods transform the multi-label problem into a single-class classification problem by considering all the possible combinations. HOMER is a labels-set tree structure built by clustering algorithm. RAkEL improves the computational efficiency of LP by decomposing the original labels set into k smaller random subsets.

Whereas, the algorithm adaptation method is to directly process multi-label data after extending the existing learning algorithm. Typical representatives have Multi-Label k-Nearest Neighbor (ML-kNN) [31], Multi-Label Decision Tree (ML-DT) [32], Ranking Support Vector Machine (Rank-SVM) [33] and Backpropagation for Multi-Label Learning (BP-MLL) [34].

B. CONTRIBUTION

Medical data on diseases are often complex and multi-source. The diversified data sources often cannot guarantee the independent and identical distribution of data. Thus, the neural network method can learn the true characteristics of complex data more than traditional machine learning classification methods. Currently, neural networks have become the most popular machine learning technologies, which have outstanding advantages in hierarchical feature description and complex function mapping [35]. Using neural network architecture to solve the multi-label learning problem has natural advantages. The contributions of this work can be summarized as follows:

(1) A novel multi-label neural network method (ML-NN) is proposed combining neural network and multi-label learning technology. Neural network itself is a multiple-input multiple-output system [36, 37] which is ideal for multi-label learning problem.

(2) The model shows strong recklessness and can search for optimal solutions at high speed. In this study, the knowledge background of multi-source data is unclear and the reasoning rules are unclear. The neural network has the functions of self-learning, fault tolerance and associative memory. It does not impose any restrictions on the input and residual distribution, and has the ability to learn and construct nonlinear complex correlation of this chronic disease dataset.

(3) The activation function designed based on cross entropy lost function and backward propagation algorithm has two obvious advantages: it captures the non-linear correlation between the inputs. And it helps transform the input into a more useful output. Then, it can quickly and accurately give the diagnosis of the patient's disease.

(4) Meanwhile, a detailed comparative analysis also showed the advantage of our method compared with many common multi-label learning algorithms. Experiments established that the proposed method could achieve better prediction accuracy and operating efficiency, so as to better assist doctors in the effective diagnosis and treatment for chronic disease patients. Therefore, the multi-label neural network method we designed can better solve the multi-source data problem of chronic diseases.

C. STRUCTURE

The rest of the paper is organized as follows. Section II describes a comprehensible process of data analysis and processing. In this Section, a multi-label neural network method is proposed to solve the problem of chronic diseases. In Section III, the experimental results of different multi-label methods are analyzed and compared. The experimental results confirmed the advantage of our method. Finally, in Section IV, some brief conclusions and open problems are drawn.

II. MATERIALS AND METHODS

A. DATASET

The initial disease dataset in this study were derived from MIMIC-II, published by Beth Israel Deaconess Medical Center (BIDMC) [38]. MIMIC is a public database that was began to collect patient information in 2001 and continued for seven years. The original database contains approximately 33,000 patients. Among them, the clinical data includes laboratory tests and medical records. Each patient's medical record generally includes a number of results from heterogeneous examinations, such as body fluid examinations, physiological measurements and rigorous scores of some vital functions.

B. FEATURE EXTRACTION AND STANDARDIZATION

After the chronic disease dataset was integrated and created, a series of operation are processed such as missing attribute values, features extraction and standardization. This paper mainly analyzes and compares with some common multi-label classification algorithms after preprocessing the raw dataset, including record extraction and missing value processing. The adult population (above 16 years old) after removing newborns and children contains about 24,000 patients. And then, after removing individuals without chronic diseases, 19733 patients were finally obtained. Their average age was 67 years. And the proportion of men and women was 56% and 44%, respectively [19].

Laboratory records and medical history records are classified according to different measured values. For numerical variables such as biochemical criterion, blood pressure and temperature, they are a part of the feature vector if they appear only one time. Otherwise, the features are represented by their mean, median, standard etc. if they appear more than once. For categorical variables, their features are coded as binarization if there is only one observation. If a variable has multiple observations, it is discretized into a mutex class. Each class is characterized by its frequency of occurrence.

