IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 23, 2020, accepted March 3, 2020, date of current version March 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2979218

Diagnosis of Chronic Obstructive Pulmonary Disease Based on Transfer Learning

QIAN WANG[1], HONG WANG[1,2], LUTONG WANG[1], AND FENGPING YU[1]
[1]School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China
[2]Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, Shandong Normal University, Jinan 250358, China

Corresponding author: Hong Wang (wanghong106@163.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61672329 and Grant 81273704, in part by the Project of the Shandong Provincial Project of Education Scientific Plan under Grant SDYY18058, and in part by the Clinical Medical Science Data Center for providing COPD data.

ABSTRACT Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease that seriously endangers human health and has high incidence and mortality worldwide. Therefore, an effective predictive model is required for COPD diagnosis. Given the limited data samples available in current COPD studies, we propose a method for diagnosing COPD based on transfer learning called balanced probability distribution (BPD) algorithm; this algorithm integrates instance- and feature-based transfers to improve the prediction accuracy of the model. First, instance-based cascaded transfer learning was used to initialize the weight distribution of the training data and obtain instances closer to the target domain. Second, the cross-domain feature filtering algorithm was adopted to filter irrelevant features, eliminate redundant features, and obtain the co-occurrence features of the source and target domains. Moreover, the remaining features were assigned different weights and transformed into the same space to reduce the distribution difference between the domains. Third, the BPD algorithm was used to balance the examples and the co-occurrence features from multiple disease source domains and construct a more suitable classification model of the target domain. Finally, the elastic network was used to further improve the generalization performance of the model. The experimental results show that the prediction effect of the BPD model is better than that of state-of-the-art methods and has strong generalization ability and robustness. We proved that our proposed BPD method works well in the COPD prediction model when the sample size is small.

INDEX TERMS Balanced probability distribution (BPD) algorithm, chronic obstructive pulmonary disease (COPD), feature extraction, few-shot learning, transfer learning.

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a preventable and treatable common disease characterized by persistent respiratory symptoms and restricted airflow. According to the global initiative for COPD (hereinafter GOLD), this sickness is often associated with airway or alveolar abnormalities caused by significant exposure to toxic particles or gases. COPD is a chronic respiratory disease that seriously threatens people's health and has a high incidence and mortality globally. The World Health Organization has reported that COPD has become the third leading cause of deaths globally and will become one of the leading respiratory diseases in China by 2020 [1]. In China, the prevalence of COPD in people over 40 years of age is 9.9%. The morbidity and mortality associated with COPD are often underestimated because of the differences in the diagnostic criteria and an insufficient understanding of the disease; the rate of missed diagnosis is as high as 70% [2]. Therefore, a predictive data mining model with good clinical reliability is an important requirement for diagnosis, treatment, and self-management of COPD [3].

In recent years, different machine-learning models have been used for predicting COPD [1], which can be characterized as either non-deep (i.e., traditional) or deep [4]. A traditional model typically comprises two major steps: feature engineering [5] and model building [6]. Feature engineering extracts the ''good'' features that are effective for constructing the model. Unlike traditional methods, the deep learning model [7] has an end-to-end learning mechanism in which the feature engineering part is implicitly integrated into the learning pipeline. Deep learning has attracted the attention

The associate editor coordinating the review of this manuscript and approving it for publication was Jing Bi.

of researchers in various fields because of its superior performance. In deep learning, it is possible to learn advanced features using big data; therefore, it is a representation-learning algorithm based on large-scale data used in machine learning. However, in some bioinformatics fields, such as COPD diagnosis, it is very difficult to construct large-scale well-labeled datasets because of the high cost of data collection and labeling; this limits the development of deep learning in these fields.

If only a small amount of data is used for training, it will lead to the problem of model overfitting and low reliability [8]. Moreover, imbalances in medical data make it difficult to obtain efficient disease prediction models. Transfer learning [9] relaxes the assumption that training data must be independent and identically distributed (i.i.d) with the test data. In addition, transfer learning applies the knowledge learned from existing tasks to new models or fields to improve the learning performance and obtain efficient prediction models greatly. In fact, transfer learning has become a new learning framework to solve many knowledge transfer problems [10]. In this work, it has a great positive impact on performance improvement despite limited training data.

There have been numerous studies on COPD [11] since the 1980s. However, the existing treatment methods for COPD do not consider the possible complications and lack a systematic approach to exploring the interactions between COPD and the complications. Clinically, the pathogenesis, clinical manifestations, diagnosis, treatment, and management of COPD complications and simple COPD+ complications will have varying degrees of differences. As stated by GOLD in 2017, the natural course of COPD develops from complex systematic outcomes and complications, which are the main characteristics of clinical COPD [1]. The systemic outcome of COPD refers to the direct pulmonary phenotype and the non-pulmonary manifestations caused by COPD.

Therefore, in this paper, we propose a novel transfer learning method based on instance-based and feature-based transfer called the balanced probability distribution (BPD) method, which solves the problem of having a small sample size. The main contributions of this research are as follows:

(1) We propose a transfer learning method called BPD, which integrates instance transfer and feature transfer; this method includes the advantages of both these methods but avoids their disadvantages. BPD is effective in the prediction model of COPD when there are only a limited number of samples. BPD can not only improve the accuracy of COPD prediction, but it can also improve the reusability and robustness of the model.

(2) By adopting cascading instance-based transfer in our BPD, we solved the problem of sparse data in the COPD prediction because the effective examples learned from multiple disease source domains were retained.

(3) By adopting a cross-domain feature filtering method in BPD to obtain co-occurrence features and increasing their weight in the COPD prediction model, we avoided overfitting and other problems caused by the direct training of COPD domain samples. At the same time, the regularization constraints of the elastic network were used to optimize the learning performance of the model further.

(4) We conducted numerous experiments to verify the effectiveness of the BPD method in this paper. In comparison with other methods, BPD has a good effect on all indicators. The BPD prediction effect is better than other methods, and it also has a strong generalization ability, which finally proves that the BPD method improves the prediction ability for small sample data.

II. RELATED WORK

In recent years, with the rapid development of computer technology and the widespread application of medical data systems, an increasing amount of disease-related data has become readily available. Based on these data, researchers have established many disease diagnoses, predictions, and classification models and obtained good results.

A. DISEASE CLASSIFICATION

The correct classification of diseases is the basis of disease diagnosis. Most disease prediction models based on machine learning formalize the disease prediction problems into classification problems [12]. Brisimi *et al.* [13] used a joint clustering and classification method to explore and predict the number of people hospitalized for heart disease and diabetes. Han *et al.* [14] proposed a probabilistic path score method to distinguish between two main types of inflammatory bowel diseases. After the classification, the probability graph model was used to include gene interactions to obtain better performance. Besides, ensemble learning has dramatically contributed to the development of disease classification. Yosipof *et al.* [15] combined multiple machine learning models to form an integrated learning method called AL Boost, which achieved the goal of classifying tumor compounds and neurological diseases.

Presently, neural network techniques are being gradually applied to disease prediction models. To prove the effectiveness of neural networks, Lipton *et al.* [16] modeled and analyzed the multi-label medical records, constructed a disease prediction model based on LSTM, and performed experiments according to the changes in a patient's medical characteristics. Moreover, Anthimopoulos *et al.* [17] used deep convolution neural networks to classify interstitial pulmonary diseases into pulmonary patterns, achieving a close match between the classification results and basic facts. Li *et al.* [18] used natural language processing and deep learning technology to automatically learn expert doctor diagnostic modes from historical medical record data to form intelligent assistant diagnostic models.

B. COPD DIAGNOSIS

COPD is a common, preventable, and treatable disease. It is characterized by persistent respiratory symptoms and

restricted airflow because of abnormalities in the respiratory tract and/or alveoli. This is usually caused by excessive exposure to harmful particles or gases [1]. Previous studies have focused on early diagnosis of mild COPD, severity assessment, differential diagnosis of high-risk groups, and so on. Demographic studies in Denmark [19] and Finland [20] found significant correlations between the educational levels, household incomes, and COPD prevalence. A Swedish study reported that in addition to the educational level, COPD prevalence was related to age, smoking, and a history of tuberculosis [21]. Another urban population survey report in Sweden used the Pearson correlation coefficient to assess the correlation between the socio-economic variables and risk factors on the incidence of COPD [22]. In China, there has also been a precedent for exploring the relationship between stress hormone levels and prognosis in elderly patients with COPD [23].

