# Hierarchical Domain Adaptation Projective Dictionary Pair Learning Model for EEG Classification in IoMT Systems

Weiwei Cai, *Member, IEEE*, Ming Gao, Yizhang Jiang, *Senior Member, IEEE*, Xiaoqing Gu, Xin Ning, *Member, IEEE*, Pengjiang Qian, *Senior Member, IEEE*, and Tongguang Ni

***Abstract*— Epilepsy recognition based on electroencephalogram (EEG) and artificial intelligence technology is the main tool of health analysis and diagnosis in Internet of medical things (IoMT). As a distributed learning framework, federated learning can train a shared model from multiple independent edge nodes using local data, which has greatly promoted the development of IoMT. One of the main challenges of EEG-based epilepsy recognition in IoMT is that EEG records show varying distributions in different devices, different times, and different people. This nonstationary characteristic of EEG reduces the accuracy of the recognition model. To improve the classification performance in IoMT, a hierarchical domain adaptation projective dictionary pair learning (HDA-PDPL) model is developed in the study. HDA-PDPL integrates EEG signals from different domains (person, edge nodes, devices, etc.) into a set of hierarchical subspace and simultaneously learns synthesis and analysis dictionary pairs in each layer. Specifically, a nonlinear transform function is introduced to seek hierarchical feature projection. The domain adaptation term on sparse coding builds a connection between different domains. Thus, the shared synthesis and analysis dictionaries can encode domain-invariant representation and discrimination knowledge from different domains. Besides, the local preserved term of projective codes is introduced to capture the potential discriminative local structures of samples. The experimental results on two EEG epilepsy classifications verified that the HDA-PDPL model can outperform other comparisons by utilizing more shared knowledge of different domains.**

***Index Terms*— Domain adaptation, electroencephalogram (EEG) signals, epilepsy classification, Internet of medical things (IoMT), projective dictionary pair.**

Manuscript received March 14, 2022; revised May 3, 2022; accepted May 17, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62171203 and Grant 62172192, in part by the Natural Science Foundation of Jiangsu Province under Grant BK 20211333, in part by the Science and Technology Project of Changzhou City under Grant CE20215032, in part by the Jiangsu Future Network Scientific Research Fund Project under Grant FNSRFP-2021-YB-36, and in part by the Fundamental Research Funds for the Central Universities under Grant JUSRP122035. *(Corresponding author: Tongguang Ni.)*

Weiwei Cai, Yizhang Jiang, and Pengjiang Qian are with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214000, China (e-mail: vivitsai@ieee.org; yzjiang@jiangnan.edu.cn; qianpjiang@jiangnan.edu.cn).

Ming Gao is with the College of Sports Science and Technology, Wuhan Sports University, Wuhan 430079, China (e-mail: gaoming418@whsu.edu.cn).

Xiaoqing Gu and Tongguang Ni are with the School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China (e-mail: guxq@cczu.edu.cn; ntg@cczu.edu.cn).

Xin Ning is with the Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100084, China (e-mail: ningxin@semi.ac.cn).

Digital Object Identifier 10.1109/TCSS.2022.3176656

## I. Introduction

THE Internet of medical things (IoMT) applies Internet of Things (IoT) technology to the healthcare industry and employs big data and edge computing technologies to realize the intelligence and personalization of medical processes. It establishes all feasible network services to connect available medical resources and various medical services [1]. With the increasing application of IoMT, it involves many aspects, such as disease prediction, intervention, disease diagnosis and treatment, and medical rehabilitation. For example, the mobile health information system designed based on IoMT can not only meet the needs of patients to pay attention to their own health status but also realize the collection and sharing of patients' medical information, thereby improving the work efficiency of medical staff [2].

Federated learning is increasingly used in IoMT due to the widespread use of mobile and edge devices. Federated learning uses a distributed learning model. The raw data collected from multiple edge nodes is directly trained on the node, and then, the model is gradually optimized through the interaction between the node and the cloud server. Thus, federated learning can collaboratively train a generalized shared model and replace data transmission by model transmission, avoiding the risk of user privacy leakage [3], [4].

Electroencephalogram (EEG) signals are the main source of information on brain activity and play a major role in brain–computer interface (BCI) applications. EEG signals are acquired by electrodes attached to the scalp or intracranial and displayed on a computer. In clinical medicine, EEG has great research value as an effective means of diagnosing neurological diseases, such as epilepsy and Alzheimer's disease [5], [6]. Epileptic seizures generally include preseizure, seizure, postseizure, intermission, and so on. Different episodes correspond to different characteristics of EEG waveforms. When the state of epilepsy changes, the abnormal activity of central neurons changes the brain waves and the corresponding brain electrical characteristics will also change. So far, the clinical diagnosis and localization of epilepsy by EEG mainly rely on the visual analysis of experts and physicians, which is time-consuming and can cause differences between raters.

With the widespread application of IoMT, the use of machine learning methods for EEG analysis is an inevitable

trend to respond to the rapidly growing clinical needs. Machine learning methods are more objective than visual analysis. It can reveal subtle signal features that are difficult or impossible to spot through visual inspection. Currently, dictionary learning is widely used in EEG signal recognition [7], [8]. In dictionary learning, the atoms are leaned from the training signals, and the test signals can be sparsely represented by the leaned atoms. Then, the classification tasks can be executed on sparse coefficients.

However, in the federated learning of IoMT, medical data often involve multiple different medical institutions or equipments. The source of model training data may come from some data features shared by edge devices, or from shared samples, or from similar datasets on multiple edge devices. Traditional machine learning methods need to satisfy the premise that the training domain and the test domain have the same probability distribution characteristics. Obviously, this premise cannot be satisfied in practical IoMT applications. Simultaneously, traditional classification methods rely on training data with sufficient label information. As we known, EEG signals have the characteristics of strong interclass similarity and data highly nonstationary. EEG signals are also easily affected by various interference factors, such as emotion and medical state. Therefore, traditional machine learning algorithms are not suitable for practical IoMT applications.

The domain adaptation methods can solve the problem of inconsistency in the "domain" of EEG signals by reducing the difference between domains while maintaining the discriminative ability. Domain adaptation usually aims to identify the common representation between the training domain and the test domain and use the sufficient samples in the related domain as the auxiliary tool, even if the training and test samples have different distributions or features [9]. Among varieties of domain adaptation methods, domain adaptive via subspace learning has drawn much attention. This type of method aims to extract a shared low dimensional space, and in such subspace, the data structure can be preserved. Subspace-based dictionary learning reconstructs the sparse representation of data in different domains in a subspace, thereby revealing their internal data structure. The optimal subspace and discriminative dictionary can be simultaneously learned in an objective function. For example, Yan *et al.* [10] developed an unsupervised domain adaptive dictionary in a sparse space. The learned encoding coefficients of different domains are constrained to embed each other and have similar feature distributions. To reduce cross-domain distribution, Huang *et al.* [11] developed a joint subspace and dictionary learning method. This method preserved the discriminative information between domains by a shared dictionary. Ni *et al.* [12] tried to find an optimal metric subspace and collaboratively learn a shared dictionary to bridge discriminative information between different domains.

In this study, we propose a new dictionary learning model called hierarchical domain adaptation projection dictionary pair learning (HDA-PDPL) model for federated learning in IoMT. The traditional subspace-based dictionary learning learns a linear subspace from the original feature space via a nonlinear transform function. However, the linear transform scheme cannot seek a suitable subspace to separate different types of EEG signals, and the learned dictionary cannot capture the nonlinear structure of EEG signals. Different from the traditional dictionary learning method, our HDA-PDPL model seeks a set of dictionary pairs into a multilayer learning framework via nonlinear transform function, so that the nonlinear structure of EEG signals can be exploited in the discriminative subspace. Also, the domain adaptation term on sparse coding build a connection between different domains. Thus, the shared synthesis and analysis dictionaries can synthesize more reconstruction and discriminative information to encode EEG signals. In addition, the local preserving term on projective codes is introduced by utilizing the underlying local structure information of data, which encourages similar signals to have more similar coding coefficients. Extensive experimental results on Bonn dataset [13] and the Children's Hospital Boston–Massachusetts Institute of Technology (CHB-MIT) scalp dataset [14] show that our HDA-PDPL model achieves the satisfactory classification performance in terms of accuracy, sensitivity, and specificity. The illustration of the HDA-PDPL model is shown in Fig. 1. Our main contributions are summarized as follows.

1) An HDA-PDPL model is proposed. With the simultaneous learning of synthesis dictionary and analysis dictionary, the reconstructive and discriminative ability of the learned dictionary pair is balanced in the domain adaptation scenarios.

2) The nonlinear transform function is adopted to find the appropriate subspaces to represent EEG signals. Therefore, EEG signals in different classes can be separated gradually by multilayer analysis dictionary projection.

