

Received February 16, 2022, accepted March 3, 2022, date of publication March 10, 2022, date of current version March 21, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3158348

MS4PS: A Mentor-Student Architecture for Patient-Specific Seizure Detection With Combination of Transfer Learning and Active Learning

SHUN MA¹, HAOJIE LIU^{1,2}, XIAOGANG ZHU³, YUFENG FAN¹, CAIXIA SU¹,
AND YONGFENG CAO¹

¹School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550003, China

²Huzhou Institute, Zhejiang University, Huzhou 313000, China

³Center of Neuroelectrophysiology, Affiliated Hospital of Guizhou Medical University, Guiyang 550031, China

Corresponding author: Yongfeng Cao (cyfeis@whu.edu.cn)

This work was supported by the Science and Technology Foundation of Guizhou Province under Grant [2018]1114.

ABSTRACT Privacy protection, high labeling cost, and varying characteristics of seizures among patients and at different times are the main obstacles to building seizure detection models. Considering these issues, we propose a novel Mentor-Student architecture for Patient-Specific seizure detection (MS4PS). It contains a new method of knowledge transferring called mentor-select-for-student, which exploits the knowledge of a mentor model by using this model to select data for training a student model, making it possible to avoid transferring patient data and the negative influence of transferring parameters/structures of pre-trained models. It also contains a new method of active learning, which uses both an experienced mentor model and a quick-learning student model to select high-quality samples for doctors to label. Each of the two models is coupled with a particular sample selection strategy that combines uncertainty/certainty and the distance between the unlabeled samples and labeled seizure samples. The proposed method can quickly train a suitable detector for a patient at his/her first epilepsy diagnosis with the help of: (1) an experienced mentor model that chooses the most category-certain electroencephalography (EEG) data segments; (2) a student model (detector itself) that chooses the most category-uncertain EEG data segments; (3) doctors who label these data segments selected by both the mentor model and student model. By replacing or improving the mentor model and refining the historical models of patients when they come next time, the MS4PS system can be sustainably promoted. The proposed method is tested on the CHB-MIT and NEO datasets, and the results demonstrate its effectiveness and efficiency.

INDEX TERMS Active learning, epilepsy, mentor-student architecture, patient-specific, transfer learning.

I. INTRODUCTION

Epilepsy is a chronic neurological disorder of the brain that causes death in many patients every year. If the symptoms could be detected in a timely manner and patients receive appropriate treatment, they are likely to live a normal life with a high probability. Currently, many devices can record the electroencephalography (EEG) of subjects, therefore, doctors can check whether they have epilepsy and analyze their

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

conditions using EEG. However, checking EEG signals is a time-consuming and challenging task for doctors.

To relieve the burden on doctors and improve efficiency, many machine learning-based methods have been proposed to build automatic seizure detectors for EEG, such as the dictionary-learning-based method [1], SVM-based method [2], GMM-based method [3].

In recent years, with the surge in the development of deep learning [4]–[6], many researchers [7]–[9] have attempted to employ deep learning methods to train patient-independent seizure detectors, and their works have shown great improvement and potential. However, there are at least three main

obstacles to building deep-learning-based seizure detectors: privacy protection of patient data, high labeling costs, and the varying characteristics of epilepsy EEG data.

The privacy protection of patient data and high labeling costs hinder the acquisition of sufficient training data. To reduce the need for labeled data and improve the quality of labeled data, deep transfer learning [10] and deep active learning (AL) [11] methods have been employed in patient-independent seizure detection [12]–[15] and have shown good performance.

The seizure characteristics vary among patients and even at different times for the same patient. This makes it difficult for patient-independent seizure-detection models to fit specific patients. Some research works [16]–[18] have been done and proved that building a patient-specific seizure detector for each patient seems more suitable than building a patient-independent seizure detector for all patients. Considering that it is still difficult to train a suitable detector for a specific patient who has no historically labeled data (e.g., when he/she sees a doctor for the first time) or does not have enough historically labeled data, it is reasonable to employ transfer learning. However, commonly used transfer learning methods can lead to negative transfers [18], [19] when the distributions of the source and target domains are different.

In light of these obstacles and solutions, we propose a novel Mentor-Student architecture for Patient-Specific seizure detection (MS4PS) that combines transfer learning and active learning. To the best of our knowledge, no researchers have proposed this method for patient-specific seizure detection.

Instead of transferring patients' data or copying a pre-trained model's parameters/structures and then fine-tuning [12]–[14], this method exploits the knowledge of a mentor model by using the mentor model to select data for the learning of the student model. Transferring knowledge through mentor-select-for-student makes it possible to avoid transferring patient data and the negative influence of transferring parameters/structures (according to [18], [19] and our experiments, transferring parameters/structures could be of little help or even hinder the learning process in patient-specific cases).

This method contains a new method of active learning, which uses both an experienced mentor model and a quick-learning student model to select samples for doctors to label. Each of the two models is coupled with a particular sample selection strategy that combines the uncertainty/certainty and the distance between the unlabeled samples and labeled seizure samples. The new active learning method has the potential to significantly reduce doctors' burden of searching for and labeling seizures in severely ill-balanced EEG data.

Overall, the proposed method could quickly train a good student model for a specific patient at his/her first epilepsy diagnosis with the help of: (1) an experienced mentor model that chooses the most category-certain EEG data segments; (2) a student model itself that chooses the most category-uncertain EEG data segments; (3) doctors who label

these data segments selected by both the mentor model and student model. These trained student models for specific patients would be saved, reused, and further trained when these patients have an epilepsy diagnosis next time. Moreover, the mentor model can be replaced and improved when sufficient labeled data are collected. This means the mentor-student system could work better and better.

The main contributions of this paper are as follows:

- 1) A novel method that combines transfer learning and active learning with a mentor-student architecture for patient-specific epilepsy detection is proposed. This method makes training a patient-specific seizure detector easy and quick, and the mentor-student system could be sustainably promoted.
- 2) A new knowledge transferring method, which uses a mentor model to select data for student-model's learning, is proposed for mentor-student system. It protects patients' privacy by not transferring their data and avoids the negative influences of transferring model's parameters/structures.
- 3) A novel active learning method, which uses both an experienced mentor model and a quick-learning student model to select samples for doctors to label, is proposed for the mentor-student system to relieve doctors' burden of searching and labeling seizures in severely ill-balanced EEG data.

On the CHB-MIT [20] and NEO [21] datasets, 39 different methods (including MS4PS methods, two state-of-art transfer learning methods [12]–[14], two classic active learning methods, and many other methods) are compared. It should be known that the two transfer learning methods (denoted in this paper by $S_M(\text{TF})\text{-R}$ and $S_M(\text{TAL})\text{-R}$) are not exactly the same as methods in [12]–[14] because to make the comparison among 39 methods fair, we force them all to use our model and dataset. The $S_M(\text{TF})\text{-R}$ method replaces the source-model's classifier with a task-specific classifier (in [14], it is a SVM layer; in [12], [13], and in this study, it is fully-connected and softmax layers) and then fine-tunes this task-specific classifier; the $S_M(\text{TAL})\text{-R}$ method does the same but fine-tunes all the model's layers. The two classic active learning methods are random and maximum entropy (denoted in this paper by $S(\text{TAL})\text{-R}$ and $S(\text{TAL})\text{-U}$). The top 5 of the 39 on Avg_F1 (average F1-score over given budgets and patients) and the top 1 of the 39 on Avg_seizures (average number of seizure samples selected over given budgets and patients) are all MS4PS methods. The best MS4PS method surpasses the two transfer learning methods, random and maximum entropy by 30.5%, 26.7%, 19.3% and 9% on Avg_F1, and by 39, 39, 39, and 5 on Avg_seizures. These results demonstrate the feasibility of MS4PS for patient-specific seizure detection problems.

II. RELATED WORKS

A. PATIENT-SPECIFIC SEIZURE DETECTION

There are many works that study patient-independent seizure detection, for example, [1] employs a real-time method

based on dictionary learning and sparse representation; [22] employs a method based on a deep neural network that combines a seizure representation part to eliminate inter-subject noise and an attention mechanism part to enhance interpretability.

To adapt to the variation in seizure characteristics among patients and at different times, some patient-specific seizure detection algorithms have been proposed. Most of them suppose having enough labeled EEG data and try to improve the performance of patient-specific detectors by using elaborate models and features, for example, [2] uses a group of SVMs and features extracted through empirical mode decomposition (EMD) and common space patterns (CSP); [17] uses a voting SVM system and features containing both the temporal-domain and spectral-domain information of EEG; [23] uses a RVM model and the harmonic multiresolution and self-similarity-based fractal features from EEG data; and [16] builds a predictor based on spatio-temporal-spectral hierarchical GCN with an active pre-ictal interval learning scheme (STS-HGCN-AL).

In contrast, our work supposes not having enough labeled EEG data and tries to solve it with MS4PS that combines transfer learning and active learning. To the best of our knowledge, only a few works have considered the problem of not enough labeled EEG data and try to solve it, for example, [18] employs a probabilistic framework for training a personalized neonatal seizure detector based on transfer learning and semi-supervised learning. It uses a source-task model (a patient-independent seizure detector) as the initial target-task model (personalized seizure detector) and labels specific patients' EEG data with the source-task model when doctors are absent. This work is the most relevant to ours. However, our work transfers knowledge through mentor-select-for-student rather than copying and refining model parameters/structures, and we use the pre-trained (patient-independent) model to select samples for active learning rather than to provide a hypothesis label for semi-supervised learning.

B. ACTIVE LEARNING

Active learning aims to alleviate the need for plenty of labeled data in training a model by selecting the most informative data for labeling [11]. The most commonly used active learning methods are uncertainty-based methods and representativeness-based methods.

The uncertainty-based methods select the most uncertain samples for the current model, making a quick model learning process. The representativeness-based methods select data that well represent the overall input patterns of unlabeled data, reducing the redundancy of selected data. Although they can achieve good results in many areas, they have some weak points: uncertainty-based methods rely on a not-lousy model and would select redundant samples [24]; the representativeness-based methods need more labeled data than the uncertainty-based ones to make a model converge. Therefore, methods that combine uncertainty and

representativeness usually yield better results [24], [25]. Combining different active learning strategies is popular and valuable, so the combinations of uncertainty, certainty and representativeness (measured by distance) are considered in our work.

Few works have used active learning methods in seizure detection. The work most similar to ours is the cost-sensitive deep active learning method [15]. It integrates uncertainty, misclassification cost and diversity to construct a utility function for the samples-selection strategy in the labeling process and develops a new generic double-deep neural network (double-DNN) to obtain utility. The critical difference between our work and the work in [15] is that our work uses both models to select samples instead of using only one model for selecting samples and the other for obtaining cost-sensitive utility.

C. TRANSFER LEARNING

Transfer learning aims to reduce the need for labeled data in the target domain by applying source domain data and source domain model to the target domain [10].

The most commonly used transfer learning methods in EEG data analysis are domain adaption, improved CSP algorithms, deep neural network-based (DNN-based) algorithms, and subspace learning, according to [19].

Our work employs DNN-based algorithms. There are many related works, for example, [12] uses ImageNet dataset to pre-train googlenet, resnet101 and vgg19 as source domain models, then replaces each source-domain-model's classifier layer with a task-specific classifier and uses EEG data to fine-tune only the task-specific classifier; [13] does almost the same as [12], but pre-trains different source domain models; [14] uses ImageNet dataset to pre-train ten different source domain models, then replaces each source-domain-model's classifier layer with two different kinds of task-specific classifier, and at last uses EEG data to fine-tune all the layers of the model or only fine-tune the task-specific classifier.

These works focus on reusing the feature extractor pre-trained with source domain data in the target model. Meanwhile, they suppose that the labeled target domain data are sufficient for fine-tuning. Unlike them, our work trains a target model from scratch with the assistance of the pre-trained source domain model and supposes that the labeled data are not sufficient and should be accumulated through active learning.