In addition, several studies have analyzed the features of the influencing factors and the prediction of disease risk at different stages. Most of these studies are based on data analysis methods. Marin *et al.* [24] studied the prognostic value of the BODE quartiles (i.e., the of body mass index, degree of airflow limitation, dyspnea, and exercise capacity) for the number and severity of patients with COPD who needed outpatient treatment, emergency, or hospitalization. The Cox regression models and BODE indexes predict the deterioration. Jensen *et al.* [25] built a linear regression model to distinguish the acute and non-acute phases of COPD with an accuracy rate of 73%. van der Heijden *et al.* [26] used cross-validation and ROC analysis to evaluate whether or not the patient's condition worsened with a probability model. Christopher [27] used multi-stage logistic regression to distinguish between stable and deteriorating periods.

C. TRANSFER LEARNING

Transfer learning is an essential branch of machine learning, which achieves knowledge transfer between similar tasks and has provided excellent results in the fields of medical health, computer vision, and recommendation systems. Transfer learning is divided into instance-based and feature-based transfer learning methods. The instance-based transfer learning method is based on a specific weight generation rule for reusing data samples for transfer learning. Dai *et al.* [28] proposed the TraAdaBoost plan, which applied the idea of AdaBoost to transfer learning to improve the instance weights beneficial to the target classification task and reduce the instance weights that are not conducive to the target classification task. The feature-based transfer learning methods usually assume that there are some overlapping features between the source and target domains. In the transfer component analysis (TCA) method proposed by Pan *et al.* [29], the core content is based on the maximum mean discrepancy (MMD) [30] taken as the measurement criterion to minimize the distribution differences in different data fields.

The instance-based and feature-based methods have advantages and disadvantages. The instance-based method has excellent theoretical support. By increasing the weight of the relevant samples and reducing the weight of the irrelevant samples in the target domain, knowledge becomes more suitable for migration to the target domain. However, data that are not similar to the target-domain data are always retained during the learning process. However, the feature-based methods can remove dissimilar features and share the common features between domains. A feature-based method relies on the marginal probability distribution to reduce the difference between the source and target domains, thereby reducing its generalization ability. Therefore, this paper proposes the BPD method by fusing the instance-based and feature-based transfer learning methods. It has advantages both methods, which reduces the overall errors and enhances the classification accuracy. First, the BPD model transfers knowledge step-by-step by using the same instance as the basis to obtain instances close to the target domain. Then, the cross-domain filtering feature algorithm is used to obtain the common features across the source and target domains. To handle the problem that the source domain data and target domain data may have different distributions, the features were transformed to the same space and given different weights, that is, the common features in the source and target domains had high weights and features irrelevant to the target domain possessed low weights. Furthermore, we adopted the MMD to reduce the distribution distance between the different fields. Finally, the elastic network was used for training-related instances to improve the generalization ability of the model.

III. BALANCED PROBABILITY DISTRIBUTION

The training data and test data with i.i.d are the necessary requirements for a training model that has good prediction accuracy in the test data. However, uniformly distributed data are rare in real life. For example, if a model trained on book review texts is directly applied for predicting movie review text data, it is likely that the result would be unsatisfactory because of the different data distributions. Therefore, training models with sound testing effects in multiple fields are particularly important for practical situations. Transfer learning can determine the potential relationship between the source and target domains and further build a suitable target domain model based on the learned source domain knowledge. Therefore, narrowing the data distribution distance between the fields is the key to cross-domain learning and building models. However, the data from different domains have different distributions, which makes it difficult to fit all the distributions simultaneously. Moreover, simply matching the distributions cannot guarantee the prediction effect on the target domain. Therefore, in this paper, we propose the BPD algorithm based on the instance and feature transfer methods. The BPD framework is shown in Fig. 1.

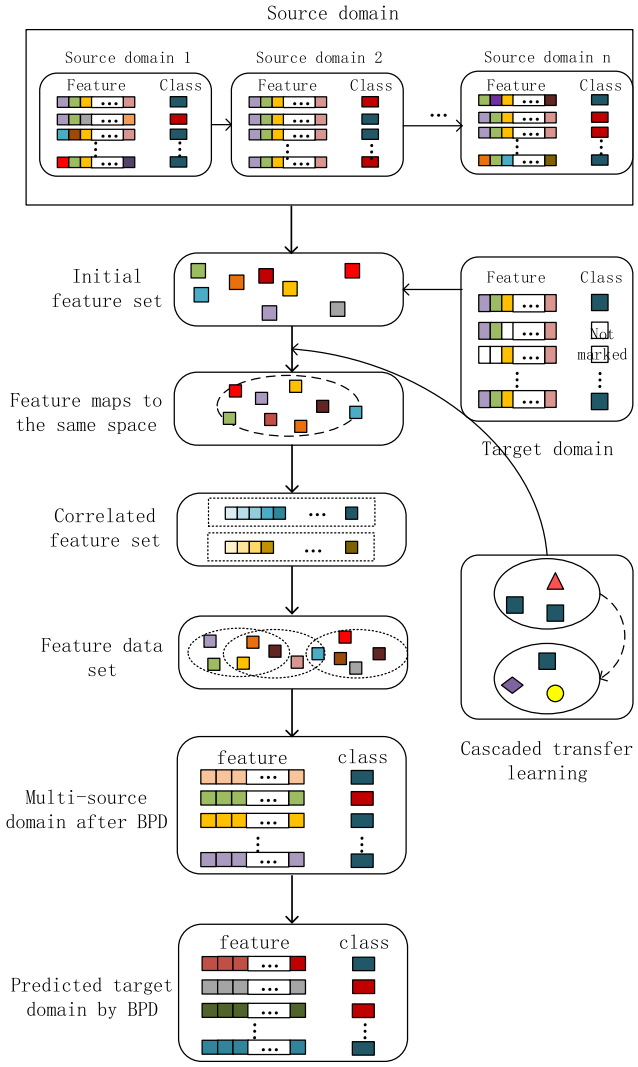


FIGURE 1. BPD framework for COPD diagnosis and prediction based on feature and instance transfer methods.

A. INSTANCE-BASED CASCADED TRANSFER LEARNING

We first consider transfer-specific prior medical knowledge in our BPD method. In fact, since the 1980s, there have been in-depth studies on COPD and its complications, which have given us considerable medical knowledge [11]. The first report about COPD and its complications was given by the McSweeny study in 1982 [1], which revealed that the incidence of COPD complications was 42%. In 2001, Laeasse found that 57% of patients with COPD also suffered from depression. In 2014, scholars from Stanford showed that COPD was accompanied by mental and sleep disorders [31]. Zhong pointed out that in addition to causing organ damage, COPD also affected the patient's daily life and led to other diseases [32]. Professor Zhang noted that COPD combined with depression produced lesions in the lung and then affected the spleen, kidney, heart, and brain [33]. GOLD 2017 [1] pointed out that COPD patients often developed a variety of complications, including heart disease, chronic respiratory failure, depression, and cognitive impairment. If these complications are found, patients should be given appropriate

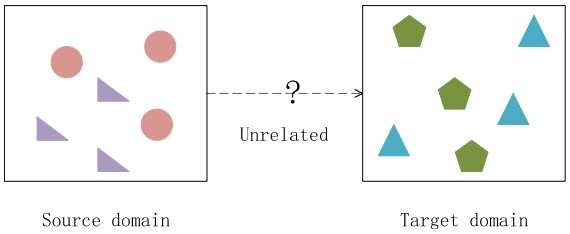


FIGURE 2. Example of the source domain and target domain with no common instance.

treatment as soon as possible. In summary, this kind of knowledge is the basis for us to realize cascaded instance-based transfer learning for improving the accuracy of the diagnosis and prediction of COPD. Here we can give a more specific example. Prior clinical knowledge shows that chronic bronchitis and emphysema are the two most common diseases that lead to COPD. If a patient has a serious disease, such as chronic bronchitis or emphysema, he can be diagnosed with COPD. At times, we could even face a situation in which there are almost no common instances between the source and target domains (see Fig. 2). However, in such situations, it is possible to realize a smooth step-by-step instance-based transfer of knowledge (see Fig. 3).

Instance-based cascaded transfer learning first initializes the weight distribution of the training instance data [34]. Then, the weighted instances are used for transfer learning. By multiple transfers of intermediate instances, we can not only obtain instances closer to the target domain but also obtain a more diversified effect [35], [36]. In particular, the first step is to find the same instances to bridge different domains; the second step is to migrate between domains. We select one or more instances to connect to a given domain and then transfer knowledge through their overlapping instances. The selection of instances is problem-specific, and different scenarios have different options. The process can be explained using the following example.