Finally, 10 chronic diseases and 76 attributes were obtained including 39 quantitative attributes (Table S I) and 37 category attributes (Table S II). The names of these chronic diseases were defined after considering the medical relevance of the dataset, the characteristics of the data and the hierarchy of the International Classification of Diseases, 9th Revision (ICD-9) coding system. In order to avoid too many missing values, only those items that are included in at least 80% of the patients are selected in the laboratory tests and medical records [39]. The information of the chronic diseases is shown in Table 1. 310 features extracted from 76 attributes were standardized by z-score.

TABLE 1. Distribution of labels of 19,773 patients with chronic diseases extracted in MIMIC-II database.

Disease categories	Number of patients	Percent (%)	ICD-9 codes
Hypertensive disease	12309	62.3	[401–405]
Fluid electrolyte disease	6177	31.2	276
Diabetes mellitus	6056	30.6	[249–250]
Lipoid metabolism disease	5965	30.2	272
Kidney disease	5828	29.5	[580–589]
COPD	4253	21.5	[490–496]
Thyroid disease	2246	11.4	[240–246]
Hypotension	1962	9.9	458
Liver disease	1088	5.5	571
Thrombosis	931	4.7	[451–453]

C. MULTI-LABEL CLASSIFIERS

For the two categories of multi-label classification methods, the compared multi-label learning algorithms include BR, HOMER, CC, RAkEL, ML-kNN, and AdaBoostMH [40]. Among them, AdaBoostMH is the multi-label adaptation of AdaBoost which belongs to problem adaptation method. ML-kNN is the multi-label adaptation of kNN. Other methods such as BR, HOMER, CC, and RAkEL belong to the problem-transformed label learning algorithm, which need to call a single label learning algorithm, such as SVM (Support Vector Machine), J48 (an implementation of a decision tree algorithm), and NB (Naive Bayes). SVM with an RBF kernel needs optimize two parameters: the coefficient of kernel and the penalty of the error term. J48 can easily explain the rules of classification and scale to large dataset. J48 also needs optimize two parameters: the confidence threshold for pruning and the minimum number of instances per leaf. NB has a simple structure and a surprising classification performance. HOMER needs one parameter to define the number of clusters for the k-means. RAkEL needs optimize two parameters: the number of models in the ensemble and the size of the subset of labels in each model of the ensemble. ML-kNN has two parameters including the number of neighbors and a smooth parameter. Therefore, there are 14 kinds of multi-label learning algorithms to be compared here. The parameters of these algorithms are optimized by grid search. The parameter tuning settings of various algorithms are consistent with the literature [19].

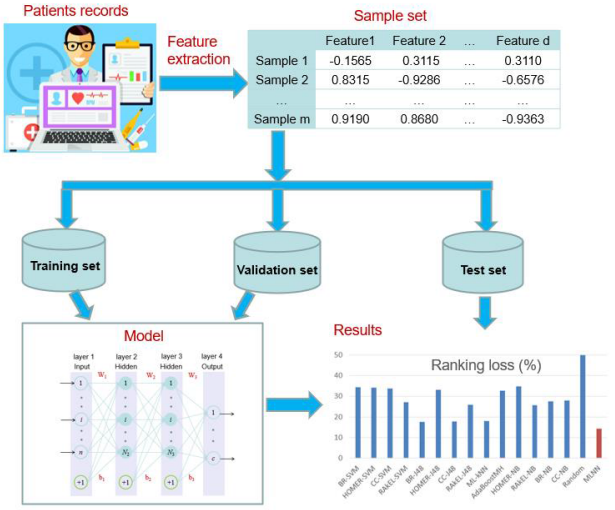

FIGURE 1. Chronic diseases prediction model based on multi-label neural network.

D. MULTI-LABEL NEURAL NETWORK METHOD

Here we demonstrate that a multi-label neural network method (ML-NN) featuring a multi-label cost function and the rectified linear unit (ReLU) activation function performs efficiently and accurately on multi-label disease diagnosis (Figure 1). After preprocessing and feature extraction of multi-source data, two-dimensional vector form is generated as the samples and features. Then the samples are divided into training, verification and test datasets. The proposed multi-label neural network method is used to train and construct the model with the first two datasets respectively. At the same time, the model is evaluated using the test dataset by several common evaluation indexes compared with a few multi-label learning algorithms.