III. METHOD

As shown in Fig. 1, the student model is a personalized model for a patient. The main process for training and using this model is as follows:

Firstly, the unlabeled EEG data segments of a specific patient are added into the unlabeled data pool after data processing (re-referencing and splitting), and if there are, the patient's historically labeled data and student model are loaded from the Database and Modelbase, respectively;

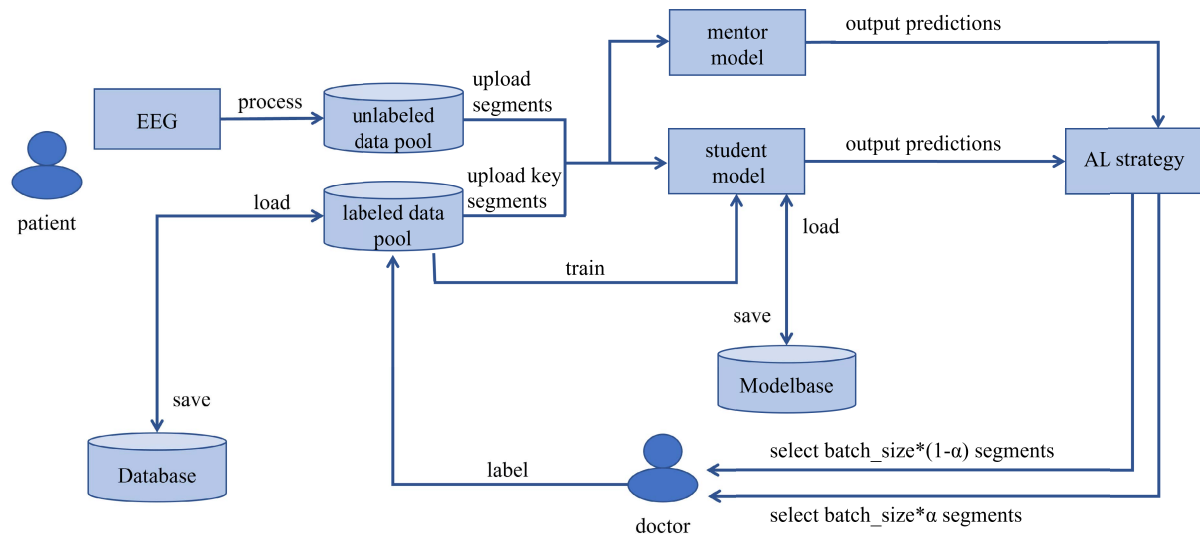


FIGURE 1. The structure of MS4PS.

Secondly, the segments from the unlabeled data pool and the key segments (labeled seizure segments of this patient) from the labeled data pool go through both the mentor model and student model to get predictions; Thirdly, a batch of EEG data segments is selected with specific active learning (AL) strategies and sent to a doctor for labeling. Among those, $\alpha * batch_size$ segments are selected basing on predictions from the mentor model, and $(1 - \alpha) * batch_size$ segments basing on predictions from the student model; Fourthly, the doctor labels these selected EEG data segments and adds them into the labeled data pool; Fifthly, all data in the labeled data pool are used for training the student model, meanwhile, updating the α according to (8). The process (including the second, third, fourth and fifth steps) keeps iterating until running out of the budget, saving the labeled data and trained student model into the Database and Modelbase, respectively. The pseudocode of the MS4PS method is presented in Algorithm 1.

When a patient comes for the first-time checking, there are no historically labeled data in the Database and trained student model in the Modelbase for him/her. However, when the patient comes for the second and the n th-time checking, there are historically labeled data in the Database and a trained student model in the Modelbase for him/her. These historical data and model can be reloaded to further fine-tune the student model and assist doctors' diagnosis.

In the following, we will describe in detail the mentor and student models, active learning strategies, knowledge transferring from mentors to students, and how to calculate the parameter α .

A. MENTOR MODEL AND STUDENT MODEL

1) MENTOR MODEL

A mentor model carries accumulated historical knowledge from the EEG data of the patients. It can be obtained from other institutions or trained with the accumulated EEG data of

Algorithm 1 MS4PS Method

Input: budget of active learning $Budget$; unlabeled data pool D_0^u ; labeled data pool D_0^l ; Mentor model M ; Student model S_0 ; AL strategy for Mentor model f_1 ; AL strategy for Student model f_2 ; $\alpha \leftarrow 0.5$; $batch_size$

Output: trained Student model S_T ($T = Budget$); labeled data pool D_T^l ($T = Budget$)

- 1: **for** $t = 0$ to $Budget$ **do**
- 2: use f_1 to select $\alpha * batch_size$ samples from D_t^u .
- 3: use f_2 to select $(1 - \alpha) * batch_size$ samples from D_t^u .
- 4: label the selected samples (denote them with D_t^s).
- 5: $D_{t+1}^l = D_t^l + D_t^s$
- 6: $D_{t+1}^u = D_t^u - D_t^s$
- 7: train S_t with D_{t+1}^l to S_{t+1}
- 8: calculate α using (8)
- 9: **end for**

this institution. It can be any useful model, and in this paper, we use a deep Convolutional Neural Network (CNN) trained with the NEO dataset [21] as the mentor model.

2) STUDENT MODEL

For each patient, a student model should be built to capture his/her personalized characteristics. When a patient comes for the first-time checking, his/her specific student model must be built from scratch or from a general model for all patients. When the patient comes for the second and the n th time, his/her historical student model could be reloaded and further fine-tuned. A student model could be any useful model, and in this paper, we use a deep CNN as the student model.

B. ACTIVE LEARNING STRATEGY

All active learning strategies try to select the most informative samples for labeling. We name these strategies with their

measurements for informativeness and these used in this paper are as follows.

1) UNCERTAINTY STRATEGY

Here the uncertainty of each sample x for the current model is measured using its entropy as

$$Uncertainty(x) = - \sum_{i=0}^1 p(c_i|x) * \log[p(c_i|x)], \quad (1)$$

where c_0 is the normal category, and c_1 is the seizure category in the case of seizure detection.

The $p(c_i|x)$ can be calculated by softmax, as

$$p(c_i|x) = \frac{e^{z_i}}{\sum_{j=0}^1 e^{z_j}}, \quad i = 0, 1, \quad (2)$$

where z_0 and z_1 are the outputs of the last fully-connected layer of the model in this paper.

The uncertainty strategy selects the samples that are the most uncertain or closest to the classification hyperplane of the current model, as shown in Fig.2 (red dotted circle). The uncertainty measurement can be calculated with either a student model or a mentor model, and we intuitively think that the model with a student model will be more useful for MS4PS.

2) CERTAINTY STRATEGY

In contrast to the uncertainty measurement, minus entropy is used to measure the certainty of each sample x for the current model, as

$$Certainty(x) = \sum_{i=0}^1 p(c_i|x) * \log[p(c_i|x)]. \quad (3)$$

3) DISTANCE STRATEGY

It is well known that EEG samples of epilepsy are ill-balanced, and seizure samples are more critical than normal samples for model learning and doctors to analyze patients' conditions. To select more seizure samples than normal, the distance strategy employs Kullback-Leibler divergence to measure the distance between EEG samples and seizure samples already labeled, and then selects samples of the minimum distance value, as shown in Fig.2 (blue dotted circle). The distance is defined as

$$D(x||DX) = \sum_{i=0}^1 p(c_i|x) \log\left[\frac{p(c_i|x)}{q(c_i|DX)}\right], \quad (4)$$

where $p(c_i|x)$ is the prediction probability of sample x belonging to class c_i . And $q(c_i|DX)$ can be calculated as

$$q(c_i|DX) = \frac{1}{|DX|} \sum_{x \in DX} p(c_i|x), \quad i = 0, 1, \quad (5)$$

where DX represents all seizure samples in the labeled data pool and $|DX|$ is the number of labeled seizure samples.

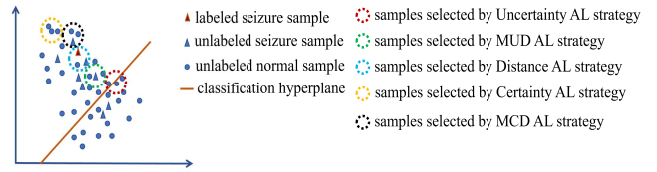


FIGURE 2. Schematic diagram of sample selection with different AL strategies. The MUD AL strategy tends to select samples that are close to both the labeled seizure samples and the classification hyperplane. The MCD AL strategy tends to select samples that are close to the labeled seizure samples and away from the classification hyperplane.

4) MIXTURE OF UNCERTAINTY/CERTAINTY STRATEGY AND DISTANCE STRATEGY

The uncertainty/certainty measurement and distance measurement can be mixed to obtain a better one as

$$MUD(x) = \gamma Uncertainty(x) + (1 - \gamma)D(x||DX) \quad (6)$$

or

$$MCD(x) = \gamma Certainty(x) + (1 - \gamma)D(x||DX), \quad (7)$$

where $\gamma \in [0, 1]$ is an empirical parameter, and we set it to 0.5 in this paper.

C. KNOWLEDGE TRANSFERRING AND PARAMETER α

Based on the mentor-student architecture, we propose a different way of knowledge transferring named mentor-select-for-student, which exploits a mentor model's knowledge by using the mentor model to select data for student-model's learning and train a student model from scratch instead of a model transferred from a mentor model.

Transferring knowledge through mentor-select-for-student (not the patients' data or model parameters/structures) makes it possible to protect the patients' privacy and avoid the negative influence of transferring parameters/structures. Institutions that own superior knowledge can safely distribute their models to help others.

Intuitively, the mentor model should select the most certain samples for training a student model, just like a mentor teaches students with his/her most certain knowledge, and the student model should select the most uncertain samples for training itself, just like a student studies the knowledge that confuses him/her for capability improvement.

In our method, a student model learns from the data chosen by a mentor model and itself, with the supervision of a doctor. The influence of the mentor model would be less and less with the performance of the student model being better and better. It is controlled by parameter α as

$$\alpha = \begin{cases} \beta & N(seizure) = 0 \\ \frac{N(seizure, M, t)}{N(seizure, M, t) + N(seizure, S, t)} & N(seizure)! = 0, \end{cases} \quad (8)$$

where $N(seizure, M, t)$ is the number of seizure samples selected by mentor model at budget t , $N(seizure, S, t)$ is the number of seizure samples selected by student model at budget t , $N(seizure) = N(seizure, M, t) +$

TABLE 1. Abbreviations.

Abbrev.	description
	models for sample selection
M	the Mentor model
S	the Student model
S_M	the Student model transferred from Mentor model
	AL strategies
R	Random active learning strategy
U	Uncertainty active learning strategy
C	Certainty active learning strategy
MUD	Mixture of Uncertainty and Distance active learning strategy
MCD	Mixture of Certainty and Distance active learning strategy
	training methods
TF	only Training the Fully-connected layers
TAL	Training All of the Layers

$N(\text{seizure}, S, t)$, and β is an empirical parameter for the sake of $N(\text{seizure}) = 0$.

IV. EXPERIMENTS

To verify the performance of MS4PS, we design experiments to simulate two kinds of scenarios:

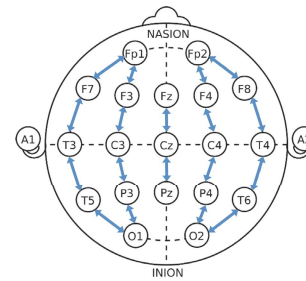
- 1) patients come for the first-time checking, in which their personalized models must be trained from scratch.
- 2) patients come for the second-time checking, in which historical models and labeled data could be used for them.

In each experiment, we try different settings to compare and find the best one. Each setting is named with a three-part pattern, “selection model(training method)-AL strategy”. For simplicity and clarity, some abbreviations are used and all of them are shown in Table1. For example, M-MCD&S(TAL)-MUD means such a setting: using both a mentor model and a student model for selecting samples, MCD AL strategy for mentor model and MUD AL strategy for the student model, and training all of the layers of student model; $S_M(\text{TF})$ -R means that using the student model, which loads the parameters of mentor model for sample selection, R AL strategy for this model, and only training its fully-connected layers.

A. DATA

1) DATA DESCRIPTION

The EEG datasets of CHB-MIT [20] and NEO [21] are used in this paper. Both of them comply with the international 10-20 system of EEG electrode positions and are sampled with 256hz. The NEO dataset is labeled by three experts, therefore, there are disagreements. The CHB-MIT dataset has definite labels. In CHB-MIT, there are 24 folders for 23 patients (the chb01 and chb21 are from the same patient. Here we regard them as different patients), and each folder contains many.edf. In NEO, there are only one.edf for each patient. In this paper,

**FIGURE 3. The structure of EEG channels. [21]****TABLE 2. Number of data segments of CHB-MIT.**

Patient Id	seizure	normal	Patient Id	seizure	normal
chb01	191	14549	chb13	217	11455
chb02	73	12582	chb14	52	9337
chb03	149	56136	chb15	909	14191
chb04	172	56136	chb16	3	6827
chb05	258	13979	chb17	134	7531
chb06	35	24001	chb18	134	12790
chb07	150	24104	chb19	105	10747
chb08	439	7105	chb20	112	9900
chb09	121	24400	chb21	45	10365
chb10	194	17959	chb22	90	11137
chb11	391	12082	chb23	182	9512
chb12	566	8358	chb24	157	3903

CHB-MIT and NEO are used for training the student model and mentor model, respectively.