3) The domain adaptation term on sparse coding is introduced into the objective function to achieve similar distributions of coding coefficient between different domains.

4) To better the discriminative information, the local preserving term on projective codes is utilized, which can exploit the local structure information of signals and improve the discriminative ability of the learned coding coefficients.

## II. Related Work

### A. IoMT and Federated Learning

IoMT provides a full range of services for healthcare. IoMT uses a cloud/edge approach to connect people (patients, users, and healthcare providers) and devices (medical devices, wearables, and sensors) through communication technologies.

The main functions of IoMT include: collecting life and health data and vital sign data through wearable sensors and IoT technology; dynamic monitoring and analysis of physiological parameters, such as blood pressure, respiration, blood oxygen, and pulse; seamless connection of the whole process management of patients' medical treatment, providing patients with professional analysis reports; and realizing the joint operation of hospitals, communities, and regional medical centers [15].

Federated learning involves a central server coordinating multiple clients to complete a learning task while keeping data private. There are numerous benefits to federated learning. First, the model training of federated learning can conduct training and real-time decision-making on edge nodes or

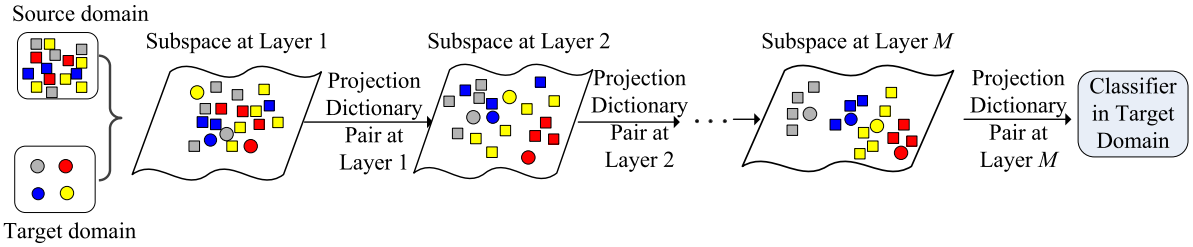



Fig. 1. Illustration of the HDA-PDPL model.

terminal devices, and the time delay will be greatly improved compared to when the decision is made in the cloud. Second, the user's original data do not need to be sent to the cloud, which avoids the possibility of leaking user privacy when the data are uploaded to the link. However, the training data of federated learning come from different data sources, and the training data and test data cannot meet the conditions of independent and identical distribution. If the data distribution of the data source is different, integrating multiple submodels become a difficult problem [16]. He *et al.* [17] used a logistic regression model as the initial model to train data from various data sources and used a neural network to integrate multiple submodels. Xu *et al.* [18] proposed a federated data-driven algorithm and applied it to multiobjective optimization, which trained a radial-basis-function-network on the client side as an agent and used a joint acquisition function on the central server to approximate the target value. Cheng *et al.* [19] proposed a client-side selection framework. During the learning process of knowledge transfer, the framework selected reliable and high-quality clients based on reinforcement learning. Liu *et al.* [20] proposed a joint transfer learning framework to improve statistical modeling under a data federation. Otoum *et al.* [21] used a deep neural network algorithm to train the network and transfer knowledge from connected edge models to build an aggregated global model.

### *B. Projective Dictionary Pair Learning*

Synthesis dictionary-based classifiers have obtained good classification results in signal processing, but it involves the problem of $l_0$ or $l_1$ norm to solve the code $\mathbf{S}$, which is computationally expensive. Instead, analysis dictionary learning finds a linear projection to obtain analytical sparse coding. Recently, the PDPL model is proposed [22]. PDPL can efficiently and linearly represent the sample signal by analyzing the atoms of sparse coding and synthesis dictionary, which can help to reduce the computational complexity. However, both the original synthesis and analysis dictionary learning are unsupervised learning problems, and more discriminative information and constraints need to be considered in classification tasks.

Suppose there is a set of $n$ labeled training EEG signals in $C$ classes $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_C] = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n] \in \mathbf{R}^{d\times n}$, $\mathbf{Y}_c$ is the signal subset of the $c$th class. The dictionary pair learning can be written as

$$\min_{\mathbf{P},\mathbf{D}} \|\mathbf{Y} - \mathbf{DPY}\|_F^2 + f_3(\mathbf{D}, \mathbf{P}, \mathbf{Y}, \mathbf{E}) \tag{1}$$

where the dictionary pair $\mathbf{D}$ and $\mathbf{P}$ is used to reconstruct $\mathbf{X}$ and analytically code $\mathbf{X}$, respectively. $\mathbf{E}$ is the class label matrix. $f_3(\mathbf{D}, \mathbf{P}, \mathbf{Y}, \mathbf{E})$ can be some discriminative functions [22]. Based on the sparse subspace clustering and class-specific dictionary technologies, the function $f_3(\mathbf{D}, \mathbf{P}, \mathbf{Y}, \mathbf{E})$ is represented as

$$f_3(\mathbf{D}, \mathbf{P}, \mathbf{Y}, \mathbf{E}) = \mathbf{P}_c\bar{\mathbf{Y}}_c \approx \mathbf{0} \tag{2}$$

where $\bar{\mathbf{Y}}_c$ is the samples from the complementary set for class $c$. A satisfactory $\mathbf{P}_c \in \mathbf{R}^{K\times d}$ can project $\mathbf{Y}_i (i \neq c)$ to a nearly null space, and then, the analysis dictionary $\mathbf{P} \in \mathbf{R}^{CK\times d}$ can make the matrix $\mathbf{PY}$ nearly block diagonal.

Considering the nonconvexity of (2), the variable matrix $\mathbf{S}$ is introduced into (1) as

$$\min_{\mathbf{P},\mathbf{D},\mathbf{S}} \sum_{c=1}^{C} \|\mathbf{Y}_c - \mathbf{D}_c\mathbf{S}_c\|_F^2 + \tau\|\mathbf{P}_c\mathbf{Y}_c - \mathbf{S}_c\|_F^2 + \lambda\|\mathbf{P}_c\bar{\mathbf{Y}}_c\|_F^2$$
$$\text{s.t. } \|d_i\| \leq 1 \tag{3}$$

where $d_i$ is the $i$th atom of $\mathbf{D} \in \mathbf{R}^{d\times CK}$. Then, all parameters can be solved by an alternative algorithm through alternatively updating $\mathbf{D}$, $\mathbf{P}$, and $\mathbf{S}$.

Finally, the class label of a test signal $\mathbf{y}$ is computed by

$$L(\mathbf{y}) = \arg\min_c \ \|\mathbf{y} - \mathbf{D}_c\mathbf{P}_c\mathbf{y}\|_2^2 \tag{4}$$

where $\mathbf{D}_c$ and $\mathbf{P}_c$ are the class-specific dictionaries on class $c$.

## III. Hierarchical Domain Adaptation Project Dictionary Pair Learning Model

The basic idea of HDA-PDPL is first introduced. Suppose the labeled source domain EEG dataset $\mathbf{Y}_s \in \mathbf{R}^{d\times n_s}$ and the labeled target domain EEG dataset $\mathbf{Y}_t \in \mathbf{R}^{d\times n_t}$, where $n_s > n_t$. Based on the multilayer hierarchical framework [23]–[25], both $\mathbf{Y}_s$ and $\mathbf{Y}_t$ are projected into a set of nonlinear space. In such subspace, the connection is established between different domains and the domain-invariant discriminative knowledge can be transformed from the source domain to the target domain.

### *A. Locality Information Preserving Term on Projective Codes*

From the point of classification, the coding of within-class signals should have similar structure. To achieve the class compactness in the subspace, the locality information preserving on projective coding is introduced by the nearest neighbor graph $\mathbf{G}$. The element $g_{i,j}$ of $\mathbf{G}$ is defined as

$$g_{i,j} = \begin{cases} \exp\left(-\left(\|\mathbf{d}_i - \mathbf{d}_j\|^2/\sigma\right), & \text{if } \mathbf{d}_j \in \text{KNN}(\mathbf{d}_i) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\sigma$ is the kernel parameter and $\mathrm{KNN}(\mathbf{d}_i)$ returns the $k$-nearest neighbors of sample $\mathbf{d}_i$. Then, the Laplace matrix $\mathbf{L}$ of graph $\mathbf{G}$ can be represented by

$$\begin{aligned}\mathbf{L} &= \mathbf{M} - \mathbf{G}\\ \mathbf{M} &= \mathrm{diag}(m_1, m_2, \ldots, m_C), m_i = \sum_{i=1}^{C} g_{i,j}.\end{aligned} \tag{6}$$

The locality constraint on the representation $\mathbf{Py}$ is to minimize the following problem:

$$\min_{\mathbf{P}} \sum_{i,j} \left\|\mathbf{P}\mathbf{y}_i - \mathbf{P}\mathbf{y}_j\right\|^2 g_{i,j} = \min_{\mathbf{P}} \; Tr\left((\mathbf{PY})^T \mathbf{L}(\mathbf{PY})\right). \tag{7}$$

The locality information preserving term on projective codes not only inherits the structure information of the signals but also emphasizes the robustness and smoothness of the model.