In our multi-label neural network, all the hidden layers use the ReLU activation function [41], and the output layer uses the Sigmoid activation function because what people usually want to know is the probability of getting a particular type of disease. The cost function of neural network for single label classification cannot be used for multi-label learning, because the output class labels are not mutually exclusive to each other. So, we employ Multi-Label Cross Entropy (MLCE) as the cost function of the neural network.

Suppose a training dataset $D = (x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})$, m is the number of training samples, $x^{(i)}$ is the i-th sample, the class label assignment of $x^{(i)}$ may be represented as a c-dimensional binary vector $y^{(i)} = [y_1^{(i)}, y_2^{(i)}, \ldots, y_c^{(i)}]$, where

$y_j^{(i)} = 1$ if the sample has the j-th label, and $y_j^{(i)} = 0$ if not. If all the weight parameters in the network are indicated as (W,b), then MLCE can be represented as:

$$J(W,b;D) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{c}\left[y_j^{(i)} log p_j^{(i)} + \left(1-y_j^{(i)}\right) log\left(1-p_j^{(i)}\right)\right] + \lambda\sum_{l=1}^{L-1}\|W_l\|_F^2, \quad (1)$$

where c is the number of possible disease labels, $p^{(i)} = [p_1^{(i)}, p_2^{(i)}, \ldots, p_c^{(i)}]$ is the actual output of sample $x^{(i)}$, $W_l (l = 1, \ldots, L-1)$ are the connection weights between layers, L is the depth of a network, λ is the weight decay parameter of regularization term, and F indicates Frobenius Norm [42]. Based on error back propagation algorithm [43], we can get the optimal parameters to make the multi-label neural network fit the training dataset very well. The regularization term is used to avoid over-fitting.

Specifically, given a training sample (x, y), the parameters in the network can be updated by the following error back-propagation procedure:

I. Calculate the activation values of all layers by forward propagation.

II. For the output layer, the error is

$$\delta_L = p - y, \quad (2)$$

III. For the hidden layer when $l = L-1, L-2, \ldots, 2$, the error is

$$\delta_l = \left(W_l^T \delta_{l+1}\right) \cdot f'(z_l), \quad (3)$$

where z_l is the input of the layer l, f is the activation function, and $\cdot$ is the element-wise product.

IV. Calculate the partial derivatives as below

$$\nabla_{Wl} J(W,b;x,y) = \delta_{l+1}(a_l)^T, \quad (4)$$

$$\nabla_{bl} J(W,b;x,y) = \delta_{l+1}, \quad (5)$$

where a_l is the activation of layer l.

Since $J(W,b) = -\frac{1}{m}\sum_{i=1}^{m} J(W,b;x,y) + \lambda\sum_{l=1}^{L-1}\|W_l\|_F^2$, it is easy to get the batch gradients $\nabla_{Wl} J(W,b)$ and $\nabla_{bl} J(W,b)$ $(l = 1, \ldots, L-1)$, and then the model parameters are optimized by the gradient descent algorithm.

E. PERFORMANCE EVALUATION

Assume that $\chi = \mathbf{R}^d$, $\Lambda = \{l_1, l_2, \ldots, l_c\}$, $S = \{(x_i, y_i)|i = 1, 2, \ldots, m'\}$ respectively represents a d-dimensional sample space, a finite label set, and a test set consisting of m' multi-label samples, where $x_i \in \chi$ is a feature vector in the sample space and $y_i \subseteq \Lambda$ is the set of labels associated with the sample. The classifier $h(\bullet)$ can predict the set $h(x) \subseteq \Lambda$ which x may have. $f(\bullet,\bullet)$ is a output function of the multi-label learner, where $r(x,l)$ represents the sorted value of label l. The following evaluation indicators can be used to measure the performance of the multi-label learner [21].