2) DATA PROCESSING

Firstly, the data of CHB-MIT and NEO are re-referenced into the same channels: Fp1-F7, F7-T3, T3-T5, T5-O1, Fp1-F3, F3-C3, C3-P3, P3-O1, Fz-Cz, Cz-Pz, Fp2-F4, F4-C4, C4-P4, P4-O2, Fp2-F8, F8-T4, T4-T6, and T6-O2, as shown in Fig.3. Secondly, the NEO dataset is relabeled with such rule: for each duration of EEG data, if two or three experts annotated it as seizure, it would be annotated with the seizure-category label and otherwise annotated with the normal-category label. Then, each.edf of CHB-MIT and NEO is split into seizure durations and normal durations. Finally, each duration of EEG data is split into 10-seconds segments. Seizure duration is split with an overlapping of 8 seconds and normal duration with no overlapping. The statistics of segments are shown in Table2 and Table3.

B. STUDENT MODEL AND MENTOR MODEL

The same model structure (Fig.5) is used for both the student model and the mentor model.

1) MODEL INPUT

In this paper, the raw data format is chosen as the input of the model, although there are other commonly used formats such as frequency features [26] and spectrograms [27]. For each

TABLE 3. Number of data segments of NEO.

seizure	normal
4842	30115

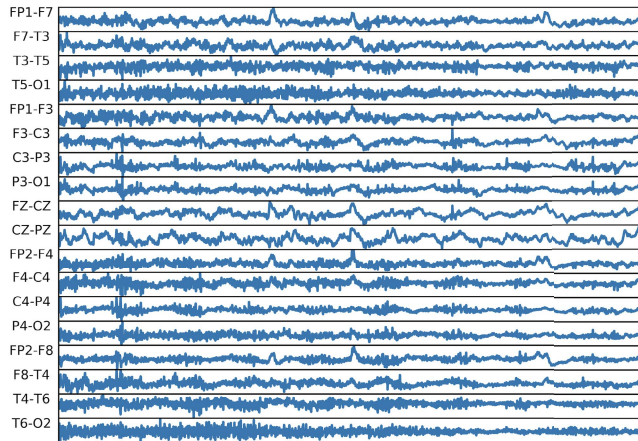


FIGURE 4. Model input schematic.

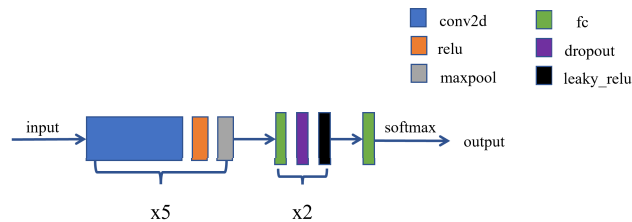


FIGURE 5. Deep neural network architecture.

EEG segment, the 18 channels of 1-D time-domain signals are stacked in a specific order to form a 2-D matrix (Fig.4) which contains the information of both the time and spatial domains. And as the input of the model used in this paper, the matrix has a shape of 18×2560 .

2) MODEL STRUCTURE

To effectively extract the features in a 2-D input format, a deep CNN [28] is used. It consists of 5 convolutional blocks, two fully-connected blocks, and a softmax block, as shown in Fig.5. The details are provided in the Appendix.

C. HOW TO GET MENTOR MODEL

A deep CNN, as shown in Fig.5, is trained with the NEO [21] dataset to be the mentor model. All seizure segments and normal segments that are randomly sampled and twice as many as the seizure segments are used for training. The model is trained until the performance could not be improved. The Adam algorithm is chosen as the optimization algorithm. The learning rate is set to 0.0001. And the cross-entropy function is selected as the loss function.

D. EVALUATION CRITERIA

Two different performance metrics that cover different demands are considered here.

1) F1-SCORE

If a relative adequate test dataset can be obtained for each patient, as for the selected patients in this paper from CHB-MIT, the F1-score is a good metric for evaluating the performance of a student model. The F1-score is a commonly used metric that well balances Recall and Precision as the harmonic mean of them, and is defined as

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision}, \tag{9}$$

where Recall = $TP/(TP+FN)$ and Precision = $TP/(TP+FP)$, TP (true positive) is the number of segments correctly detected as the seizure class, FN (false negative) is the number of segments incorrectly detected as the normal class, TN (true negative) is the number of segments correctly detected as the normal class, and FP (false positive) is the number of segments incorrectly detected as the seizure class.

2) #seizure/B

In practice, when training a patient-specific model for the one who comes for checking, it is hard to get a test dataset for calculating F1-score on it. In this case, the number of seizure samples detected in a given Budget B ($\#seizure/B$) could be a good metric to evaluate the performance of a student model. It is available during the process of a doctor’s diagnosis. Meanwhile, it is intelligible and friendly to the doctor. It could be defined as

$$\#seizure/B = \sum_{k=1}^B (\#seizure)_k, \tag{10}$$

where $(\#seizure)_k$ is the number of seizure segments detected in the kth AL loop. In this paper, B is set to 10 for the experiments of the first-time checking and to 3 for the experiments of the second-time checking.

3) AVERAGE METRICS

To alleviate the bias of performance caused by random factors, the average metrics will be employed as

$$Avg_F1 = \frac{1}{T * P * B} \sum_{i=1}^T \sum_{j=1}^P \sum_{k=1}^B F1 - score_{ijk} \tag{11}$$

and

$$Avg_seizures = \frac{1}{T * P} \sum_{i=1}^T \sum_{j=1}^P \sum_{k=1}^B (\#seizure)_{ijk}, \tag{12}$$

where T is repeat times of experiment (T = 5), P is the number of patients (P = 23 or P = 1), and B is the Budget (B = 10 for the first-time checking and B = 3 for the second-time checking).

E. THE FIRST TIME CHECKING

In this section, experiments are assumed to be in such a scenario: patients come for the first-time checking, in which their personalized models must be trained from scratch.

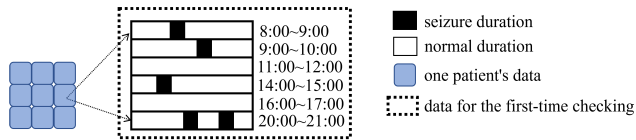


FIGURE 6. Data for the first-time checking.

1) TARGET

The experiments are designed to answer the following questions:

- 1) Comparing with training a student model from scratch, could the common knowledge transferring method, copying and refining parameters of a pre-trained model, bring improvement for patient specific seizure detection?
- 2) Of the many different AL methods composed by coupling different models (such as S, S_M , M, and their combination) with element sample selection strategies (as described in Section III.B), which is the best for MS4PS?
- 3) Does the method of transferring knowledge through mentor-select-for-student work well in training good patient-specific seizure detectors?
- 4) Is the distance strategy (proposed in Section III.B) helpful in MS4PS?

2) DATA PREPARING

The CHB-MIT dataset is used for training the student model. Because chb16 has too few seizure segments (see Table 2), it is ignored, and so at last, there are 23 patients' data for the experiments. As shown in Fig. 6, for each patient, all segments are used. In each experiment, the segments of each patient are shuffled and divided into a training dataset and a testing dataset with 80% and 20% of the total, respectively.

3) STUDENT MODELS

There are two ways to obtain an initial student model: using an empty model (denoted as S) and transferring from other pre-trained models such as mentor model (denoted as S_M). For the empty model S, we train all its layers (denoted as S(TAL)). For S_M , we fine-tune all its layers (denoted as S_M (TAL)) or fine-tune only the fully-connected layers (denoted as S_M (TF)).

4) ACTIVE LEARNING METHODS

Five different element sample selection strategies are chosen here, and they are Random active learning strategy (R), Uncertainty active learning strategy (U), Certainty active learning strategy (C), Mixture of Uncertainty and Distance active learning strategy (MUD) and Mixture of Certainty and Distance active learning strategy (MCD).

We select five different models (S(TAL), S_M (TAL), S_M (TF), M& S_M (TAL), and M&S(TAL)) to couple with the above element strategies for sample selection, getting dozens of different AL methods totally.

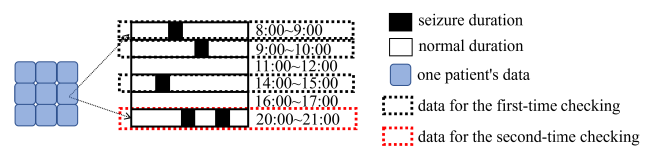


FIGURE 7. Data for the second-time checking. For each patient, only the EEG data containing seizure are used and they are chronologically divided into two groups. The first group (labeled by black dot rectangles) is used for simulating the first-time checking, and the second (labeled by red dot rectangles) for simulating the second-time checking.

5) OTHER SETTINGS

The experiments are repeated 5 times to alleviate the influence of random factors. In each experiment, the student model is randomly initialized, Budget is set to 10, batch_size is set to 32, training times (epoch) is set to 10, and the learning rate is set to 0.0001. The Adam algorithm is chosen as the optimization algorithm and cross-entropy as the loss function.

F. THE SECOND TIME CHECKING

In this section, experiments are assumed to be in such a scenario: patients come for the second-time checking, in which historical models and labeled data could be used for them.

1) TARGET

The experiments are designed to answer the following:

- 1) Does fine-tuning the historical model work better than retraining a new model from scratch for MS4PS?
- 2) Of the two methods, reusing the historically labeled segments and not using these segments, which is better for MS4PS?
- 3) Could the student model get sustained promotion through MS4PS, when patients come more times and his/her data are accumulated more and more?

2) DATA PREPARING

The CHB-MIT dataset of 23 patients (except chb16) is used here. For each patient, only the EEG data containing seizure are used, and they are chronologically divided into two groups (see Fig. 7): the first one is used for simulating the first-time checking (with the same setting as E. THE FIRST TIME CHECKING) to get a historical student model and labeled data; and the second for simulating the second-time checking. In each experiment, segments of the second group are shuffled and divided into a training dataset and a testing dataset with 80% and 20% of the total, respectively.

3) OTHER SETTINGS

The experiments are repeated 5 times to alleviate the influence of random factors. In each experiment, the student model is randomly initialized, Budget is set to 3, batch_size is set to 32, training times (epoch) is set to 7, and the learning rate is set to 0.0001. The Adam algorithm is chosen as the optimization algorithm, and cross-entropy as the loss function.

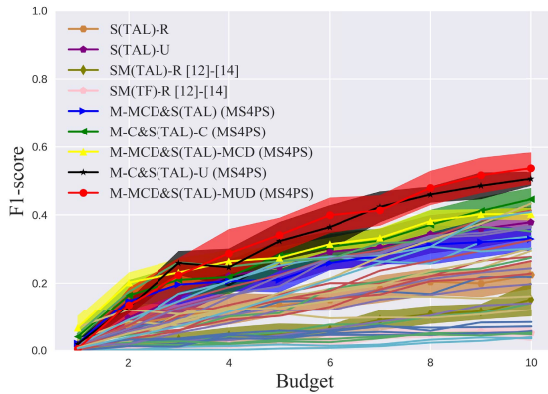


FIGURE 8. The Avg_F1 (the F1-score averaged over all patients and five independent repeats) curves of all 39 methods in Table4. The legends only show nine methods, including five MS4PS methods, two commonly used active learning methods (S(TAL)-R, S(TAL)-U), and two transfer learning methods (S_M (TAL)-R, S_M (TF)-R). The color shade shows the standard deviation over five independent repeats. It should be known that the two transfer learning methods (S_M (TAL)-R, S_M (TF)-R) are not exactly the same as methods in [12]–[14], because to make the comparison among 39 methods fair we force them all use our DNN models and dataset.

V. RESULTS AND ANALYSIS

In this section, we would analyze the results of experiments and answer the questions proposed in section IV EXPERIMENTS.

A. THE FIRST TIME CHECKING

1) Avg_F1

For each patient, the Avg_F1 of different AL methods are shown in Table10-Table15 of the Appendix, and the F1-score curves are shown in Fig.9-Fig.14 of the Appendix. The Avg_F1 is 0 for patients 4, 6, and 21 in almost every experiment. The main reason for that is that it is too hard to select seizure segments of these patients for training the model due to the severe imbalance between seizure segments and normal segments, as shown in Table2.

Table4 shows the Avg_F1 ($P=23$, $T=5$, $B=10$) and Fig.8 shows the Avg_F1 ($P=23$, $T=5$) curves of all 39 methods in Table4. It could be found that when coupling with a same AL strategy, S_M (TF) gets worse performances than S_M (TAL). It implies that the difference between the NEO dataset and CHB-MIT dataset is so significant that the feature extractor (the convolutional layers in this paper) trained with segments of the NEO dataset could not be used directly for segments of the CHB-MIT dataset. Meanwhile, when coupling with most AL strategies, the empty model S works better than S_M . Model S gets its top Avg_F1 0.241 with S(TAL)-U, and S_M gets its top Avg_F1 0.218 with S_M (TAL)-C. The above results show that transferring knowledge with the parameters of the pre-trained model (denoted as S_M) could not bring improvement for patient-specific seizure detection models in this paper. This finding is consistent with that of [18].

