



Better electrobiological markers and a improved automated diagnostic classifier for schizophrenia—based on a new EEG effective information estimation framework

Tianyu Jing¹ · Jiao Wang¹ · Zhifen Guo¹ · Fengbin Ma¹ · Xindong Xu¹ · Longyue Fu¹

Accepted: 30 June 2024 / Published online: 10 July 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Advances in AI techniques have fueled research on using EEG data for psychiatric disorder diagnosis. Despite EEG's cost-effectiveness and high temporal resolution, low Signal-to-Noise Ratio (SNR) hampers critical marker extraction and model improvement, while denoising techniques will lead to a loss of effective information in EEG. The aim of this study is to employ AI methods for the processing of raw EEG data. The primary objectives of the processing are twofold: first, to acquire more reliable markers for schizophrenia, and second, to construct a superior automatic classification for schizophrenia. To remove the noises and retain task-related (classification tasks) effective information mostly, we introduce an Effective Information Estimation Framework (EIEF) based on three key principles: the task-centered approach, leveraging 1D-CNNs' test metrics to gauge effective information proportion, and feedback. We address a theoretical foundation by integrating these principles into mathematical derivations to propose the mathematical model of EIEF. In experiments, we established a paradigm pool of 66 denoising paradigms, with EIEF successfully identifying the optimal paradigms (on two datasets) for restoring effective information. Utilizing the processed dataset, we trained a 3D-CNN for automatic schizophrenia diagnosis, achieving outstanding test accuracies of 99.94% on dataset 1 and 98.02% on dataset 2 in subject-dependent evaluations, and accuracies of 89.85% on dataset 1 and 98.02% on dataset 2 in subject-independent evaluations. Additionally, we extracted 38 features from each channel of both processed and raw datasets, revealing that 20.86% (dataset 1) of feature distribution differences between the patients and the healthy exhibited significant changes after implementing the optimal paradigm. We enhance model performance and extract more reliable electrobiological markers. These findings have promising implications for advancing the field of the clinical diagnosis and pathological analysis of Schizophrenia.

Keywords Schizophrenia · Automatic diagnosis · EEG signals · Denoising · CNN · Feature analysis

1 Introduction

Schizophrenia (SZ) is a chronic and intricate neuropsychiatric disorder characterized by symptoms such as blunted affect, hallucinations, and delusions. These debilitating symptoms lead to significant cognitive and, ultimately, social deficits [1], often resulting in individuals spending a substantial portion of their lives in psychiatric care facilities. The precise etiology of schizophrenia remains unclear, but it is believed to be influenced by a complex interplay of genetic, environmental, and psychosocial factors [2].

Electroencephalogram (EEG) is a biological signal characterized by high temporal resolution and low acquisition cost, allowing for the detection of changes in brain states [3]. Consequently, EEG has been widely utilized in the field AI for identifying various human mental states, such as drunkenness [4, 5], depression [6], and various emotions [7], etc. Clearly, the diagnosis of schizophrenia using EEG, the primary focus of this paper, is an ongoing area of research pursued by many researchers [8–10].

In the above context, when developing an automatic system for the diagnosis of SZ, we have several primary objectives. Firstly, our main goal is to create a system that can assist clinicians in diagnosing SZ effectively. Secondly, we aim to identify and validate interpretable features that have the potential to classify SZ patients accurately. Lastly, we endeavor to establish electrobiological markers, which are

Fengbin Ma, Xindong Xu and Longyue Fu contributed equally to this work.

Extended author information available on the last page of the article

essentially features, for SZ using AI techniques. This paper will address all three of these objectives, striving to provide solutions and insights into the diagnosis of SZ.

Regrettably, EEG data's inherently low SNR [11] poses a challenge to achieving the aforementioned objectives. Firstly, the presence of noise in EEG data dilutes the relevant information, consequently affecting the performance of trained classifiers. Besides, due to the limitation of the acquisition equipment, the scale of EEG datasets is generally not large. Thus, the situation easily happens that the noise components and the redundant components have different distributions in the two sub-datasets belonging to the healthy and the patients. For example, Fig. 1 shows the boxplots of the bubble entropies [12] of all the disjoint segments with 4 seconds in the dataset 1 mentioned below (the SZ segments and the healthy(HC) segments have their boxplot, respectively). It becomes evident that, in the raw data, the average and median entropy values for SZ patients are higher than those for healthy individuals. However, after the noise removal, these metrics for SZ patients drop below those for healthy individuals. Such a problem definitely hinders the achievement of identifying potential qualified features.

Moreover, training a automatic classification system using such datasets may mislead the system, causing it to extract noise as hidden layer features, ultimately leading to a decrease in the generalization ability of the system.

Because the EEG redundant signal has yet to be defined clearly, removing them is impossible at current. As for the noise, although more and more denoising approaches for EEG signals have been proposed in recent years [11, 13–15], these aforementioned issues still persist without complete resolution. Even worse, sometimes signal preprocessing operations significantly influence the outcomes of some specific classification tasks when applied to task-related datasets, potentially causing the loss of essential task-related information [16, 17].

To overcome the limitations of the aforementioned researches, this paper introduces an EEG effective information estimation framework(EIEF). EIEF is designed to be directly aligned with the classification task at hand and is tailored to a specific EEG dataset. The core mechanism of the framework is using the testing metrics of a trained end-to-end DNN to feed the stock and proportion of the effective information back to denoising approach selection. Let us take a testing metric as the objective function. Finding the optimal denoising paradigm for a specific EEG dataset and the corresponding optimal estimation of the SZ-related effective information of this dataset will become an optimization problem. Ideally, the optimal paradigm can remove a lot of noise components with the effective information retained, and enhance the reliability of the SZ electrobiological marker discovery and the generalization capacity of SZ classification systems based on the optimal dataset.

