Inference of Brain States Under Anesthesia With Meta Learning Based Deep Learning Models

Qihang Wang, Feng Liu, Guihong Wan, and Ying Chen

***Abstract*— Monitoring the depth of unconsciousness during anesthesia is beneficial in both clinical settings and neuroscience investigations to understand brain mechanisms. Electroencephalogram (EEG) has been used as an objective means of characterizing brain altered arousal and/or cognition states induced by anesthetics in real-time. Different general anesthetics affect cerebral electrical activities in different ways. However, the performance of conventional machine learning models on EEG data is unsatisfactory due to the low Signal to Noise Ratio (SNR) in the EEG signals, especially in the office-based anesthesia EEG setting. Deep learning models have been used widely in the field of Brain Computer Interface (BCI) to perform classification and pattern recognition tasks due to their capability of good generalization and handling noises. Compared to other BCI applications, where deep learning has demonstrated encouraging results, the deep learning approach for classifying different brain consciousness states under anesthesia has been much less investigated. In this paper, we propose a new framework based on meta-learning using deep neural networks, named *Anes-MetaNet*, to classify brain states under anesthetics. The Anes-MetaNet is composed of Convolutional Neural Networks (CNN) to extract power spectrum features, and a time consequence model based on Long Short-Term Memory (LSTM) networks to capture the temporal dependencies, and a meta-learning framework to handle large cross-subject variability. We use a multi-stage training paradigm to improve the performance, which is justified by visualizing the high-level feature mapping. Experiments on the office-based anesthesia EEG dataset demonstrate the effectiveness of our proposed Anes-MetaNet by comparison of existing methods.**

***Index Terms*— Brain state estimation, meta learning, anesthesia EEG, deep learning.**

Manuscript received December 31, 2021; revised March 9, 2022; accepted April 4, 2022. Date of publication April 11, 2022; date of current version April 29, 2022. The work of Qihang Wang and Ying Chen was supported in part by the National Natural Science Foundation of China under Grant 72101066, Grant 72131005, and Grant 91846301. *(Qihang Wang and Feng Liu contributed equally to this work.) (Corresponding author: Ying Chen.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Qihang Wang and Ying Chen are with the School of Management, Harbin Institute of Technology, Harbin 150001, China (e-mail: yingchen@hit.edu.cn).

Feng Liu is with the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030 USA.

Guihong Wan is with the Department of Dermatology and the Department of Biomedical Informatics, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02445 USA.

Digital Object Identifier 10.1109/TNSRE.2022.3166517

I. Introduction

PROVIDING the right dose of anesthetic drugs to a patient is of paramount importance for safely and humanely performing most surgical and many nonsurgical procedures [1]. Underdoses of anesthetic drugs will make patients wake up during surgery, while an overdose of drugs increases the risk of post-operative delirium [2], [3]. The inference of brain states provides a unique opportunity for precise administration of drugs on different subjects based on real-time closed-loop anesthesia delivery system [4]. According to the Richmond Agitation-Sedation Scale (RASS), anesthesiologists mostly estimate the depth of anesthesia (DoA) manually at intervals; however, this would require considerable attention with a complicated stimuli-response procedure during the surgery [5]. Measurements through the observation of heart rate, breathing pattern, blood pressure and other factors, have been used to measure DoA [6]; however, these physiological signals are the secondary measurements of DoA, showing large cross-subject variability that requires significant experience from anesthesiologists to decode the DoA. In recent years, EEG as an invaluable modality of recording brain activity has attracted more and more attention among anesthesiologists [7]. Compared to the secondary measurements, EEG signal is a direct measurement for brain states as the brain cognitive process relies on communications between neuronal populations through electrical signal [8]. Importantly, it has been found that DoA is associated with EEG signatures given different drugs [3], [9]. Currently, the Bispectral Index (BIS, Aspect Medical Systems, USA) is the most widely used DoA monitor index in clinical practice; however, previous researchers have reported a poor connection between the DoA and the BIS [10], which requires more accurate means to estimate the DoA in clinical setting.

Existing work: Numerous EEG-based monitoring frameworks for the DoA have been developed using statistical or machine learning models [11]–[13]. Specifically, Shalbaf *et al.* monitored the DoA by extracting entropy features from EEG and using artificial neural networks for classification of DoA [14]. Peker *et al.* estimated the DoA by combining ReliefF feature selection and random forest algorithm [15]. Jahanseir *et al.* estimated the DoA with a multi-output least-square support vector regression method [16]; Saadeh *et al.* assessed the DoA with a machine learning fine decision tree classifier for DoA classification of 4 states (deep, moderate, and light DoA versus awake state) [17]. In addition to the above classical machine learning models, recently, many researchers employed deep learning as classification models for monitoring the DoA. Compared to the traditional machine

learning models, deep learning has demonstrated outstanding performance on many computer vision and natural language processing tasks, among other interesting applications [18]–[22]. As for the DoA task, only a few researches based on deep learning have been conducted. For instance, Park *et al.* proposed a real-time DoA monitoring system based on a convolutional neural network framework [23]; Li *et al.* introduced a novel method based on hybrid features including sample entropy, permutation entropy, spectra, and alpha-ratio and performed classification using recurrent neural network [24]; Afshar *et al.* developed a combinatorial deep learning structure involving convolutional neural networks, bidirectional long short-term memory, and an attention layer to estimate the DoA precisely from EEG signals [25].

Knowledge gap and our contribution: Although deep learning frameworks proposed in the previous literature typically outperform the existing classical machine learning algorithms, these studies are designed to analyze the anesthetic EEG data collected from the hospital-based environment. Compared to hospital-based anesthesia (HBA), office-based anesthesia (OBA) is conducted in an outpatient setting. The advantages of OBA include easy scheduling, lower cost, and improved patient privacy, etc. However, it also has some disadvantages, which include slow recognition or response to emergent situations and lack of experience intensively monitoring of drug administration. It can result in a high level of motion artifacts and low signal to noise ratio (SNR) in the collected EEG data, which makes the EEG data from OBA more challenging to analyze. However, to the best of our knowledge, there have been no previous studies that build classification models for OBA with deep learning frameworks. To bridge the research gap, in this study, we propose a deep-learning classification framework for monitoring brain states under anesthetics in the OBA context. However, two issues should be taken into account for OBA data analysis. The first issue we try to address is the cross-subject variability within the EEG data. The EEG signal across subjects can have high intra-class variability. Applying some signal processing methods, such as wavelet transformation [26], [27], Fourier transformation [28], [29], multitaper analysis [30] and artifacts removal algorithms [31], can reduce the impact of noise. However, there is no principled way to address the intra-subject variability. As demonstrated in the Ref. [32]–[34], convolutional neural networks (CNN) can extract features of EEG data and achieve outstanding performance in the hospital-based DoA classification. However, the CNN model alone faces the challenges of achieving a good classification result in cross-subject testing. In other words, the direct application of CNN on the EEG data from anesthetic patients might result in unsatisfactory performance on the testing data, though the training performance can be good. Therefore, how to address the inter-subject variability under OBA becomes the main focus of this study. Secondly, given the different contexts of OBA and HBA, classifying the DoA based on OBA can be challenging due to the high level of noise in OBA. Moreover, the OBA dataset to analysis was under the condition of a mixed infusion of propofol, ketamine, dexmedetomidine and lidocaine, contributing significantly to the intra-class variability. We propose to exploit the temporal dependency of spectrum features using long short-term memory (LSTM) networks to mitigate the high level of artifact EEG noise. Our assumption is the brain only maintains one state given a short period of time. As for subject variability, inspired by the great success of meta-learning in dealing with cross-subject variability [35], we propose to incorporate the deep learning framework into the meta-learning paradigm. The essence of using meta-learning is to increase the generalization ability of the learner in multiple tasks [36]. Hence, addressing these two issues motivates us to develop a robust DoA estimation framework based on meta-learning with deep learning models. Given the unbalanced samples from different classes, we introduce a sequential classification method by first classifying the classes with better separability and then classifying the remaining classes. The experiments are conducted on the subjects with quality EEG data, and eight methods or models are used as benchmarks including two latest models (one is from [24] and the other is from [25]), to demonstrate the advantages of the proposed model.

We name the proposed network employing CNN, LSTM with sequential classification framework under the framework of meta-learning as *Anes-MetaNet*. To summarize, our contributions can be listed as:

- We propose to use deep learning to build the classification framework to infer DoA using EEG data from OBA setting.
- We incorporate a meta-learning framework to address cross-subject variability from the OBA.
- We introduce a sequential classification method to reduce the impact of the imbalanced classes.
- Our proposed Anes-MetaNet achieves the best performance against the other baseline methods.

The remainder of this paper is organized as follows: Section II introduces the proposed method and baseline methods; Section III presents the experimental results and discussion, followed by conclusion in Section IV.