Of all the cases, the top 5 Avg_F1 performances (0.334, 0.322, 0.287, 0.278, and 0.263) are obtained by M-MCD&S(TAL)-MUD, M-C&S(TAL)-U, M-MCD&S(TAL)-MCD, M-C&S(TAL)-C and M-MCD&S(TAL),

respectively. These top 5 settings all use our new knowledge transferring method, mentor-select-for-student. It proves that the proposed knowledge transferring method of MS4PS really works well.

It could be found that coupling model S with U/MUD gets better performance than with other AL strategies, and coupling model M with C/MCD gets better performance than with other AL strategies. That is consistent with our intuition (see III. Method).

In addition, it is hard to say the distance strategy helps all models, and it is difficult and meaningless to analyze the results one patient by one patient. Nevertheless, M-MCD&S(TAL)-MUD and M-MCD&S(TAL)-MCD outperform the M-C&S(TAL)-U and M-C&S(TAL)-C with an excess of 0.012 and 0.009 on Avg_F1, respectively, suggesting that the distance strategy really does some help in these cases.

2) Avg_seizures

Table5 shows the Avg_seizures of different AL methods. It could be found that the M-MCD&S(TAL)-MUD selects the most, 45, seizure segments. If we regard the AL methods with R as a basic-level doctor who tries to find seizure segments, and other AL methods as higher-level doctors, then from Table6, it could be found that the M-MCD&S(TAL)-MUD could promote the efficiency of diagnosis with 7.5 times of that of a basic-level doctor.

From the above results and analysis of the first-time checking, it could be verified that the M-MCD&S(TAL)-MUD is the best in this paper, not only on Avg_F1 but also on Avg_seizures. Furthermore, because M-MCD&S(TAL)-MUD does select more seizure segments than M-C&S(TAL)-U, it could be confirmed that the distance strategy makes sense.

B. THE SECOND TIME CHECKING

Note that only the top 1 AL method of the first-time checking is employed in the second-time checking.

1) Avg_F1

Table7 shows the Avg_F1 of the second-time checking. It is evident that the reloaded student model gets better performance than the empty student model in the same conditions. When the segments of the first-time checking are used to assist the second-time checking, there are performance promotions of 0.059 and 0.262 for the reloaded student model and the empty student model, respectively.

2) Avg_seizures

Table8 shows Avg_seizures of the second-time checking. The M-MCD&S(TAL)-MUD with the reloaded student model and the segments of the first-time checking selects 24 seizure segments, getting the best performance.

M-MCD&S(Fixed)-MUD means that the student model would not be trained in the second-time checking, simulating an actual scenario where there is not enough computing

TABLE 4. Avg_F1 ($P = 23, T = 5, B = 10$) of the first-time checking. The standard deviation information are calculated over T(independent repeats).

AL method	Avg_F1	AL method	Avg_F1	AL method	Avg_F1	AL method	Avg_F1	AL method	Avg_F1
S(TAL)-R	0.141 ± 0.027	S(TAL)-U	0.241 ± 0.041	S(TAL)-C	0.09 ± 0.009	S(TAL)-MUD	0.214 ± 0.025	S(TAL)-MCD	0.133 ± 0.016
S_M (TAL)-R	0.067 ± 0.016	S_M (TAL)-U	0.157 ± 0.024	S_M (TAL)-C	0.218 ± 0.003	S_M (TAL)-MUD	0.158 ± 0.029	S_M (TAL)-MCD	0.217 ± 0.012
S_M (TF)-R	0.029 ± 0.006	S_M (TF)-U	0.018 ± 0.016	S_M (TF)-C	0.051 ± 0.004	S_M (TF)-MUD	0.012 ± 0.009	S_M (TF)-MCD	0.052 ± 0.008
M-R&S(TAL)	0.134 ± 0.012	M-U&S(TAL)	0.036 ± 0.004	M-C&S(TAL)	0.228 ± 0.023	M-MUD&S(TAL)	0.035 ± 0.003	M-MCD&S(TAL)	0.263 ± 0.023
M-R& S_M (TAL)	0.065 ± 0.014	M-U& S_M (TAL)	0.03 ± 0.005	M-C& S_M (TAL)	0.147 ± 0.028	M-MUD& S_M (TAL)	0.033 ± 0.008	M-MCD& S_M (TAL)	0.135 ± 0.022
M-R&S(TAL)-R	0.145 ± 0.025	M-U&S(TAL)-U	0.16 ± 0.029	M-C&S(TAL)-C	0.278 ± 0.031	M-MUD&S(TAL)-MUD	0.166 ± 0.018	M-MCD&S(TAL)-MCD	0.287 ± 0.018
M-R& S_M (TAL)-R	0.066 ± 0.012	M-U& S_M (TAL)-U	0.137 ± 0.029	M-C& S_M (TAL)-C	0.236 ± 0.02	M-MUD& S_M (TAL)-MUD	0.11 ± 0.02	M-MCD& S_M (TAL)-MCD	0.215 ± 0.014
M-C&S(TAL)-U	0.322 ± 0.028	M-MCD&S(TAL)-MUD	0.334 ± 0.04	M-C& S_M (TAL)-U	0.205 ± 0.026	M-MCD& S_M (TAL)-MUD	0.196 ± 0.022		

TABLE 5. Avg_seizures ($P = 23, T = 5, B = 10$) of the first-time checking (Avg_S stands for Avg_seizures). The standard deviation information are calculated over T (independent repeats).

AL method	Avg_S	AL method	Avg_S	AL method	Avg_S	AL method	Avg_S	AL method	Avg_S
S(TAL)-R	6 ± 0	S(TAL)-U	40 ± 3	S(TAL)-C	6 ± 4	S(TAL)-MUD	42 ± 2	S(TAL)-MCD	23 ± 2
S_M (TAL)-R	6 ± 0	S_M (TAL)-U	27 ± 4	S_M (TAL)-C	21 ± 1	S_M (TAL)-MUD	30 ± 5	S_M (TAL)-MCD	21 ± 1
S_M (TF)-R	6 ± 0	S_M (TF)-U	9 ± 4	S_M (TF)-C	36 ± 3	S_M (TF)-MUD	9 ± 3	S_M (TF)-MCD	37 ± 1
M-R&S(TAL)	6 ± 0	M-U&S(TAL)	4 ± 0	M-C&S(TAL)	11 ± 0	M-MUD&S(TAL)	4 ± 0	M-MCD&S(TAL)	11 ± 0
M-R& S_M (TAL)	6 ± 0	M-U& S_M (TAL)	4 ± 0	M-C& S_M (TAL)	11 ± 0	M-MUD& S_M (TAL)	3 ± 0	M-MCD& S_M (TAL)	10 ± 0
M-R&S(TAL)-R	6 ± 0	M-U&S(TAL)-U	27 ± 5	M-C&S(TAL)-C	21 ± 2	M-MUD&S(TAL)-MUD	33 ± 3	M-MCD&S(TAL)-MCD	24 ± 2
M-R& S_M (TAL)-R	6 ± 0	M-U& S_M (TAL)-U	24 ± 6	M-C& S_M (TAL)-C	18 ± 1	M-MUD& S_M (TAL)-MUD	20 ± 4	M-MCD& S_M (TAL)-MCD	30 ± 2
M-C&S(TAL)-U	36 ± 5	M-MCD&S(TAL)-MUD	45 ± 4	M-C& S_M (TAL)-U	26 ± 4	M-MCD& S_M (TAL)-MUD	30 ± 5		

TABLE 6. The efficiency of diagnosis. Regarding the AL methods with R as basic-level doctors with efficiency ratio 1.

AL method	ratio	AL method	ratio	AL method	ratio	AL method	ratio	AL method	ratio
S(TAL)-R	1	S(TAL)-U	6.67	S(TAL)-C	1	S(TAL)-MUD	7	S(TAL)-MCD	3.83
S_M (TAL)-R	1	S_M (TAL)-U	4.5	S_M (TAL)-C	3.5	S_M (TAL)-MUD	5	S_M (TAL)-MCD	3.5
S_M (TF)-R	1	S_M (TF)-U	1.5	S_M (TF)-C	6	S_M (TF)-MUD	1.5	S_M (TF)-MCD	6.17
M-R&S(TAL)	1	M-U&S(TAL)	0.67	M-C&S(TAL)	1.83	M-MUD&S(TAL)	0.67	M-MCD&S(TAL)	1.83
M-R& S_M (TAL)	1	M-U& S_M (TAL)	0.67	M-C& S_M (TAL)	1.83	M-MUD& S_M (TAL)	0.5	M-MCD& S_M (TAL)	1.67
M-R&S(TAL)-R	1	M-U&S(TAL)-U	4.5	M-C&S(TAL)-C	3.5	M-MUD&S(TAL)-MUD	5.5	M-MCD&S(TAL)-MCD	4
M-R& S_M (TAL)-R	1	M-U& S_M (TAL)-U	4	M-C& S_M (TAL)-C	3	M-MUD& S_M (TAL)-MUD	3.33	M-MCD& S_M (TAL)-MCD	5
M-C&S(TAL)-U	6	M-MCD&S(TAL)-MUD	7.5	M-C& S_M (TAL)-U	4.33	M-MCD& S_M (TAL)-MUD	5		

TABLE 7. Avg_F1 of the second-time checking ($T = 5, P = 23, B = 3$). The field name "segment" means using historical segments, "non-segment" means not using historical segments. "Reloaded" means using the reloaded student model. "Empty" means using an empty student model.

	segment		non-segment	
	Reloaded	Empty	Reloaded	Empty
M-MCD&S(TAL)-MUD	0.69	0.679	0.631	0.417

resource for training a model online. In that case, a slight performance degradation (in Table8, 20 instead of 24 seizure segments are selected) would happen.

Meanwhile, it could be found that using historically labeled segments of the first-time checking always gets a larger or equal number of seizure segments than without using them.

From the above results and analysis of the second-time checking, it could be verified that the reloaded student model is better than the empty student model, and reusing historically labeled segments would help improve performance, revealing that the patient-specific models would get sustained promotion through MS4PS when patients come more times and their data are accumulated more and more.

TABLE 8. Avg_seizures of the second-time checking. The field name "segment" means using historical segments, "non-segment" means not using historical segments. "Reloaded" means using the reloaded student model. "Empty" means using an empty student model. M-MCD&S(Fixed)-MUD means that the student model would not be trained.

	segment		non-segment	
	Reloaded	Empty	Reloaded	Empty
M-MCD&S(TAL)-MUD	24	18	21	18
M-MCD&S(Fixed)-MUD	20	11	18	11

C. FURTHER DISCUSSION

Considering the complexity of the method we proposed, we would make a further discussion to explain how and why MS4PS could effectively help to obtain advances and better effects, and support the reasoning with the results shown in Table4-5 and Fig.8. We think there are several factors that help MS4PS to obtain advances.

- 1) The knowledge transferring method named mentor-select-for-student. Transferring knowledge through mentor-select-for-student makes it possible to avoid the negative transfer [19] of transferring parameters/structures. In our experiments (see Table4), the negative transfer is that the S_M (TAL)-R and S_M (TF)-R

TABLE 9. The details of the deep neural network used in this paper as student model and mentor model.

Operation	Input shape	parameters	Output shape
Conv2D	(1, 18, 2560)	$kernel_size = (3, 3) stride = (1, 1) padding = (1, 1)$	(32, 18, 2560)
relu	(32, 18, 2560)		(32, 18, 2560)
maxpool2d	(32, 18, 2560)		$kernel_size = (2, 2)$
Conv2D	(32, 9, 1280)	$kernel_size = (3, 3) stride = (1, 1) padding = (1, 1)$	(64, 9, 1280)
relu	(64, 9, 1280)		(64, 9, 1280)
maxpool2d	(64, 9, 1280)		$kernel_size = (2, 2)$
Conv2D	(64, 4, 640)	$kernel_size = (3, 3) stride = (1, 1) padding = (1, 1)$	(128, 4, 640)
relu	(128, 4, 640)		(128, 4, 640)
maxpool2d	(128, 4, 640)		$kernel_size = (2, 2)$
Conv2D	(128, 2, 320)	$kernel_size = (3, 3) stride = (1, 1) padding = (1, 1)$	(256, 2, 320)
relu	(256, 2, 320)		(256, 2, 320)
maxpool2d	(128, 2, 320)		$kernel_size = (2, 2)$
Conv2D	(256, 1, 160)	$kernel_size = (1, 3) stride = (1, 1) padding = (1, 1)$	(256, 2, 160)
relu	(256, 2, 160)		(256, 2, 160)
maxpool2d	(256, 2, 160)		$kernel_size = (2, 2)$
Linear	(1, 20480)	$p = 0.5$ $slope = 0.2$	(1, 512)
Dropout	(1, 512)		(1, 512)
LeakyRelu	(1, 512)		(1, 512)
Linear	(1, 512)	$p = 0.5$ $slope = 0.2$	(1, 64)
Dropout	(1, 64)		(1, 64)
LeakyRelu	(1, 64)		(1, 64)
Linear	(1, 64)		(1, 2)

TABLE 10. Avg_F1 (P = 1, T = 5, B = 10) of the first-time checking.