In this paper, the framework utilizes a 1D-CNN as the core component of the DNN, as introduced in [18]. But when applying the framework's result to construct an automatic SZ diagnosis system, we opt for the 3D-CNN proposed by [19] with a more substantial scale and increased depth to enhance system performance.

To sum up, this paper focuses on the establishment and validation of EIEF, the assessment of changes in EEG feature disparities between patients and healthy individuals before and after implementing the optimal denoising paradigm, and the construction of an automated SZ diagnosis system by EIEF. The paper is structured into three main parts: theory and methodology, experiments, and verifications. In the part of the theory and methodology, starting with discussing the properties of the EEG effective information of classification tasks, then relying on the estimation for that information, we propose EIEF, comprising the objective function, constraints and solution methods. As for the method, according to the conditions and solution of EIEF, after specifying the

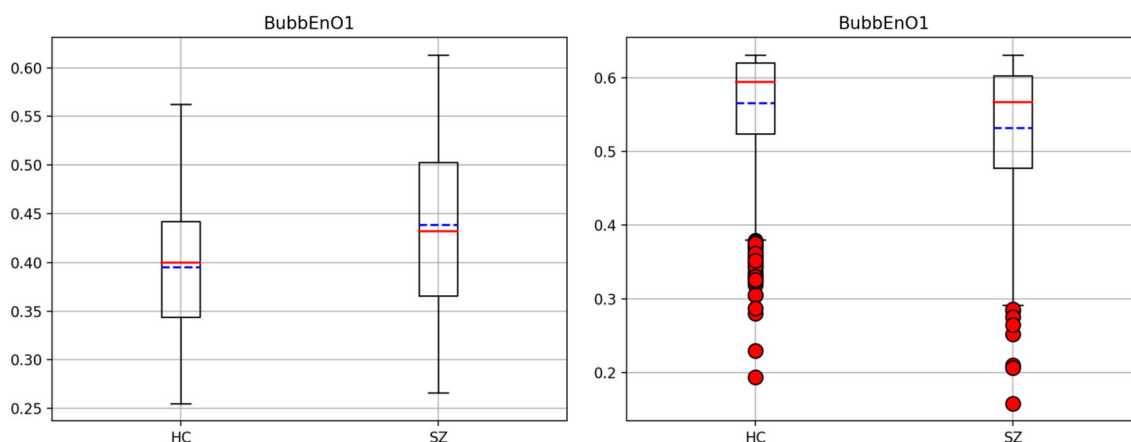


Fig. 1 The boxplots of the bubble entropies based on the dataset after removing line noise (left). The boxplots of the bubble entropies based on the original dataset(right). **SZ** schizophrenia patients. **HC** healthy control

two EEG datasets used, the research will create a paradigm pool consisting of 66 denoising paradigms, which is followed by the description of the two used CNNs. In the part of the experiments, for each dataset processed by each paradigm, we input segments of the processed dataset into the 1D-CNN for training and testing, using subject-independent cross-validation. We aim to find the two optimal paradigm for the two datasets within the pool, with test accuracy as the optimization objective. Subsequently, using the datasets processed by the optimal paradigms, we develop an automatic SZ diagnostic system based on the 3D-CNN and compare it with S.O.T.A. In the part of the verifications. Making the 3D-CNN replace the 1D-CNN, we will implement another search to verify the stability of the optimal paradigm and the strength of putting the 1D-CNN into the framework. Furthermore, the change in the features disparities between the two groups of individuals before and after the implementation is evaluated to elucidate the significance of the denoising paradigm.

2 Related Work

2.1 EEG noise removal techniques

EEG has a high resolution and the signals are prone to unwanted noise pollution, resulting in various artifacts [20]. Eye movement, blinking, heart activity, and muscle activity in EEG signals are the main physiological artifact types. Besides, there are many extrinsic artifacts existing, like the line noise and volume conduct artifact [11]. Traditional artifact removal techniques encompass regression methods, wavelet transformations, blind signal separation (BSS) methods, filtering methods, among others. In recent years, several new approaches, such as AI-based methods and hybrid methods, have emerged, demonstrating improved performance and reduced computational demands. Sweeney, Ward, and McLoone assumed that each channel was the accumulation of pure EEG data and a certain proportion of artifacts. The estimated artifacts were then subtracted from the EEG [21]. Gianluca Di Flumeri proposed the regression-based eye correction algorithm (REBLINCA) with a higher ability to retain the EEG signal in the no-eye movement part. Besides, the method does not require an EOG channel compared to other regression-based methods. Compared to ICA-based algorithms, this requires fewer channels and facilitates the calculation [14]. Bigdely-Shamlo et al. Introduced a robust referrals algorithm that attempted to estimate the actual average of EEG channels after removing bad channel contamination. Their efforts were to develop a standardized early-stage preprocessing pipeline (the PREP pipeline) that

detects and removes certain experimentally generated artifacts, such as eye blinks or muscle activations [13]. Banghua Yang proposed a novel blind source separation method called CCA-EEMD to remove EOG artifacts automatically as well as reserve more valuable information from raw EEG. A distinctive aspect of this method is that the identified EOG component is not removed directly but used to extract neural EEG data, which would keep more effective information [15]. Sadiq et al. used multiscale principal component analysis to decompose EEG signals, and employed Kaiser rule to select principal components to remove the noises [22]. Morteza Zangeneh Soroush introduced a novel method to detect artifactual components estimated by second-order blind identification (SOBI). Artifacts are detected using a mixture of well-established conventional classifiers and were removed employing stationary wavelet transform (SWT) to reserve neural information. This method combines signal processing techniques and machine learning algorithms, yielding significant results across various scenarios [23]. However, all the above methods are general and aren't task-centered. In other words, they are open-loop methods without feedback.