II. Methodology

In this section, we first introduce the OBA EEG dataset, and the corresponding preprocessing procedure. Then, we discuss the detailed implementation of different models for classification, including CNN, meta-learning CNN (MCNN), LSTM network, as well as our proposed Anes-MetaNet.

A. Data Description

We use a public office-based anesthesia EEG dataset [37] for this study. The dataset provides EEG recordings of patients during general anesthesia. The induction and maintenance of sedation is done by administering low-dose propofol ramp, intermitted by boluses of propofol, ketamine, dexmedetomidine and lidocaine. Table I lists the provided attributes of the dataset. The RASS score is from a manual assessment of the degree of sedation for patients, ranging from 0 to -5, representing the degree of sedation from consciousness (CON, score 0) to complete loss of consciousness (LOC, score -5). EEG data contains recordings from five channels. In our

TABLE I
VARIABLE DESCRIPTION OF THE ANESTHESIA DATA SET USED IN THE EXPERIMENT

Variables	Description
EEG	Raw five channel EEG (bandpass filtered between 0.1-55 Hz). Each row corresponds to channel names in Channelname variable
Channelname	EEG channel labels
EEG time	Continuous EEG time stamps
Fs	Sampling frequency
Patient char	Patient characteristics: Procedure type, sex (m/f), age, height (cm), weight (kg)
RASS	RASS annotations: 0,-1,-2,-3,-4, and -5
RASS time	Time of RASS assessment

paper, we test the data of FPZ, FP1 and FP2 channels for the experiment. Due to high motion artifacts during EEG recordings, there are significant noises in the data for a large number of patients.

1) Brain State Labels: After visualizing the spectrogram of the EEG data, the SNR of some subjects is quite low, apparently worse than the HBA data used in [25] and [24]. Hence, it is difficult to classify the OBA data into multiple states as conducted in the previous works. Despite this, the transitioned state between CON and LOC states is critically important for the anesthetists to control the doses of drugs. We then determine to classify the DoA into LOC, CON and the transitioning states in this study. Specifically, we use the EEG epochs with RASS = 0 and -1 as the CON state, the EEG epochs with RASS $= -2$ and -3 as the semi-consciousness (semi-CON) state, and the EEG epochs with RASS $= -4$ and -5 as LOC state, denoted as *class 0 (CON)*, *class 1 (semi-CON)*, and *class 2 (LOC)*.

2) Power Spectrum Analysis: The spectrogram of each subject can be obtained by using multitaper analysis implemented in Chronux Toolbox [38]. The horizontal axis x_n of the spectrogram image represents time in seconds; the vertical axis y_m of the spectrogram represents the frequency in Hz. The value of the ordinate is the signal power in decibel (DB) of x_i time and at y_j frequency. The spectrogram images of 4 subjects are shown in Fig. 1. From the spectrogram images, we visually picked relative high-quality EEG recordings. For example, from Fig. 1, we can find that subjects *a* and *b* have good quality EEG data, while the EEG recordings of subjects *c* and *d* contain a lot of noises or artifacts.

a) Subjects a and b: The middle part of the recording is usually the LOC stage, and the beginning and end parts are usually the CON and semi-CON stages. We name the combination of CON and semi-CON stages as the pre-LOC stage. From subjects *a* and *b*, we can observe that the EEG data of LOC state has higher energy in the range of 10-20 Hz. And the EEG data of the pre-LOC state has higher energy within 10 Hz, and lower energy in other frequency ranges. The spectrograms of LOC period and pre-LOC period are significantly different, and satisfactory results can be achieved by classifying LOC and pre-LOC states. However, in the pre-LOC state, there is no obvious difference between the CON and semi-CON states in the spectrogram. As a result, the classification of the CON and semi-CON states is challenging, which motivates us to design a two-stage classification framework.

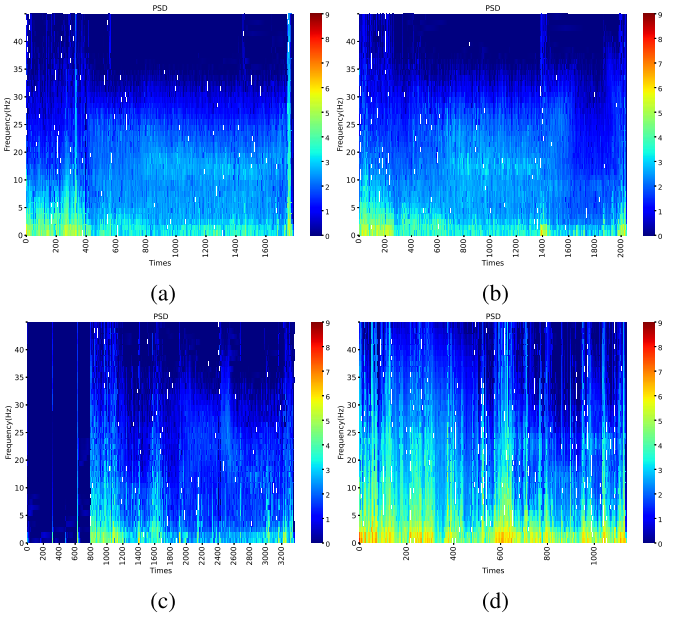

Fig. 1. Spectrogram images of 4 subjects: *a*-*d*. Compared to subjects *a* and *b*, subjects *c* and *d* contain significant motion artifacts and noises.

b) Subject c: From the spectrogram image, we can clearly see that the beginning part of the data of subject c is close to 0. This is caused by technical problems in the recording process of EEG data, and then we exclude this subject.

c) Subject d: The spectrogram image of subject d shows low SNR of EEG data due to very strong artifacts, hence, we exclude this subject as well.

By observing the multitaper spectrogram of raw EEG data, we select 13 patients with relatively good data for the following classification experiment.

B. Data Preprocessing

We use a bandpass filter to filter the EEG data with a frequency band of 0.1 to 40 Hz. The RASS scores are aggregated into classification labels for patients' anesthesia states. The time intervals between two RASS assessments are not constant, ranging from a few seconds to a few minutes. We add labels for each timestamp of EEG data using the following steps: 1) Find the intermediate time points between all adjacent assessment time periods. 2) If two labels at the beginning time point and ending time point share the same label, then label all the time stamps between two RASS scores as the same label. 3) If the two labels are distinct, label the left half of EEG data within the current interval using the same label as the assessment on the left; label the right half of EEG data with the same label as the assessment on the right. It is worth noting that the labels in 3) are not perfectly accurate. However, we only use them in the training process, and use the labels determined by 2), as the testing data.

C. Sequential Classification Framework

We use a multi-stage sequential classification framework to address the imbalanced data distribution. Taking a three-way classification as an example, in the first step, we observe

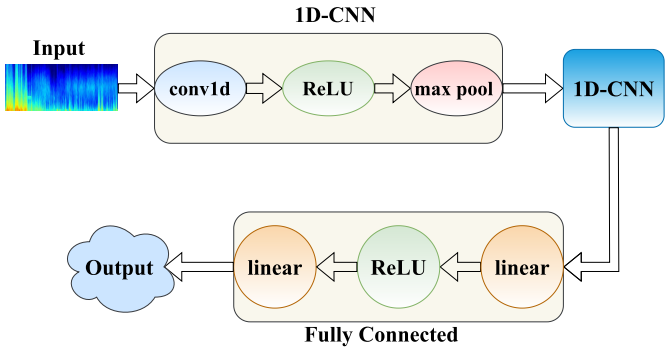

Fig. 2. Structure diagram of CNN model.

the data distribution to identify which labeled stage has the largest amount of data and allocate all the rest data to the other class. By using a classification model, such as MCNN, we can obtain a binary classification. In the second step, following the first-step implementation of classification models on the remaining classes, we can obtain the remaining binary classification result. Given the characteristics of the OBA EEG data, we determine to classify the DoA into three states (CON, semi-CON, and LOC) as discussed above. Hence, this classification framework consists of two stages: stage 1: classify pre-LOC vs LOC, and stage 2: classify CON vs semi-CON.

D. Convolutional Neural Networks

Since the power spectrum features at one time point are a one-dimensional vector, the one-dimensional convolutional neural network (1D-CNN) is adopted to extract the features. Unlike the traditional neural networks, convolutional layers in CNN are not fully connected, which allows CNNs to have fewer parameters to adjust, and to be more computationally efficient. As seen in Fig. 2, 1D-CNN includes a convolutional layer, rectified linear unit (ReLU) and a one-dimensional maximum pooling operation. For the classification of DoA states, our CNN model is illustrated in Fig. 2, which includes two 1D-CNN blocks and one fully connected block. For the classification task, we use cross-entropy as the loss function, and Adam as the optimization algorithm to search the optimal parameters in CNN model. Such a loss function and optimization algorithm are also exploited in all the other deep learning models of this study.