Suppose patient A suffers from chronic bronchitis, gastritis, diarrhea, and heart disease; patient B suffers from chronic bronchitis, upper airway obstruction syndrome, arthritis and diabetes; and patient C suffers from upper airway obstruction syndrome, coronary heart disease, and diabetes. We cannot transfer knowledge from patient A to patient C because there is no common disease instance between them. Therefore, we take the disease of patient B as a path to transfer instances from patients A to C. Both patients A and B suffer from chronic bronchitis. Therefore, for migration, we assign a high weight to patient A for chronic bronchitis. It is the same as patients B and C. The common diseases between B and C are upper airway obstruction syndrome and diabetes; we assign a high weight to the upper airway obstruction syndrome. If there are multiple candidate instances, we transfer them together. However, we cannot determine whether there are common instances between the source and final target domains; therefore, we incorporate feature-based transfer to reduce the overall errors and achieve high classification accuracy.

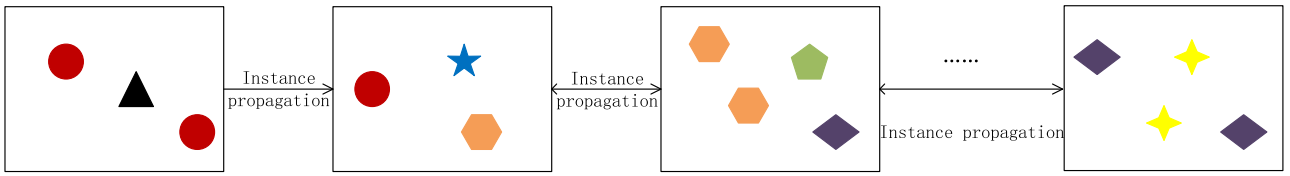


FIGURE 3. Diagram of smooth step-by-step.

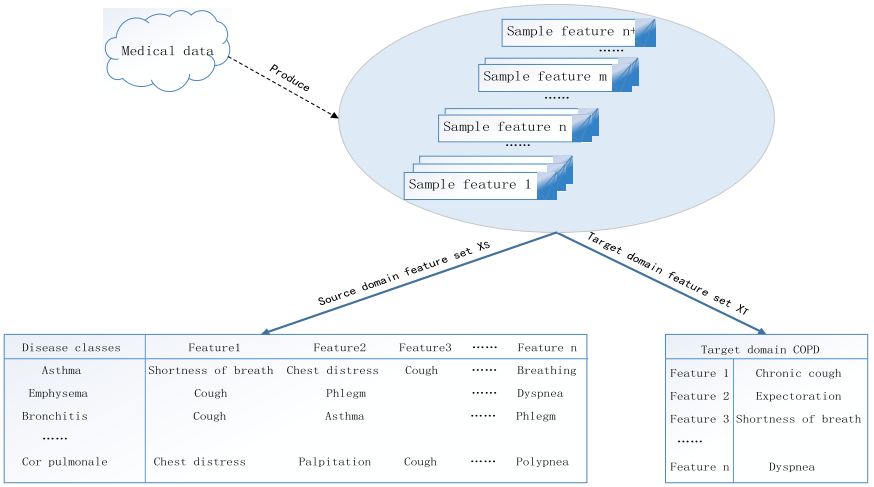


FIGURE 4. Example of feature representation in the original feature space.

B. CROSS-DOMAIN FEATURE FILTERING

We use the feature filtering algorithm to obtain common features between the source and target domains. After the common features are obtained, they are mapped to the same space. Then, the irrelevant features are filtered, and the redundant characteristics are eliminated.

The steps of the cross-domain feature filtering algorithm are as follows:

Step 1: Map features to a common feature space using the multidimensional scaling (MDS) method. In this space, the degree of difference between the features can be retained, and the relationship between the features can be found.

Step 2: Filter the irrelevant and redundant features and obtain the relevant features with the approximate Markov blanket defined by the symmetric uncertainty (SU).

The detailed algorithm is as follows:

In step 1, we use (MDS) [37] to map the features in the different domains to a unified feature space. In fact, there are other methods, such as principal component analysis (PCA), t-distributed neighborhood embedding (t-SNE), and singular value decomposition (SVD), which can achieve the goal of feature mapping. Our experiments show that the effects of these methods are lower than that of MDS. The comparison results are given in Table 1.

The MDS process is as follows. Assuming that the dataset X contains $n + m$ features, the dataset is represented as

$$X = (X_S, X_T) = (x_{s1}, x_{s2}, \ldots, x_{sn}, x_{t1}, x_{t2}, \ldots x_{tm}) \quad (1)$$

TABLE 1. Comparison of different feature mapping methods.

Method	Average accuracy (%)	Standard deviation(%)	Dimensionality reduction (%)
Initial	78.6	0.30	0
PCA	82.3	0.32	41.7
t-SNE	80.9	0.31	50.6
SVD	83.5	0.30	48.3
MDS	85.1	0.31	66.9

where X_S is the source domain feature set, X_T is the target domain feature set, n is the number of features in the source domain, and m is the number of features in the target domain. A concrete example is shown in Fig. 4.

We assume that the distance matrix of $(n + m)$ features in the original space is D, and the element D_{ij} in the ith row and jth column is the distance from the sample x_i to x_j. The goal is to obtain the representation Z of the samples in the common feature space, with the Euclidean distance of any two samples in the space being equal to the distance in the original space, that is, $||z_i - z_j|| = D_{ij}$. From (2), the MDS plots the similarity or the distance between multiple features in a low-dimensional Euclidean space. Further, we determined the relationship between the various features. We assumed that B is the inner product matrix mapped to the feature space of the common samples. Let $B = Z^T Z$, and $b_{ij} = z_i^T z_j$, then

$$\left(D_{ij}^X\right)^2 = \left(x_i - x_j\right)^T \left(x_i - x_j\right) = \|x_i\|^2 - 2x_i^T x_j + \|x_j\|^2 \quad (2)$$

The sample of the common feature space is $\sum_{i=0}^{n+m} z_i = 0$ by dimension reduction and mean subtraction. Obviously, the sum of the rows and columns of matrix B is 0, that is, $\sum_{i=1}^{n+m} b_{ij} = \sum_{j=1}^{n+m} b_{ij} = 0$.

We can see that

$$\sum_{i=1}^{n+m} D_{ij}^2 = tr(B) + (n+m) b_{jj} \tag{3}$$

$$\sum_{j=1}^{n+m} D_{ij}^2 = tr(B) + (n+m) b_{ii} \tag{4}$$

$$\sum_{i=1}^{n+m} \sum_{j=1}^{n+m} D_{ij}^2 = 2(n+m) tr(B) \tag{5}$$

Here, tr(·) represents the trace of the matrix $tr(B) = \sum_{i=1}^{m} \|z_i\|^2$. Let

$$D_{i\cdot}^2 = \frac{1}{n+m} \sum_{j=1}^{n+m} D_{ij}^2 \tag{6}$$

$$D_{\cdot j}^2 = \frac{1}{n+m} \sum_{i=1}^{n+m} D_{ij}^2 \tag{7}$$

$$D_{\cdot\cdot}^2 = \frac{1}{(n+m)^2} \sum_{i=1}^{n+m} \sum_{j=1}^{n+m} D_{ij}^2 \tag{8}$$

From (2), (3), and (8), we obtain

$$b_{ij} = -\frac{1}{2} \left(D_{ij}^2 - D_{i\cdot}^2 - D_{\cdot j}^2 + D_{\cdot\cdot}^2 \right) \tag{9}$$

Thus, we can obtain the inner product matrix B by mapping to the distance matrix D that stays constant before and after the eigenspace.

Eigenvalue decomposition of the matrix B, $B = V \wedge V^T$, where $\wedge = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$ is the diagonal matrix comprising the eigenvalues, and V is the eigenvector matrix. Assuming that d∗ of the nonzero eigenvalues constitutes a diagonal matrix $\wedge_* = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{d*})$, V_* represents the corresponding eigenvector matrix. We anticipate that the distance mapped to the feature space is as close as possible to the distance in the original space without having to be strictly equal. Then Z can be expressed as

$$Z = \Lambda_*^{1/2} V_*^T \tag{10}$$

Z is the representation of the sample in the new feature space.

Subsequently, considering the correlation between the features and categories in the case of a small data sample, we use the SU method to perform correlation analysis to measure the contribution of various features to the different classes.

In step 2, the irrelevant features were filtered by using the SU method, and the redundant features were removed using the approximate Markov blankets. Thus, the relevant features were selected, which were the common features.