2.2 EEG-based SZ classification

From the raw EEG data to the final classification results, researchers' processing can be broadly categorized into the following steps. Firstly, data extraction is conducted, where commonly encountered EEG signals for disease classification include resting-state signals and task-related signals. Subsequently, data denoising is performed, as mentioned in the preceding section. Then, signal analysis, such as feature extraction, nonlinear signal decomposition, spectral analysis, and so forth. Finally, based on the analyzed data, classifiers are trained. Therefore, in this subsection, we will emphasize how researchers conduct signal analysis and classifier design in those prominent achievements.

More than a decade ago, Sabeti et al. [24] extracted Shannon entropy, spectral entropy, approximate entropy, Lempel-Ziv complexity and Higuchi fractal dimension from an EEG dataset (recorded data in resting-state with eyes opened), and achieved a classification accuracy of 86% and 90% obtained by LDA and Adaboost respectively. Parvinnia et al. [25] also used resting-state EEG signals with eyes opened to conduct research. After extracting fractal dimension, band power and autoregressive (AR) model, they applied weighted distance nearest neighbor (WDNN) for classification. And the accuracy was 95.3%. Murphy et al. [26] collected a task-related dataset from duration deviant MMN tasks, and found adolescents with psychotic symptoms were characterised by a reduction in MMN amplitude at frontal and temporal regions

compared to the controls through statistical analysis. During that period, researchers were directly extracting a few features from raw data and then using them to perform simply statistical analysis or train classical machine learning classifiers.

As research progresses, researchers are increasingly incorporating new technologies into the feature extraction process. For example, researchers can simultaneously extract features of several dozen different types at once and then use feature selection techniques to select a suitable subset, and use this subset to train a classifier. Jahmunah et al. [27] total mined 157 features from the dataset, and select 14 features using Student's t-test. Based on these feature, they implemented classification practice with various ML classifiers, DT, LD, KNN, PNN, and SVM with various kernels. And the average performance value is 92.91%. Prabhakar et al. [28] first extracted 9 nonlinear features and then optimized the selection of the features by Artificial Flora (AF) optimization, Glowworm Search (GS) optimization, Black Hole (BH) optimization, and Monkey Search (MS) optimization. They also trained several classifiers by the optimized features and found SVM-RBF can reach the best performance of 97.54% (for normal cases) and 92.17% (for schizophrenia cases).

In recent years, various signal decomposition techniques have been widely employed for state recognition based on EEG. Sadiq et al. achieved a sensitivity, specificity and classification accuracy of 93%, 92.1% and 91.4%, respectively, on a motor imagery dataset by utilizing a robust and simple automated multivariate empirical wavelet transform (MEWT) to obtain joint instantaneous amplitude and frequency components [29]. And achieved an average classification accuracy of 99.8% by employing a multivariate variational mode decomposition (MVMD) method to obtain joint modes in frequency scale across all channels [30]. In the realm of schizophrenia recognition, researchers have also begun innovating feature engineering from the perspective of signal decomposition. Krishnan et al. [31] used Multivariate Empirical Mode Decomposition (MEMD) to decompose the EEG data into Intrinsic Mode Functions (IMF) signal. Then five entropy measures were measured from the IMF signals. And the subset of features was selected by Recursive Feature Elimination. Based on Radial Basis Function (SVM-RBF), they achieved the highest accuracy and F1-score of 93% with 95 features and obtained an AUC of 0.9831. Baygin [32] conducted feature extraction from 19-channel EEG signals with healthy and schizophrenia classes, using Tunable Q-Factor Wavelet Transform (TQWT) and statistical moment methods, and selected feature subset by the ReliefF method. He chose KNN to be the classifier and achieved an accuracy of 99.12%. Khare et al. [33] used the Fisher score method to select the most discriminant channel, then used

flexible tunable Q wavelet transform (F-TQWT) to decompose the EEG signal. After the decomposition, similar to the aforementioned researches, they extracted five features and employed the Kruskal-Wallis test to select a subset of features. Subsequently, this subset was fed into an flexible least square support vector machine (F-LSSVM) classifier. In their paper, a more innovative approach involved utilizing the grey wolf optimization algorithm to incorporate feedback from SVM results into the selection of Q-wavelets. An accuracy of 91.39%, sensitivity, specificity, precision, F-1 measure, false positive rate and error of 92.65%, 93.22%, 95.57%, 0.9306, 6.78% and 8.61% was achieved.