E. Meta-Learning CNN

EEG signals vary significantly from subject to subject. The classification function f obtained using CNN training from EEG data of some subjects may not be applicable to other subjects. Meta-learning is used in this paper to address this problem. As the meta-learning model is capable of adapting and generalizing to new tasks and new environments that have never been encountered during training time. The difference between meta-learning and supervised learning is that the function of supervised learning directly finds the mapping between features and labels, but the meta-learning function $F(\cdot)$ is to find a new f, so that the new f is suitable for the new machine learning task [35]. The algorithm flow of MCNN is shown as follows:

Step 1: Prepare N training tasks. Prepare a test task to evaluate the effects of the parameters learned by MCNN. Prepare a support set and a query set for each task.

Step 2: Define the network structure of CNN and initialize a meta-network parameter ϕ^0. The meta-learning network is the network that is eventually applied to the new test task.

Step 3: Start performing iterative pretraining:

a) For task i, assign parameter ϕ^0 of the meta-learning network to the CNN to get $\hat{\theta}^i$.
b) The support set of task i is used to optimize and update $\hat{\theta}^i$ based on the learning rate α^i of task i.
c) Calculate loss $l^i\ (\hat{\theta}^i)$ for task i using the query set, and compute the gradient of $l^i\ (\hat{\theta}^i)$ with respect to $\hat{\theta}^i$.
d) The gradient is multiplied by the learning rate α^{meta} of the meta-learning network and update ϕ^0 to ϕ^1.
e) Repeat a)-d) on training tasks.

Step 4: Parameters of the meta-learning network are obtained by step 3 and used for testing. Fine-tune the network with the support set of the test task.

Step 5: Evaluate the effectiveness of the MCNN using the query set of the test task.

The structure of the MCNN model is shown in Fig. 3. As noted, the CNN block shown in Fig. 3 is the CNN model presented in Fig. 2. After training, the MCNN model finds a function $F(\cdot)$. Whenever a new test task is entered, $F(\cdot)$ can always find a set of appropriate parameters to achieve an improved classification performance of CNN model.

F. Long Short-Term Memory

Recurrent neural networks (RNN) is good at processing sequential data. LSTM network is a special instance of RNN which mitigates the problems of gradient vanishing and gradient explosion in long sequence training of RNN. LSTM can effectively extract the time characteristics of EEG data, and combine them with the frequency domain features extracted by CNN to achieve high-accuracy classification. Each LSTM model is comprised of three gates, namely input gate, forget gate and output gate. In addition, it also includes a cell state that has long-short term memory. The operation among these four components are as follows. First, the input gate updates the cell state with a specific mechanism after receiving the input information. And then, the forget gate determines what input information will be discarded. Last, the output gate will generate the results based on the input information and the cell state. Let x_t denote the input vector at time t, h_t indicate the output of hidden layer, c'_t connote the preliminary information that will transmit to cell layer, c_t stand for the cell memory, i_t, f_t and o_t represent the output of input gate, forget gate and output gate, respectively. The relationships among these variables are shown below (W and b are the weight and bias in the operation of model training, separately):

$$i_t = s(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

$$f_t = s(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

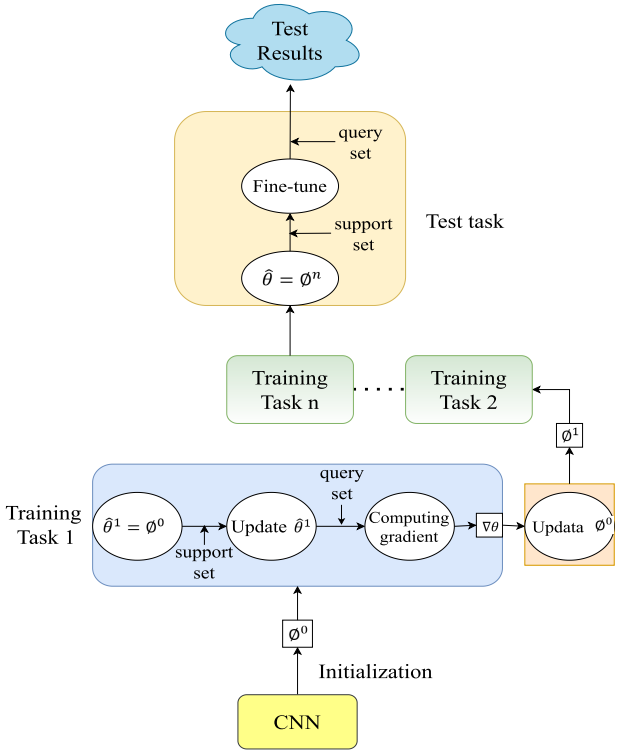

Fig. 3. Structure diagram of MCNN model.

$$c_t = i_t \cdot c_t' + f_t \cdot c_{t-1} \tag{3}$$

$$o_t = s(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{4}$$

$$h_t = o_t \cdot tanh(c_t) \tag{5}$$

Fig. 4. presents the LSTM model used in this study, which consists of two LSTM layers and two fully connected layers. Particularly, in order to prevent over-fitting during neural network training, dropout operations are added to the fully connected block. The parameter of dropout is set to 0.5, which means that neurons at this layer have a random 50% chance of being dropped and not participating in the training during each iteration.

G. Proposed Anes-MetaNet

Anes-MetaNet uses CNN and LSTM to capture power spectrum density (PSD) features with the consideration of temporal dependency under the meta-learning training scheme. In Anes-MetaNet, CNN based meta-learning can be trained with multiple stages, depending on the number of classifications. The detailed structure of Anes-MetaNet is shown in Fig. 5. In this study, we classify the EEG anesthetic data into three classes so that two stages are enough. In order to classify CON, semi-CON and LOC states, in the first stage of training, labels of EEG data are divided into two categories: Pre-LOC and LOC. We perform the second-stage classification on the correctly classified Pre-LOC data. For Pre-LOC data that is misclassified in the first stage, we eliminate them before the second-stage classification. In the second stage of training for Anes-MetaNet, CON and semi-CON states are further classified. We save the features extracted by MCNN as the input of LSTM. The features corresponding to the LOC state are the features extracted from the first trained MCNN model. The features of CON and semi-CON states are the features extracted from the second trained MCNN model. We use the features of 20 time windows extracted by CNN as the input of LSTM. One time window is 3 seconds in size and moves in a stride of 1 second. The output is the predicted state of the 20th time window. We believe the length of 20 time windows is good since the anesthesia data has a slow change pattern within a short time, and LSTM is likely to capture the good features for such an input according to its internal mechanism. As noted, the meta-learning CNN and LSTM blocks in Fig. 5, are from Fig. 3 and from Fig. 4, respectively.

III. Experiments and Discussion

We conduct nine experiments to demonstrate the effectiveness of our proposed method. Among the 13 subjects, we implement a cross-subject validation by using the data of 10 subjects as the training set, the EEG data of 3 remaining subjects is used as the test data. We take each subject in turn as the training set and the test set for cross validation. The dataset is split into training dataset and testing dataset, each with 50% of the total samples, and we use the training dataset to train the MCNN model. For each type of label, we randomly extract the same amount of data for training, which can effectively prevent the overfitting problem caused by the problem of class imbalances.

A. Experiments Results

We conduct experiments on three channels of the EEG data. We compare our proposed network with other baseline models, including support vector machines (SVM), random forest (RF), CNN, LSTM, CNN with LSTM (CLSTM), and MCNN. In addition, we select two latest deep learning models from literature as benchmarks, CLA [25] and ALSTM [24]. Table II shows the experimental results of classification accuracy and standard deviation based on FPZ channel.

1) Experiments on SVM and RF: Table II shows that the average classification accuracy of SVM and RF is lower than the deep learning method, which demonstrates the superiority of handling large noise under deep learning paradigm. The average classification accuracy of SVM is 72.8%, and the average classification accuracy of RF is 69.3%. SVM uses "gridsearch" algorithm to optimize its hyperparameters, and its classification accuracy is slightly higher than RF's. The confusion matrices of the classification results of SVM and RF are shown in Fig. 6. As seen in Fig. 6(a), an obvious disadvantage of SVM is its bad performance in classifying the semi-CON state. Fig. 6(b) illustrates that although the RF model can classify the semi-CON state, the accuracy of the classification is less than 30%. Moreover, the classification accuracy of the RF model for the CON state is only about 50%. The above results showcase the difficulty of classifying DoA using traditional machine learning models.