### *B. Domain Adaptation Term on Sparse Coding*

The domain adaptation term in HDA-PDPL is expected to build a connection between different domains. When obtained the optimal shared common dictionary $\mathbf{D}$, the sparse coding coefficients $\mathbf{S}_s \in \mathbf{R}^{CK\times n_s}$ and $\mathbf{S}_t \in \mathbf{R}^{CK\times n_t}$ may share similar distributions. Using the representation coefficient $\mathbf{Q}_{st} \in \mathbf{R}^{n_t\times n_s}$ from source to target domains and the representation coefficient $\mathbf{Q}_{ts} \in \mathbf{R}^{n_s\times n_t}$ from target to source domains, the domain adaptation term can be written as

$$\min_{\mathbf{S}_s,\mathbf{S}_t\mathbf{Q}_{ts},\mathbf{Q}_{st}} \|\mathbf{S}_s - \mathbf{S}_t\mathbf{Q}_{ts}\|_F^2 + \|\mathbf{S}_t - \mathbf{S}_s\mathbf{Q}_{st}\|_F^2. \tag{8}$$

It is noted that $\mathbf{PY}_t \approx \mathbf{S}_t$ and $\mathbf{PY}_s \approx \mathbf{S}_s$ in analysis dictionary learning. Substituting these two approximate equalities into (8), the domain adaptation term can be rewritten as

$$\min_{\mathbf{P},\mathbf{S}_s,\mathbf{S}_t\mathbf{Q}_{ts},\mathbf{Q}_{st}} \|\mathbf{S}_s - \mathbf{PY}_t\mathbf{Q}_{ts}\|_F^2 + \|\mathbf{S}_t - \mathbf{PY}_s\mathbf{Q}_{st}\|_F^2. \tag{9}$$

### *C. Objective Function*

Suppose a multilayer hierarchical framework of projection pair dictionary with $M$ layers, the original signals data $\mathbf{Y}_s^{(1)}$ and $\mathbf{Y}_t^{(1)}$ can be first decomposed by dictionaries $\mathbf{P}^{(1)}$ and $\mathbf{D}^{(1)}$. This process is essential that $\mathbf{Y}_s^{(1)}$ and $\mathbf{Y}_t^{(1)}$ are projected into the subspace by $\mathbf{P}^{(1)}$ and simultaneously reconstructed by $\mathbf{D}^{(1)}$. Then, taking $\mathbf{Y}_s^{(1)}$ as the example, the coding vectors as the input of the second layer can be represented as $\mathbf{Y}_s^{(2)} = \Theta(\mathbf{P}^{(1)}\mathbf{Y}_s^{(1)})$, where $\Theta(\cdot)$ is a transform function. Then, taking $\mathbf{Y}_s^{(2)}$ as the input, one can cascade this model by imposing a hierarchical representation by $\mathbf{Y}_s^{(3)}$ with the corresponding dictionaries $\mathbf{P}^{(2)}$ and $\mathbf{D}^{(2)}$, and the coding vectors $\mathbf{Y}_s^{(3)}$ is represented as $\mathbf{Y}_s^{(3)} = \Theta(\mathbf{P}^{(2)}\mathbf{Y}_s^{(2)})$. Finally, the coding vectors of $\mathbf{Y}_s^{(1)}$ at layer $M$ are represented as $\mathbf{Y}_s^{(M)} = \Theta(\mathbf{P}^{(M-1)}\mathbf{Y}_s^{(M-1)})$. Following this, one can obtain the objective function of HDA-PDPL at the $m$th $(1 \le m \le M)$ layer

$$\begin{aligned}\min_{\mathbf{P},\mathbf{D},\mathbf{S}_s,\mathbf{S}_t,\mathbf{Q}_{ts},\mathbf{Q}_{st}} \; & J = J_1 + \beta_2 J_2 + \beta_3 J_3\\ = \sum_{c=1}^{C}\Bigg(&\left\|\left[\mathbf{Y}_{s,c}^{(m)}, \mathbf{Y}_{t,c}^{(m)}\right] - \mathbf{D}_c^{(m)}\left[\mathbf{S}_{s,c}^{(m)}, \mathbf{S}_{t,c}^{(m)}\right]\right\|_F^2\\ &+ \alpha_1\left\|\mathbf{P}_c^{(m)}\left[\mathbf{Y}_{s,c}^{(m)}, \mathbf{Y}_{t,c}^{(m)}\right] - \left[\mathbf{S}_{s,c}^{(m)}, \mathbf{S}_{t,c}^{(m)}\right]\right\|_F^2\\ &+ \alpha_2\left\|\mathbf{P}_c^{(m)}\overline{\mathbf{Y}_c^{(m)}}\right\|_F^2\Bigg)\\ &+ \beta_2 Tr\left(\left(\mathbf{P}^{(m)}\mathbf{Y}^{(m)}\right)^T\mathbf{L}^{(m)}\left(\mathbf{P}^{(m)}\mathbf{Y}^{(m)}\right)\right)\\ &+ \beta_3\Bigg(\left\|\mathbf{S}_s^{(m)} - \mathbf{P}^{(m)}\mathbf{Y}_t^{(m)}\mathbf{Q}_{ts}^{(m)}\right\|_F^2\\ &+ \left\|\mathbf{S}_t^{(m)} - \mathbf{P}^{(m)}\mathbf{Y}_s^{(m)}\mathbf{Q}_{st}^{(m)}\right\|_F^2\Bigg)\\ \text{s.t.} \; &\left\|\mathbf{d}_j^{(m)}\right\| \le 1 \quad \forall j\end{aligned} \tag{10}$$

where $\mathbf{Y}_c^{(m)} = [\mathbf{Y}_{s,c}^{(m)}, \mathbf{Y}_{t,c}^{(m)}]$, $\mathbf{Y}^{(m)} = [\mathbf{Y}_1^{(m)}, \mathbf{Y}_2^{(m)}, \ldots, \mathbf{Y}_c^{(m)}]$, and $\overline{\mathbf{Y}_c^{(m)}}$ is the samples from the complementary set for class $c$ at the layer $m$. $\mathbf{S}_s^{(m)} = [\mathbf{S}_{s,1}^{(m)}, \ldots, \mathbf{S}_{s,C}^{(m)}]$ and $\mathbf{S}_t^{(m)} = [\mathbf{S}_{t,1}^{(m)}, \ldots, \mathbf{S}_{t,C}^{(m)}]$. $\alpha_1$, $\alpha_2$, $\beta_2$, and $\beta_3$ are the regularization parameters.

The first term $J_1$ in HDA-PDPL is the domain adaptation term at the layer $m$. It inherits the advantages of the PDPL method and pursuits the discriminative synthesis dictionary and analysis dictionary in domain adaptation scenarios. The second term $J_2$ is the local preserved term on projective codes at the layer $m$ and the third term $J_3$ is the domain adaptation term on sparse coding at the layer $m$. It can be easily seen in the objective function of HDA-PDPL that the synthesis dictionary establishes the connection between different domains and transfers the discriminative information from the source domain to the target domain at each layer of hierarchical framework. The analysis dictionary also strengthens the discrimination ability of the model and it encourages similar signals to have similar coding coefficients.

To make the model more discriminative, the nonlinear transform function used in this study is the absolute value extension function

$$\Theta(\mathbf{y}) = \left[\max\left(\mathbf{y}^T, 0\right), \left|\min\left(\mathbf{y}^T, 0\right)\right|\right]^T. \tag{11}$$

### *D. Optimization*

All the parameters in the objective function are updated in an alternating iteration scheme. The parameters are divided into groups of $\mathbf{D}_c^{(m)}$, $\mathbf{P}$,$(\mathbf{S}_s^{(m)}, \mathbf{S}_t^{(m)})$, and $(\mathbf{Q}_{st}^{(m)}, \mathbf{Q}_{ts}^{(m)})$. Such scheme can optimize these parameters in the closed solutions.

1) Update $\mathbf{D}_c^{(m)}$. The $F$-norm regularization term on $\mathbf{D}_c^{(m)}$ is introduced into the objective function to relax the constraint on $\mathbf{D}_c^{(m)}$. By fixing other parameters, the objective function of HDA-PDPL becomes

$$\min_{\mathbf{D}} \; = \sum_{c=1}^{C}\left(\left\|\mathbf{Y}_c^{(m)} - \mathbf{D}_c^{(m)}\mathbf{S}_c^{(m)}\right\|_F^2 + \delta_0\left\|\mathbf{D}_c^{(m)}\right\|_F^2\right) \tag{12}$$

where $\delta_0$ is the balance parameter.