In addition to intensive research in feature engineering, with the continuous breakthroughs in deep learning technology, researchers have also begun to utilize various types of DNNs for schizophrenia recognition. One category of research involves directly feeding continuous or segmented EEG signals into the network, for example: Oh et al. [18] introduced a 1D-CNN model designed to analyze signals, automatically extract salient features, and perform classification. This model achieved a classification accuracy of 98.07% for subject-dependent (SD) evaluation and 81.26% for subject-independent(SI) evaluation. Sharma et al. [34] proposed a schizophrenia hybrid neural network (SzHNN), which is a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM). They divided the original data of two EEG datasets into non-overlapping segments and used these segments to train the SzHNN. The performance is an accuracy of 99.9% on dataset 1 and an accuracy of 99.5% on dataset 2. Another category of research involves the fusion of feature engineering with DNNs. Leveraging the scale of DNNs, such papers often yield large-sized feature sets in their feature engineering, such as the visualized image of EEG signals. Shen et al. [35] developed an image feature, functional brain network, using a multivariate autoregressive model and coherence connectivity algorithm. And they used 3D-CNN to classify the SZ patients. The proposed 3D-CNN method achieved the performance of a $98.47 \pm 1.47\%$ in accuracy, $99.26 \pm 1.07\%$ in sensitivity, and $97.23 \pm 3.76\%$ in specificity. Similarly, Khare et al. [36] captured the instantaneous information of EEG signals in the time-frequency domain using MH-TFD, converted the information to two-dimensional plots, and fed the plots to the developed CNN model. The developed CNN is SchizoNET model. And the proposed model achieved an accuracy of 97.4%, 99.74%, and 96.35% on the three datasets, respectively. Zülfiqar et al. [37] integrated Empirical Mode Decomposition (EMD) with the VGG16 pre-trained CNN. HS (Hilbert Spectrum) images of the first four Intrinsic Mode Functions (IMF) components obtained by applying EMD to EEG signals were fed into several famous CNN.

They obtained the classification performance of 98.2% for Dataset I and 96.02% for Dataset II, using VGG16 network.

The innovation in the aforementioned articles includes the introduction of new features and new methods of feature acquisition (such as signal decomposition), the introduction of new feature subset selection methods, the introduction of new classifiers, etc. However, none of these articles focus on improving the reliability of existing features, which is the focal point of our research.

3 Theory and method

The primary objective of this chapter is to leverage mathematics to introduce an *effective information estimation framework* that can get the optimum among a given series of denoising paradigms, based on any end-to-end classification model and a certain dataset. Then according to the conditions of EIEF, after specifying the EEG dataset used, the research will create a paradigm pool consisting of 60 denoising paradigms for the following grid search for the optimum. And, the two classifiers inside EIEF and outside EIEF will be detailed, plus the classifier inside serves as the foundation for assessing the metrics used in the paradigm search, and the classifier outside is responsible for constructing the automated SZ diagnosis system.

Here, the flow chart from EIEF's work to the construction and the use of the diagnosis system is illustrated in Fig. 2. And Table 1 is the list of symbols used in the thory part.

3.1 Hypotheses on property of effective information

Let $Z \in \mathbb{R}^{c \times t}$ denote the $c \times t$ -dimensional matrix variable, and $X \in \Theta \subset \mathbb{R}^{c \times t}$ denote the observed EEG signal sample, where c is the number of recorded channels, t is the sample size and Θ is the sample space. The purpose of this subsection is to describe a common property of EEG effective information for any classification task with mathematic. So first, in order to denote the effective information, we shall explain and denote EEG signal objective components.

Here, we denote an EEG signal objective component as an $i \in I_s$ which must satisfy the condition: $\forall X, \exists S^i \in \Theta^i \subset \mathbb{R}^{c \times t}$, where S^i can be called the sample of component i (or sample of t), I_s is the tag set, Θ^i is the component space generated by Θ . For now, the elements that are objective components inside I_s are not all clear, but it is clear that I_s is a finite set. Then denote a sample of the real signal under any reference λ as $S^r \in \Theta^r$, and a sample of the real noise is $S^{rn} \in \Theta^{rn}$, plus $\forall X, \exists S^r, S^{rn} : X = S^r + S^{rn}$. When the classification task is $T \in \Gamma$ (Γ is EEG classification task set), a sample of the effective information of T is S^{iT} , the redundant information S^{iTm} , plus $\forall S^r, \exists S^{iT}, S^{iTm} : S^r = S^{iT} + S^{iTm}$.