- Hamming Loss

$$HammingLoss = \frac{1}{m'}\sum_{i=1}^{m'}\frac{|h(x_i)\Delta y_i|}{c}, \quad (6)$$

where Δ represents the symmetric difference between two sets, and $|\bullet|$ is used to find the potential of the set (the number of elements). Hamming Loss indicates the percentage of all predictors that are wrong. The smaller the Hamming Loss, the better. The optimal value is 0.

- One-error

$$OneError = \frac{1}{m'}\sum_{i=1}^{m'} I([argmax_{l\in\mathcal{L}} f(x_i,l)] \notin y_i), \quad (7)$$

One-error represents the proportion of samples in which the label corresponding to the top-rank output value of all samples does not belong to the relevant label set.

- Coverage

$$Coverage = \frac{1}{m'}\sum_{i=1}^{m'} max_{l\in y_i} r(x_i,l) - 1, \quad (8)$$

If depth is used to represent the maximum sorted value of all relevant labels for a sample, Coverage is the average depth of all samples.

- Ranking Loss

$$RankingLoss = \frac{1}{m'}\sum_{i=1}^{m'}\frac{1}{|y_i|\,|\bar{y}_i|} \times \left|\left\{(l,l')|f(x_i,l) \le f(x_i,l'), (l,l') \in y_i \times \bar{y}_i\right\}\right|, \quad (9)$$

where $\bar{y}_i$ is the complement of y_i, that is, the set of irrelevant labels for sample x_i. Ranking Loss is the average ranking loss of all samples. For a single sample, its ranking loss is the percentage of ranking errors (the output value of the relevant label is less than the output of the unrelated label) among all the pairs of related labels and unrelated labels.

- Average Precision

$$AveragePrecision = \frac{1}{m'}\sum_{i=1}^{m'}\frac{1}{|y_i|}\sum_{l\in y_i}\frac{\left|\left\{,l'|r(x_i,,l') \le r(x_i,l),,l' \in y_i\right\}\right|}{r(x_i,l)}, \quad (10)$$

Average Precision evaluates the average fraction of relevant labels ranked higher than a particular label $l \in Y_i$.

III. RESULTS AND DISCUSSION

Just like the data preprocessing in [19], the experimental dataset was randomly divided equally into three subsets according to the standard machine learning algorithm, in which the training set was used for model learning, the verification set for parameter adjustment and model selection, and the test set for results comparison between models [44]. Meanwhile, the results of all methods are obtained after parameter optimization.

Similar to other neural network models, hyperparameters have a great impact on the performance of the proposed

TABLE 2. Performance comparison of different multi-label learning algorithms in independent test set.

Method /Metric	Hamming loss (%)↓	Ranking loss (%)↓	Average precision (%)↑	One-error (%)↓	Coverage↓
BR-SVM	16.94±0.12	34.47±0.64	61.85±0.56	35.17±0.67	5.35±0.05
HOMER-SVM	16.97±0.11	34.18±0.61	62.01±0.56	35.34±0.74	5.33±0.05
CC-SVM	17.01±0.14	33.67±0.70	62.58±0.59	35.11±0.80	5.28±0.06
RAkEL-SVM	17.18±0.07	27.08±0.45	68.32±0.54	30.68±0.81	4.72±0.04
BR-J48	17.63±0.13	17.70±0.31	72.23±0.55	31.15±0.70	3.50±0.01
HOMER-J48	17.75±0.17	33.06±1.05	62.45±0.89	34.74±0.87	5.31±0.10
CC-J48	17.83±0.15	17.76±0.40	72.14±0.79	31.48±1.32	3.50±0.02
RAkEL-J48	18.17±0.09	25.88±1.32	68.53±0.99	32.28±0.66	4.57±0.13
ML-kNN	18.91±0.16	18.10±0.17	71.37±0.22	32.05±0.36	3.52±0.02
AdaBoostMH	21.23±0.09	32.81±0.13	57.65±0.19	37.75±0.29	4.89±0.01
HOMER-NB	21.78±0.30	34.76±0.47	57.61±0.99	44.35±2.89	5.29±0.03
RAkEL-NB	24.88±0.40	25.63±0.39	64.07±0.76	45.93±2.19	4.37±0.03
BR-NB	28.40±0.34	27.62±0.50	60.32±0.70	54.80±1.37	4.42±0.03
CC-NB	28.61±0.36	27.92±0.51	59.96±0.71	55.55±1.40	4.44±0.03
Random	49.89±0.19	49.99±0.43	40.13±0.40	76.03±0.54	6.30±0.03
ML-NN	**16.59±0.09**	**14.30±0.27**	**76.25±0.46**	**26.74±0.73**	**3.12±0.01**

Best result is shown in bold. ↓ : the lower the better. ↑ : the higher the better.