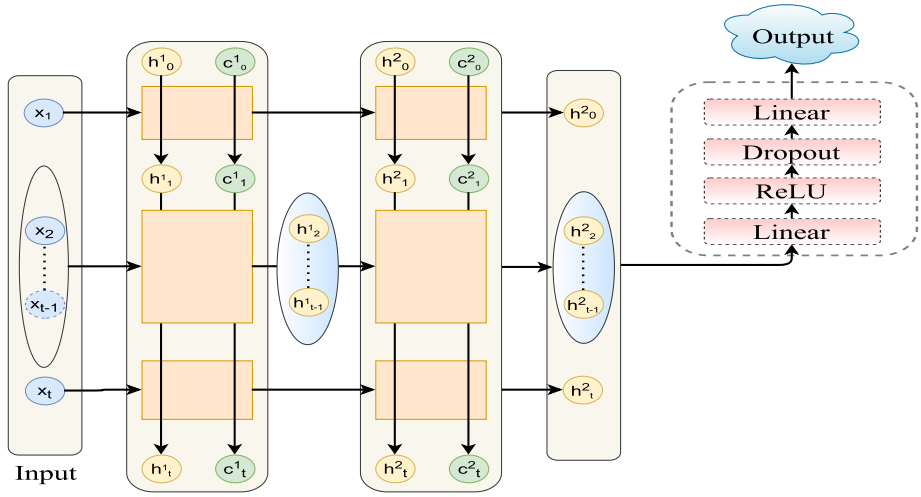

Fig. 4. Structure diagram of LSTM model.

TABLE II
CLASSIFICATION RESULTS OF EEG DATA FROM FPZ CHANNEL

subject	SVM	RF	CNN	LSTM	CLSTM	MCNN	CLA	ALSTM	Anes-MetaNet
1	75.9(±0.3)	73.1(±1.1)	76.3(±0.2)	77.0(±0.2)	77.9(±0.5)	79.8(±0.6)	82.1(±0.2)	78.6(±0.2)	83.6(±4.0)
2	65.7(±0.3)	62.6(±0.4)	76.8(±0.1)	76.2(±0.2)	75.3(±2.0)	75.9(±0.3)	77.3(±0.3)	64.7(±4.1)	79.2(±0.2)
3	75.8(±0.4)	72.5(±0.3)	69.5(±0.1)	72.3(±5.4)	70.0(±6.2)	75.3(±0.1)	74.6(±0.5)	74.8(±4.7)	82.5(±1.8)
4	81.6(±0.2)	80.5(±0.5)	84.0(±0.2)	86.5(±1.5)	83.8(±0.2)	84.6(±0.2)	88.9(±0.3)	84.2(±0.3)	90.5(±5.4)
5	73.8(±0.2)	73.2(±0.2)	73.6(±0.3)	74.4(±1.0)	69.4(±0.8)	78.5(±0.5)	76.1(±0.3)	72.0(±5.9)	83.3(±3.6)
6	80.0(±0.2)	57.1(±2.8)	81.5(±0.5)	84.0(±0.5)	80.2(±0.6)	80.6(±0.2)	81.4(±0.2)	82.7(±1.5)	91.5(±0.5)
7	73.1(±0.4)	71.5(±0.5)	71.2(±0.2)	70.1(±0.1)	64.2(±9.8)	70.3(±0.1)	71.4(±0.1)	73.9(±0.7)	82.3(±1.5)
8	84.1(±0.8)	82.7(±0.3)	83.9(±0.5)	80.1(±1.2)	38.1(±12.0)	83.8(±0.1)	85.9(±0.4)	76.0(±0.5)	64.1(±6.1)
9	53.8(±0.2)	54.1(±0.4)	53.1(±0.3)	53.1(±1.0)	57.1(±0.9)	63.3(±2.3)	53.1(±0.1)	61.3(±0.5)	73.9(±1.6)
10	62.7(±1,1)	56.8(±0.4)	61.6(±0.4)	66.3(±0.6)	57.4(±0.4)	63.6(±2.7)	58.6(±1.3)	52.6(±7.2)	82.3(±3.7)
11	72.2(±0.2)	72.0(±0.5)	75.2(±1.2)	70.5(±1.4)	68.0(±0.5)	78.5(±0.5)	68.7(±0.3)	78.6(±0.2)	78.7(±0.8)
12	72.4(±0.1)	72.7(±0.3)	71.8(±0.5)	69.6(±0.4)	47.5(±4.9)	70.1(±1.0)	72.0(±0.6)	70.3(±2.4)	83.9(±0.7)
13	75.7(±0.2)	72.5(±0.5)	76.6(±0.1)	79.6(±0.7)	44.6(±19.0)	75.1(±0.4)	80.1(±0.2)	77.0(±1.3)	87.9(±3.9)
AVG	72.8	69.3	73.4	73.8	64.1	75.3	74.6	72.8	81.8

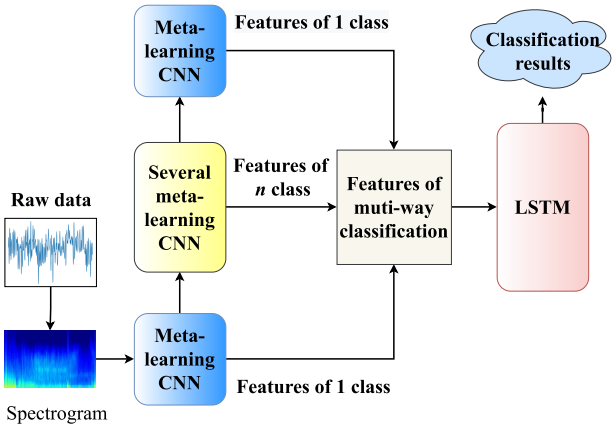

Fig. 5. Structure diagram of Anes-MetaNet.

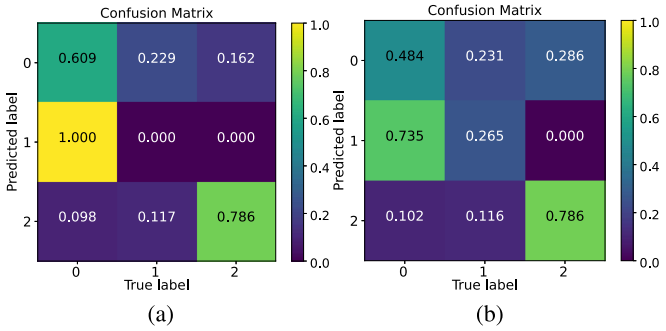

Fig. 6. Confusion matrix of the classification results of SVM and RF, where (a) is the confusion matrix of the classification results of SVM, and (b) is the confusion matrix of the classification results of RF. The experimental data is the EEG data of the FP2 channel of subject 1.

2) Experiment on CNN Model: In Table II, the average classification accuracy of CNN model has reached 73.4%, which is slightly higher than traditional machine learning methods (SVM and RF), without significant improvement. Fig. 7(a) shows the confusion matrix of the classification results of the CNN model. It can be seen that CNN model also has poor classification performance on semi-CON state. In this experiment, we use the exponential decay method to adjust the learning rate of the CNN model. Fig. 7(b) shows the convergence of loss function of CNN.

3) Experiments on LSTM Model: In Table II, the results of LSTM model are superior a little to CNN model, with an average classification accuracy of 73.8%. Fig. 8(a) shows that

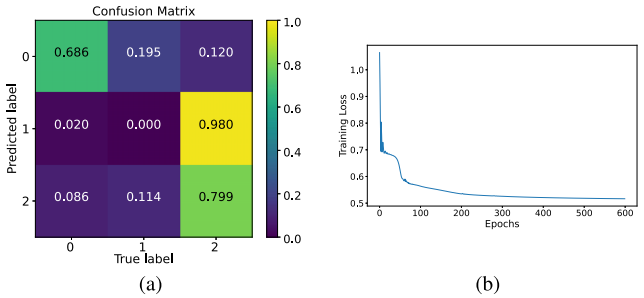

Fig. 7. Classification result of CNN model, (a) Confusion matrix diagram of the classification, (b) training loss of CNN.

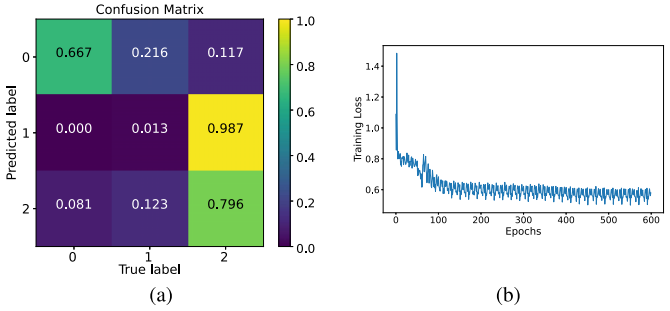

Fig. 8. Classification result of LSTM model, (a) Confusion matrix diagram of the classification, (b) training loss of LSTM.

although the accuracy of LSTM model is improved compared with those of SVM and RF models, an obvious shortcoming of LSTM model is that it also cannot classify semi-CON states well. Moreover, Table II illustrates that compared with SVM, RF and CNN models, LSTM model has lower classification stability. Fig. 8(b) shows the loss time series of the LSTM model. It can be seen that the loss curve is gradually stable, which indicates that the model has converged.