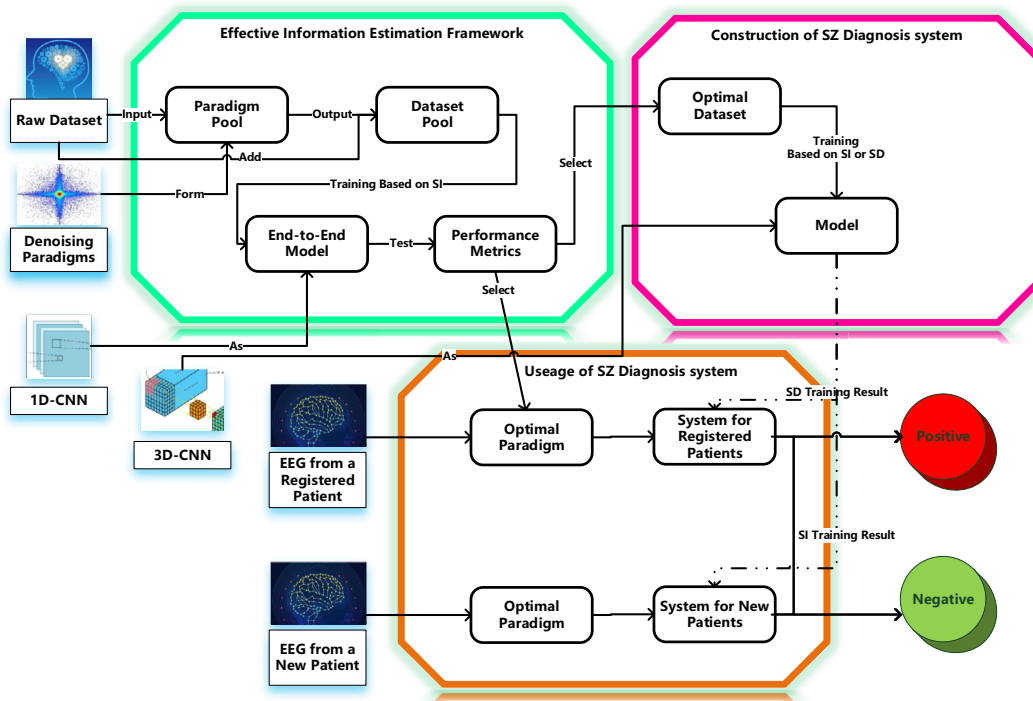


Fig. 2 The flow chart from EIEF to the diagnosis system

Table 1 List of Symbols

Type	Symbol	Explanation	Type	Symbol	Explanation
Variable	c	Recorded channel number	Set	Θ	Sample space
	t	EEG sample size		Θ^i	Component space of an i
	Z	$c \times t$ -dimensional matrix		Γ	Classification task set
	X	Observed EEG signal sample		G_i	Set of all g_i based on an i
	i	Tag of a objective component		G	Set of all g
	S^i	Sample of an i		$\bar{\Theta}_{g_i}$	Complete observed sample space based on a g_i
	λ	Signal reference		$\bar{\Theta}$	Family of all $\bar{\Theta}_{g_i}$
	T	Classification task		F^i	Set of all f_{g_i} based on an i
	i^T	Tag of the effective Information of a T		F	Set of all f
	S^{iT}	Sample of the effective information of a T		H	Classifier hypothesis space
	S^{iTm}	Sample of the redundant information		H_{re}	Set of all realizable hypotheses
	C	Parameter matrix of the prior information		B_j	Set of all b of a j
	\bar{i}	Sample size of a \bar{Z} inputted to classifiers		B	Set of all b
	\bar{Z}	Block of a corresponding Z		H_j	Set of all h corresponding to a B_j
	cl	Number of the classes decided by the T		I_b	Tag set containing all j
	\hat{y}	Output of classifiers		I_s	Tag set containing all i
	b	Classification model		Y^T	Classification label space of a T
	j	Tag of a classification model category		I_r	Tag set containing all i of real signals
	θ_j	Parameters of models in a B_j		I_{rn}	Tag set containing all i of real noises
	ω_j	Number of the parameters of a B_j		I_{else}	Tag set containing all i : $i \in I_s \wedge i \notin I_r \wedge i \notin I_{rn}$
	y	Real label vector of a sample		P	Function space of p
	k	Index of a block inputted to classifiers		A	Set of all a
	v	Index of a classification label vector		A^i	Set of all a qualified to estimating the f_i based on an i
	a	EEG processing algorithm		P^i	Set of all p corresponding to an A^i
	n	Tag of a processing method category		A_n^i	Set of all a belonging to an n
	θ_n	Parameters of algorithms in an A_n^i		I_a^i	Tag set of all n qualified to estimating an i
	ω_n	Number of the parameters of an A_n^i		P_n^i	Set of all p corresponding to an A_n^i
	r	Real signal under a λ , type of objective components		Θ^*	Sample set of a research
	rn	Real noise under a λ , type of objective components		Y^{T*}	Classification label set of a research
	δ	Traning termination condition		E^*	Example set of a research
$\theta_{j^*}^*$	Optimal parameters of the h_{j^*} inside EIEF	Θ^{iT*}	Research's component set of an i		
$\theta_{\hat{f}_{iT}}$	Optimal parameters of the p inside EIEF	E_v^*	Example subset of a research		
Map	g_i	Constructed function based on an i	Distribution	E_{test}^*	Example train set of a research
	f_{g_i}	g_i -based restoration function for an i		E_{train}^*	Example test set of a research
	h	Classifier		D^g	Distribution of an example space $\bar{\Theta}_{g_i} \times Y^T$
	\mathcal{P}	Probability density function of a D^g		D_i^T	Distribution of an example space of an i and a T
	p	Processing method of EEG		$D^g(y = y_v)$	Distribution of an example subspace
	L	Loss function			
	\mathcal{M}	Metric function, such as loss function			
	\hat{f}_{iT}	Optimal estimation of an f_{iT}			
f_{iT}	Restoration function for an i^T				

Note that $\forall i, \exists X$ which has different corresponding S^i , but in one real sampling, there is only one S^i for any possible sampled X . That's because the prior information during the sampling, outside the Θ , decides the unique S^i . Conversely, if having enough prior information, we can construct a X using $\forall S^i$. In here, a mathematical model will be proposed to demonstrate the prior information and to achieve the construction.

Definition 1 $\forall i$, define a mapping and matrix function

$$g_i(\in G_i \subset G) : \mathbb{R}^{c \times t} \rightarrow \mathbb{R}^{c \times t}, C \rightarrow C_{g_i} = g_i(C) \quad (1)$$

as a *constructed function* based on i if $\forall X, \exists S^i, C : X = S^i + g_i(C)$ and $\forall S^i, \exists X, C : X = S^i + g_i(C)$, where C is called the parameter matrix of the prior information of X expect S^i , G_i is the function set containing all qualified *constructed functions* based on i , and G is the function set containing all g . Typically, $\forall i$, a identical mapping must be a *constructed function*. And in this case, since $\forall X, \exists S^i, C$ shall be $X - S^i$. It is easy to use any *constructed function* to define a new sample space.