	S(TAL)-R	S_M (TAL)-R	S_M (TF)-R	M-R&S(TAL)	M-R& S_M (TAL)	M-R&S(TAL)-R	M-R& S_M (TAL)-R
chb01	0.154	0.047	0.002	0.074	0.02	0.136	0.038
chb02	0.016	0.03	0.0	0.005	0.014	0.016	0.002
chb03	0.008	0.007	0.001	0.009	0.003	0.006	0.004
chb04	0.0	0.0	0.0	0.0	0.0	0.003	0.0
chb05	0.246	0.034	0.01	0.276	0.122	0.211	0.123
chb06	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb07	0.027	0.022	0.0	0.008	0.01	0.0	0.001
chb08	0.282	0.145	0.052	0.358	0.099	0.361	0.087
chb09	0.019	0.0	0.0	0.087	0.004	0.007	0.009
chb10	0.09	0.023	0.001	0.11	0.164	0.169	0.122
chb11	0.634	0.268	0.336	0.588	0.256	0.53	0.3
chb12	0.229	0.061	0.028	0.262	0.032	0.238	0.059
chb13	0.171	0.036	0.001	0.066	0.032	0.183	0.009
chb14	0.097	0.224	0.003	0.059	0.175	0.074	0.153
chb15	0.433	0.258	0.151	0.498	0.236	0.489	0.223
chb17	0.095	0.03	0.002	0.017	0.02	0.085	0.02
chb18	0.003	0.006	0.0	0.028	0.013	0.023	0.004
chb19	0.182	0.129	0.01	0.134	0.039	0.165	0.15
chb20	0.008	0.001	0.0	0.023	0.006	0.01	0.014
chb21	0.001	0.001	0.0	0.002	0.0	0.0	0.001
chb22	0.118	0.004	0.0	0.193	0.058	0.153	0.003
chb23	0.118	0.03	0.006	0.103	0.027	0.234	0.003
chb24	0.313	0.178	0.071	0.175	0.169	0.254	0.185
average	0.141	0.067	0.029	0.134	0.065	0.145	0.066

TABLE 11. Avg_F1 (P = 1, T = 5, B = 10) of the first-time checking.

	S(TAL)-U	S_M (TAL)-U	S_M (TF)-U	M-U&S(TAL)	M-U& S_M (TAL)	M-U&S(TAL)-U	M-U& S_M (TAL)-U
chb01	0.125	0.001	0.0	0.0	0.003	0.0	0.081
chb02	0.004	0.135	0.0	0.0	0.075	0.0	0.094
chb03	0.087	0.006	0.0	0.014	0.005	0.025	0.01
chb04	0.0	0.001	0.0	0.0	0.001	0.0	0.001
chb05	0.651	0.096	0.0	0.0	0.001	0.408	0.071
chb06	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb07	0.0	0.003	0.0	0.012	0.004	0.073	0.035
chb08	0.541	0.131	0.0	0.424	0.262	0.377	0.323
chb09	0.058	0.001	0.0	0.0	0.003	0.026	0.001
chb10	0.0	0.338	0.0	0.0	0.0	0.145	0.0
chb11	0.772	0.401	0.145	0.0	0.004	0.573	0.424
chb12	0.468	0.466	0.138	0.0	0.001	0.497	0.33
chb13	0.399	0.101	0.0	0.0	0.005	0.105	0.085
chb14	0.615	0.715	0.061	0.338	0.304	0.599	0.709
chb15	0.59	0.616	0.006	0.0	0.002	0.412	0.582
chb17	0.0	0.002	0.001	0.0	0.002	0.0	0.001
chb18	0.087	0.044	0.0	0.0	0.004	0.001	0.003
chb19	0.394	0.014	0.0	0.004	0.01	0.112	0.166
chb20	0.007	0.073	0.0	0.026	0.012	0.05	0.028
chb21	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb22	0.288	0.089	0.012	0.0	0.0	0.19	0.0
chb23	0.318	0.051	0.0	0.0	0.0	0.053	0.0
chb24	0.128	0.335	0.042	0.0	0.0	0.04	0.218
average	0.241	0.157	0.018	0.036	0.03	0.16	0.137

TABLE 12. Avg_F1 (P = 1, T = 5, B = 10) of the first-time checking.

	S(TAL)-C	S_M (TAL)-C	S_M (TF)-C	M-C&S(TAL)	M-C& S_M (TAL)	M-C&S(TAL)-C	M-C& S_M (TAL)-C
chb01	0.113	0.476	0.05	0.131	0.016	0.182	0.338
chb02	0.001	0.038	0.006	0.0	0.0	0.001	0.0
chb03	0.013	0.018	0.001	0.093	0.007	0.005	0.031
chb04	0.0	0.003	0.0	0.0	0.004	0.0	0.007
chb05	0.1	0.469	0.077	0.28	0.117	0.513	0.44
chb06	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb07	0.021	0.121	0.009	0.057	0.008	0.284	0.156
chb08	0.175	0.301	0.143	0.403	0.172	0.46	0.391
chb09	0.036	0.089	0.006	0.057	0.002	0.165	0.128
chb10	0.039	0.505	0.078	0.557	0.5	0.5	0.596
chb11	0.216	0.503	0.266	0.716	0.545	0.771	0.59
chb12	0.026	0.083	0.016	0.345	0.112	0.366	0.174
chb13	0.001	0.025	0.0	0.33	0.178	0.144	0.07
chb14	0.0	0.047	0.0	0.022	0.069	0.167	0.084
chb15	0.137	0.054	0.028	0.516	0.374	0.495	0.332
chb17	0.185	0.193	0.016	0.2	0.182	0.138	0.172
chb18	0.014	0.125	0.002	0.096	0.019	0.052	0.116
chb19	0.259	0.45	0.048	0.422	0.192	0.559	0.462
chb20	0.068	0.166	0.02	0.018	0.01	0.124	0.058
chb21	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb22	0.179	0.527	0.167	0.161	0.201	0.42	0.431
chb23	0.325	0.416	0.125	0.389	0.271	0.559	0.41
chb24	0.152	0.412	0.11	0.458	0.401	0.467	0.433
average	0.09	0.218	0.051	0.228	0.147	0.278	0.236

TABLE 13. Avg_F1 (P = 1, T = 5, B = 10) of the first-time checking.

	S(TAL)-MUD	S_M (TAL)-MUD	S_M (TF)-MUD	M-MUD&S(TAL)	M-MUD& S_M (TAL)	M-MUD&S(TAL)-MUD	M-MUD& S_M (TAL)-MUD
chb01	0.178	0.021	0.0	0.0	0.0	0.31	0.113
chb02	0.012	0.182	0.0	0.0	0.05	0.0	0.104
chb03	0.001	0.004	0.0	0.011	0.012	0.005	0.005
chb04	0.0	0.002	0.0	0.0	0.002	0.0	0.001
chb05	0.484	0.103	0.0	0.0	0.001	0.262	0.243
chb06	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb07	0.315	0.003	0.0	0.028	0.031	0.0	0.027
chb08	0.53	0.084	0.0	0.352	0.15	0.351	0.222
chb09	0.0	0.012	0.0	0.0	0.003	0.082	0.001
chb10	0.221	0.013	0.0	0.0	0.0	0.181	0.126
chb11	0.155	0.703	0.084	0.0	0.02	0.295	0.242
chb12	0.394	0.422	0.053	0.0	0.0	0.41	0.167
chb13	0.157	0.059	0.0	0.0	0.005	0.107	0.07
chb14	0.544	0.734	0.053	0.351	0.444	0.509	0.573
chb15	0.586	0.586	0.063	0.0	0.001	0.395	0.576
chb17	0.113	0.008	0.0	0.0	0.001	0.11	0.001
chb18	0.0	0.011	0.0	0.0	0.003	0.084	0.003
chb19	0.362	0.095	0.0	0.007	0.003	0.261	0.003
chb20	0.138	0.06	0.0	0.05	0.015	0.035	0.041
chb21	0.0	0.0	0.0	0.0	0.016	0.0	0.0
chb22	0.062	0.213	0.0	0.0	0.0	0.209	0.0
chb23	0.452	0.076	0.0	0.0	0.0	0.01	0.001
chb24	0.228	0.236	0.025	0.0	0.002	0.208	0.011
average	0.214	0.158	0.012	0.035	0.033	0.166	0.11

TABLE 14. Avg_F1 (P = 1, T = 5, B = 10) of the first-time checking.

	S(TAL)-MCD	S_M (TAL)-MCD	S_M (TF)-MCD	M-MCD&S(TAL)	M-MCD& S_M (TAL)	M-MCD&S(TAL)-MCD	M-MCD& S_M (TAL)-MCD
chb01	0.214	0.421	0.074	0.126	0.077	0.277	0.311
chb02	0.0	0.002	0.0	0.0	0.0	0.005	0.018
chb03	0.016	0.029	0.005	0.11	0.007	0.02	0.033
chb04	0.002	0.01	0.0	0.004	0.001	0.0	0.01
chb05	0.27	0.528	0.092	0.424	0.255	0.47	0.419
chb06	0.0	0.0	0.0	0.0	0.0	0.0	0.0
chb07	0.019	0.148	0.007	0.101	0.005	0.308	0.084
chb08	0.21	0.259	0.145	0.504	0.202	0.451	0.218
chb09	0.048	0.157	0.002	0.057	0.0	0.196	0.244
chb10	0.27	0.571	0.075	0.505	0.245	0.479	0.481
chb11	0.339	0.531	0.3	0.814	0.454	0.827	0.498
chb12	0.226	0.116	0.017	0.31	0.176	0.404	0.139
chb13	0.059	0.023	0.0	0.28	0.02	0.06	0.049
chb14	0.0	0.015	0.0	0.041	0.0	0.058	0.087
chb15	0.079	0.063	0.006	0.602	0.383	0.539	0.369
chb17	0.147	0.293	0.007	0.125	0.198	0.24	0.167
chb18	0.028	0.157	0.007	0.12	0.047	0.209	0.097
chb19	0.396	0.313	0.056	0.46	0.195	0.596	0.415
chb20	0.043	0.047	0.042	0.019	0.011	0.074	0.087
chb21	0.002	0.0	0.0	0.007	0.0	0.0	0.0
chb22	0.231	0.423	0.122	0.527	0.184	0.437	0.333
chb23	0.277	0.398	0.106	0.358	0.222	0.479	0.429
chb24	0.176	0.49	0.122	0.549	0.463	0.483	0.466
average	0.133	0.217	0.052	0.263	0.135	0.287	0.215