4) Experiments on CLSTM Model: In this paper, we use CLSTM to represent the model combining CNN and LSTM models. Note that the CLSTM model consists of two convolutional layers and two LSTM layers as well as two fully connected layers. Table II shows that the classification accuracy of CLSTM is only 64.1%, which is far lower than traditional machine learning models and deep learning models. After analysis, we believe that the poor classification effect of CLSTM is caused by the poor classification results of CNN model. Even though CLSTM model can achieve a higher accuracy on semi-CON state than SVM, CNN and LSTM models; however, the accuracy is only 20.7%. Nevertheless, we still believe CLSTM has a potential to reach a higher classification accuracy once adding some other mechanisms into it. This is confirmed by the following comparative experiments.

5) Experiments on MCNN Model: Table II shows that the classification accuracy of the MCNN model has reached 75.3%, which is higher than all traditional machine learning and deep learning models. For poor-quality data such as subject 9, the classification accuracy of MCNN is significantly improved compared to SVM and CNN models. MCNN is trained in two stages, and each stage conducts a binary classification, following the sequential classification method.

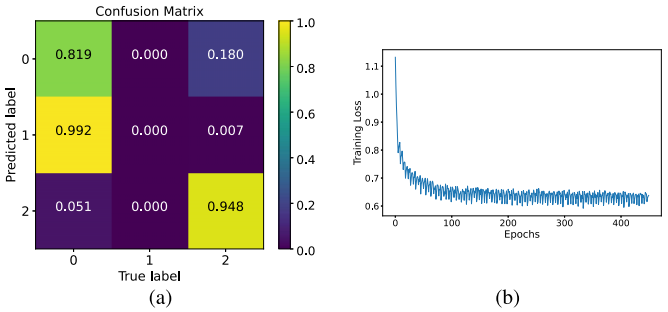

Fig. 9. Classification result of CLSTM model, (a) Confusion matrix of the classification by CLSTM, (b) training loss of CLSTM.

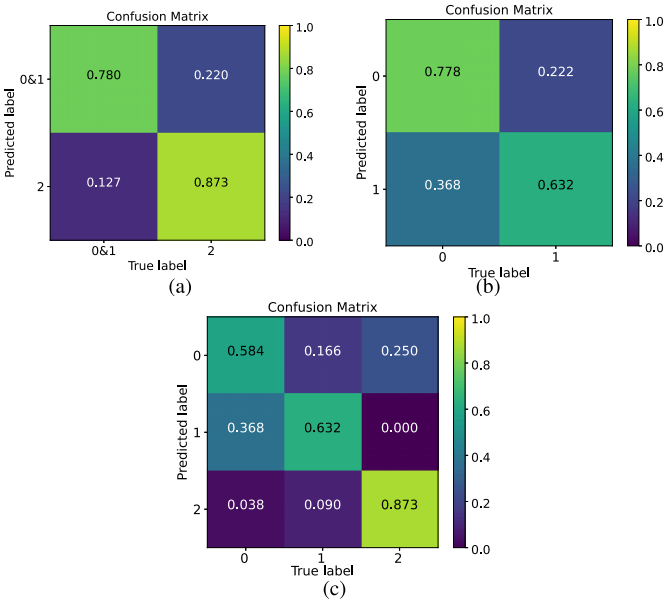

Fig. 10. Classification result of MCNN model, (a) confusion matrix of MCNN in the first stage, (b) confusion matrix of MCNN in the second stage, (c) final confusion matrix.

Fig. 10(a) shows the confusion matrix of the classification results of the first stage. The classification accuracy of the LOC state is about 87.3%, and the classification accuracy of the pre-LOC state is about 78.0%. Fig. 10(b) shows the confusion matrix of the classification results of the second stage. The classification accuracy of the CON state is about 77.8%, and the classification accuracy of the semi-CON state is about 63.2%. Fig. 10(c) shows the confusion matrix of the final classification results of MCNN. The classification accuracy of the CON state is about 58.4%, the classification accuracy of the semi-CON state is about 63.2%, and the classification accuracy of the LOC state is about 87.3%. Fig. 10(c) illustrates that MCNN not only has a high overall classification accuracy, but also can improve the classification accuracy of the CON and semi-CON states.

6) Experiments on CLA Model: In [25], Afshar *et al.* proposed a combinatorial deep learning model, which uses inception, Bi-LSTM and attention mechanism. We name this model as CLA in this study. For the specific OBA data used in this study, besides using the attention mechanism, we set this model includes two inception layers and two Bi-LSTM layers.

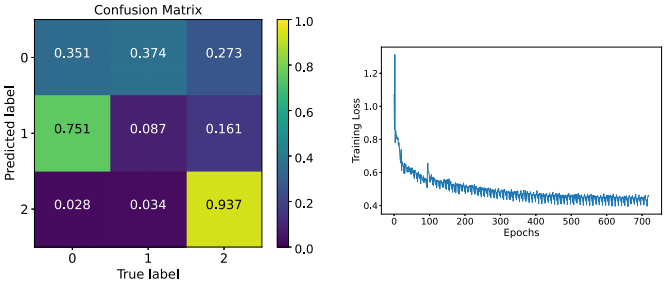

Fig. 11. Classification result of CLA model, (a) confusion matrix diagram of the classification, (b) training loss of CLA.

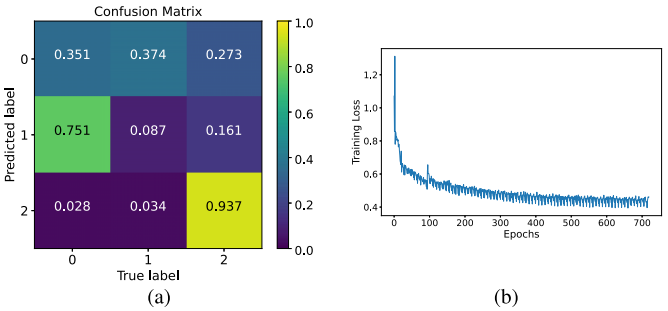

Fig. 12. Classification result of ALSTM model, (a) confusion matrix diagram of the classification, (b) training loss of ALSTM.

Even though this model is targeted to classify the DoA under the HBA condition, we exploit it as another benchmark to demonstrate the effectiveness of the proposed model. Table II shows that the average accuracy is 74.6%, which is slightly lower than that using MCNN model. Compared to MCNN, this model also has a quite poor performance in the Subject 9. Fig. 11(a) presents the confusion matrix of the classification accuracy using this model. As seen, this model cannot classify the semi-CON state completely, which is the same as that in above SVM and CNN models. Even though the CLA model performs much better than the other existing models in [25] when applying it to an HBA dataset, such results indicate this model is not suitable for the OBA condition.

7) Experiments on ALSTM Model: In [24], Li *et al.* introduced a novel method based on LSTM and a sparse denoising autoencoder (SDAE) to monitor the DoA for the HBA condition. We name this model as ALSTM in our study. Consistent with above experiments, the input data for ALSTM is also the spectrogram. We follow Li *et al.* to train the SDAEs one by one, two times in total. The output of the second SDAE's encoder is used as the input of the two-layer LSTM for the DoA prediction training. As seen in Table II, the classification accuracy is only 72.8%, which is the same as that of SVM. Even though the training loss is converged in Fig. 12(b) and the LOC state has a high classification accuracy, *i.e.,* 82.6%, the classification on the first two states has a quite low accuracy. Such a result denotes that this ALSTM model is also not appropriate to be employed for estimating the DoA under the OBA condition.