Definition 2 $\forall i, \forall g_i$, the space $\bar{\Theta}_{g_i} \subset (\mathbb{R}^{c \times t}, \mathbb{R}^{c \times t})$ is the complete observed sample space based on g_i if $\forall (Z_1, Z_2) \in \bar{\Theta}_{g_i} : Z_1 \in \Theta$ and $\forall (X, C) \in \bar{\Theta}_{g_i}, \exists S^i : S^i = X - g_i(C)$. Note that $\bar{\Theta}_{g_i} \in \bar{\Theta}$, where $\bar{\Theta}$ is the family of all $\bar{\Theta}_{g_i}$.

Through complete observed sample spaces, getting the effective information of $\forall T$ will be implemented by a mapping.

Definition 3 $\forall i, \forall g_i$, define the mapping

$$f_{g_i}(\in F^i \subset F) : \bar{\Theta}_{g_i} \rightarrow \mathbb{R}^{c \times t}, (X, C) \rightarrow S^i = f_{g_i}(X, C) = X - g_i(C) \quad (2)$$

as the g_i -based *restoration function* for i , where F^i is the function set containing all f_{g_i} for component i based on all possible g_i , and F is the set of all f for all components. Obviously, $\forall T, f_{g_{iT}}$ shall be the *restoration function* for i^T . So far, $\forall T$, through $f_{g_{iT}}$, any S^{iT} can be gotten from Θ . But the property we propose to describe is related to classifiers, which requires the definition of classifiers.

Definition 4 In this paper, a *classifier* for EEG is

$$h(\in H_{re} \subset H) : \mathbb{R}^{c \times \bar{t}} \rightarrow \mathbb{R}^{\bar{c}}, \bar{Z} \rightarrow \hat{y} = h(\bar{Z}), \quad (3)$$

where H is the hypothesis space, $\bar{Z} \in \mathbb{R}^{c \times \bar{t}}$ is a block of the corresponding Z , or a subsample of the corresponding X , \bar{t} is the sample size of the subsample ($\bar{t} \leq t$), \bar{c} is the number of the classes decided by task T , and \hat{y} is the output vector describing the probabilities that \bar{Z} belongs to each class.

There is a fact that h demands the segmented EEG epochs to be the input, which tells us the *classifiers* here are not the traditional classification models but actually cover the EEG signal process, feature extraction and selection, and feature dimensionality transformation. Thus, H_{re} is the hypothesis set containing all realizable hypotheses, and these hypotheses must come from models where the above procedures are all learnable (end-to-end models). For instances, an SVM classifier based on manual feature extraction doesn't belong to H_{re} , and a CNN classifier belongs to H_{re} if taking the segmented data as the input and the probability vector as the output. Such a definition could avoid setting priori hypotheses in classifier construction as much as possible, since the priori hypotheses will definitely influence classifiers' sensitivity to the component change of an EEG trainset.

Before the proposal of the hypothesis, there are some final explanations. Assume a *classifier* which can be represented by any of model $b \in B_j \subset B$ is $h_j(\cdot; \theta_j) \in H_j \subset H_{re}$ ($j \in I_b$ and $\forall j \in I_b : ((\exists! B_j \subset B) \wedge (\exists! H_j \subset H_{re}))$), where B_j is the set containing all the models of a specific model category (like 3D-CNN with definite hyper-parameters), B is the set containing all the end-to-end models for classification, and H_j is the set of *classifier* corresponding to B_j , $\theta_j \in \mathbb{R}^{1 \times \omega_j}$ is the parameter vector of models belong to B_j where ω_j represent the number of the parameters.

$\forall T \in \Gamma, \forall g_{iT} \in G_{iT}$, Let $\bar{\Theta}_{g_{iT}} \times Y^T$ represents the complete observed example space based on g_{iT} , where $Y^T \subset \mathbb{R}^{\bar{c}}$ is the label space of T , and $\forall ((X, C), y) \in \bar{\Theta}_{g_{iT}} \times Y^T$ represents a observed example and a random variable from the space, where y is the real label of the sample. Thus, we can use $\forall ((X, C), y) : ((X, C), y) \sim D^S$ to demonstrate that $\bar{\Theta}_{g_{iT}} \times Y^T$ follows a unknown prior distribution. Besides, $\forall i \in I_s$, there is a component i example space $\Theta^i \times Y^T$, and $\forall (S^i, y) : (S^i, y) \sim D_i^T$.

Then, let $\{I_r, I_{rn}, I_{else}\}$ denotes a partition of I_s , which $\forall i \in I_r$ must represents a objective component of real signal and $\forall i \in I_{rn}$ must represents a objective component of real noise. Now, it's time to present the hypothesis.