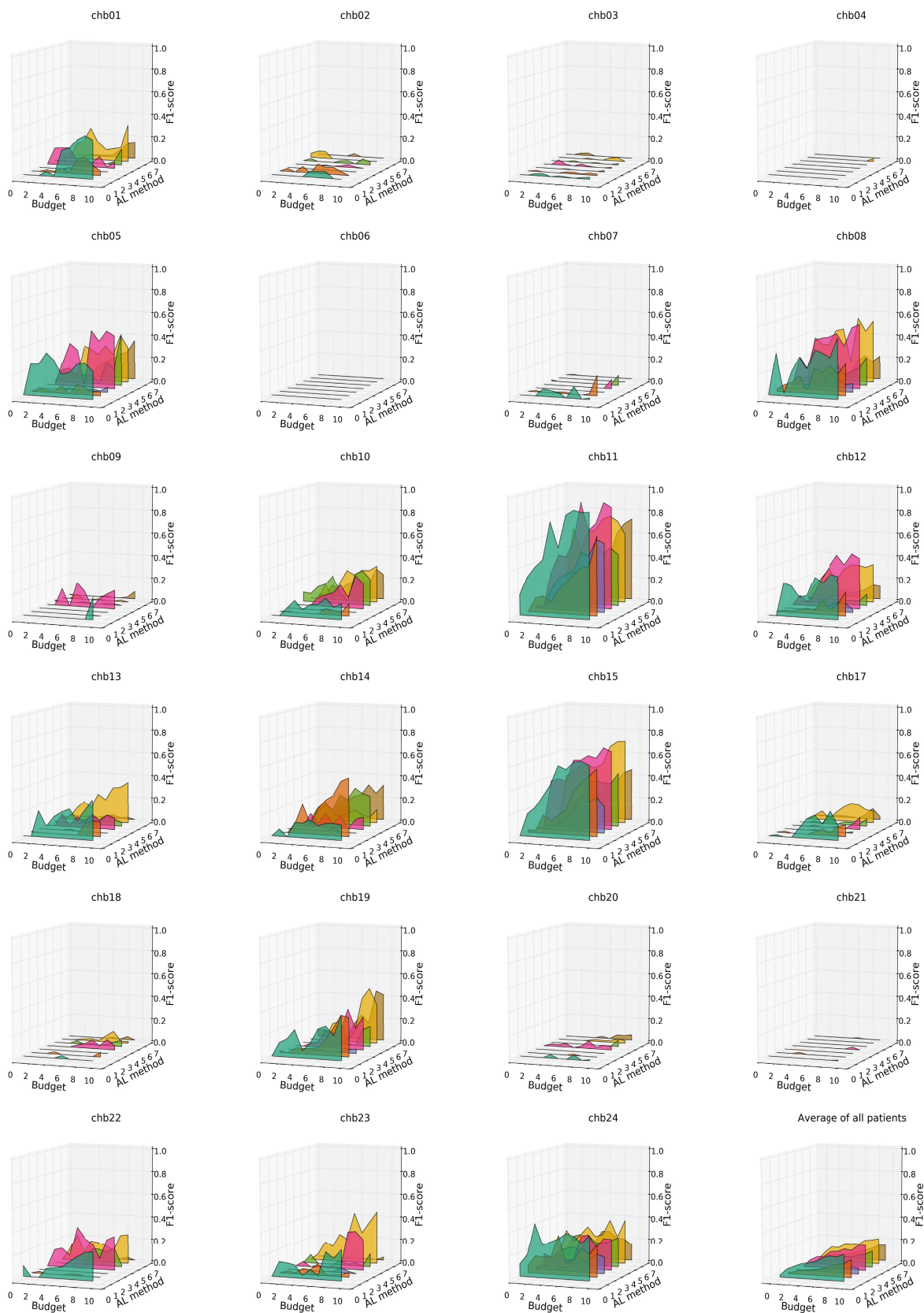


FIGURE 9. F1-score curves of different AL methods for different patients in their first-time checking. For AL method, 1 is S(TAL)-R, 2 is S_M (TAL)-R, 3 is S_M (TF)-R, 4 is M-R&S(TAL), 5 is M-R& S_M (TAL), 6 is M-R&S(TAL)-R, 7 is M-R& S_M (TAL)-R.

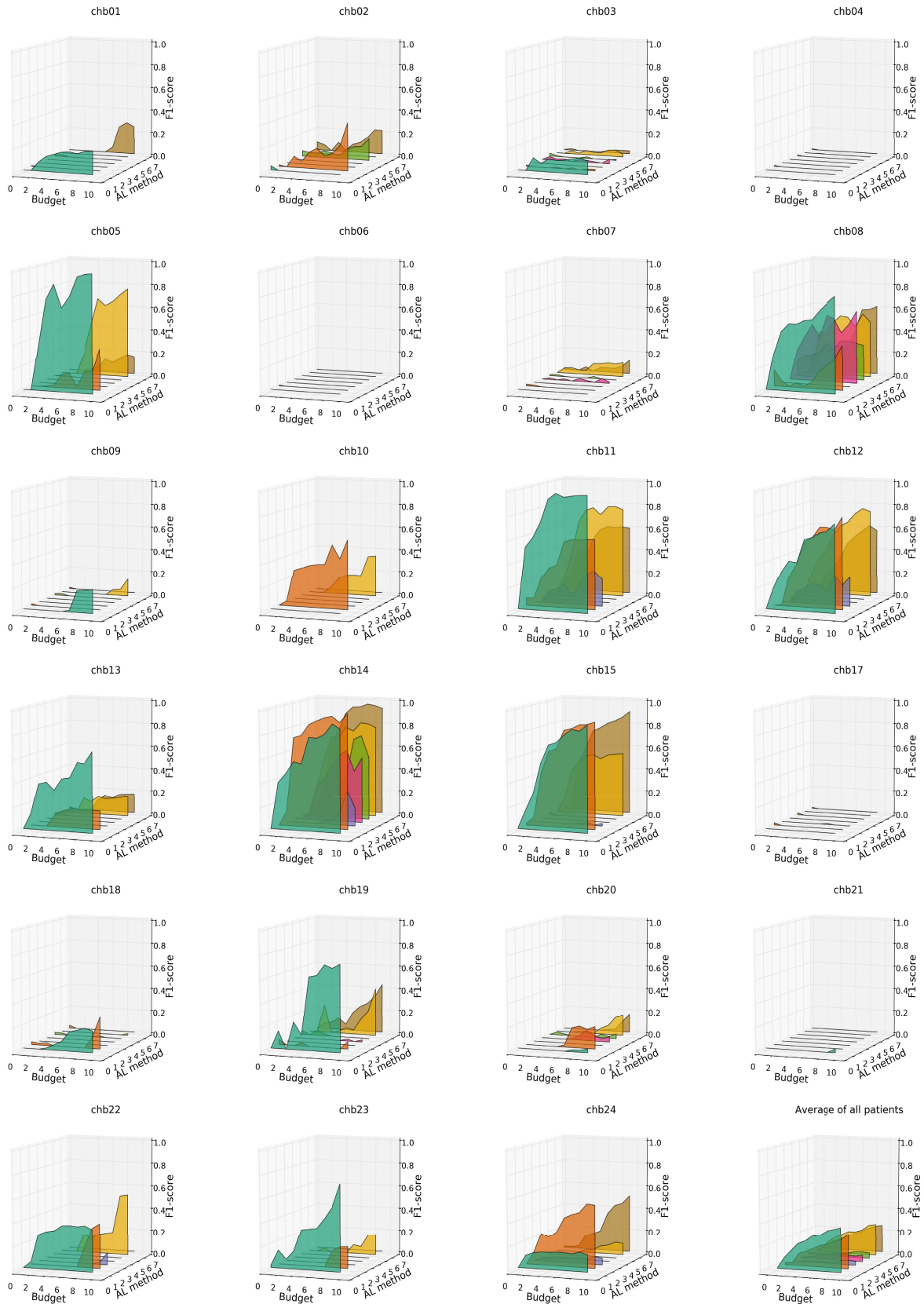


FIGURE 10. F1-score curves of different AL methods for different patients in their first-time checking. For AL method, 1 is $S(TAL)-U$, 2 is $S_M(TAL)-U$, 3 is $S_M(TF)-U$, 4 is $M-U&S(TAL)$, 5 is $M-U&S_M(TAL)$, 6 is $M-U&S(TAL)-U$, 7 is $M-U&S_M(TAL)-U$.

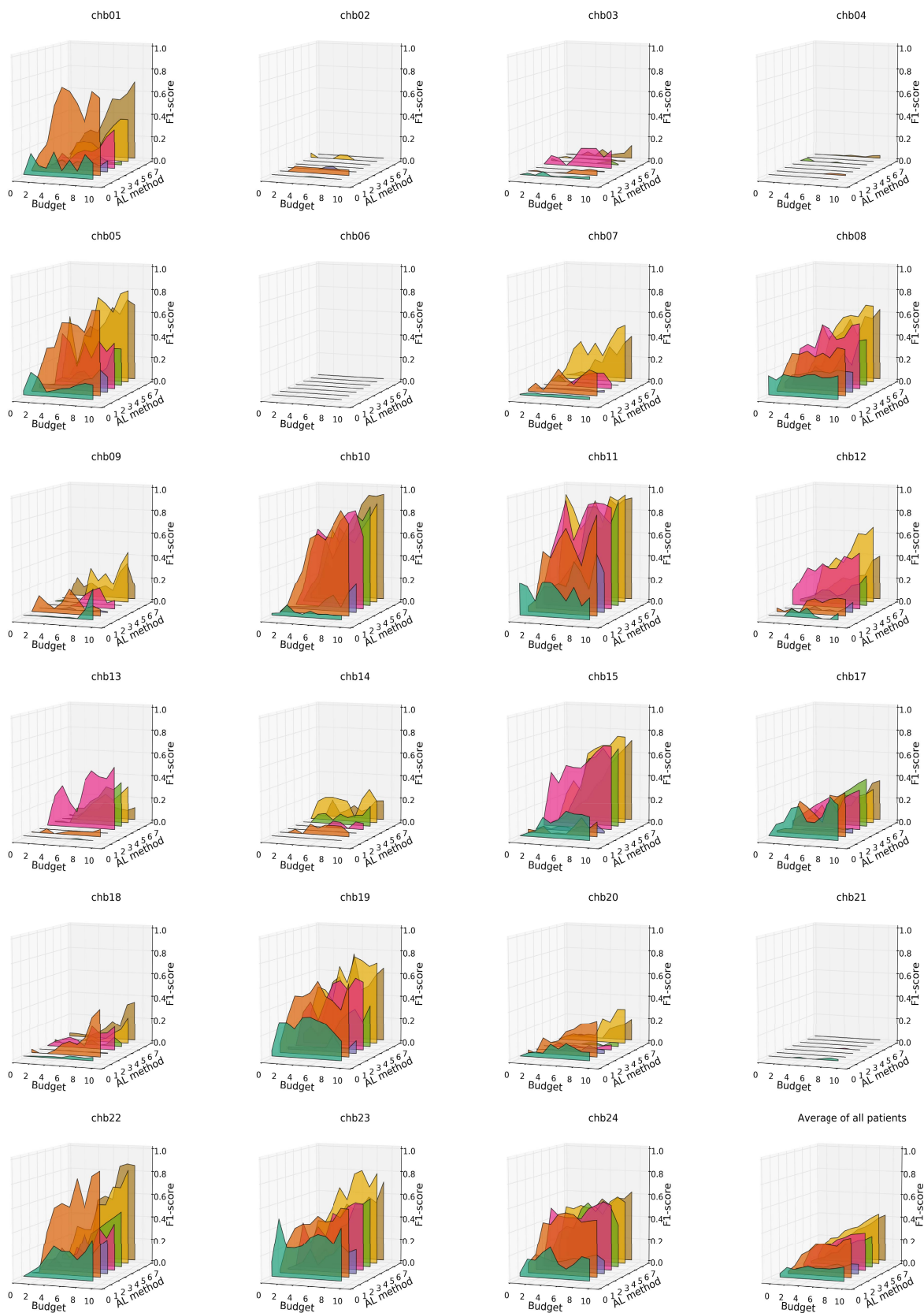


FIGURE 11. F1-score curves of different AL methods for different patients in their first-time checking. For AL method, 1 is S(TAL)-C, 2 is S_M (TAL)-C, 3 is S_M (TF)-C, 4 is M-C&S(TAL), 5 is M-C& S_M (TAL), 6 is M-C&S(TAL)-C, 7 is M-C& S_M (TAL)-C.