SU is a non-linear related information measure based on the definition of information entropy [38], and the information entropy E(x) of the feature x is defined as follows:

$$E(x) = -\sum_i P(x_i) \log_2 P(x_i) \tag{11}$$

The information entropy E (y) of the class y is defined as follows:

$$E(y) = -\sum_j P(y_j) \log_2 P(y_j) \tag{12}$$

The conditional entropy E ($x|y$) is given as follows:

$$E(x|y) = -\sum_j P(y_i) \sum_i P(x_i|y_j) \log_2 P(x_i|y_j) \tag{13}$$

Based on the above equations, we define the mutual information (MI) between feature x and class y. MI describes how much the information uncertainty of class y is reduced if and only if the information in the feature space x is determined. Obviously, different features have different MI with class y. The definition of the MI between the feature x and class y is as follows:

$$MI(y|x) = E(x) - E(x|y) \tag{14}$$

As mentioned, the features with high MI have stronger classification capabilities; that is, they have more critical features for classification. Thus, we define the maximum MI as follows:

$$MI_{\max}(y|x) = \frac{MI(x|y)}{\min\{E(x), E(y)\}} \tag{15}$$

Definition 1: Symmetric uncertainty

The SU calculated for feature x and class y is defined as follows:

$$SU(x|y) = \frac{2MI_{\max}(y|x)}{E(x) + E(y)} \tag{16}$$

It is known that the value of SU is between [0,1]. When the SU value is close to 1, the correlation between x and y will be high. When the SU value is close to 0, the relationship between x and y will be small. In extreme cases, if the SU value is 0, then x and y will be completely uncorrelated; therefore, x will be removed as an unrelated feature.

Now, all the relevant features remain. However, not all these features are necessary for classification because redundant features were also included, which will reduce the classification accuracy. Therefore, we use the heuristic method [40] to approximate the Markov blankets called approximate Markov blankets to retain relatively weak correlation characteristics through SU.

Definition 2: Approximate Markov blankets.

Suppose x_i and x_j are features, and y is a class. Feature x_i is an approximate Markov blanket of feature x_j if SU (x_i, y) ≥ SU (x_j, y), and SU (x_i, x_j) ≥ SU (x_j, y).

TABLE 2. Effect of approximate Markov blankets.

Step	Average accuracy (%)	Standard deviation	Dimensionality reduction (%)
Step 1: MDS	85.1	0.31	66.9
Step 2: SU	86.7	0.32	72.0
Step 3: Approximate Markov blanket	88.2	0.30	81.6

The idea of approximate Markov blankets is based on the Markov blanket [39]. The Markov blanket contains information on the features and classes. It is used to eliminate the features because of redundancy and retain features without any redundancy. It has been proved that features highly relevant to the class are not redundant; therefore, they will not be removed at any stage. This meets the requirements of our research. However, we need not only some highly relevant features but also some weakly relevant features. Therefore, we had to adopt approximate Markov blankets to retain relatively weak correlation characteristics because traditional Markov blanket cannot achieve this. Therefore, using the approximate Markov blanket, the irrelevant features and some redundant features were removed; this helped us obtain the updated feature set.

The experimental results proved that the approximate Markov blanket effectively removes the redundant features and improves the dimensionality reduction rate (Table 2).

In summary, the cross-domain feature filtering algorithm can effectively filter irrelevant features and eliminate redundant features. Therefore, the co-occurrence features of the source and target domains can be found, which helps to establish a classification model in the target domain. Fig. 5 shows the co-occurring features in multiple areas such as COPD, emphysema, and bronchitis, which were found by the cross-domain feature filtering algorithm.

C. BALANCED PROBABILITY DISTRIBUTION MODEL

1) JOINT MATCHING MARGINAL PROBABILITY DISTRIBUTION AND CONDITIONAL PROBABILITY DISTRIBUTION

In machine learning, we assume that both the training and test data obey the same distribution, and under this assumption, the predictive risk of the model can be controlled. Transfer learning does not comply with this assumption. In other words, both marginal distribution differences and the conditional distribution differences need to be considered. Marginal distribution refers to the distribution of the generated data, and conditional distribution refers to the distribution between the data and labels. Our goal is to reduce the distance between the marginal probability distribution and conditional probability distribution of the source domain and the target domain, thereby completing the transfer learning. The joint matching marginal probability distribution and conditional probability distribution can combine the advantages of the two and compensate for their shortcomings; this will greatly improve the training effect.

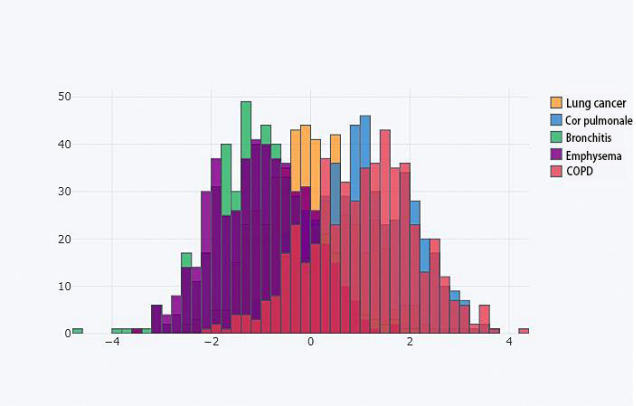


FIGURE 5. Co-occurrence features found by cross-domain feature filtering algorithm.

Marginal probability distribution and conditional probability distribution were not equally important in practical applications when the BPD method fused the instances and features; therefore, we introduced the balance parameter λ to adjust the two probability distributions dynamically. The probability distribution distance is given as follows:

$$D\left(D_s, D_t\right) = \lambda MMD\left(P_s, P_t\right) + (1-\lambda)\sum_{y=1}^{Y} MMD\left(Q\left(y_s\left|x_s\right.\right), Q\left(y_t\left|x_t\right.\right)\right) \quad (17)$$

where $P(x_s)$ and $P(x_t)$ are the marginal probability distributions and $Q(y_s|x_s)$ and $Q(y_t|x_t)$ are the conditional probability distributions. The equilibrium parameter $\lambda \in [0, 1]$ acts as the weight adjustment of the marginal probability distribution and the conditional probability distribution.

In (17), $MMD(P_s, P_t)$ is the marginal probability distribution of the source and target domains. After introducing the kernel mapping, we have

$$\begin{aligned} MMD(P_s, P_t) &= \left\| \frac{1}{n}\sum_{s=1}^{n}\phi\left(x_s\right) - \frac{1}{m}\sum_{t=1}^{m}\phi\left(x_t\right) \right\| \\ &= \left\| \frac{1}{n}\sum_{s=1}^{n} W^T x_s - \frac{1}{m}\sum_{t=1}^{m} W^T x_j \right\| \\ &= tr\left(W^T X M_0 X^T W\right) \end{aligned} \quad (18)$$

In (18), n represents the number of features in the source domain. The total number of features in the source domain D_s is $s = 1, 2, \ldots, n$, and m is the number of features in the target domain. The total number of features in the target domain D_t is $t = 1, 2, \ldots, m$. M_0 is the MMD matrix as given below.

$$(M_0)_{st} = \begin{cases} \frac{1}{n^2}, & x_s \in D_s, x_t \in D_t \\ \frac{1}{m^2}, & x_s \in D_s, x_t \in D_t \\ -\frac{1}{mn}, & else \end{cases} \quad (19)$$

In (17), $MMD(Q(y_s|x_s), Q(y_t|x_t))$ is the conditional probability distribution of the source and target domains. After

introducing the kernel mapping, we have

$$\begin{aligned} &MMD\left(Q\left(y_s\left|x_s\right.\right), Q\left(y_t\left|y_t\right.\right)\right) \\ &= \sum_{y=1}^{Y}\left\| \frac{1}{n^{(c)}} \sum_{x_s \in D_s^{(c)}} \phi\left(x_s\right) - \frac{1}{m^{(c)}} \sum_{x_t \in D_t^{(c)}} \phi\left(x_t\right) \right\| \\ &= \sum_{y=1}^{Y}\left\| \frac{1}{n^{(y)}} \sum_{x_s \in D_s^{(y)}} W^T x_s - \frac{1}{m^{(y)}} \sum_{x_t \in D_t^{(y)}} W^T x_j \right\| \\ &= \sum_{y=1}^{Y} tr\left(W^T X M_c X^T W\right) \end{aligned} \tag{20}$$

where $n(y)$ represents the number of y-class features in the source domain, and $m(y)$ represents the number of y-class features in the target domain. $D_s^{(y)}$ represents the feature set that belongs to class y in the source domain, and $D_t^{(y)}$ represents the feature set belonging to class y in the target domain. We calculate the M_c of the adaptive class matrix as follows:

$$\left(M_c\right)_{st} = \begin{cases} \frac{1}{n_c^2}, & x_s, x_t \in D_s^{(y)} \\ \frac{1}{m_c^2}, & x_s, x_t \in D_t^{(y)} \\ -\frac{1}{m_c n_c}, & \begin{cases} x_s \in D_s^{(y)}, & x_t \in D_t^{(y)} \\ x_s \in D_t^{(y)}, & x_t \in D_s^{(y)} \end{cases} \\ 0, & else \end{cases} \tag{21}$$

Substituting (18) and (20) into (17), we obtain the following:

$$\begin{aligned} &D\left(D_s, D_t\right) \\ &= \lambda MMD\left(P_s, P_t\right) + (1-\lambda) \sum_{y=1}^{Y} MMD\left(Q\left(y_s\left|x_s\right.\right), Q\left(y_t\left|x_t\right.\right)\right) \\ &= \lambda tr\left(W^T X M_0 X^T W\right) + (1-\lambda) \sum_{y=1}^{Y} tr\left(W^T X M_c X^T W\right) \end{aligned} \tag{22}$$

When the equilibrium parameter λ approaches 1, the result is mainly the result of the marginal probability distribution. Similarly, when λ approaches 0, the result is mainly the result of the conditional probability distribution. To find the best value of λ, we performed experiments 10 times with the COPD data. The results are shown in Fig. 6; the x-axis represents the value of λ, and the y-axis represents the prediction accuracy. We can see that the best prediction accuracy is 89.5% when $\lambda = 0.6$.