TABLE 3. Performance comparison of neural networks with different hidden layers in ML-NN model.

Layers	Hamming loss (%)↓	Ranking loss (%)↓	Average precision (%)↑	One-error (%)↓	Coverage↓
2	16.59±0.09	14.30±0.27	76.25±0.46	26.74±0.73	3.12±0.01
3	16.63±0.08	14.40±0.11	76.08±0.34	27.10±0.66	3.15±0.01
4	16.59±0.08	14.31±0.25	76.26±0.43	26.74±0.67	3.12±0.02
5	16.93±0.07	14.93±0.18	75.61±0.39	27.41±0.68	3.20±0.02
6	17.21±0.09	15.39±0.13	75.24±0.25	27.64±0.45	3.25±0.01

Best result is shown in bold. ↓ : the lower the better. ↑ : the higher the better.

multi- label neural network (ML-NN), such as loss function, learning rate, regularization weight, the number of hidden layers and nodes, etc. [45] The experimental network consists of two hidden layers with 100 nodes in each layer. The learning rate, the regularization weight, and the maximum number of iterations is set to 0.01, 0.001 and 1000 respectively.

The experimental results of the optimal models of various multi-label learning algorithms are shown in Table 2 for the test dataset. From the results, decision trees can generally achieve good classification performance. For the methods based on SVM classification, their indexes of hamming loss are better, but the indexes of ranking loss are worse. ML-kNN has poor scalability and performance comparable to decision trees. AdaBoostMH cannot obtain competitive results because its optimized objective is mainly the hamming loss rather than the ranking loss. For this multi-source dataset, there are many correlations between features. So, the classification effect of NB method is not significant. A random multi-label classification method is experimented to know the upper and lower bounds of classification. For the transformation method CC, its effect of classification is not improved significantly compared with the BR method. HOMER also does not obtain an obvious improvement compared with other methods except for NB as the basic classifier. RAkEL with an SVM classifier obtains good performance with respect to other methods especially in average precision and ranking loss metrics.

Thus, the proposed multi-label neural network model (ML-NN) in this paper is superior to traditional multi-label learning algorithms in various indicators [19]. In terms of average results, only Hamming loss has a slight (approximately 2% decrease) advantage over the second-ranked algorithm BR-SVM. The advantages of other indicators are obvious. The Ranking loss is approximately 19% lower than the second-ranked algorithm BR-J48. One-error is reduced by about 12% compared with the second-ranked algorithm RAkEL-SVM. Coverage is about 11 percent lower than second-place BR-J48. Average precision was about 6 percent higher than the second-place CC-J48.

The amount of this dataset in this study is not very large, so it is not suitable for the neural network method with deep network layers. Since the dataset is relatively small, we have tried several shallow neural network models. Performance comparison of neural networks with different hidden layers was shown in detail in Table 3. The results show that the neural networks with two hidden layers, three hidden layers and four hidden layers have almost the same classification effect. As the number of network layers increases, the classification effect becomes worse. Therefore, the two-layer neural network is effective based on Occam's razor principle [46]. In addition, the method proposed in this paper also has the general disadvantages of neural network and its interpretability needs to be improved.

Furthermore, this paper also compares the training time (accurate to minutes) of various multi-label learning algorithms, as shown in Table 4. The ML-NN is tested in personal computer (Intel Core i3@3.40GHz, 4GB RAM, Matlab R2013b). And other methods refer to the test of Damien Zufferey et al in the JAVA software package Mulan 1.4 environment (Intel Core i7@2.93GHz, 16GB RAM) [40]. Due to the large samples of chronic diseases, the methods using SVM as the base classifier are very slow because of the grid searches for optimal parameters, especially when combined with the RAkEL method which adopts integrated

TABLE 4. Comparison of training time for the different multi-label learning algorithms.