The optimal solution of $\mathbf{D}_c^{(m)}$ can be obtained by

$$\mathbf{D}_c^{(m)} = \mathbf{Y}_c^{(m)}\mathbf{S}_c^{(m)T}\left(\mathbf{S}_c^{(m)}\mathbf{S}_c^{(m)T} + \delta_0\mathbf{I}\right)^{-1}. \tag{13}$$

2) Update $\mathbf{P}_c^{(m)}$. To update $\mathbf{P}_c^{(m)}$ by fixing other parameters, the objective function of HDA-PDPL becomes

$$\min_{\mathbf{P}^{(m)}} \sum_{c=1}^{C}\left\{\alpha_1\left\|\mathbf{P}_c^{(m)}\left[\mathbf{Y}_{s,c}^{(m)}, \mathbf{Y}_{t,c}^{(m)}\right] - \left[\mathbf{S}_{s,c}^{(m)}, \mathbf{S}_{t,c}^{(m)}\right]\right\|_F^2\right.$$

$$+\alpha_2\left\|\mathbf{P}_c^{(m)}\overline{\mathbf{Y}_c^{(m)}}\right\|_F^2 + \beta_2 Tr\left(\left(\mathbf{P}_c^{(m)}\mathbf{Y}_c^{(m)}\right)^T\mathbf{L}_c^{(m)}\left(\mathbf{P}_c^{(m)}\mathbf{Y}_c^{(m)}\right)\right) + \beta_3\left(\left\|\mathbf{S}_{s,c}^{(m)} - \mathbf{P}_c^{(m)}\mathbf{Y}_{t,c}^{(m)}\mathbf{Q}_{ts,c}^{(m)}\right\|_F^2 + \left\|\mathbf{S}_{t,c}^{(m)} - \mathbf{P}_c^{(m)}\mathbf{Y}_{s,c}^{(m)}\mathbf{Q}_{st,c}^{(m)}\right\|_F^2\right)\Bigg\} \quad (14)$$

where $\mathbf{L}_c^{(m)}$ is the graph Laplacian matrix of synthesis dictionary atoms in class $c$ at the $m$th layer.

The optimal solution of $\mathbf{P}_c^{(m)}$ can be obtained by

$$\mathbf{P}_c^{(m)} = \left(\alpha_1\mathbf{S}_c^{(m)}\left(\mathbf{Y}_c^{(m)}\right)^T + \beta_3\left(\mathbf{S}_{s,c}^{(m)}\left(\mathbf{Q}_{ts,c}^{(m)}\right)^T\left(\mathbf{Y}_{t,c}^{(m)}\right)^T + \mathbf{S}_{t,c}^{(m)}\left(\mathbf{Q}_{st,c}^{(m)}\right)^T\left(\mathbf{Y}_{s,c}^{(m)}\right)^T\right)\right) \times \left\{\alpha_1\mathbf{Y}_c^{(m)}\left(\mathbf{Y}_c^{(m)}\right)^T + \alpha_2\overline{\left[\mathbf{Y}_c^{(m)}\right]\left[\mathbf{Y}_c^{(m)}\right]}^T + \beta_2\mathbf{Y}_c^{(m)}\mathbf{L}_c^{(m)}\left(\mathbf{Y}_c^{(m)}\right)^T + \beta_3\left(\mathbf{Y}_{t,c}^{(m)}\mathbf{Q}_{ts,c}^{(m)}\left(\mathbf{Q}_{ts,c}^{(m)}\right)^T\left(\mathbf{Y}_{t,c}^{(m)}\right)^T + \mathbf{Y}_{s,c}^{(m)}\mathbf{Q}_{st,c}^{(m)}\left(\mathbf{Q}_{st,c}^{(m)}\right)^T\left(\mathbf{Y}_{s,c}^{(m)}\right)^T\right) + \delta_1\mathbf{I}\right\}^{-1} \quad (15)$$

where $\delta_1$ is a very small positive parameter.

3) Update $\mathbf{Q}_{st,c}^{(m)}$ and $\mathbf{Q}_{ts,c}^{(m)}$. To update $\mathbf{Q}_{st,c}^{(m)}$ and $\mathbf{Q}_{ts,c}^{(m)}$ by fixing other parameters, the objective function of HDA-PDPL becomes

$$\min_{\mathbf{Q}_{ts,c}^{(m)},\mathbf{Q}_{st,c}^{(m)}}\sum_{c=1}^{C}\left(\left\|\mathbf{S}_{s,c}^{(m)} - \mathbf{P}_c^{(m)}\mathbf{Y}_{t,c}^{(m)}\mathbf{Q}_{ts,c}^{(m)}\right\|_F^2 + \left\|\mathbf{S}_{t,c}^{(m)} - \mathbf{P}_c^{(m)}\mathbf{Y}_{s,c}^{(m)}\mathbf{Q}_{st,c}^{(m)}\right\|_F^2\right). \quad (16)$$

The optimal solution of $\mathbf{Q}_{st,c}^{(m)}$ and $\mathbf{Q}_{ts,c}^{(m)}$ can be obtained by

$$\mathbf{Q}_{ts,c}^{(m)} = \left(\left(\mathbf{Y}_{t,c}^{(m)}\right)^T\left(\mathbf{P}_c^{(m)}\right)^T\mathbf{P}_c^{(m)}\mathbf{S}_{t,c}^{(m)} + \delta_2\mathbf{I}\right)^{-1} \times \left(\mathbf{Y}_{t,c}^{(m)}\right)^T\left(\mathbf{P}_c^{(m)}\right)^T\mathbf{S}_{s,c}^{(m)} \quad (17)$$

$$\mathbf{Q}_{st,c}^{(m)} = \left(\left(\mathbf{Y}_{s,c}^{(m)}\right)^T\left(\mathbf{P}_c^{(m)}\right)^T\mathbf{P}_c^{(m)}\mathbf{S}_{s,c}^{(m)} + \delta_2\mathbf{I}\right)^{-1} \times \left(\mathbf{Y}_{s,c}^{(m)}\right)^T\left(\mathbf{P}_c^{(m)}\right)^T\mathbf{S}_{t,c}^{(m)} \quad (18)$$

where $\delta_2$ is a very small positive parameter.

4) Update $\mathbf{S}_s^{(m)}$ and $\mathbf{S}_t^{(m)}$. To update $\mathbf{S}_s^{(m)}$ by fixing other parameters, the objective function of HDA-PDPL becomes

$$\left\|\mathbf{Y}_t^{(m)} - \mathbf{D}^{(m)}\mathbf{S}_s^{(m)}\right\|_F^2 + \alpha_1\left\|\mathbf{P}^{(m)}\mathbf{Y}_s^{(m)} - \mathbf{S}_s^{(m)}\right\|_F^2 + \beta_3\left\|\mathbf{S}_s^{(m)} - \mathbf{P}^{(m)}\mathbf{Y}_t^{(m)}\mathbf{Q}_{ts}^{(m)}\right\|_F^2. \quad (19)$$

The optimal solution of $\mathbf{S}_s^{(m)}$ can be obtained by

$$\mathbf{S}_s^{(m)} = \left(\left(\mathbf{D}^{(m)}\right)^T\mathbf{D}^{(m)} + (\alpha_1+\beta_3)\mathbf{I}\right)^{-1} \times\left(\left(\mathbf{D}^{(m)}\right)^T\mathbf{Y}_t^{(m)} + \alpha_1\mathbf{P}^{(m)}\mathbf{Y}_s^{(m)} + \beta_3\mathbf{P}^{(m)}\mathbf{Y}_s^{(m)}\mathbf{Q}_{st}^{(m)}\right). \quad (20)$$

Similarly, the optimal solution of $\mathbf{S}_t^{(m)}$ can be obtained by

$$\mathbf{S}_t^{(m)} = \left(\left(\mathbf{D}^{(m)}\right)^T\mathbf{D}^{(m)} + (\alpha_1+\beta_3)\mathbf{I}\right)^{-1} \times\left(\left(\mathbf{D}^{(m)}\right)^T\mathbf{Y}_s^{(m)} + \alpha_1\mathbf{P}^{(m)}\mathbf{Y}_t^{(m)} + \beta_3\mathbf{P}^{(m)}\mathbf{Y}_t^{(m)}\mathbf{Q}_{ts}^{(m)}\right). \quad (21)$$

After multiple-layer dictionary pair learning, the test signal $\mathbf{y}$ is represented by $\mathbf{y}^{(M)}$. Then, the classifier in the target domain can be built using the learned dictionary pair $\mathbf{D}^{(M)}$ and $\mathbf{P}^{(M)}$, and the class label of the signal $\mathbf{y}$ is computed by

$$L(\mathbf{y}) = \arg\min_c \left\|\mathbf{y}^{(M)} - \mathbf{D}_c^{(M)}\mathbf{P}_c^{(M)}\mathbf{y}^{(M)}\right\|_2^2. \quad (22)$$

The optimization steps of HDA-PDPL can be summarized in Algorithm 1.