2) MODEL OPTIMIZATION

In the process of model optimization, we will inevitably encounter the problem of determining super parameters, and common solutions are fixed using manual designs. The elastic model [41] can dynamically learn from the data without any extra calculation costs; the elastic model is universal to all network structures and can be seamlessly embedded. In addition, we added the norm constraint performance of L1 and L2 in the traditional elastic network. The improved elastic network has many advantages, such as learning from data instead of manual designing and changing with the change in data. Finally, the improved elastic network can be applied to the classification model of COPD, which can further improve the diagnosis accuracy and prediction of COPD. The objective function of the elastic network is as follows:

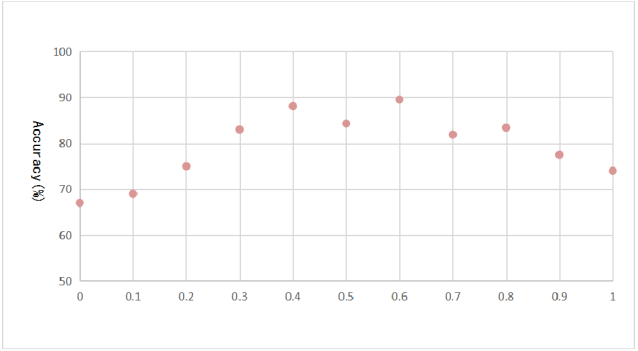


FIGURE 6. Prediction accuracy with different values of λ.

$$f = \min_{\alpha_0, \alpha}\left[\sum_{i=1}^{j}\left(y_i - \alpha_0 - x_i^T\right)^2 + \mu P_\beta\left(\alpha\right)\right] \tag{23}$$

Here, y_i represents the prediction result of the i disease class, and x_i is the feature of the ith disease; α is the estimated regression coefficient, and μ is the minimum mean square error. The regularization term $P_\beta(\alpha)$ is given as

$$P_\beta\left(\alpha\right) = \sum_{j=1}^{|x_i|}\left(\frac{1-\beta}{2}\alpha_j^2 + \beta\left|\alpha_j\right|\right) \tag{24}$$

where $\beta \in [0, 1]$. When $\beta = 0$, the term is expressed as ridge regression, and when $\beta = 1$, the term is expressed as the least absolute shrinkage and selection operator. Here α is determined by cross-validation.

The class label obtained previously is used as the pseudo-identity label for the next identification, and the features involved in the transfer will not change. After iterating t times, the function converges to its minimum. We thus achieve model optimization.

D. ALGORITHMS

1) CROSS-DOMAIN FEATURE FILTERING ALGORITHM

The cross-domain feature filtering algorithm first uses the MDS method to map features to the same dimension and then constructs a feature space.

In the space, the diversity between features is retained, and the relationship between the features is also determined (see steps (2) to (4)). Subsequently, we use the approximate Markov blanket defined by the SU to filter the irrelevant and redundant features and select the relevant feature set

(see steps (5) to (10)). The cross-domain feature filtering algorithm is as follows:

Algorithm 1 Cross-domain Feature Filtering Algorithm

1. X≠ø;
2. **for** $i = 1$ to $n + m$ //See Algorithm 2 for details
3. compute $\left(D_{ij}^{X}\right)^2$;
4. **end for**
5. **for** $i = 1$ to $n+m$ //See Algorithm 3 for details
6. compute SU;
7. **for** j = i + 1 to n+m **do**
8. SU(x_i, x_j) ≥SU(x_j, y)
9. $X = X - \{x_i\}$;
10. **end for**
11. **end**

MDS uses space and distance to reflect the relationship between features, and finally this algorithm obtains the low-dimensional position structure relationship that contains all the features in a unified space. The MDS algorithm is given below.

Algorithm 2 MDS Algorithm

Input: Distance matrix D, whose element D_{ij} is the distance from sample x_i to x_j

Output: The representation matrix of the sample in the new feature space is $\wedge_*^{1/2} V T_*$

1. Calculate $D_{i\cdot}^2$, $D_{\cdot j}^2$, $D_{\cdot\cdot}^2$ according to (6)–(8) respectively;
2. Calculate matrix B according to (9);
3. Eigenvalue decomposition of matrix B;
4. $\wedge = \mathrm{diag}(\lambda_1, \lambda_2, \ldots \lambda_d)$ is the diagonal matrix formed by the eigenvalues. V is the eigenvector matrix.

The unsymmetrical features are filtered using SU. We remove the redundant features using the approximate Markov blankets; the relevant feature sets can be selected, and the standard features can be found. The specific algorithm is as follows:

2) ALGORITHM STEPS OF BALANCED PROBABILITY DISTRIBUTION MODEL

The specific stages of the BPD model algorithm based on the instance and feature transfers are as follows:

Algorithm 3 Approximate Markov Blanket Algorithm with the Symmetric Uncertainty

Input: low-dimensional features x, class y in the same feature space

Output: optimal feature dataset

1. Calculate the information entropy E(x) of feature x, the information entropy E (y) of class y, and the conditional entropy E ($x|y$) according to (10)–(12);
2. Calculate the maximum mutual information as $\mathrm{MI}_{\max}$ ($y|x$), the feature x, and the SU of class y according (14)–(15);
3. Take the i feature x_i from X;
4. Extract the j feature x_j from X;
5. If SU(x_i, x_j) ≥SU(x_j, y)
6. $X = X - \{x_i\}$

Algorithm 4 BPD Algorithm

Input: Feature dataset X, weight balance parameter λ of marginal distribution and conditional distribution, estimated regression coefficient α;

Output: Target domain prediction class y

1. $x = x_0, y = y(x)$ //Initial state, precision
2. $K = \varphi(x)^T \varphi(x)$, $W^T W = E$
   //compute kernel matrix K,transformation matrix W
3. $MMD(P_s, P_t) = tr(W^T X M_0 X^T W)$
   //compute $MMD(P_s, P_t)$, $(M_0)_{st}$
4. $MMD(Q(y_s|x_s), Q(y_t|x_t)) = \sum_{y=1}^{Y} tr(W^T X M_c X^T W)$
   //compute $MMD(Q(y_s|x_s), Q(y_t|x_t))$,$(M_c)_{st}$
5. $\lambda = \lambda_0$ //Initial λ
6. $D(D_s, D_t) = \lambda MMD(P_s, P_t) + (1-\lambda) \sum_{y=1}^{Y} MMD(Q(y_s|x_s), Q(y_t|x_t))$
   //compute $D(D_s, D_t)$
7. $\alpha = \alpha_0$;$f_{final} = f$; //Initial α, f_{final} is state.
8. $t = 1$;
9. if ($f_{final} = f$){
10. $f = f(\alpha)$; }
11. else {$t++$}
12. return x_0, $y(x)$

IV. EXPERIMENT AND ANALYSIS

We used medical test results and symptoms to predict whether patients have COPD and to classify the diseases. To verify the effectiveness of the BPD algorithm based on the instance and feature transfers, we performed experiments on the following two datasets: the COPD dataset provided by the Clinical Medical Science Data Center and the COPD dataset extracted from the electronic medical records of a partner medical system [42]. The test set was the COPD dataset obtained from the electronic medical document of the partner medical system. The data in this dataset were screened after at least five years of observations of patients.

A. EXPERIMENTAL DATASET

The COPD dataset provided by the Clinical Medical Science Data Center included relevant demographic information, electronic medical record information, examination results, health self-scores, and follow-up information over five years; more than 360 types of features were used. The dataset contained 1999 samples, including the data obtained from 829 patients with COPD, 1021 non-COPD patients, and 149 undiagnosed patients. The COPD dataset also included

TABLE 3. Contents of the COPD dataset.