Multi-label learning algorithms	Training time (h:hours, m:minutes)
BR-NB	1 m
CC-NB	1 m
HOMER-NB	1 m
BR-J48	4 m
CC-J48	4 m
HOMER-J48	4 m
AdaBoostMH	6 m
RAkEL-NB	7 m
RAkEL-J48	15 m
ML-kNN	35 m
BR-SVM	3 h 32 m
CC-SVM	3 h 33 m
HOMER-SVM	4 h 11 m
RAkEL-SVM	28 h 13 m
ML-NN	7 m

learning strategy. Among them, ML-kNN is a method based on nearest neighbor search, which has poor scalability to the data scale. So the running time of ML-kNN is relatively long. The training time of other methods is at the level of minutes, which is relatively fast. Although ML-NN uses a 4-layer neural network structure, it only takes 7 minutes to train the dataset after 1000 iterations because of the high efficiency of the ReLU activation function, which brings great convenience for network optimization and hyperparameter adjustment. As we all know, the operating efficiency of Matlab language is generally not as high as that of Java language. Perhaps the Java version of ML-NN will further reduce the training time.

IV. CONCLUSION

Chronic patients tend to have a variety of complications. For these patients, multi-label learning can be used to identify their complications. Comparative experiments on 10 chronic diseases show that the proposed multi-label neural network algorithm (ML-NN) in this paper is significantly better than the traditional multi-label learning algorithms. Maybe this dataset is relatively large and the neural network can fit it well even if there are relevant features. What's more, the proposed method is also very competitive in running time, which is mainly due to the efficiency of multi-label loss function. Therefore, the proposed method can effectively assist doctors in the diagnosis and treatment of patients with chronic diseases. In the future, the interpretability of neural networks will be analyzed in order to make new discoveries and deepen the understanding of chronic diseases.

APPENDIX

Table S I 39 quantitative attributes in chronic diseases dataset.

Table S II 37 category attributes in chronic diseases dataset.

ACKNOWLEDGMENT

(Ruiquan Ge and Renfeng Zhang contributed equally to this work.) This article was presented in part at the 2020 International Conference on Biomedicine, Bioinformation and Intelligent Computing (BBIC2020).