**Algorithm 1** The HDA-PDPL Model

Input: Training dataset (source domain $\mathbf{Y}^s$and target domain $\mathbf{Y}^t$) and its class label matrix

Output: $\mathbf{D}^{(1)},\ldots,\mathbf{D}^{(M)}$, and $\mathbf{P}^{(1)},\ldots,\mathbf{P}^{(M)}$.

Initialize $\mathbf{D}^{(1)}$and $\mathbf{P}^{(1)}$by using K-SVD method [11];

While not converge do

  For $m = 1$ to $M$ do

    Update $\mathbf{D}_c^{(m)}$using Eq.(16);

    Construct graph Laplacian matrix $\mathbf{L}^{(m)}$ using Eqs.(7)-(8);

    Update $\mathbf{P}_c^{(m)}$using Eq.(17);

    Update $\mathbf{Q}_{st}^{(m)}$and$\mathbf{Q}_{ts}^{(m)}$using Eqs.(19)-(20);

    Update $\mathbf{S}_s^{(m)}$ and $\mathbf{S}_t^{(m)}$using Eqs.(22)-(23);

  endfor

Return $\mathbf{D}^{(1)},\ldots,\mathbf{D}^{(M)}$, and $\mathbf{P}^{(1)},\ldots,\mathbf{P}^{(M)}$

## IV. Experiment

### A. Datasets and Experimental Setting

Two public epilepsy EEG datasets Bonn University dataset and CHB-MIT scalp EEG dataset are studied in the experiment. The Bonn dataset is a multiclass epilepsy classification dataset. The Bonn dataset contains five EEG signal subsets named as A–E. Each segment has 4096 sampling points. Subsets A and B are collected from healthy volunteers. Subsets C and D are recorded during the seizure-free stage. Subset E is epileptic seizure activities. Eight domain adaptation cases are designed on the Bonn dataset, as shown in Table I.

CHB-MIT EEG database consists of 23 pediatric epilepsy cases (22 subjects, indices 01 predates indices 21 by 1.5 years from the same patient) of intractable seizure records. Each data file contains data over 23 or more channel. There are data files with missing data on some channels, and some data files have different channel montages from other files. These files are removed in the experiment. A total of 23 cases including 665 seizures are investigated, and 665 nonseizure segments are collected for evaluation over a balanced classification problem.

TABLE I
EEG EPILEPSY CLASSIFICATION TASKS ON THE BONN DATASET

| Cases | Source domain | Target domain | Number of classes |
|---|---|---|---|
| C1 | A, E | A, C | Binary |
| C2 | A, E | B, D | Binary |
| C3 | B, E | A, C | Binary |
| C4 | B, E | A, D | Binary |
| C5 | A, D, E | B, D, E | Three class |
| C6 | B, C, E | A, D, E | Three class |
| C7 | A, D, E | B, C, E | Three class |
| C8 | B, D, E | A, C, E | Three class |

TABLE II
CLASSIFICATION ACCURACIES ON THE BONN DATASET

| Case | PDPL | TCA | JDA | LTCCSP | CDC-KUM | TSMDL | HDA-PDPL |
|---|---|---|---|---|---|---|---|
| C1 | 90.73 | 95.27 | 95.59 | 95.82 | 97.10 | 97.48 | **98.27** |
| | (1.26) | (1.15) | (1.65) | (1.32) | (1.88) | (2.26) | (1.28) |
| C2 | 90.45 | 95.30 | 95.30 | 95.42 | 96.43 | 96.62 | **97.82** |
| | (2.14) | (3.00) | (2.58) | (1.86) | (2.72) | (3.01) | (1.86) |
| C3 | 90.77 | 95.45 | 95.70 | 95.99 | 96.82 | 96.96 | **97.98** |
| | (3.04) | (2.78) | (3.17) | (2.21) | (2.01) | (2.31) | (1.70) |
| C4 | 90.39 | 95.74 | 95.83 | 95.97 | 96.32 | 96.53 | **98.41** |
| | (2.03) | (3.06) | (2.25) | (3.01) | (3.11) | (2.97) | (2.22) |
| C5 | 88.86 | 94.08 | 94.04 | 94.25 | 95.17 | 95.35 | **97.63** |
| | (2.32) | (2.14) | (1.99) | (2.07) | (2.53) | (1.23) | (1.11) |
| C6 | 88.74 | 94.11 | 94.37 | 94.71 | 95.44 | 95.69 | **97.79** |
| | (2.07) | (2.18) | (2.26) | (3.00) | (3.15) | (2.30) | (1.98) |
| C7 | 88.89 | 93.55 | 94.03 | 94.09 | 94.34 | 94.78 | **97.25** |
| | (2.15) | (2.23) | (2.68) | (3.02) | (2.40) | (2.57) | (1.99) |
| C8 | 88.48 | 93.08 | 93.40 | 93.71 | 94.02 | 94.32 | **97.51** |
| | (1.91) | (2.69) | (2.70) | (2.23) | (1.88) | (2.31) | (1.91) |
| Mean | 89.66 | 94.57 | 94.78 | 95.00 | 95.71 | 95.97 | **97.83** |
| | (2.12) | (2.40) | (2.41) | (3.21) | (2.90) | (2.25) | (2.01) |

We perform our model in the VGG-net model, a well-known convolutional neural network (CNN) architecture in EEG signal processing [26]. The number of layers $M$ is set to 3 on both Bonn dataset and CHB-MIT dataset to achieve a balance between classification performance and efficiency. The size of subdictionary pairs $\{\mathbf{D}_c^{(m)}, \mathbf{P}_c^{(m)}\}$ for each class in each layer is empirically set to 50. The regularization parameters are set in the grid $\{0.01, 0.05, 0.1, 0.5, 1, 1.5, 2\}$. In the experiment, HDA-PDPL is compared with two baseline classification and five domain adaptation methods. As the baseline methods, PDPL is compared. Five domain adaptation methods include transfer component analysis (TCA) [27], joint distribution adaptation (JDA) [28], local temporal correlation common spatial pattern (LTCCSP) [29], cross-domain classification model with knowledge utilization maximization (CDC-KUM) [7], and transfer model based on supervised multiple dictionary learning (TSMDL) [30]. The exploited features from VGG-net are applied to these methods. To reduce the data dimension, the feature dimensions are reduced to 250 by principal component analysis (PCA). In this study, three criteria including accuracy, sensitivity, and specificity are adopted for classification performance evaluation. The experiments are conducted on a computer with Intel Xeon Processor E5-2620 v4 and 64-GB RAM. All methods are implemented in MATLAB.

### *B. Experimental Results on Bonn Dataset*

Tables II–IV present a comparison between the classification results of the HDA-PDPL and those of comparison methods on the Bonn dataset in terms of accuracy, sensitivity, and specificity. The best results are highlighted bold. The HDA-PDPL model is an improvement of PDPL in domain adaptation scenarios. As shown in Tables II–IV, the accuracy, sensitivity, and specificity of the HDA-PDPL are improved by 8.17%, 7.96%, and 8.21% compared with PDPL. These results indicate that the HDA-PDPL model is suitable for EEG epilepsy recognition scenarios when the training signals and test signals have different distributions. All domain adaptation methods perform better than PDPL, but HDA-PDPL still shows better performance. For example, the average classification accuracy of HDA-PDPL is higher 3.26%, 3.05%, 2.83%, 2.12%, and 1.86% than TCA, JDA, LTCCSP, CDC-KUM, and TSMDL, respectively.