COPD dataset of the Clinical Medical Science Data Center	
Sex	
Male	998 (49.92%)
Female	1001 (50.08%)
BMI	
<18.5	thin 43 (2.15%)
18.5–23.9	normal 437 (21.86%)
> 24	overweight 1519 (75.99%)
GOLD level	
1	167 (8.35%)
2	440 (22.01%)
3	308 (15.41%)
4	165 (8.25%)
0	737 (36.87%)
–1	182 (9.10%)
Related common diseases	
Asthma	387 (19.36%)
Bronchitis	914 (45.72%)
Pneumonia or bronchial pneumonia	846 (42.32%)
Chronic bronchitis	433 (21.66%)
Emphysema	582 (29.11%)
Smoking stage	
Ever smoked	1999 (100%)
Smoking now	0 (0%)

TABLE 4. Contents of the COPD dataset of the electronic medical records.

Feature	Eigenvalues
F0: Sex	Male,female
F1: FEV1/FVC	0 - 1
F2: Labor	0, 1, 2, 3, 4, 5
F3: Runny	0, 1
F4: Age	0, 1, 2, 3, 4, 5
F5: Fatigue	0, 1
F6: Sweat	0, 1
F7: Expectorant	0, 1, 2, 3, 4, 5
F8: Sleep	0, 1, 2, 3, 4, 5
F9: Smoking	0, 1
F10: Weight	0, 1, 2, 3, 4, 5
F11: Constipation	0, 1
F12: mMRC	0, 1, 2, 3, 4, 5
F13: Dry throat	0, 1
F14: Cough	0, 1, 2, 3, 4, 5
F15: Fever	0, 1
F16: Chest pain	0, 1
F17: Chest tightness	0, 1, 2, 3, 4, 5
F18: Flustered	0, 1
F19: Confidence	0, 1, 2, 3, 4, 5
F20: Energy	0, 1, 2, 3, 4, 5
F21: Fever	0, 1
F22: Itching	0, 1
F23: Puffiness	0, 1
F24: Tongue coating	0, 1
F25: Aster	0, 1

data on heart diseases, asthma, emphysema, and other common diseases associated with COPD (Table 3).

A total of 1200 pieces of data were extracted from the COPD dataset from the electronic medical records of the partner healthcare system [42]; this included two classes of 750 COPD patients and 450 non-COPD patients who had symptoms similar to COPD patients. Table 4 gives the original 26 feature descriptions extracted from the electronic medical record.

B. EXPERIMENTAL SETUP AND EVALUATION

In medical datasets, the evaluation criterion for the multi-class performance of algorithms is usually accuracy, which is calculated as follows:

$$accuracy = \frac{1}{j}\sum_{i=1}^{j} |z(x) = y(x)| \tag{25}$$

where y is the class space set of diseases with a total of j disease classes, $y(x)$ is the predicted class of feature x, and $z(x)$ is the correct class of feature x.

It would be too one-sided to use accuracy as the evaluation index; therefore, in this paper, we introduce precision, recall, and F1 values as indicators to evaluate the model. Taking COPD as an example, there were four outputs of any sample in the target domain after using the BPD model based on the instance and feature transfers. The COPD samples were correctly predicted as the COPD disease and recorded as TP; the non-COPD samples were erroneously predicted as COPD diseases and were recorded as FP. The COPD samples were incorrectly predicted as non-COPD diseases, and they were recorded as FN. The non-COPD samples were correctly classified as non-COPD samples, and they were marked as TN.

Precision refers to the proportion of correctly divided samples among all the samples predicted as COPD.

$$precision = TP/(TP + FP) \tag{26}$$

Recall refers to the proportion of correctly predicted samples among all COPD samples.

$$recall = TP/(TP + FN) \tag{27}$$

Using accuracy and recall, F1 can fully reflect the pros and cons of the actual performance of the method.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{28}$$

C. EXPERIMENTAL RESULTS

In this paper, we propose a cross-domain feature filtering algorithm and a BPD model based on the instance and feature transfers; this algorithm achieved excellent results in COPD prediction. Accuracy and F1 values play an essential role in the evaluation of model performance. To verify the effectiveness of the model proposed in this paper, we first compare the BPD algorithm with the TraAdaBoost algorithm, TCA algorithm, and multi-task learning (MTL) algorithm based on the classical transfer learning method in terms of accuracy and F1 values.

As shown in Figs. 7 and 8, on the COPD dataset provided by the Clinical Medical Science Data Center, the accuracy of the TraAdaBoost algorithm is 88.9%, while the accuracies of the MTL and TCA algorithms are 80% and 90.8%, respectively. The accuracy of the BPD algorithm reached 92.1%, whereas the F1 value reached 88.7%.

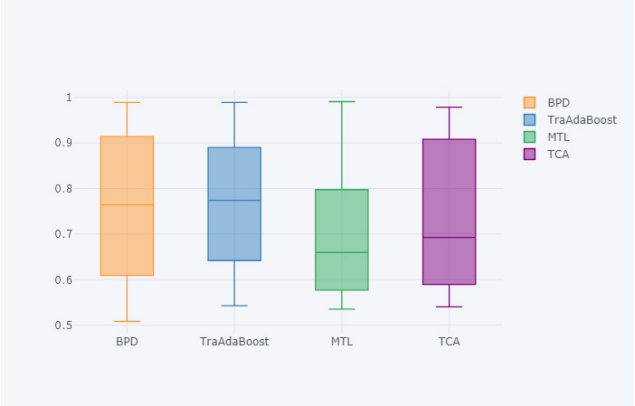


FIGURE 7. Comparison of the accuracies of different algorithms on the COPD dataset provided by the Clinical Medical Science Data Center.

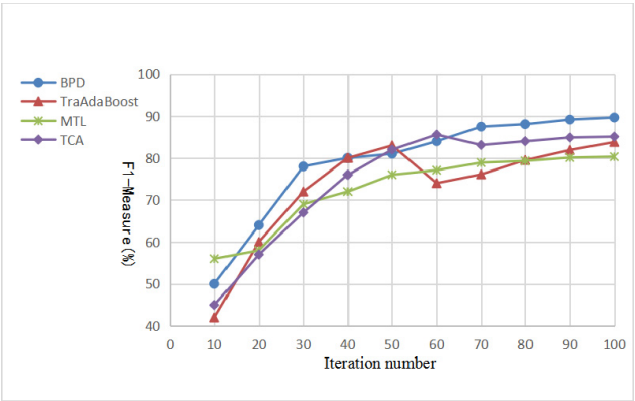


FIGURE 8. Comparison of F1 of the COPD dataset in the clinical medical science data center.

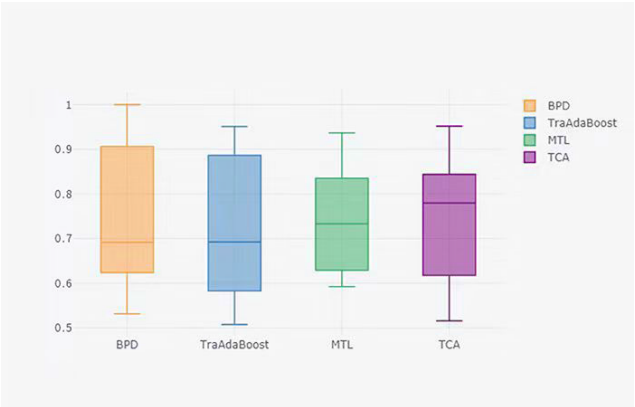


FIGURE 9. Comparison of the accuracies of different algorithms on the COPD dataset of the electronic medical records of the partner medical system.

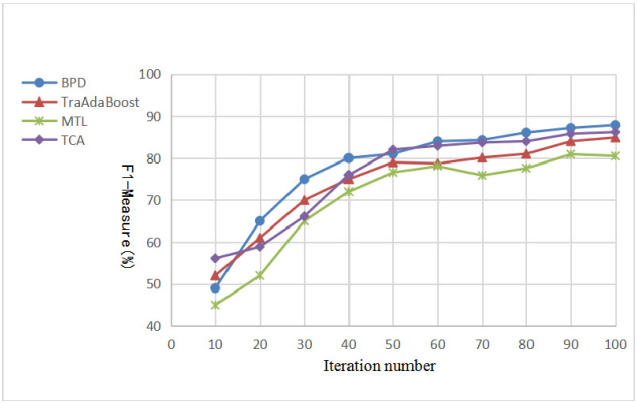


FIGURE 10. Comparison of different algorithms on electronic medical record COPD dataset F1 of the partner medical system.