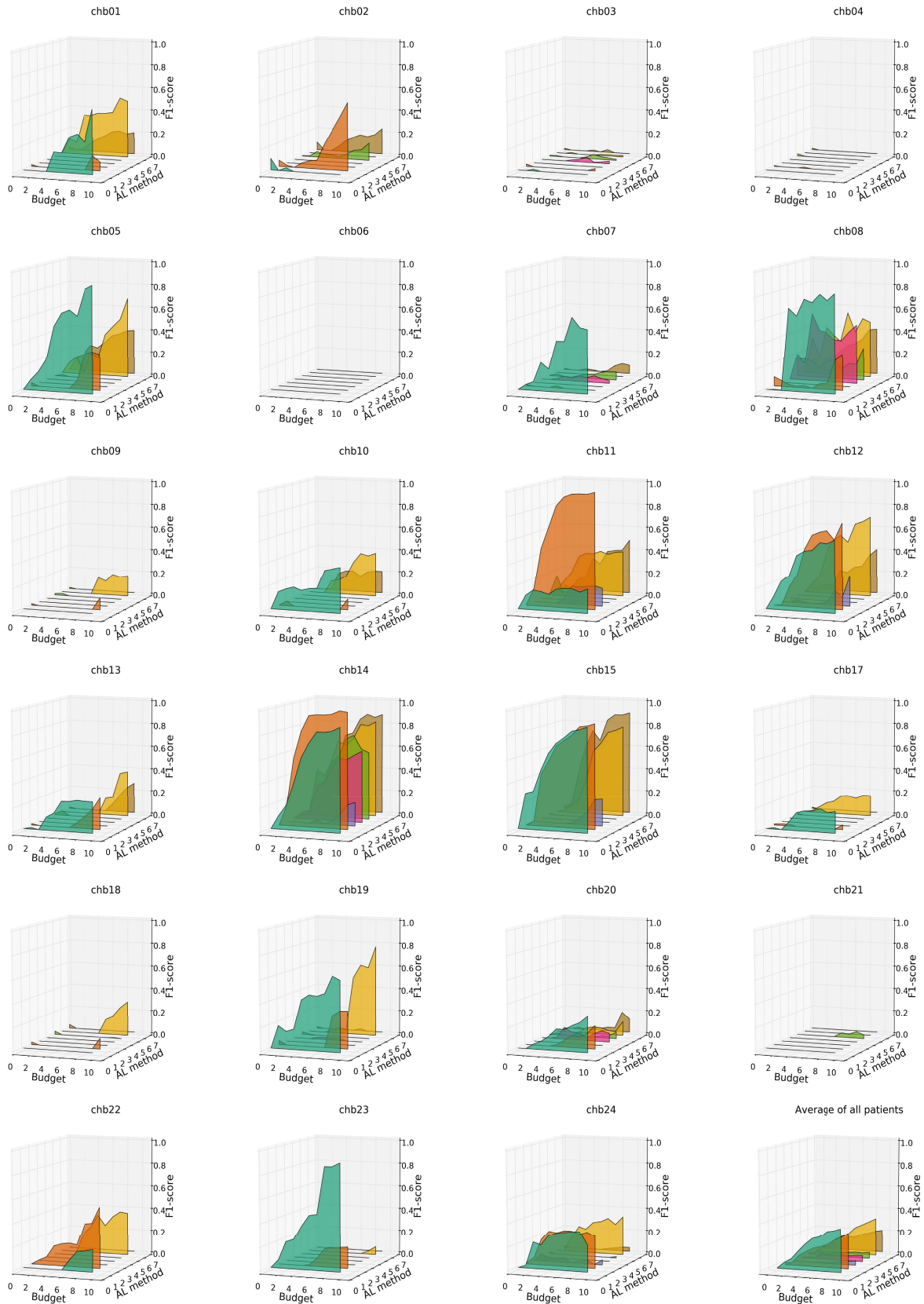


FIGURE 12. F1-score curves of different AL methods for different patients in their first-time checking. For AL method, 1 is $S(TAL)$ -MUD, 2 is $S_M(TAL)$ -MUD, 3 is $S_M(TF)$ -MUD, 4 is M -MUD& $S(TAL)$, 5 is M -MUD& $S_M(TAL)$, 6 is M -MUD& $S(TAL)$ -MUD, 7 is M -MUD& $S_M(TAL)$ -MUD.

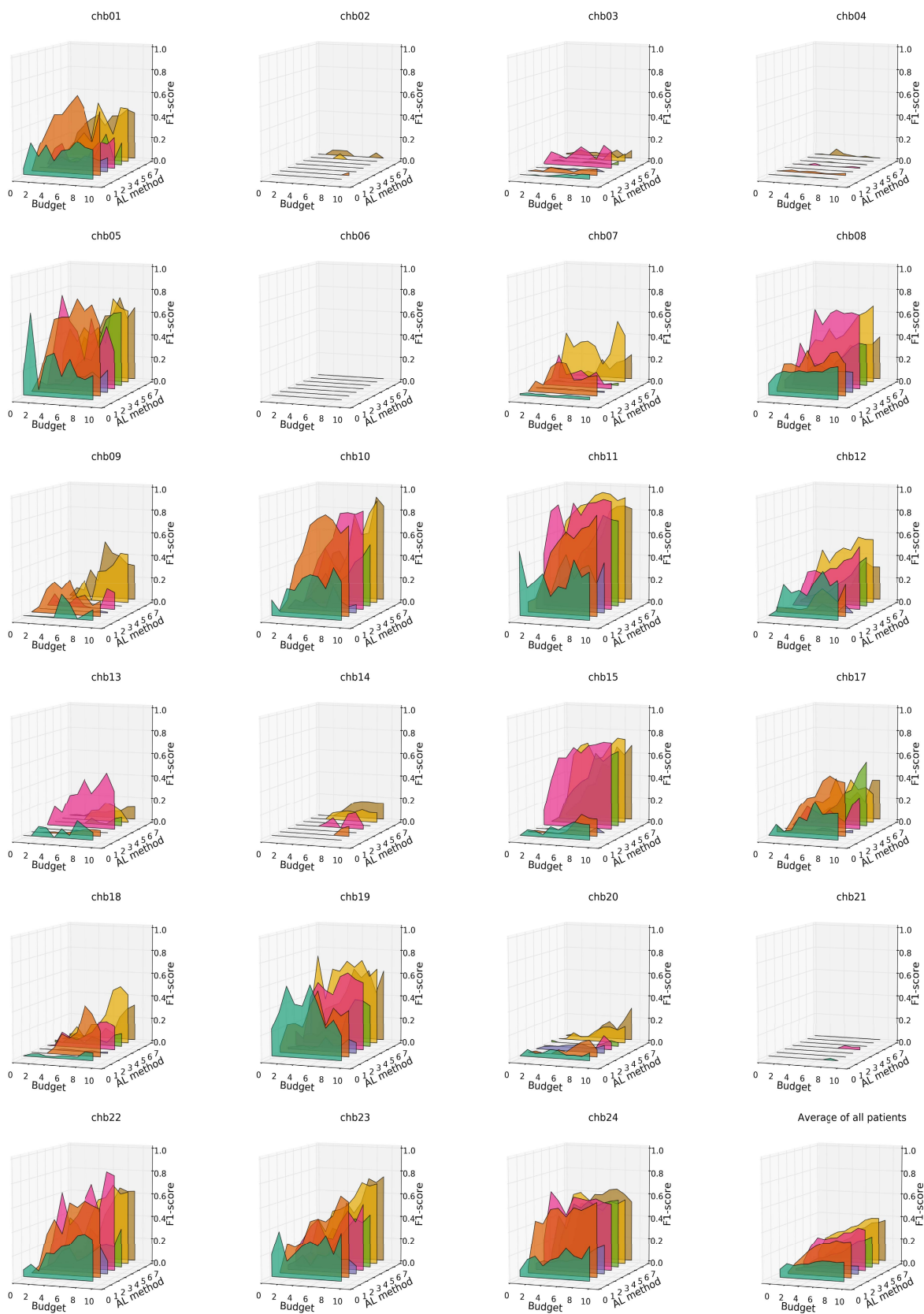


FIGURE 13. F1-score curves of different AL methods for different patients in their first-time checking. For AL method, 1 is $S(TAL)$ -MCD, 2 is $S_M(TAL)$ -MCD, 3 is $S_M(TF)$ -MCD, 4 is M-MCD& $S(TAL)$, 5 is M-MCD& $S_M(TAL)$, 6 is M-MCD& $S(TAL)$ -MCD, 7 is M-MCD& $S_M(TAL)$ -MCD.

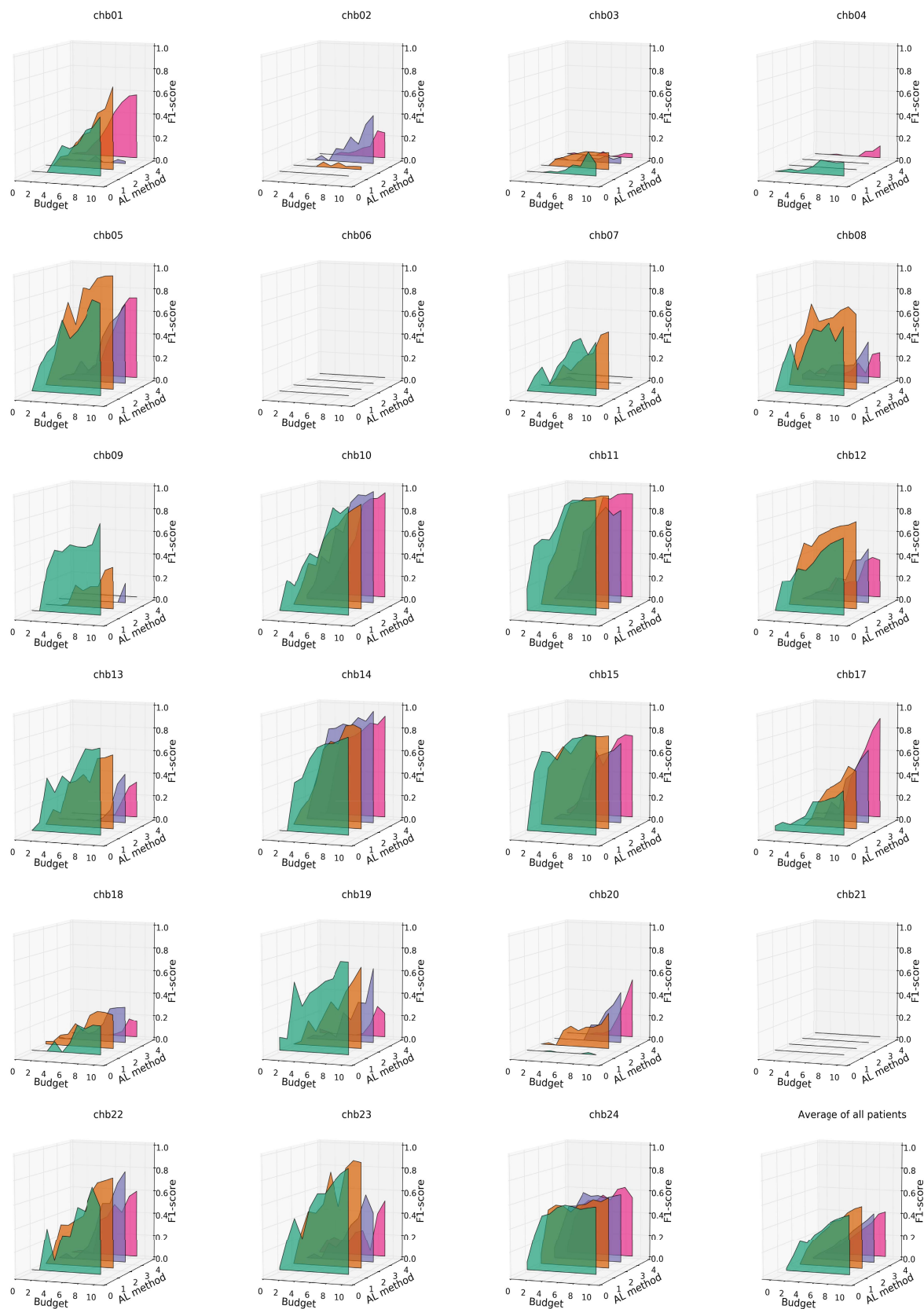


FIGURE 14. F1-score curves of different AL methods for different patients in their first-time checking. For AL method, 1 is M-C&S(TAL)-U, 2 is M-MCD&S(TAL)-MUD, 3 is M-C&S_M(TAL)-U, 4 is M-MCD&S_M(TAL)-MUD.

TABLE 15. Avg_F1 (P = 1, T = 5, B = 10) of the first-time checking.

	M-C&S(TAL)-U	M-MCD&S(TAL)-MUD	M-C& S_M (TAL)-U	M-MCD& S_M (TAL)-MUD
chb01	0.209	0.27	0.012	0.258
chb02	0.0	0.018	0.141	0.073
chb03	0.051	0.094	0.029	0.014
chb04	0.061	0.0	0.0	0.022
chb05	0.482	0.657	0.271	0.3
chb06	0.0	0.0	0.0	0.0
chb07	0.267	0.217	0.003	0.001
chb08	0.396	0.534	0.141	0.075
chb09	0.452	0.14	0.017	0.0
chb10	0.52	0.542	0.589	0.488
chb11	0.747	0.718	0.496	0.617
chb12	0.41	0.556	0.191	0.162
chb13	0.452	0.369	0.092	0.073
chb14	0.535	0.547	0.69	0.577
chb15	0.649	0.63	0.349	0.415
chb17	0.193	0.24	0.231	0.31
chb18	0.111	0.177	0.134	0.036
chb19	0.506	0.312	0.195	0.067
chb20	0.002	0.133	0.133	0.109
chb21	0.0	0.0	0.0	0.0
chb22	0.347	0.455	0.301	0.27
chb23	0.54	0.576	0.223	0.188
chb24	0.471	0.494	0.205	0.458
average	0.322	0.334	0.205	0.196

work worse than S(TAL)-R, implying that the transferred model S_M is of no help and even hinders the learning process in the patient-specific case.