REFERENCES

[1] R. Nugent, "Preventing and managing chronic diseases," *BMJ*, vol. 364, p. l459, Jan. 2019.
[2] S. Mendis, S. Davis, and B. Norrving, "Organizational update: The world health organization global status report on noncommunicable diseases 2014; one more landmark step in the combat against stroke and vascular disease," *Stroke*, vol. 46, no. 5, pp. e121–e122, May 2015.
[3] V. J. Wirtz, W. A. Kaplan, G. F. Kwan, and R. O. Laing, "Access to medications for cardiovascular diseases in low-and middle-income countries," *Circulation*, vol. 133, pp. 2076–2085, May 24 2016.
[4] G. B. D. D. Collaborators, "Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017," *Lancet*, vol. 393, no. 10184, pp. 1958–1972, May 11 2019.
[5] I. C. Jeong, D. Bychkov, and P. C. Searson, "Wearable devices for precision medicine and health state monitoring," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1242–1258, May 2019.
[6] S. J. Kabadi and A. Krishnaraj, "Strategies for improving the value of the radiology report: A retrospective analysis of errors in formally over-read studies," *J. Amer. College Radiol.*, vol. 14, no. 4, pp. 459–466, Apr. 2017.
[7] J. Béal, A. Montagud, P. Traynard, E. Barillot, and L. Calzone, "Personalization of logical models with multi-omics data allows clinical stratification of patients," *Frontiers Physiol.*, vol. 9, p. 1965, Jan. 2019.
[8] X. Guo, F. Liu, Y. Ju, Z. Wang, and C. Wang, "Human protein subcellular localization with integrated source and multi-label ensemble classifier," *Sci. Rep.*, vol. 6, no. 1, p. 28087, Jun. 2016.
[9] S. Wan, M.-W. Mak, and S.-Y. Kung, "Sparse regressions for predicting and interpreting subcellular localization of multi-label proteins," *BMC Bioinf.*, vol. 17, no. 1, p. 97, Feb. 2016.
[10] X. Wang, W. Zhang, Q. Zhang, and G.-Z. Li, "MultiP-SChlo: Multi-label protein subchloroplast localization prediction with Chou's pseudo amino acid composition and a novel multi-label classifier," *Bioinformatics*, vol. 31, no. 16, pp. 2639–2645, Aug. 2015.
[11] S. Wan, M.-W. Mak, and S.-Y. Kung, "MPLR-loc: An adaptive decision multi-label classifier based on penalized logistic regression for protein subcellular localization prediction," *Anal. Biochem.*, vol. 473, pp. 14–27, Mar. 2015.
[12] S. Wan, M.-W. Mak, and S.-Y. Kung, "Transductive learning for multi-label protein subchloroplast localization prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 1, pp. 212–224, Jan. 2017.
[13] Y. Che, Y. Ju, P. Xuan, R. Long, and F. Xing, "Identification of multi-functional enzyme with multi-label classifier," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0153503.
[14] L. De Ferrari and J. B. Mitchell, "From sequence to enzyme mechanism using multi-label machine learning," *BMC Bioinf.*, vol. 15, no. 1, p. 150, May 2014.
[15] S. Wan, M.-W. Mak, and S.-Y. Kung, "Mem-ADSVM: A two-layer multi-label predictor for identifying multi-functional types of membrane proteins," *J. Theor. Biol.*, vol. 398, pp. 32–42, Jun. 2016.
[16] X. Xiao, H.-L. Zou, and W.-Z. Lin, "iMem-Seq: A multi-label learning classifier for predicting membrane proteins types," *J. Membrane Biol.*, vol. 248, pp. 745–752, Aug. 2015.
[17] Y. Xu, H. Min, H. Song, and Q. Wu, "Multi-instance multi-label distance metric learning for genome-wide protein function prediction," *Comput. Biol. Chem.*, vol. 63, pp. 30–40, Aug. 2016.
[18] J.-S. Wu, S.-J. Huang, and Z.-H. Zhou, "Genome-wide protein function prediction through multi-instance multi-label learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 11, no. 5, pp. 891–902, Sep. 2014.
[19] D. Zufferey, T. Hofer, J. Hennebert, M. Schumacher, R. Ingold, and S. Bromuri, "Performance comparison of multi-label learning algorithms on clinical data for chronic diseases," *Comput. Biol. Med.*, vol. 65, pp. 34–43, Oct. 2015.
[20] S. Bromuri, D. Zufferey, J. Hennebert, and M. Schumacher, "Multi-label classification of chronically ill patients with bag of words and supervised dimensionality reduction algorithms," *J. Biomed. Inform.*, vol. 51, pp. 165–175, Oct. 2014.
[21] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.
[22] B. Al-Salemi, S. A. Mohd Noah, and M. J. Ab Aziz, "RFBoost: An improved multi-label boosting algorithm and its application to text categorisation," *Knowl.-Based Syst.*, vol. 103, pp. 104–117, Jul. 2016.
[23] Y. C. Wei, W. Xia, M. Lin, J. Huang, and B. Ni, "HCP: A flexible CNN framework for multi-label image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1901–1907, Sep. 2016.