Hypothesis 1 $\forall T$, the i^T named effective information for T has a property that $\forall j \in I_b$,

$$i^T \in \arg \min_{i \in I_r} \min_{h_j \in H_j} \int_{\bar{\Theta}_{g_i} \times Y^T} \left(\frac{1}{n} \sum_{j=1}^n L(h([f_{g_i}(X, C)]_k), y) \right) \mathcal{P}((X, C), y) d(X, C) dy, \quad (4)$$

where $\bar{\Theta}_{g_i}$ is a complete observed example space based on $\forall g_i \in G_i$, $[f_{g_i}(X, C)]_k$ is the k -th block of $f_{g_i}(X, C)$ (from left to right, if not belonging to $\mathbb{R}^{c \times \bar{t}}$, the last will not be involved), n is the number of the blocks belonging to

$f_{g_i}(X, C)$, L is the loss function, $\mathcal{P}((X, C), y)$ is the probability of sampling $((X, C), y)$ from D^g . Equation (4) is essentially a variant of the formula used to calculate the generalization error of a classification model [38]. This hypothesis means, among the classifiers with the best generalization abilities that are found based on all EEG components belonging to I_r , the best classifier based on the effective information for T must be the best of the best, no matter which kind of model H_j with definite hyper-parameters is chosen for classification.

To extend *Hypothesis 1* to $\forall i \in I_s$, there are still some works requiring to be done. Another hypothesis is the next.

Hypothesis 2 There is an intuitive and important view that $\forall i \in I_{else}$ shall be composed by an $i_1 \in I_r$ and an $i_2 \in I_{rn}$ at least.

Keep going, $\forall T \in \Gamma$, $\forall g_{iT} \in G_{iT}$, when the label of $\bar{\Theta}_{g_{iT}} \times Y^T$ is controlled to a specific label, the complete observed example space degenerates to the complete observed example subspace which is denoted as $\bar{\Theta}_{g_{iT}} \times \{y_v\}$ ($v \in \{1, 2, \dots, cl\}$), and $\forall v : \mathcal{V}((X, C), y_v) \in \bar{\Theta}_{g_{iT}} \times \{y_v\}$. Similarly, use $((X, C), y_v) \sim D^g (|y = y_v)$ to demonstrate that $\bar{\Theta}_{g_{iT}} \times \{y_v\}$ follows a unknown prior distribution, and use $\mathcal{V}(S^i, y_v) : (S^i, y_v) \sim D_i^T (|y = y_v)$ for the same reason.

Finally, based on the two Hypotheses, a corollary is proposed to show that, in some conditions, The restriction in *Hypothesis 1* that i^T is the best only among I_r could be removed.

Corollary 1 $\forall T \in \Gamma, i \in I_{rn}, \alpha, \beta \in \{1, 2, \dots, cl\}$, if the following condition holds:

$$D_i^T (|y = y_\alpha) = D_i^T (|y = y_\beta), \tag{5}$$

then for $\forall j \in I_b$, the following formula holds:

$$F^{i^T} \subset \arg \min_{f \in F} \min_{h_j \in H_j} \int_{\bar{\Theta}_g \times Y^T} \left(\frac{1}{n} \sum_{j=1}^n L(h([f(X, C)]_k), y) \right) \mathcal{P}((X, C), y) d(X, C) dy, \tag{6}$$

where $\bar{\Theta}_g$ is the complete observed example space based on the unique g corresponding to the f . Since the paper is mainly focused on the application, here we just give the idea of the proof of the corollary.

Idea of proof: First, it is obvious that the generalization error of $f \in F^i$ won't be the minimum if $i \in I_{rn}$, because, for classifiers, there is no way to effectively distinguish

the classification of the data which has the same distribution between the different example subspaces. Second, if $i \in I_{else}$, assuming $\exists \bar{i}$ which lets $f \in F^{\bar{i}}$ reach the minimum, then the $f_{\bar{i}} \in F^{\bar{i}}$ (\bar{i} represents the component that satisfies for $\bar{i}, \exists i^{rn} \in I_{rn}, \forall X \in \Theta : f_i(X) = f_{\bar{i}}(X) + f_{i^{rn}}(X)$) and \bar{i} belongs to I_r . *Hypothesis 2* ensures the exist of \bar{i}) will definitely reach that minimum or a lower generalization error, which means that i must reach the minimum together with its \bar{i} . Thus, it is clear that the generalization error of f_i won't less than that of f_{i^T} . Then finally, according to hypothesis 1, the minimum will be obtained when $f \in F^{i^T}$ ($f = f_{i^T}$) if $i \in I_r$.

3.2 EEG effective information estimation framework

From the above deduction, it is assumed that, $\forall T, i^T$ shall be the component that could make every end-to-end classifier trained to get the strongest generalization ability. Therefore, for a specific task and a specific dataset, if finding a way to seek out an EEG processing method that leads an end-to-end classifier to get the optimal test result, we shall consider that method as the optimal estimation of f_{i^T} , and the corresponding processed dataset as the optimal estimation of the effective information for that task. Based on the above, we propose the *effective information estimation framework*.

Making some necessary preparations is still the first.

Definition 5 Define a mapping and matrix function

$$p(\in P) : \mathbb{R}^{c \times t} \rightarrow \mathbb{R}^{c \times t}, Z \rightarrow Z_p = p(Z) \tag{7}$$

as a *processing method* of EEG data, where Z_p is the image of Z under p and P is the function space. obviously, different *processing methods* can be composed, and denoted as $p^* = p_1 \circ p_2$.