- 2) The AL method using both a mentor model and a student model to select samples. As shown in Table4 and Fig.8, the top 5 of 39 methods are all using both a mentor model and a student model, implying that using the double-model has more potential of performance improvement than using only one model. In the top 2 MS4PS methods, M-MCD&S(TAL)-MUD and M-C&S(TAL)-U, the mentor model is coupled with certainty-based AL methods, and the student model is coupled with uncertainty-based AL methods. That is consistent with our intuition that the mentor model should select the most certain samples for training the student model, just like a mentor teaches students with his/her most certain knowledge, and the student model should select the most uncertain samples for training itself, just like a student studies the knowledge that confuses him/her for capability improvement.
- 3) The distance strategy for solving the problem of ill-balanced EEG samples. The EEG samples of epilepsy are ill-balanced, and seizure samples are more critical than normal samples for model learning and for doctors to analyze patients' conditions. To select more seizure samples than normal, we introduce the distance strategy. As shown in Table4 and Table5, M-MCD&S(TAL)-MUD gets a higher Avg_F1 and

selects more seizure samples than M-C&S(TAL)-U. So, the distance strategy does make sense.

As shown in Fig.8, the top 5 of 39 methods in Table4 are all MS4PS methods. The more factors are included, the better performance the MS4PS methods will get. Combining all the above factors makes the best MS4PS method, M-MCD&S(TAL)-MUD.

VI. CONCLUSION

The main obstacles to building good seizure detection models are privacy protection, high labeling costs, and the varying characteristics of seizures among patients and at different times.

In this paper, we have proposed a novel mentor-student architecture for patient-specific seizure detection. It contains a new way of knowledge transferring named mentor-select-for-student, which exploits mentor-model's knowledge by using a mentor model to select data for student-model's learning, making it possible to protect the data of patients and avoid the negative influence of transferring parameters/structures of pre-trained models. It also contains a new way of active learning, which uses both an experienced mentor model and a quick-learning student model to select samples for doctors to label, and each of these with a particular sample selection strategy that combines uncertainty/certainty and the distance between unlabeled samples and labeled seizure samples.

The proposed method could quickly train a suitable detector for a patient at his/her first epilepsy diagnosis, with the help of: (1) an experienced mentor model that chooses the most category-certain EEG data segments; (2) a student model (detector itself) that chooses the most category-uncertain EEG data segments; (3) the doctors who label these data segments selected by both the mentor model and student model. By replacing or improving the mentor model and refining the historical models of patients when they come next time, the MS4PS system could be sustainably promoted.

Except for the potential advantages in protecting the privacy of patients, getting good performance for patient-specific seizure detection, and sustainably capability promoting, the MS4PS has another merit, it could be easily generalized to any other object-specific problems that suffer similar obstacles to patient-specific seizure detection, for example, patient-specific depression detection.

There are still some insufficiencies in this paper, such as (1) the labeled seizure segments are too few for some patients in the CHB-MIT, which causes zero performance for them in almost every experiment, dragging down the overall performance of MS4PS. (2) The primary attention is paid to verifying the feasibility of MS4PS rather than improving seizure detection performance. Further performance improvement could be made by elaborately designing deep neural networks and feature inputs for seizure detection, which will be one of our future works.

APPENDIX

In this section, we will explain why the parameters Budget and β are set such values in this paper, and show the further details of the DNN model and results.

A. BUDGET

We set the Budget to a not big number, 10, in THE FIRST TIME CHECKING. There are two main reasons for this: (1) most patients in the CHB-MIT have a few of seizure segments (see Table2). (2) A good student model (seizure detector) should select seizure samples out of EEG data as fast as possible. We set the Budget to 3 in THE SECOND TIME CHECKING, mainly because that there is only a smaller number of segments left for simulating the second time checking, after using some for simulating the first time checking (see Fig.7).

B. β

In our works, we set the β to 0.5. This is because when neither the mentor model nor the student model could select seizure segments, we should let both of them have an equal opportunity to make a tentative choice.

C. DETAILS OF DNN

The details of the deep neural network are shown in Table9.

D. DETAILS OF RESULTS

Table4 has shown the Avg_F1 averaged over 23 patients, 5 repeat times and 10 budgets, for 39 different AL methods. Here, the 39 different AL methods are divided into 6 sub-groups. Table10-Table15 show the Avg_F1 averaged over each patient, 5 repeat times and 10 budgets. And Fig.9-Fig.14 show F1-score curves along Budget, averaged over each patient and 5 repeat times.

REFERENCES

- [1] H. Peng, C. Li, J. Chao, T. Wang, C. Zhao, X. Huo, and B. Hu, "A novel automatic classification detection for epileptic seizure based on dictionary learning and sparse representation," *Neurocomputing*, vol. 424, pp. 179–192, Feb. 2021, doi: [10.1016/j.neucom.2019.12.010](https://doi.org/10.1016/j.neucom.2019.12.010).
- [2] C. Li, W. Zhou, G. Liu, Y. Zhang, M. Geng, Z. Liu, S. Wang, and W. Shang, "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 458–467, Jan. 2021, doi: [10.1109/TNSRE.2021.3055276](https://doi.org/10.1109/TNSRE.2021.3055276).
- [3] M. Z. Ahmad, M. Saeed, S. Saleem, and A. M. Kambh, "Seizure detection using EEG: A survey of different techniques," in *Proc. Int. Conf. Emerg. Technol. (ICET)*, Islamabad, Pakistan, Oct. 2016, pp. 1–6.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [5] H. Liu, S. Ma, D. Xia, and S. Li, "SFANet: A spectrum-aware feature augmentation network for visible-infrared person reidentification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 31, 2021, doi: [10.1109/TNNLS.2021.3105702](https://doi.org/10.1109/TNNLS.2021.3105702).
- [6] D. Xia, H. Liu, L. Xu, and L. Wang, "Visible-infrared person reidentification with data augmentation via cycle-consistent adversarial network," *Neurocomputing*, vol. 443, pp. 35–46, Jul. 2021, doi: [10.1016/j.neucom.2021.02.088](https://doi.org/10.1016/j.neucom.2021.02.088).
- [7] R. Hussein, H. Palangi, R. Ward, and Z. Jane Wang, "Epileptic seizure detection: A deep learning approach," 2018, *arXiv:1803.09848*.
- [8] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," 2016, *arXiv:1608.00220*.
- [9] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 83–94, Jan. 2019, doi: [10.1109/JBHI.2018.2871678](https://doi.org/10.1109/JBHI.2018.2871678).
- [10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009, doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191).
- [11] B. Settles, "Active learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. TR1648, 2009.
- [12] A. Agrawal, G. C. Jana, and P. Gupta, "A deep transfer learning approach for seizure detection using RGB features of epileptic electroencephalogram signals," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci. (CloudCom)*, Sydney, NSW, Australia, Dec. 2019, pp. 367–373.
- [13] B. Zhang, W. Wang, Y. Xiao, S. Xiao, S. Chen, S. Chen, G. Xu, and W. Che, "Cross-subject seizure detection in EEGs using deep transfer learning," *Comput. Math. Methods Med.*, vol. 2020, 2020, Art. no. 7902072, doi: [10.1155/2020/7902072](https://doi.org/10.1155/2020/7902072).
- [14] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Netw.*, vol. 124, pp. 202–212, Apr. 2020, doi: [10.1016/j.neunet.2020.01.017](https://doi.org/10.1016/j.neunet.2020.01.017).
- [15] X. Chen, J. Ji, T. Ji, and P. Li, "Cost-sensitive deep active learning for epileptic seizure detection," in *Proc. ACM Int. Conf. Bioinform. Comput. Biol. Health Inform.*, Washington, DC, USA, 2018, pp. 226–235, doi: [10.1145/3233547.3233566](https://doi.org/10.1145/3233547.3233566).
- [16] Y. Li, Y. Liu, Y.-Z. Guo, X.-F. Liao, B. Hu, and T. Yu, "Spatio-temporal-spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction," *IEEE Trans. Cybern.*, early access, May 25, 2021, doi: [10.1109/TCYB.2021.3071860](https://doi.org/10.1109/TCYB.2021.3071860).
- [17] F.-G. Tang, Y. Liu, Y. Li, and Z.-W. Peng, "A unified multi-level spectral-temporal feature learning framework for patient-specific seizure onset detection in EEG signals," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106152, doi: [10.1016/j.knsys.2020.106152](https://doi.org/10.1016/j.knsys.2020.106152).
- [18] A. Temko, A. K. Sarkar, G. B. Boylan, S. Mathieson, W. P. Marnane, and G. Lightbody, "Toward a personalized real-time diagnosis in neonatal seizure detection," *IEEE J. Transl. Eng. Health Med.*, vol. 5, 2017, Art. no. 2800414, doi: [10.1109/JTEHM.2017.2737992](https://doi.org/10.1109/JTEHM.2017.2737992).

- [19] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021, doi: [10.1016/j.neucom.2020.09.017](https://doi.org/10.1016/j.neucom.2020.09.017).
- [20] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, MIT Division Health Sci. Technol., Harvard Univ., Cambridge, MA, USA, 2009.
- [21] N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal EEG recordings with seizure annotations," *Sci. Data*, vol. 6, no. 1, pp. 1–8, Mar. 2019, doi: [10.1038/sdata.2019.39](https://doi.org/10.1038/sdata.2019.39).
- [22] X. Zhang, L. Yao, M. Dong, Z. Liu, Y. Zhang, and Y. Li, "Adversarial representation learning for robust patient-independent epileptic seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2852–2859, Oct. 2020, doi: [10.1109/JBHI.2020.2971610](https://doi.org/10.1109/JBHI.2020.2971610).
- [23] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-time epileptic seizure detection using EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 2146–2156, Nov. 2017, doi: [10.1109/TNSRE.2017.2697920](https://doi.org/10.1109/TNSRE.2017.2697920).
- [24] S.-J. Huang, R. Jin, and Z.-H. Zhou, "Active learning by querying informative and representative examples," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1936–1949, Feb. 2014, doi: [10.1109/TPAMI.2014.2307881](https://doi.org/10.1109/TPAMI.2014.2307881).
- [25] Z. Du, Y. Tan, and L. Yang, "Cost-effective active learning with informativeness and representativeness for semantic segmentation," in *Proc. 12th Int. Symp. Comput. Intell. Design (ISCID)*, Hangzhou, China, Dec. 2019, pp. 37–40.
- [26] L. Frassinetti, D. Ermini, R. Fabbri, and C. Manfredi, "Neonatal seizures detection using stationary wavelet transform and deep neural networks: Preliminary results," in *Proc. IEEE 20th Medit. Electrotechn. Conf.*, Palermo, Italy, Jun. 2020, pp. 344–349.
- [27] R. Ramos-Aguilar, J. A. Olvera-López, I. Olmos-Pineda, and S. Sánchez-Urrieta, "Feature extraction from EEG spectrograms for epileptic seizure detection," *Pattern Recognit. Lett.*, vol. 133, pp. 202–209, May 2020, doi: [10.1016/j.patrec.2020.03.006](https://doi.org/10.1016/j.patrec.2020.03.006).
- [28] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Apr. 2019, Art. no. 031001.



XIAOGANG ZHU was born in Guiyang, Guizhou, China, in 1986. He received the B.S. (Clinical Medicine) degree in medicine from Guizhou Medical University, in 2009. He is currently working with the Neuroelectrophysiology Center, Affiliated Hospital of Guizhou Medical University, mainly responsible for EEG and EMG clinical examination.



YUFENG FAN received the M.S. degree in computer science and technology from Guizhou Normal University, China, in 2016, where he is currently pursuing the Ph.D. degree in computational mathematics. His research interests include image interpretation, data analysis, and machine learning.



CAIXIA SU was born in 1983. She received the master's degree from the North China University of Water Resources and Electric Power, China. She is currently a Lecturer and a Research Assistant with the Image Processing and Machine Vision Laboratory, School of Big Data and Computer Science, Guizhou Normal University, China. Her research interests include image processing and machine learning.



SHUN MA received the B.S. degree from the Wanfang Institute of Technology, Henan Polytechnic University, China, in 2018. He is currently pursuing the master's degree with the School of Big Data and Computer Science, Guizhou Normal University. His research interests include deep learning and reinforcement learning.



HAOJIE LIU received the B.S. degree from the School of Information Engineering, Yancheng Institute of Technology, Yancheng, China, in 2019, and the M.S. degree from the School of Big Data and Computer Science, Guizhou Normal University, Guiyang, China, in 2022. He joins, as an Algorithm Engineer, the Huzhou Institute, Zhejiang University, Huzhou, China, in 2022. His main research interests include person re-identification and object detection.



YONGFENG CAO received the B.S. and M.S. degrees in electronic engineering and the Ph.D. degree in communication and information system from Wuhan University, China, in 1999, 2001, and 2004, respectively. He is currently a Professor with the Image Processing and Machine Vision Laboratory, School of Big Data and Computer Science, Guizhou Normal University, China. His research interests include signal processing, image interpretation, pattern recognition, and machine learning.