8) Experiments on Anes-MetaNet: Anes-MetaNet achieves an average classification accuracy of 81.8% in the experiments

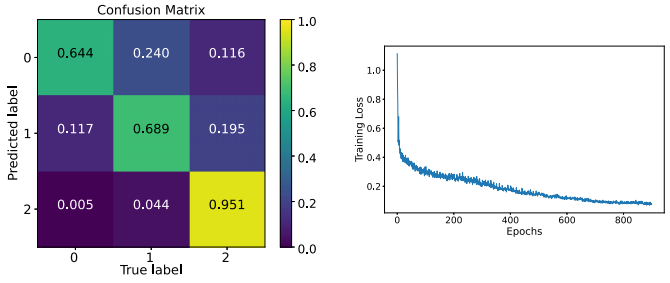

Fig. 13. Classification result of Anes-MetaNet model, (a) confusion matrix diagram of the classification, (b) training loss of Anes-MetaNet.

shown in Table II, which is the only model with an average accuracy higher than 80%. For poor quality data such as subjects 9 and 10, the accuracy of Anes-MetaNet is more than 10% higher than other models. Fig. 13(a) shows the confusion matrix of the classification result of Anes-MetaNet. The classification accuracy of CON state is about 64.4%, the classification accuracy of semi-CON state is about 68.9% and LOC state is about 95.1%. Compared to the traditional machine learning models and deep learning models, the proposed Anes-MetaNet can obtain a much higher accuracy on semi-CON state. Furthermore, compared to MCNN model, Anes-MetaNet can have a better classification accuracy on each class. Fig. 13(b) shows the loss curve of Anes-MetaNet during the training process. The curve fluctuates slightly during the training process. This is because we use the dropout layer to prevent the network from overfitting. The loss curve eventually stabilizes, indicating that the network converges. The experimental results of Anes-MetaNet prove that the classification accuracy of the CNN model is an important factor affecting the classification accuracy of the CLSTM model. If the accuracy of the CNN model is low, combining LSTM would not further improve the classification accuracy (see the results using CLSTM). Anes-MetaNet uses meta-learning to improve CNN's classification ability, which greatly improved LSTM's classification accuracy.

We use t-SNE map to visualize high-level features before the final fully connected layer of CNN, and the result is illustrated in Fig. 14. As seen, class 2 (LOC) has a more pronounced difference from other classes (0: CON and 1: semi-CON), which justifies our previous choice of using a two-stage training scheme to first distinguish LOC from pre-LOC (combining CON and semi-CON). Besides this, we can see that class 0 and class 1 are mixed together closely, which explains why the traditional machine learning models and deep learning models cannot classify the semi-CON state well.

Table III and Table IV respectively show the experimental results of the EEG data recorded by the FP1 and FP2 channels of 13 subjects. By comparing the results in Tables II-IV, we can draw the following conclusions:

1) As the OBA EEG data contains a high level of noises, traditional machine learning and deep learning models including two latest models from the literature, usually cannot perform well; however, meta-learning is a good method to reduce the

TABLE III
EXPERIMENTAL RESULTS OF EEG DATA FROM FP1 CHANNEL

Subject	SVM	RF	CNN	LSTM	CLSTM	MCNN	CLA	ALSTM	Anes-MetaNet
1	80.0(±0.1)	80.5(±0.1)	81.2(±0.2)	87.5(±0.5)	86.4(±1.3)	86.2(±0.9)	88.2(±0.2)	79.1(±0.3)	86.4(±1.0)
2	75.5(±0.1)	69.9(±0.1)	76.3(±0.1)	66.2(±2.9)	70.5(±2.7)	77.6(±0.2)	78.7(±0.1)	74.2(±0.2)	81.3(±0.2)
3	71.6(±0.2)	69.4(±0.1)	68.9(±1.5)	79.8(±0.1)	70.4(±2.2)	78.5(±0.9)	77.2(±0.1)	83.5(±0.5)	78.5(±2.8)
4	88.9(±0.1)	91.1(±0.1)	86.4(±2.7)	90.5(±0.5)	87.9(±1.4)	89.8(±0.8)	92.8(±0.3)	92.4(±0.2)	92.8(±3.3)
5	72.0(±0.4)	72.4(±0.1)	73.6(±1.9)	75.3(±0.1)	72.4(±0.4)	78.7(±0.2)	74.0(±0.5)	76.9(±3.0)	79.3(±1.9)
6	82.2(±0.2)	80.6(±0.1)	81.8(±0.1)	84.9(±0.3)	83.4(±2.1)	84.0(±0.3)	82.4(±0.2)	78.0(±4.6)	90.5(±0.1)
7	73.0(±0.1)	72.5(±0.1)	74.0(±0.1)	76.5(±0.1)	76.6(±0.2)	74.8(±0.3)	77.6(±0.1)	74.5(±0.9)	89.5(±0.8)
8	85.4(±0.6)	83.5(±0.1)	84.0(±1.2)	82.9(±2.4)	55.1(±15.9)	84.0(±0.4)	88.9(±0.8)	83.3(±2.1)	86.8(±0.9)
9	52.1(±0.3)	54.7(±0.3)	54.0(±0.7)	64.5(±1.3)	62.7(±2.1)	63.6(±0.6)	61,0(±0.4)	59.3(±0.4)	71.9(±0.8)
10	58.5(±0.3)	59.0(±0.5)	58.8(±0.8)	49.7(±2.7)	54.3(±0.2)	57.1(±0.4)	55.1(±1.4)	61.3(±2.3)	74.1(±0.8)
11	76.3(±0.2)	77.1(±0.1)	75.3(±0.2)	64.4(±0.2)	66.2(±1.7)	81.7(±1.1)	66.1(±0.2)	67.4(±2.8)	84.7(±2.1)
12	76.4(±0.2)	76.1(±0.1)	75.3(±1.0)	79.1(±1.3)	75.5(±0.5)	74.7(±0.3)	62.4(±0.4)	79.0(±0.2)	83.9(±0.5)
13	76.2(±0.1)	76.2(±0.2)	75.6(±0.1)	75.6(±1.7)	77.2(±0.4)	75.8(±0.3)	79.5(±0.3)	70.8(±7.0)	81.6(±2.9)
AVG	74.4	74.0	74.2	75.1	72.2	77.4	75.6	75.3	83.1

TABLE IV
EXPERIMENTAL RESULTS BASE ON EEG DATA FROM FP2 CHANNEL

Subject	SVM	RF	CNN	LSTM	CLSTM	MCNN	CLA	ALSTM	Anes-MetaNet
1	79.2(±0.4)	76.9(±0.3)	83.5(±0.8)	80.6(±0.6)	84.9(±1.1)	87.5(±0.1)	86.3(±0.1)	79.3(±0.1)	93.3(±3.4)
2	77.1(±0.2)	73.7(±0.2)	76.2(±0.5)	78.4(±1.1)	63.6(±14.6)	72.4(±0.4)	78.4(±0.1)	78.1(±0.5)	81.0(±2.3)
3	74.4(±0.3)	69.0(±0.1)	72.1(±0.1)	76.7(±0.6)	72.6(±3.5)	79.7(±0.5)	76.1(±0.4)	76.5(±0.4)	86.5(±2.6)
4	91.3(±0.2)	92.1(±0.1)	90.9(±0.7)	91.6(±2.4)	93.2(±0.6)	91.2(±0.5)	92.7(±0.4)	87.4(±5.9)	93.5(±1.9)
5	74.9(±0.1)	74.8(±0.1)	74.1(±0.1)	78.3(±0.1)	75.3(±0.3)	78.0(±1.0)	79.7(±0.3)	79.6(±0.3)	80.8(±3.0)
6	81.5(±0.2)	81.7(±0.1)	81.9(±0.2)	82.6(±0.1)	83.2(±0.1)	83.9(±0.3)	82.9(±0.1)	80.7(±1.1)	80.5(±2.8)
7	74.5(±0.3)	74.2(±0.3)	74.0(±0.1)	75.1(±0.1)	73.7(±0.5)	76.1(±0.3)	74.5(±0.5)	74.6(±0.6)	83.1(±5.9)
8	80.7(±1.0)	83.2(±0.4)	79.3(±0.8)	87.3(±0.3)	59.7(±10.1)	84.2(±0.2)	86.7(±0.3)	88.5(±0.4)	64.8(±3.7)
9	49.2(±0.1)	51.0(±0.4)	50.5(±0.2)	65.0(±13.3)	46.7(±3.4)	56.6(±0.4)	52.7(±0.1)	53.8(±0.4)	77.7(±0.6)
10	56.5(±1.6)	52.8(±0.6)	60.6(±0.4)	49.7(±0.2)	47.4(±1.6)	61.8(±2.3)	51.2(±1.8)	50.5(±1.7)	79.2(±2.2)
11	75.2(±0.2)	76.5(±0.2)	75.3(±0.3)	68.1(±1.7)	66.0(±0.3)	76.7(±0.1)	66.5(±0.4)	73.7(±4.3)	67.1(±0.8)
12	76.7(±0.2)	76.0(±0.4)	77.7(±0.1)	78.2(±1.7)	76.7(±1.5)	74.7(±0.6)	79.0(±0.1)	75.5(±1.1)	83.7(±0.6)
13	77.9(±0.1)	78.4(±0.3)	78.0(±0.5)	80.3(±0.3)	79.1(±0.6)	76.5(±1.3)	79.9(±0.4)	72.7(±5.4)	86.3(±0.8)
AVG	74.5	73.8	74.9	76.3	70.9	76.8	75.8	74.6	81.3

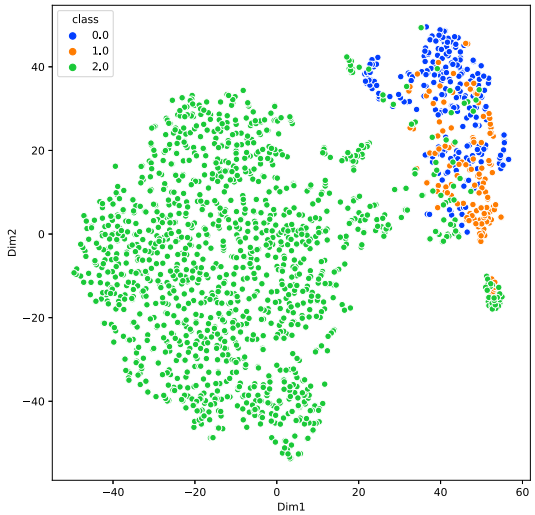

Fig. 14. Visualization of t-SNE maps.

cross-subject noises and improve the classification accuracy by combining with deep learning methods.