[24] J. Du, Q. Chen, Y. Peng, Y. Xiang, C. Tao, and Z. Lu, "ML-net: Multi-label classification of biomedical texts with deep neural networks," *J. Amer. Med. Inform. Assoc.*, vol. 26, no. 11, pp. 1279–1285, Nov. 2019.
[25] P. Szymanski and T. Kajdanowicz, "Scikit-multilearn: A scikit-based Python environment for performing multi-label classification," *J. Mach. Learn. Res.*, vol. 20, Jan. 2019.
[26] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognit.*, vol. 37, no. 9, pp. 1757–1771, Sep. 2004.
[27] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Mach. Learn.*, vol. 85, no. 3, pp. 333–359, Dec. 2011.
[28] J. Read, "A pruned problem transformation method for multi-label classification," in *Proc. New Zealand Comput. Sci. Res. Student Conf.*, 2008, pp. 143–150.
[29] I. K. Grigorios Tsoumakas and I. Vlahavas, "Effective and efficient multilabel classification in domains with large number of labels," in *Proc. ECML/PKDD Workshop Mining Multidimensional Data (MMD)*, pp. 30–44, 2008.
[30] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-Labelsets for multilabel classification," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 7, pp. 1079–1089, Jul. 2011.
[31] M. L. Zhang and Z. H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, pp. 2038–2048, Jul. 2007.
[32] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data," in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery*, vol. 2168, 2002, pp. 42–53.
[33] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2002, pp. 681–687.
[34] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1338–1351, Oct. 2006.
[35] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[36] A. Maxwell, R. Li, B. Yang, H. Weng, A. Ou, H. Hong, Z. Zhou, P. Gong, and C. Zhang, "Deep learning architectures for multi-label classification of intelligent health risk prediction," *BMC Bioinf.*, vol. 18, no. S14, p. 523, Dec. 2017.
[37] X. Zhang, H. Zhao, S. Zhang, and R. Li, "A novel deep neural network model for multi-label chronic disease prediction," *Frontiers Genet.*, vol. 10, p. 351, Apr. 2019.
[38] M. Saeed, M. Villarroel, A. T. Reisner, and G. Clifford, "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, May 2011.
[39] R. J. Little, R. D'Agostino, M. L. Cohen, K. Dickersin, S. S. Emerson, J. T. Farrar, and C. Frangakis, "The prevention and treatment of missing data in clinical trials," *New England J. Med.*, vol. 367, pp. 1355–1360, Oct. 2012.
[40] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, and I. Vlahavas, "Mulan: A java library for multi-label learning," *J. Mach. Learn. Res.*, vol. 12, pp. 2411–2414, Jun. 2011.
[41] P. Wang, R. Ge, X. Xiao, Y. Cai, G. Wang, and F. Zhou, "Rectified-Linear-Unit-Based deep learning for biomedical multi-label data," *Interdiscipl. Sci., Comput. Life Sci.*, vol. 9, no. 3, pp. 419–422, Sep. 2017.
[42] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear-norm and Frobenius-Norm-Based representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.
[43] V. V. Phansalkar and P. S. Sastry, "Analysis of the back-propagation algorithm with momentum," *IEEE Trans. Neural Netw.*, vol. 5, no. 3, pp. 505–506, May 1994.
[44] M. Huang, F. Zhuang, X. Zhang, X. Ao, Z. Niu, M.-L. Zhang, and Q. He, "Supervised representation learning for multi-label classification," *Mach. Learn.*, vol. 108, no. 5, pp. 747–763, May 2019.
[45] C. P. George and H. Doss, "Principled selection of hyperparameters in the latent Dirichlet allocation model," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 5937–5974, 2018.
[46] J. Schaffer, "What not to multiply without necessity," *Australas. J. Philosophy*, vol. 93, no. 4, pp. 644–664, Oct. 2015.

RUIQUAN GE received the Ph.D. degree in computer application technology from the Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences, in 2017. From February 2019 to June 2020, he was a Visiting Scholar with the Toyota Technological Institute at Chicago (TTIC). He is currently a Teacher with the School of Computer Science and Technology, Hangzhou Dianzi University. His current research interests include machine learning and bioinformatics.

RENFENG ZHANG received the Ph.D. degree from the School of Medicine, Shanghai Jiaotong University, in 2015. He is currently a Master's Supervisor with Shandong University. He is also an Associate Chief Technician with the Shandong Provincial Hospital, China. His current research interests include tumor immunology and clinical microbiology.

PU WANG received the Ph.D. degree from the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, in 2017. He is currently an Associate Professor with the Computer School, Hubei University of Arts and Science, Xiangyang, China. His research interests include machine learning and bioinformatics.