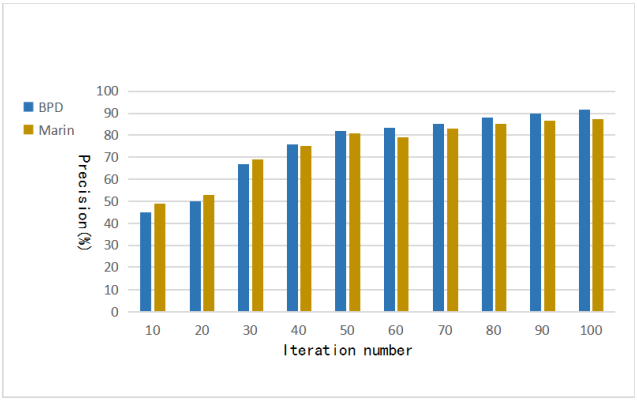


FIGURE 11. Comparison of accuracies of the proposed BPD method and the method proposed by Marin *et al.* [24].

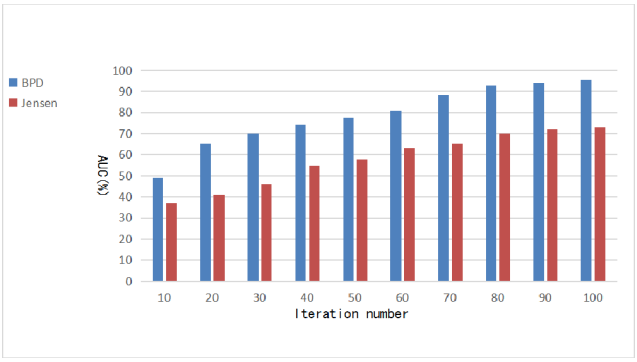


FIGURE 12. AUC comparison chart.

In the COPD dataset extracted from the electronic medical records of the partner medical system, the accuracies of the TraAdaBoost, MTL, TCA, and BPD algorithms were 87.8%, 83.6%; 85%, and 90.5%, respectively, whereas the F1 value reached 88.7% (see Figs. 9 and 10).

The experiment results of the two datasets show that the proposed BPD method is superior to the feature-based transfer learning method TCA and the instance-based transfer learning method TrAdaBoost. In addition to comparing with other transfer learning algorithms, we compared the proposed BPD method with the methods proposed by Marin *et al.* [24] and Jensen *et al.* [25] (Fig. 11 and Fig. 12, respectively).

Marin *et al.* [24] studied the BODE index (i.e., the body mass index, degree of airflow limitation, dyspnea, and exercise capacity) to predict that the accuracy rate of the COPD patients admitted to hospitals was 87.3%. The accuracy rate for predicting COPD patients using a cross-domain feature

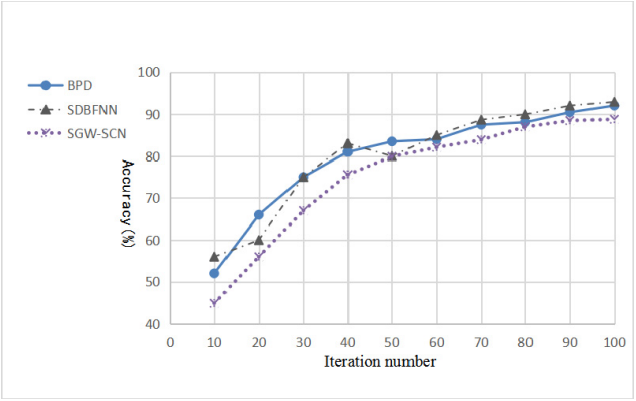


FIGURE 13. Comparison of the accuracies of BPD, SDBFNN, and SGW–SCN.

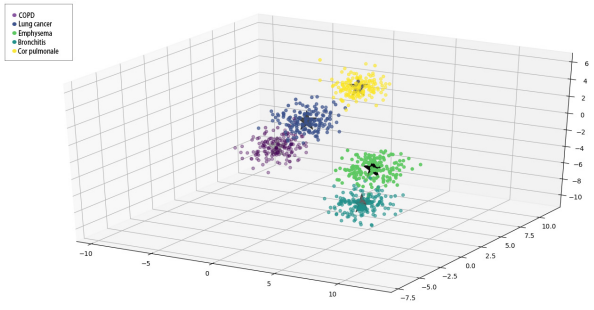


FIGURE 14. Disease classification.

filtering algorithm was 92.7%. Jensen *et al.* [25] used a linear regression model to distinguish the acute and non-acute phases of COPD, and the AUC value reached 73%. The AUC value of the BPD method based on the instance and feature transfers proposed in this paper was 95.2%, which is quite robust.

There are many other advanced methods for prediction. We also compared the method proposed in this paper with advanced methods such as sparse deep belief network with fuzzy neural network (SDBFNN) proposed by Wang *et al.* [43] and Savitzky–Golay wavelet-supported stochastic configuration network (hereinafter SGW–SCN) proposed by Bi *et al.* [44] by using the evaluation standard of accuracy. As shown in Fig. 13, the accuracy of the SDBFNN method was 93%, which is slightly higher than that of the proposed BPD method (92.1%). The accuracy of SGW–SCN was 88.7%, which is lower than the proposed BPD method. To summarize, our experiments prove that the BPD method is more effective than most other methods. From the obtained results, it is clear that the method proposed in this paper is effective.

In this research, we not only applied the knowledge transfer learned from multi-source domains to the COPD field but also achieved excellent results for the COPD identification and classification. For the dataset provided by the Clinical Medical Science Data Center, we distinguished between bronchitis, emphysema, COPD, and their comorbidities (i.e., lung cancer and pulmonary heart disease). As shown in Fig. 14, a BPD algorithm that fuses the instance-based and feature-based transfers can effectively distinguish among the different conditions showing similar symptoms.

V. CONCLUSION

For the prediction of few-shot learning in COPD studies, we propose a method for the diagnosis of COPD based on transfer learning, i.e., the BPD algorithm. BPD first uses instance-based cascaded transfer learning to obtain instances close to the target domain. Then, it uses the cross-domain feature filtering algorithm to obtain the co-occurrence features of source domains and target domains. The transference of learning from a multi-disease source domain to the COPD domain is realized through instances and co-occurrence features to construct the classification model of the target domain. Next, an elastic network is used to further improve the generalization performance of the model. The BPD method integrates the instance-based transfer and feature-based transfer, and our multiple experiments prove that the BPD method can obtain accurate prediction results.

Our experimental results also show that the BPD method not only improves COPD identification but it can also effectively distinguish among different diseases having similar symptoms. In subsequent studies, we propose to use different diagnostic approaches to simulate medical diagnostic thinking to obtain more accurate and interpretable disease prediction models.

REFERENCES

[1] C. F. Vogelmeier *et al.*, "Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 Report. GOLD executive summary," *Amer. J. Respiratory Crit. Care Med.*, vol. 195, no. 5, pp. 557–582, Mar. 2017.

[2] A. Quach, J. Giovannelli, N. Chérot-Kornobis, A. Ciuchete, G. Clément, R. Matran, P. Amouyel, J.-L. Edmé, and L. Dauchet, "Prevalence and underdiagnosis of airway obstruction among middle-aged adults in Northern France: The ELISABET study 2011–2013," *Respiratory Med.*, vol. 109, no. 12, pp. 1553–1561, Dec. 2015.

[3] D. Sanchez-Morillo, M. A. Fernandez-Granero, and A. Leon-Jimenez, "Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: A systematic review," *Chronic Respiratory Disease*, vol. 13, no. 3, pp. 264–283, Aug. 2016.

[4] X. Min, B. Yu, and F. Wang, "Predictive modeling of the hospital readmission risk from Patients' claims data using machine learning: A case study on COPD," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.

[5] R. Liu, H. Wang, and X. Yu, "Shared-nearest-neighbor-based clustering by fast search and find of density peaks," *Inf. Sci.*, vol. 450, pp. 200–226, Jun. 2018.

[6] B. Hu, H. Wang, X. Yu, W. Yuan, and T. He, "Sparse network embedding for community detection and sign prediction in signed social networks," *J. Ambient Intell. Hum. Comput.*, vol. 10, no. 1, pp. 175–186, Jan. 2019.

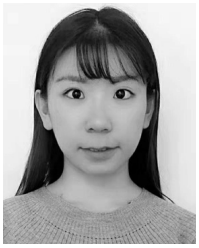
[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] F. Shen, J. Chao, and J. Zhao, "Forecasting exchange rate using deep belief networks and conjugate gradient method," *Neurocomputing*, vol. 167, pp. 243–253, Nov. 2015.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[10] G. Wang, J. Qiao, J. Bi, W. Li, and M. Zhou, "TL-GDBN: Growing deep belief network with transfer learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 874–885, Apr. 2019.