TABLE III
SENSITIVITY RESULTS ON THE BONN DATASET

| Case | PDPL | TCA | JDA | LTCCSP | CDC-KUM | TSMDL | HDA-PDPL |
|---|---|---|---|---|---|---|---|
| C1 | 91.06 | 95.84 | 95.71 | 96.09 | 97.56 | 97.97 | **98.83** |
| | (2.03) | (2.37) | (3.15) | (3.17) | (3.07) | (2.29) | (2.12) |
| C2 | 90.79 | 95.89 | 96.00 | 95.77 | 96.88 | 97.21 | **98.31** |
| | (2.60) | (2.63) | (1.64) | (2.30) | (2.35) | (2.12) | (2.15) |
| C3 | 91.37 | 95.79 | 96.03 | 96.33 | 97.21 | 97.30 | **98.13** |
| | (2.23) | (2.11) | (1.50) | (1.92) | (1.65) | (1.34) | (1.36) |
| C4 | 90.57 | 96.07 | 96.45 | 96.55 | 96.68 | 96.83 | **98.58** |
| | (1.50) | (2.16) | (2.33) | (2.54) | (1.52) | (2.02) | (1.88) |
| C5 | 88.95 | 94.91 | 94.72 | 94.46 | 95.91 | 95.82 | **96.63** |
| | (1.78) | (1.39) | (2.26) | (2.41) | (1.61) | (2.17) | (2.06) |
| C6 | 89.50 | 94.62 | 94.73 | 94.88 | 95.92 | 95.89 | **98.27** |
| | (1.35) | (1.42) | (2.22) | (1.79) | (2.06) | (2.15) | (1.81) |
| C7 | 89.03 | 93.56 | 94.06 | 94.47 | 94.94 | 95.09 | **97.50** |
| | (1.94) | (2.20) | (2.74) | (3.16) | (2.41) | (2.55) | (2.24) |
| C8 | 88.88 | 93.24 | 93.40 | 94.05 | 94.56 | 94.85 | **97.58** |
| | (2.36) | (3.03) | (1.97) | (3.05) | (3.18) | (2.32) | (2.05) |
| Mean | 90.02 | 94.99 | 95.14 | 95.33 | 96.21 | 96.37 | **97.98** |
| | (1.97) | (2.16) | (2.22) | (2.54) | (2.23) | (2.12) | (1.97) |

TABLE IV
SPECIFICITY RESULTS ON THE BONN DATASET

| Case | PDPL | TCA | JDA | LTCCSP | CDC-KUM | TSMDL | HDA-PDPL |
|---|---|---|---|---|---|---|---|
| C1 | 90.40 | 94.70 | 95.46 | 95.55 | 96.64 | 96.99 | **97.70** |
| | (2.90) | (2.25) | (2.25) | (1.75) | (2.33) | (1.72) | (1.46) |
| C2 | 90.10 | 94.71 | 94.59 | 95.07 | 95.99 | 96.04 | **97.33** |
| | (2.54) | (2.49) | (3.08) | (2.93) | (2.57) | (2.29) | (2.19) |
| C3 | 90.17 | 95.11 | 95.37 | 95.65 | 96.43 | 96.61 | **97.82** |
| | (2.82) | (2.90) | (3.07) | (3.00) | (2.09) | (2.59) | (2.01) |
| C4 | 90.21 | 95.41 | 95.21 | 95.39 | 95.96 | 96.23 | **98.24** |
| | (2.33) | (2.11) | (2.20) | (1.96) | (3.05) | (2.39) | (1.86) |
| C5 | 88.78 | 93.24 | 93.35 | 94.04 | 94.43 | 94.87 | **96.64** |
| | (3.06) | (2.25) | (2.43) | (2.48) | (2.23) | (1.92) | (1.88) |
| C6 | 87.98 | 93.60 | 94.01 | 94.54 | 94.96 | 95.49 | **97.31** |
| | (1.91) | (3.04) | (2.39) | (2.39) | (2.58) | (2.45) | (2.30) |
| C7 | 88.76 | 93.54 | 94.00 | 93.71 | 93.74 | 94.46 | **96.99** |
| | (2.19) | (2.16) | (2.31) | (1.80) | (2.55) | (2.27) | (2.04) |
| C8 | 87.28 | 92.92 | 93.41 | 93.36 | 93.49 | 93.80 | **97.43** |
| | (2.44) | (2.21) | (2.36) | (2.04) | (2.42) | (1.76) | (2.50) |
| Mean | 89.21 | 94.15 | 94.43 | 94.66 | 95.21 | 95.56 | **97.42** |
| | (2.52) | (2.42) | (2.51) | (2.29) | (2.48) | (2.17) | (2.03) |

TCA learns certain common components between different domains by maximum mean distance strategy. LTCCSP utilizes the cross-domain information of local temporal correlation common spatial pattern. Although TCA minimizes the discrepancy between two dictionaries, it cannot exploit sufficient knowledge across domains. CDC-KUM uses the pairwise constraint information and soft-partition clustering as auxiliary information to build the classifier in the kernel space. However, the kernel-based methods easily suffer from the scalability problem. The experimental results indicate that our HDA-PDPL is capable of learning more discriminative EEG representation and discriminative classifier in the hierarchical dictionary pair framework via nonlinear transform function. Furthermore, the iterative learning strategy combined with pair dictionary learning and subspace ensures the simultaneous optimization of all parameters.

### *C. Experimental Results on CHB-MIT Dataset*

CHB-MIT dataset is a subject-independent EEG signal dataset for epilepsy identification. There are in total 23 subjects, since two cases 01 and 21 are recorded from the same person. In the CHB-MIT dataset, the "leave-one-subject-out"

TABLE V
COMPARISON OF ACCURACY RESULTS ON THE CHB-MIT DATASET

| Subject | PDPL | TCA | JDA | LTCCSP | CDC-KUM | TSMDL | HDA-PDPL |
|---|---|---|---|---|---|---|---|
| S1, S21 | 69.97 | 76.19 | 77.97 | 79.05 | 80.38 | 78.14 | **84.40** |
| S2 | 75.36 | 83.37 | 83.49 | 85.04 | 86.34 | 85.18 | **89.53** |
| S3 | 76.62 | 85.02 | 87.35 | 87.15 | 86.42 | 88.78 | **90.81** |
| S4 | 76.30 | 80.77 | 83.70 | 83.90 | 88.14 | 88.78 | **88.88** |
| S5 | 63.52 | 66.02 | 66.54 | 76.67 | 74.25 | 75.23 | **81.80** |
| S6 | 67.44 | 70.25 | 76.69 | 76.79 | 78.10 | 77.70 | **79.49** |
| S7 | 79.43 | 79.08 | 87.05 | 88.10 | 89.41 | 89.61 | **91.30** |
| S8 | 67.48 | 75.06 | 84.42 | 80.62 | 80.55 | 80.40 | **85.73** |
| S9 | 74.59 | 76.77 | 80.67 | 79.50 | 82.76 | 81.46 | **82.46** |
| S10 | 81.57 | 85.22 | 89.19 | 89.57 | 89.15 | 88.15 | **90.88** |
| S11 | 80.55 | 86.12 | 86.04 | 89.15 | 88.92 | 89.07 | **90.05** |
| S12 | 60.72 | 64.80 | 65.53 | 69.72 | 70.95 | 71.51 | **72.36** |
| S13 | 63.51 | 60.81 | 65.42 | 70.21 | 73.03 | 75.17 | **75.34** |
| S14 | 68.37 | 68.69 | 70.12 | 72.18 | 72.69 | 75.71 | **75.96** |
| S15 | 69.42 | 74.88 | 76.37 | 77.26 | 79.68 | 79.01 | **82.19** |
| S16 | 55.09 | 59.50 | 64.08 | 64.37 | 65.09 | 64.47 | **67.33** |
| S17 | 73.14 | 84.24 | 83.90 | 86.36 | **88.01** | 87.53 | 87.87 |
| S18 | 81.46 | 84.91 | 86.60 | 85.98 | 88.84 | 89.31 | **89.97** |
| S19 | 77.70 | 86.77 | 86.97 | 83.25 | 84.88 | 85.26 | **88.63** |
| S20 | 71.05 | 83.18 | 80.90 | 79.08 | 79.10 | 81.03 | **83.44** |
| S22 | 80.10 | 87.84 | 81.82 | 81.77 | 85.31 | 86.93 | **90.24** |
| S23 | 77.73 | 93.38 | 89.67 | 93.07 | 91.47 | 92.20 | **96.31** |
| S24 | 74.44 | 73.03 | 78.62 | 76.98 | 73.20 | 75.64 | **79.19** |
| Mean | 72.41 | 77.65 | 79.78 | 80.68 | 81.59 | 82.01 | **84.53** |