2) When using single-channel EEG data, the data of FP1 and FP2 channels are better for the classification usage of brain states than the data of FPZ channel (see the average results of all methods in Tables II-IV). As noted, Anes-MetaNet achieves the highest accuracy for classifying the DoA using FP1 channel of the EEG data.

3) LSTM is easy to cause the instability for the classification results (see the results using LSTM, CLSTM, ALSTM and Anes-MetaNet). Nevertheless, LSTM improve the classification accuracy by comparing the results between using MCNN and Anes-MetaNet.

From the above experimental results, it can be seen that the classification results of CNN and LSTM are better than SVM and RF. However, the result of CLSTM is not as good as CNN or LSTM. The reason is that the classification results using CNN are not good enough (no better or just slightly better than a random guess); thus, adding LSTM makes the prediction further away from the true label given too many wrongly classified samples by CNN. The training of LSTM needs better represented features and predicted labels rendered by CNN. The wrong features and labels lead to the inability of LSTM to train an accurate model. The classification results of MCNN are not only better than CNN in the final accuracy, but also much better than CNN in the classification of the more difficult task (CON vs semi-CON). From the confusion matrix, we can see that MCNN achieves a high accuracy in the classification results of each class, while CNN only has high accuracy in the classification of LOC state and CON state, and the accuracy in the classification of semi-CON state is about 0. Such a phenomenon can be explained that meta-learning and two-stage classification mechanism in MCNN are much helpful to improve the classification capability of CNN, so that the features extracted by the CNN model can more

TABLE V
EXPERIMENTAL RESULTS BASED ON AN HBA EEG DATASET

Subject	SVM	RF	CNN	LSTM	CLSTM	MCNN	CLA	ALSTM	Anes-MetaNet
1	84.8(±0.3)	85.1(±0.4)	84.0(±0.2)	86.1(±1.4)	84.4(±0.2)	84.0(±0.4)	83.3(±0.3)	82.0(±2.8)	88.9(±3.7)
2	90.7(±1.2)	92.5(±0.1)	90.5(±0.5)	93.7(±1.4)	93.7(±0.2)	93.3(±0.1)	94.5(±0.2)	94.1(±0.4)	94.5(±0.2)
3	82.4(±0.5)	80.3(±0.5)	81.2(±0.4)	82.1(±1.5)	80.0(±0.5)	83.8(±0.2)	80.5(±2.7)	80.0(±0.5)	80.5(±2.7)
4	83.3(±0.7)	75.0(±0.2)	82.8(±0.6)	67.9(±3.2)	72.3(±0.2)	82.2(±3.0)	81.3(±4.9)	75.8(±1.2)	81.3(±4.9)
5	94.1(±0.5)	93.5(±0.2)	94.3(±0.2)	90.7(±1.9)	91.0(±1.6)	91.7(±2.2)	93.8(±1.7)	93.7(±2.9)	93.0(±0.6)
6	84.6(±0.1)	84.1(±0.1)	84.7(±0.1)	82.1(±4.1)	87.3(±2.5)	79.6(±1.8)	74.7(±0.3)	87.8(±1.4)	88.5(±0.5)
7	83.8(±0.4)	83.9(±0.1)	81.9(±0.4)	84.7(±1.6)	89.6(±0.6)	84.9(±0.4)	83.4(±1.0)	82.8(±0.2)	87.1(±0.5)
8	90.7(±0.3)	89.8(±0.1)	87.5(±0.2)	92.0(±1.9)	92.7(±1.7)	86.3(±0.1)	92.5(±1.4)	90.1(±0.9)	93.1(±0.2)
9	91.3(±0.4)	91.5(±0.1)	91.8(±1.0)	95.4(±0.4)	95.5(±0.7)	92.0(±1.5)	94.9(±1.5)	93.2(±0.1)	96.3(±0.8)
10	90.8(±0.5)	86.3(±0.1)	87.3(±0.7)	84.6(±0.8)	84.4(±0.9)	85.7(±0.2)	89.3(±1.9)	86.8(±0.1)	89.3(±1.8)
AVG	87.6	86.2	86.6	85.9	87.0	85.1	86.8	86.6	89.5

TABLE VI
BI-CLASSIFICATION EXPERIMENTAL RESULTS BASED ON THE OBA EEG DATASET FROM FP1 CHANNEL

Subject	SVM	RF	CNN	LSTM	CLSTM	MCNN	CLA	ALSTM	Anes-MetaNet
1	84.5(±0.7)	90.6(±0.2)	91.7(±0.8)	92.0(±2.2)	92.3(±1.7)	92.0(±0.1)	95.4(±0.1)	94.5(±0.5)	95.4(±0.3)
2	78.0(±1.4)	73.2(±0.4)	75.5(±2.2)	70.4(±5.3)	68.0(±4.8)	79.3(±0.3)	81.5(±0.1)	67.6(±1.5)	81.8(±2.1)
3	72.5(±0.5)	80.6(±0.4)	81.5(±0.6)	83.3(±1.6)	84.1(±0.5)	80.1(±0.2)	82.0(±0.4)	84.9(±0.5)	84.4(±0.9)
4	88.5(±0.5)	89.7(±0.4)	88.4(±1.0)	86.8(±1.4)	87.8(±1.6)	88.0(±1.0)	90.8(±1.0)	85.9(±0.1)	88.2(±0.2)
5	73.0(±0.1)	83.1(±0.3)	82.3(±0.4)	81.4(±1.6)	83.9(±0.6)	80.8(±1.3)	85.8(±1.2)	84.3(±1.5)	83.5(±1.3)
6	84.7(±0.3)	84.9(±0.2)	87.4(±0.6)	86.1(±1.9)	80.5(±2.4)	89.1(±0.1)	86.4(±1.4)	82.5(±0.1)	93.5(±1.4)
7	76.9(±0.1)	75.3(±0.3)	78.5(±0.5)	76.2(±1.3)	78.1(±0.1)	78.0(±0.4)	77.2(±0.5)	75.7(±1.7)	79.2(±0.6)
8	87.4(±0.4)	77.7(±0.5)	82.5(±3.4)	83.0(±5.6)	79.3(±9.2)	85.8(±1.5)	89.8(±0.4)	90.8(±1.0)	87.2(±0.5)
9	54.9(±0.1)	64.8(±0.1)	65.6(±0.5)	63.6(±0.2)	65.2(±1.2)	66.5(±0.1)	62.6(±0.6)	62.6(±1.1)	65.6(±1.3)
10	59.1(±0.2)	58.6(±0.1)	60.2(±0.5)	62.9(±0.5)	62.3(±2.5)	58.8(±1.4)	59.8(±0.8)	61.4(±1.9)	67.8(±0.5)
11	80.2(±0.1)	75.0(±0.2)	74.3(±0.4)	75.2(±0.1)	73.7(±0.1)	72.5(±0.6)	71.2(±2.6)	75.3(±0.7)	75.5(±2.2)
12	78.2(±0.1)	75.3(±0.2)	76.8(±0.2)	76.4(±1.1)	76.0(±0.2)	74.8(±1.5)	81.0(±1.0)	77.3(±0.1)	78.5(±1.6)
13	80.5(±0.5)	76.9(±0.3)	78.3(±0.2)	82.8(±1.2)	78.8(±1.2)	78.3(±0.3)	81.9(±1.6)	77.8(±0.2)	82.7(±1.6)
AVG	76.8	77.3	78.6	78.4	77.6	78.7	80.4	78.5	81.7

accurately represent the DoA. Compared CLSTM to Anes-MetaNet, accurately representing the input features for LSTM is important for the further feature learning in improving the classification accuracy. Compared CLA and ALSTM models to Anes-MetaNet, the well-performed models in the HBA conditions are not appropriate to estimate the DoA in the OBA conditions.