[11] G. Lu, D. Li, and L. Zhang, "Clinical investigation of depression in elderly patients with chronic obstructive pulmonary disease," *China Med. Pharmacy*, vol. 3, no. 1, pp. 12–14, 2013.
[12] J. Bi, T. Feng, and H. Yuan, "Real-time and short-term anomaly detection for GWAC light curves," *Comput. Ind.*, vol. 97, pp. 76–84, May 2018.
[13] T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams, and I. C. Paschalidis, "Predicting chronic disease hospitalizations from electronic health records: An interpretable classification approach," *Proc. IEEE*, vol. 106, no. 4, pp. 690–707, Apr. 2018.
[14] L. Han, M. Maciejewski, C. Brockel, W. Gordon, S. B. Snapper, J. R. Korzenik, L. Afzelius, and R. B. Altman, "A probabilistic pathway score (PROPS) for classification with applications to inflammatory bowel disease," *Bioinformatics*, vol. 34, no. 6, pp. 985–993, Mar. 2018.
[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Frontiers Chem.*, vol. 6, p. 162, May 2018.
[16] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," 2015, *arXiv:1511.03677*. [Online]. Available: http://arxiv.org/abs/1511.03677
[17] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.
[18] X. Li, H. Wang, Y. Xiong, Y. Zeng, T. He, J. Wu, and J. Chen, "Application of convolutional neural network in pediatric disease prediction," *China Digit. Med.*, vol. 13, no. 10, pp. 11–13, 2018.
[19] E. Prescott, P. Lange, and J. Vestbo, "Socioeconomic status, lung function and admission to hospital for COPD: Results from the Copenhagen city heart study," *Eur. Respiratory J.*, vol. 13, no. 5, pp. 1109–1114, May 1999.
[20] M. Kanervisto, T. Vasankari, T. Laitinen, M. Heliövaara, P. Jousilahti, and S. Saarelainen, "Low socioeconomic status is associated with chronic obstructive airway diseases," *Respiratory Med.*, vol. 105, no. 8, pp. 1140–1146, Aug. 2011.
[21] P. Danielsson, I. S. Ólafsdóttir, B. Benediktsdóttir, T. Gíslason, and C. Janson, "The prevalence of chronic obstructive pulmonary disease in Uppsala, Sweden—The burden of obstructive lung disease (BOLD) study: Cross-sectional population-based study," *Clin. Respiratory J.*, vol. 6, no. 2, pp. 120–127, Apr. 2012.
[22] Y. Borné, W. Ashraf, S. Zaigham, and S. Frantz, "Socioeconomic circumstances and incidence of chronic obstructive pulmonary disease (COPD) in an urban population in sweden," *COPD: J. Chronic Obstructive Pulmonary Disease*, vol. 16, no. 1, pp. 51–57, Jan. 2019.
[23] H. Zhao *et al.*, "Relationship between stress hormone levels and prognosis in elderly patients with COPD," *J. Guangxi Univ. Traditional Chin. Med.*, vol. 22, no. 2, pp. 26–28, 2019.
[24] J. M. Marin, S. J. Carrizo, C. Casanova, P. Martinez-Camblor, J. B. Soriano, A. G. N. Agusti, and B. R. Celli, "Prediction of risk of COPD exacerbations by the BODE index," *Respiratory Med.*, vol. 103, no. 3, pp. 373–378, Mar. 2009.
[25] M. H. Jensen, S. L. Cichosz, B. Dinesen, and O. K. Hejlesen, "Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare," *J. Telemed. Telecare*, vol. 18, no. 2, pp. 99–103, Mar. 2012.
[26] M. van der Heijden, P. J. F. Lucas, B. Lijnse, Y. F. Heijdra, and T. R. J. Schermer, "An autonomous mobile system for the management of COPD," *J. Biomed. Informat.*, vol. 46, no. 3, pp. 458–469, Jun. 2013.
[27] C. Burton, H. Pinnock, and B. McKinstry, "Changes in telemonitored physiological variables and symptoms prior to exacerbations of chronic obstructive pulmonary disease," *J. Telemed. Telecare*, vol. 21, no. 1, pp. 29–36, Jan. 2015.
[28] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, 2007, pp. 193–200.
[29] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
[30] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.
[31] M. M. Ohayon, "Chronic obstructive pulmonary disease and its association with sleep and mental disorders in the general population," *J. Psychiatric Res.*, vol. 54, pp. 79–84, Jul. 2014.
[32] N. Zhong and C. Cai, "Relationship between chronic obstructive pulmonary disease and anxiety depression," *Cont. Med. Educ.*, vol. 21, no. 16, pp. 17–19, 2006.
[33] Y. Zhu and W. Zhang, "Discussion on the treatment of chronic obstructive pulmonary disease from lung, spleen and kidney," *Shanxi J. Traditional Chin. Med.*, vol. 31, no. 7, pp. 1–2, 2015.
[34] J. Bi, H. Yuan, M. Zhou, and Q. Liu, "Time-dependent cloud workload forecasting via multi-task learning," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2401–2406, Jul. 2019.
[35] X. Guo, M. Zhou, S. Liu, and L. Qi, "Lexicographic multiobjective scatter search for the optimization of sequence-dependent selective disassembly subject to multiresource constraints," *IEEE Trans. Cybern.*, to be published.
[36] X. Luo, M. Zhou, Z. Wang, Y. Xia, and Q. Zhu, "An effective scheme for QoS estimation via alternating direction method-based matrix factorization," *IEEE Trans. Services Comput.*, vol. 12, no. 4, pp. 503–518, Jul./Aug. 2019.
[37] R. Veevers and S. Hayward, "Morphing and docking visualisation of biomolecular structures using multi-dimensional scaling," *J. Mol. Graph. Model.*, vol. 82, pp. 108–116, Jun. 2018.
[38] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance andredundancy," *J. Mach. Learn. Res.*, vol. 5, no. Oct, pp. 1205–1224, 2004.
[39] K. Passi, A. Nour, and C. K. Jain, "Markov blanket: Efficient strategy for feature subset selection method for high dimensional microarray cancer datasets," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 1864–1871.
[40] J. Bi, H. Yuan, and M. Zhou, "Temporal prediction of multiapplication consolidated workloads in distributed clouds," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1763–1773, Oct. 2019.
[41] H. Wang, A. Kembhavi, A. Farhadi, A. L. Yuille, and M. Rastegari, "ELASTIC: Improving CNNs with dynamic scaling policies," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2258–2267.
[42] Y. Fang, H. Wang, L. Wang, R. Di, and Y. Song, "Feature-maximum-dependency-based fusion diagnosis method for COPD," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 1–18, 2019.
[43] G. Wang, Q.-S. Jia, J. Qiao, J. Bi, and C. Liu, "A sparse deep belief network with efficient fuzzy learning framework," *Neural Netw.*, vol. 121, pp. 430–440, Jan. 2020.
[44] J. Bi, H. Yuan, L. Zhang, and J. Zhang, "SGW-SCN: An integrated machine learning approach for workload forecasting in geo-distributed cloud data centers," *Inf. Sci.*, vol. 481, pp. 57–68, May 2019.



QIAN WANG was born in 1995. She received the bachelor's degree in information science and engineering from the Shandong Youth Political Institute, Jinan, China, in 2017. She is currently pursuing the master's degree in machine learning and data mining application technology with the School of Information Science and Engineering, Shandong Normal University. Her research interests include machine learning and data mining.

HONG WANG received the master's degree in software from the Department of Computer Science and Technology, Tianjin University, in 1991, the Ph.D. degree in computer applications from the Computing Institute of the Chinese Academy of Sciences, in 2002. She was a Postdoctoral Researcher at the Postdoctoral Station of Computer Science and Technology, Shandong University, in 2009. She is currently a Ph.D. Professor and a Doctoral Tutor with the School of Information Science and Engineering, Shandong Normal University. She has published more than 60 articles in academic journals at home and abroad. As the person in charge of the project, she presided over the completion of the Shandong Young Scientist Award Fund, the Postdoctoral Grant Selection Program in Shandong, and the Natural Science Foundation of Shandong. Her research interests are in the area of machine learning, deep learning, graph learning, data mining, bioinformatics, and personalized recommendation.

LUTONG WANG was born in 1995. She received the bachelor's degree in information science and engineering from the School of Information Science and Engineering, Shandong Normal University, Jinan, China, in 2017, where she is currently pursuing the master's degree in machine learning and data mining. Her research interests include machine learning and bioinformatics.

FENGPING YU was born in 1996. She is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University. Her research interests include machine learning, data mining, and so on.