TABLE VI
COMPARISON OF SENSITIVITY RESULTS ON THE CHB-MIT DATASET

| Subject | PDPL | TCA | JDA | LTCCSP | CDC-KUM | TSMDL | HDA-PDPL |
|---|---|---|---|---|---|---|---|
| S1, S21 | 71.62 | 76.31 | 75.86 | 78.01 | 81.31 | 80.49 | **88.01** |
| S2 | 67.89 | 80.45 | 80.42 | 81.07 | 82.45 | 81.99 | **83.99** |
| S3 | 72.66 | 82.89 | 81.01 | 83.40 | 81.12 | 85.16 | **86.08** |
| S4 | 78.22 | 86.90 | 87.14 | 89.60 | 92.58 | 94.72 | **95.64** |
| S5 | 80.75 | 89.94 | 89.29 | 89.26 | 90.64 | 92.74 | **95.73** |
| S6 | 67.37 | 65.65 | 72.80 | 73.16 | 75.46 | 74.86 | **76.00** |
| S7 | 80.91 | 82.29 | 89.92 | 92.87 | 90.30 | 91.99 | **93.40** |
| S8 | 67.80 | 76.63 | 87.88 | 84.83 | 82.95 | 82.39 | **90.30** |
| S9 | 80.20 | 81.69 | 91.24 | 87.30 | 91.97 | 92.26 | **92.41** |
| S10 | 87.13 | 90.06 | 90.39 | 91.25 | 91.41 | 90.04 | **91.98** |
| S11 | 83.07 | 91.46 | 87.71 | 91.62 | 90.21 | 90.99 | **92.90** |
| S12 | 50.72 | 51.51 | 52.45 | 57.37 | 60.33 | **60.83** | 60.17 |
| S13 | 51.37 | 53.43 | 58.14 | 60.56 | 62.47 | 63.93 | **63.95** |
| S14 | 61.36 | 63.28 | 64.59 | 64.21 | 62.96 | 67.80 | **68.34** |
| S15 | 62.00 | 69.85 | 65.75 | 67.25 | 70.25 | 70.76 | **71.96** |
| S16 | 58.24 | 55.98 | 66.78 | 67.16 | 69.79 | 68.34 | **70.84** |
| S17 | 78.63 | 88.28 | 88.26 | 91.34 | **93.96** | 93.95 | 93.47 |
| S18 | 81.55 | 86.53 | 85.36 | 87.31 | 87.11 | **89.33** | 89.14 |
| S19 | 72.94 | 83.36 | **87.56** | 78.36 | 78.82 | 78.83 | 80.87 |
| S20 | 60.14 | 69.72 | **78.45** | 68.58 | 67.11 | 71.79 | 74.01 |
| S22 | 86.79 | 87.54 | 74.49 | 74.89 | 86.17 | 89.12 | **91.30** |
| S23 | 80.83 | 92.67 | 88.38 | 92.68 | 92.34 | 92.57 | **94.73** |
| S24 | 70.66 | 65.78 | 70.19 | **73.06** | 69.12 | 70.63 | 71.23 |
| Mean | 71.86 | 77.05 | 79.01 | 79.35 | 80.47 | 81.54 | **83.32** |

verification scheme is adopted. In each process, the EEG signals of one subject are regarded as the target domain, while the EEG signals of remaining 22 subjects are regarded as the source domain; 20% of EEG signals in the target domain participate in the training dataset, and the classification performance is then evaluated on the remaining signals in the target domain.

Tables V–VII present the comparison between the classification performance results of the HDA-PDPL and those of common methods on the CHB-MIT dataset in terms of accuracy, sensitivity, and specificity, respectively. The best results are highlighted bold. Our proposed model obtains better results than the three nondomain adaptation methods. Meanwhile, in most cases, HDA-PDPL model achieves better accuracy, sensitivity, and specificity than the three-domain adaptation comparison methods. HDA-PDPL is based on shared dictionaries in discriminative subspaces, and such shared dictionaries build a connection between different domains and may synthesize more discriminative information to encode EEG signals. Thus, the strategy of combining projection subspace

TABLE VII
COMPARISON OF SPECIFICITY RESULTS ON THE CHB-MIT DATASET

| Subject | PDPL | TCA | JDA | LTCCSP | CDC-KUM | TSMDL | HDA-PDPL |
|---|---|---|---|---|---|---|---|
| S1, S21 | 68.31 | 76.07 | 80.08 | 80.08 | 79.45 | 75.78 | **80.78** |
| S2 | 82.82 | 86.29 | 86.55 | 89.00 | 90.22 | 88.37 | **95.06** |
| S3 | 80.57 | 87.15 | 93.69 | 90.90 | 91.71 | 92.40 | **95.53** |
| S4 | 74.38 | 74.64 | 80.25 | 78.20 | **83.69** | 82.83 | 82.11 |
| S5 | 46.29 | 42.09 | 43.79 | 64.08 | 57.85 | 57.72 | **67.86** |
| S6 | 67.51 | 74.84 | 80.58 | 80.42 | 80.73 | 80.54 | **82.98** |
| S7 | 77.95 | 75.87 | 84.18 | 83.32 | 88.52 | 87.22 | **89.20** |
| S8 | 67.15 | 73.48 | 80.95 | 76.40 | 78.15 | 78.40 | **81.16** |
| S9 | 68.97 | 71.84 | 70.09 | 71.69 | **73.55** | 70.65 | 72.50 |
| S10 | 76.00 | 80.38 | 87.98 | 87.88 | 86.89 | 86.26 | **89.78** |
| S11 | 78.03 | 80.77 | 84.36 | 86.68 | **87.63** | 87.15 | 87.19 |
| S12 | 70.72 | 78.09 | 78.60 | 82.07 | 81.56 | 82.19 | **84.55** |
| S13 | 75.64 | 68.18 | 72.69 | 79.86 | 83.59 | 86.41 | **86.73** |
| S14 | 75.38 | 74.10 | 75.65 | 80.14 | 82.42 | **83.61** | 83.57 |
| S15 | 76.83 | 79.91 | 86.98 | 87.27 | 89.10 | 87.25 | **92.41** |
| S16 | 51.94 | 63.02 | 61.38 | 61.58 | 60.38 | 60.60 | **63.82** |
| S17 | 67.65 | 80.20 | 79.53 | 81.37 | 82.06 | 81.10 | **82.26** |
| S18 | 81.37 | 83.29 | 87.83 | 84.64 | 90.56 | 89.29 | **90.80** |
| S19 | 82.45 | 90.17 | 86.38 | 88.13 | 90.93 | 91.68 | **96.39** |
| S20 | 81.96 | **96.64** | 83.34 | 89.57 | 91.09 | 90.27 | 92.87 |
| S22 | 73.41 | 88.14 | 89.14 | 88.64 | 84.45 | 84.74 | **89.18** |
| S23 | 74.63 | 94.08 | 90.96 | 93.45 | 90.59 | 91.83 | **97.88** |
| S24 | 78.22 | 80.28 | 87.04 | 80.90 | 77.27 | 80.64 | **87.15** |
| Mean | 72.96 | 78.24 | 80.54 | 82.01 | 82.71 | 82.48 | **85.73** |

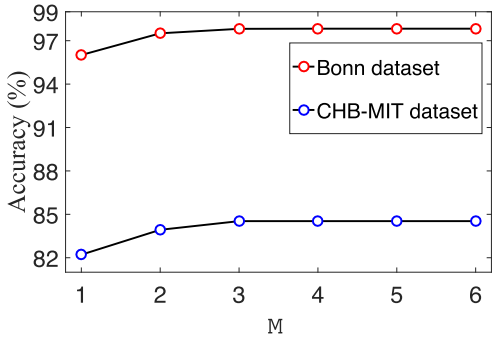



Fig. 2. Classification accuracies with different $M$ values.

and hierarchical dictionary learning adopted in this study is effective and may have an important role in the application of this model in actual clinical diagnosis problems.

### *D. Parameter Analysis*

To discuss the sensitivity of layer number in the HDA-PDPL model, we conduct experiments on Bonn and CHB-MIT datasets with different layers from 2 to 6. The variation of classification accuracies with layer parameter $M$ is shown in Fig. 2. The variation of classification accuracies is very small when the number of layers reaches 3. The classification accuracies are in an upswing when $M$ is relatively small ($M = 1$ or 2). When $M$ is larger than 3, the performance

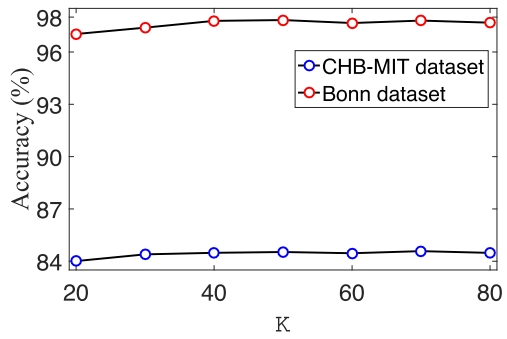



Fig. 3. Classification accuracies with different $K$ values.

increase is small. On the other hand, too many layers can increase computation. To balance the model efficiency and performance, the three-layer HDA-PDPL model is suitable for EEG diagnosis of epilepsy. Then, we conduct experiments to discuss the sensitivity of dictionary size. The variation of classification accuracies with dictionary size parameter $K$ is shown in Fig. 3. The classification performance improves gradually when $K$ varies from 20 to 40. When its size is larger than 40, our proposed model becomes insensitive to the size of the class-specific dictionary.

## V. Conclusion

For smart medical in IoMT, due to the strong individual characteristics of EEG data and the inconsistent standards of collection equipment, domain adaptation methods have drawn more and more attention in across domain EEG processing and recognition. In this article, a hierarchical domain adaptation dictionary pair learning model is proposed, which joints multi-layer common synthesis dictionaries, domain-specific analysis dictionaries, and deep learning architecture into an optimization procedure. In such a framework, the common knowledge across the source and target domains is hierarchically utilized by multiple dictionary pairs. In particular, the nonlinear feature transforms are introduced, more internal structure information of data can be exploited. Besides, the local preserved term promotes the model's discriminative ability. We applied our model to the EEG diagnosis of epilepsy. The experimental results on Bonn and CHB-MIT datasets are promising and superior to the compared methods. Since the extracted features are CNN features, in the next stage, we will further apply HDA-PDPL on a more large scale or noisy EEG datasets. A challenging task is to test our model in a cross-dataset context. Besides, we will try to embed our model into more CNN models, such as AlexNet and GoogLeNet.