9) Experiments on an HBA Dataset: As stated above, OBA data is more complex with lower SNR than HBA data. In order to further demonstrate the effectiveness of the proposed model, we apply the Anes-MetaNet to an HBA dataset from Abel *et al.* [39]. This HBA dataset contains 10 subjects with two states, pre-LOC and LOC. Following the above experiments, we also implement a cross-subject validation by using the data of seven subjects as the training dataset, and the data from the remaining subjects as the test dataset. The other settings are the same as those in the above OBA experiments. As noted, this dataset is collected from FP1 channel. The classification results are presented in Table V. As seen, for such a bi-classification problem, SVM performs better than the deep learning based baseline models including the two recent developed models, CLA and ALSTM, from the literature. Despite this, our proposed model still performs slightly better than SVM. Note that since there are only two states in this dataset, the two-stage process in the implementation of Anes-MetaNet is omitted. In addition, for a fair comparison, we conduct another bi-classification experiment with the OBA data collecting from FP1 channel, and display the results in Table VI. As we can observe, Anes-MetaNet still achieves the best performance. An interesting finding is that in Table V, all the classification accuracy is between 85.1% and 89.5% for the HBA dataset, while in Table VI, all the classification accuracy is just between 76.8% and 81.7%. Such a comparison clearly indicates that the data obtained from OBA conditions have much lower SNR than that acquired from HBA conditions, which further demonstrates the meaning of introducing the Anes-MetaNet to the literature.

IV. CONCLUSION

In this study, we propose an Anes-MetaNet to solve the DoA classification problem in office-based environments, in which, studies on the classification of anesthesia EEG data have not been well investigated. The model we develop, provides good performance for the classification of brain states using anesthesia EEG data. The proposed Anes-MetaNet can mitigate the individual differences among subjects. We train the model in a sequential way to reach a better classification accuracy. The t-SNE visualization validates our two-stage training mechanism. The comparison results using an HBA dataset and an OBA dataset, showcase how complex the OBA data is and the necessity of developing the Anes-MetaNet. In the future, a systematic comparison of OBA EEG data and HBA EEG data under different anesthetics need to be conducted with regards to model accuracy and interpretability.

REFERENCES

[1] E. N. Brown, P. L. Purdon, and C. J. Van Dort, "General anesthesia and altered states of arousal: A systems neuroscience analysis," *Annu. Rev. Neurosci.*, vol. 34, no. 1, pp. 601–628, Jul. 2011.

[2] B. A. Fritz *et al.*, "Intraoperative electroencephalogram suppression predicts postoperative delirium," *Anesthesia Analgesia*, vol. 122, no. 1, p. 234, 2016.

[3] P. L. Purdon *et al.*, "Electroencephalogram signatures of loss and recovery of consciousness from propofol," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 12, pp. E1142–E1151, 2013.

[4] S. Chakravarty *et al.*, "Closed-loop control of anesthetic state in non-human primates," *bioRxiv*, 2021, doi: 10.1101/2021.09.12.459958.

[5] C. N. Sessler *et al.*, "The Richmond agitation–sedation scale: Validity and reliability in adult intensive care unit patients," *Amer. J. Respiratory Crit. Care Med.*, vol. 166, no. 10, pp. 1338–1344, 2002.

[6] T. A. Bowdle, "Depth of anesthesia monitoring," *Anesthesiol. Clinics North Amer.*, vol. 24, no. 4, pp. 793–822, 2006.

[7] X.-S. Zhang, R. J. Roy, and E. W. Jensen, "EEG complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1424–1433, Dec. 2001.

[8] F. L. da Silva, "EEG and MEG: Relevance to neuroscience," *Neuron*, vol. 80, no. 5, pp. 1112–1128, 2013.

[9] O. Akeju *et al.*, "Electroencephalogram signatures of ketamine anesthesia-induced unconsciousness," *Clin. Neurophysiol.*, vol. 127, no. 6, pp. 2414–2422, Jun. 2016.

[10] H. W. Shin *et al.*, "Monitoring of anesthetic depth and EEG band power using phase lag entropy during propofol anesthesia," *BMC Anesthesiol.*, vol. 20, no. 1, pp. 1–10, Dec. 2020.

[11] R. Shalbaf, H. Behnam, and H. J. Moghadam, "Monitoring depth of anesthesia using combination of EEG measure and hemodynamic variables," *Cognit. Neurodyn.*, vol. 9, no. 1, pp. 41–51, 2015.

[12] M. Jospin *et al.*, "Detrended fluctuation analysis of EEG as a measure of depth of anesthesia," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 840–846, May 2007.

[13] T. Nguyen-Ky, P. F. Wen, and Y. Li, "Consciousness and depth of anesthesia assessment based on Bayesian analysis of EEG signals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1488–1498, Jun. 2013.

[14] R. Shalbaf, H. Behnam, J. W. Sleigh, A. Steyn-Ross, and L. J. Voss, "Monitoring the depth of anesthesia using entropy features and an artificial neural network," *J. Neurosci. Methods*, vol. 218, no. 1, pp. 17–24, 2013.

[15] M. Peker, A. Arslan, B. Şen, F. V. Çelebi, and A. But, "A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+ RF)," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, 2015, pp. 1–8.

[16] M. Jahanseir, K. Setarehdan, and S. Momenzadeh, "Estimation of the depth of anesthesia by using a multioutput least-square support vector regression," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 26, no. 6, pp. 2792–2801, 2018.

[17] W. Saadeh, F. H. Khan, and M. A. B. Altaf, "Design and implementation of a machine learning based EEG processor for accurate estimation of depth of anesthesia," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 4, pp. 658–669, Aug. 2019.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.

[19] S. Zhou, X. Li, Y. Chen, S. T. Chandrasekaran, and A. Sanyal, "Temporal-coded deep spiking neural network with easy training and robust performance," 2019, *arXiv:1909.10837*.

[20] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[21] Y. Fujisawa *et al.*, "Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis," *Brit. J. Dermatol.*, vol. 180, no. 2, pp. 373–381, Feb. 2019.

[22] M. Jiao, D. Wang, Y. Yang, and F. Liu, "More intelligent and robust estimation of battery state-of-charge with an improved regularized extreme learning machine," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104407.

[23] Y. Park, S.-H. Han, W. Byun, J.-H. Kim, H.-C. Lee, and S.-J. Kim, "A real-time depth of anesthesia monitoring system based on deep neural network with large EDO tolerant EEG analog front-end," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 4, pp. 825–837, Aug. 2020.

[24] R. Li, Q. Wu, J. Liu, Q. Wu, C. Li, and Q. Zhao, "Monitoring depth of anesthesia based on hybrid features and recurrent neural network," *Frontiers Neurosci.*, vol. 14, p. 26, Feb. 2020.

[25] S. Afshar, R. Boostani, and S. Sanei, "A combinatorial deep learning structure for precise depth of anesthesia estimation from EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3408–3415, Sep. 2021.

[26] T. Zoughi and R. Boostani, "Presenting a combinatorial feature to estimate depth of anesthesia," *Int. J. Signal Process.*, vol. 6, no. 2, pp. 10–14, 2010.

[27] S. Afrasiabi, R. Boostani, S. Koochaki, and F. Zand, "Presenting an effective EEG-based index to monitor the depth of anesthesia," in *Proc. 16th CSI Int. Symp. Artif. Intell. Signal Process. (AISP)*, May 2012, pp. 557–562.

[28] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, 2007.

[29] H. Cecotti and A. Graeser, "Convolutional neural network with embedded Fourier transform for EEG classification," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[30] B. Babadi and E. N. Brown, "A review of multitaper spectral analysis," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1555–1564, May 2014.

[31] B. Noureddin, P. D. Lawrence, and G. E. Birch, "Online removal of eye movement and blink EEG artifacts using a high-speed eye tracker," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2103–2110, Aug. 2012.

[32] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2017.

[33] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[34] Z. Jiao, X. Gao, Y. Wang, J. Li, and H. Xu, "Deep convolutional neural networks for mental load classification based on EEG data," *Pattern Recognit.*, vol. 76, pp. 582–595, Apr. 2018.

[35] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artif. Intell. Rev.*, vol. 18, no. 2, pp. 77–95, Jun. 2002.

[36] J. Vanschoren, "Meta-learning: A survey," 2018, *arXiv:1810.03548*.

[37] S. B. Nagaraj, P. L. Purdon, F. Shapiro, and B. Westover, "Electroencephalogram monitoring of depth of anesthesia during office-based anesthesia," *bioRxiv*, 2020, doi: 10.1101/2020.10.27.356592.

[38] H. Bokil, P. Andrews, J. E. Kulkarni, S. Mehta, and P. P. Mitra, "Chronux: A platform for analyzing neural signals," *J. Neurosci. Methods*, vol. 192, no. 1, pp. 146–151, Sep. 2010.

[39] J. Abel *et al.*, "Multitaper spectra recorded during GABAergic anesthetic unconsciousness," *PhysioNet*, 2021, doi: 10.13026/m792-h077.