## References

[1] M. Kumar and S. Chand, "A secure and efficient cloud-centric internet-of-medical-things-enabled smart healthcare system with public verifiability," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10650–10659, Oct. 2020.

[2] P. Zeng, Z. Zhang, R. Lu, and K.-K.-R. Choo, "Efficient policy-hiding and large universe attribute-based encryption with public traceability for internet of medical things," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10963–10972, Jul. 2021.

[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, Mar. 2019.

[4] J. Xu, W. Du, Y. Jin, W. He, and R. Cheng, "Ternary compression for communication-efficient federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1162–1176, Mar. 2022.

[5] P. Mathur, V. K. Chakka, and S. B. Shah, "Ramanujan periodic subspace based epileptic EEG signals classification," *IEEE Sensors Lett.*, vol. 5, no. 7, pp. 1–4, Jul. 2021.

[6] B. P. Prathaban, R. Balasubramanian, and R. Kalpana, "A wearable ForeSeiz headband for forecasting real-time epileptic seizures," *IEEE Sensors J.*, vol. 21, no. 23, pp. 26892–26901, Dec. 2021.

[7] M. Kashefpoor, H. Rabbani, and M. Barekatain, "Supervised dictionary learning of EEG signals for mild cognitive impairment diagnosis," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101559.

[8] X. Gu, C. Zhang, and T. Ni, "A hierarchical discriminative sparse representation classifier for EEG signal detection," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 5, pp. 1679–1687, Sep. 2021.

[9] Z. Jiang, F.-L. Chung, and S. Wang, "Recognition of multiclass epileptic EEG signals based on knowledge and label space inductive transfer," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 630–642, Apr. 2019.

[10] K. Y. Yan, W. M. Zheng, Z. Cui, Y. Zong, T. Zhang, and C. G. Tang, "Unsupervised facial expression recognition using domain adaptation based dictionary learning approach," *Neurocomputing*, vol. 319, no. 30, pp. 84–89, 2018.

[11] K. Huang, H. Wen, C. Zhou, C. Yang, and W. Gui, "Transfer dictionary learning method for cross-domain multimode process monitoring and fault isolation," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8713–8724, Nov. 2020.

[12] T. Ni, C. Zhang, and X. Gu, "Transfer model collaborating metric learning and dictionary learning for cross-domain facial expression recognition," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 5, pp. 1213–1222, Oct. 2021.

[13] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, Nov. 2001, Art. no. 061907.

[14] (2010). *CHB-MIT Scalp EEG Database*. [Online]. Available: http://www.physionet.org/pn6/chbmit/

[15] Y. Sun, F. Lo, and B. Lo, "Security and privacy for the internet of medical things enabled healthcare systems: A survey," *IEEE Access*, vol. 7, pp. 183339–183355, 2019.

[16] Y. Liu *et al.*, "Federated forest," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 843–854, Jun. 2022.

[17] D. He, R. Du, S. Zhu, M. Zhang, K. Liang, and S. Chan, "Secure logistic regression for vertical federated learning," *IEEE Internet Comput.*, vol. 26, no. 2, pp. 61–68, Mar. 2022.

[18] J. Xu, Y. Jin, and W. Du, "A federated data-driven evolutionary algorithm for expensive multi-/many-objective optimization," *Complex Intell. Syst.*, vol. 7, no. 6, pp. 3093–3109, Dec. 2021.

[19] Y. Cheng, J. Lu, D. Niyato, B. Lyu, J. Kang, and S. Zhu, "Federated transfer learning with client selection for intrusion detection in mobile edge computing," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 552–556, Mar. 2022.

[20] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, "A secure federated transfer learning framework," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 70–82, Jul./Aug. 2020.

[21] Y. Otoum, Y. Wan, and A. Nayak, "Federated transfer learning-based IDS for the internet of medical things (IoMT)," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[22] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, 2014, pp. 793–801.

[23] J. Song, X. Xie, G. Shi, and W. Dong, "Multi-layer discriminative dictionary learning with locality constraint for image classification," *Pattern Recognit.*, vol. 91, pp. 135–146, Jul. 2019.

[24] Y. Peng, S. Liu, X. Wang, and X. Wu, "Joint local constraint and Fisher discrimination based dictionary learning for image classification," *Neurocomputing*, vol. 398, pp. 505–519, Jul. 2020.

[25] T. Ni, Y. Ding, J. Xue, K. Xia, X. Gu, and Y. Jiang, "Local constraint and label embedding multi-layer dictionary learning for sperm head classification," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 3s, pp. 1–16, Oct. 2021.

[26] P. Z. Yan, F. Wang, N. Kwok, B. B. Allen, S. Keros, and Z. Grinspan, "Automated spectrographic seizure detection using convolutional neural networks," *Seizure*, vol. 71, pp. 124–131, Oct. 2019.

[27] 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.
[28] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.
[29] S. Hatamikia and A. M. Nasrabadi, "Subject transfer BCI based on composite local temporal correlation common spatial pattern," *Comput. Biol. Med.*, vol. 64, pp. 1–11, Sep. 2015.
[30] Y. Gu and K. Li, "A transfer model based on supervised multi-layer dictionary learning for brain tumor MRI image recognition," *Frontiers Neurosci.*, vol. 15, May 2021, Art. no. 687496.

**Weiwei Cai** (Member, IEEE) is with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China, and also with the Graduate School, Northern Arizona University, Flagstaff, AZ, USA. Prior to that, he worked with IT industry for more than ten years in the roles of a system architect and the program manager. His research interests include machine learning, deep learning, and computer vision.

Mr. Cai is an Academic Editor of *Journal of Healthcare Engineering* and *Wireless Communications and Mobile Computing*. He has also served as a Guest Editor for the *CMES—Computer Modeling in Engineering & Sciences* and *Information*.

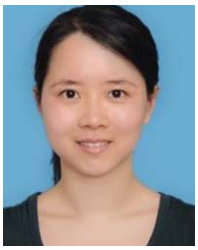
**Ming Gao** received the B.S. degree in sports human science from the Department of Sports Human Science and Psychology, Wuhan Institute of Physical Education, Wuhan, China, in 2009.

He is currently an Associate Professor with the College of Sports Science and Technology, Wuhan Sports University. He has published over ten articles in international/national authoritative journals. His current research interests include data mining and pattern recognition.

**Yizhang Jiang** (Senior Member, IEEE) received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2015.

He has also been a Research Assistant with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, for two years. He is the author/coauthor of more than 70 research articles in international/national journals, including IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and *Information Sciences*. His research interests include pattern recognition, intelligent computation, and their applications.

Dr. Jiang is an Associate Editor of IEEE ACCESS in 2019. He has served as a Reviewer or Co-Reviewer for several international conferences and journals, such as ICDM, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Pattern Recognition*, *Neurocomputing*, and *Neural Computing & Applications*. He has also served as a Leader Guest Editor or Guest Editor for several international journals, such as *Journal of Ambient Intelligence and Humanized Computing*, *Computational and Mathematical Methods in Medicine*, and *Frontiers in Neuroscience*.



**Xiaoqing Gu** received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2017.

She is currently an Associate Professor with the School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China. She has published over 50 articles in international/national authoritative journals. Her current research interests include pattern recognition and machine learning.

**Xin Ning** (Member, IEEE) received the B.S. degree in software engineering and the Ph.D. degree in electronic circuit and system from the University of Chinese Academy of Sciences, Beijing, China, in 2012 and 2017, respectively.

He is currently an Associate Professor with the Laboratory of Artificial Neural Networks and High Speed Circuits, Institute of Semiconductors, Chinese Academy of Sciences, Beijing. He has published by first or corresponding author more than 50 papers in journals and refereed conferences. His current research interests include pattern recognition, computer vision, and image processing.

**Pengjiang Qian** (Senior Member, IEEE) received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2011.

He is currently a Full Professor with the School of Artificial Intelligence and Computer Science, Jiangnan University. He has authored or coauthored more than 70 papers published in international/national journals and conferences, such as IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS—PART B: CYBERNETICS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Pattern Recognition*, *Information Sciences*, and *Knowledge-Based Systems*. His research interests include data mining, pattern recognition, bioinformatics, and their applications, such as analysis and processing for medical imaging, intelligent traffic dispatching, and advanced business intelligence in logistics.

**Tongguang Ni** received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2015.

He is currently an Associate Professor with the School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China. He has published over 50 articles in international/national authoritative journals. His current research interests include pattern recognition, intelligent computation, and their application.