Assuming an algorithm that can be applied to estimating the *restoration function* for $\forall i (\forall f)$ is $a \in A^i$, there is a true proposition denoted as $\forall i \in I_s : ((\exists! \Theta^i) \wedge (\exists! F^i \subset F) \wedge (\exists! A^i \subset A) \wedge (\exists! P^i \subset P))$, where A is the set composed of all the algorithms which can be applied to processing EEG signal, A^i is the set composed of all the algorithms which can be applied to estimating f_i , and P^i is the function set corresponding to A^i . Next, for $\forall i$, let $p_n^i(; \theta_n) \in P_n^i, n \in I_a^i$ denote a *processing method* that can be represented by any $a \in A_n^i$, where A_n^i is the set containing all the algorithms of a certain algorithm category (like re-reference), I_a^i is the index set that $\forall n \in I_a^i : ((\exists! A_n^i \subset A^i) \wedge (\exists! P_n^i \subset P^i))$, $\theta_n \in \mathbb{R}^{1 \times \omega_n}$ is the parameter vector of the algorithms belonging to A_n^i where ω_n represent the number of the parameters, P_n^i is the set of *processing methods* corresponding to A_n^i (per $a \in A_n^i$ has its paired θ_n).

For a EEG based research, if the following conditions hold:

1. The research contains a classification task $T \in \Gamma$
2. The research has a specific and finite observed sample set $\Theta^* \subset \Theta$ and a label multiset Y^{T*} corresponding to Θ^* ($\forall y \in Y^{T*} : y \in Y^T$) which compose an example set $E^* \subset \Theta \times Y^T$. The data process of this research aims to draw $S^{iT} \in \Theta^{iT*}$ for $\forall X \in \Theta^*$.
3. The number of the kinds of the selected algorithms (A_n^{iT}), and the composition order of the *processing methods* corresponding to selected algorithms are clear, which means the composite *processing methods* is denoted as

$$p(; \theta_p) = p_{n_q}^{iT} (; \theta_{n_q}) \circ p_{n_{q-1}}^{iT} (; \theta_{n_{q-1}}) \circ \dots \circ p_{n_1}^{iT} (; \theta_{n_1}), \tag{8}$$

where $p_{n_m}^{iT} (; \theta_{n_m})$ is a *processing methods* that is represented by an a with specific θ_{n_m} uniquely, $\theta_p = (\theta_{n_1}, \theta_{n_2}, \dots, \theta_{n_q}) \in \mathbb{R}^{1 \times \sum_{m=1}^q \omega_{n_m}}$ where q is the total number of the kinds of the selected algorithms.

4. The selected model category $B_j \subset B$ (and its paired $h_j (; \theta_j)$) for classification is clear, and denoted by $B_{j^*} (h_{j^*} (; \theta_{j^*}))$.
5. As much as possible, all kinds of noises ($i \in I_{rn}$) have the same distribution between the different example subsets. If not, the distribution difference of any noise must not cause the classifier trained by the dataset consisting of that noise to perform better than the classifier trained by the dataset consisting of S^{iT} . An example subsets is denoted by $E_v^* (v \in \{1, 2, \dots, cl\})$ and $(\forall v, \forall (X_1, y_1), (X_2, y_2) \in E_v^*) : y_1 = y_2$.
6. The testset E_{test}^* and the trainset E_{train}^* are split based on SI strategy.

Then an objective function could be proposed to estimate f_{iT} :

$$(\theta_{\hat{f}_{iT}}, \theta_{j^*}^*) = \underset{(\theta_p, \theta_{j^*})}{\text{arg min}} \mathcal{M}(E_{test}^*, h_{j^*} (; \theta_{j^*}), p (; \theta_p));$$

$$s.t. \quad \sum_{(X, y) \in E_{train}^*} \left(\frac{1}{n} \sum_{j=1}^n L(h_{j^*}([p(X)]_k), y) \right) < \delta. \tag{9}$$

Note that \mathcal{M} is the metric function which is determined by what metric we emphasize, and the selected metric must relate to the loss (like accuracy, recall, etc). For example, the function value will be calculated through the formula in the above s.t. (but calculated on testset), if the selected metric function is loss function. Besides, the constraint is designed to simulate the training process. δ is the reflection of the training termination condition. $\theta_{j^*}^*$ denotes the optimal parameter vector of $h_{j^*} (; \theta_{j^*})$, and $\theta_{\hat{f}_{iT}}$ is the optimum of

$p (; \theta_p)$. Let \hat{f}_{iT} represent the optimal estimation of f_{iT} , then through (9), it is obvious that $\hat{f}_{iT} = p (; \theta_{\hat{f}_{iT}})$.

So, (9) is the objective function of EIEF, and its preconditions is the preconditions of EIEF. It shall be especially reminded that the first 5 conditions are obvious or based on the statement of the last subsection, but only condition 6 is not mentioned before and will be explained below.

In our paper, the solution method of EIEF (the solution method of (9)) is the following: At a resolution ratio, we perform grid processing on the domain of $p (; \theta_p)$, and in each grid, the problem degenerates into a *classifier* training problem which could be solved by the standard model training process and provide a best metric function value. By the comparison of the best metric function values among different grids, finding the best grid corresponding to the \hat{f}_{iT} will be easy. Note that, from *Corollary 1* we can demonstrate that under ideal conditions, the optimal processing method found through all DNNs will remain consistent. However, in real experiments, achieving ideal conditions is often challenging due to various constraints, which will be discussed in the analysis of the experimental results.

Reasonably, the following part of this chapter is going to describe the research preconditions demanded by EIEF, like the dataset, the denoising approaches considered, the approaches' composition order to form the paradigms, and the classification model.

3.3 Datasets

3.3.1 Dataset 1

The raw dataset [39] used in this experiment collected 14 paranoid SZ, 7 males and 7 females respectively, who were collected from the Institute of Psychiatry and Neurology in Warsaw, Poland. At the same time, healthy subjects of the same age and sex ratio were recruited from the same institute. Each participant provided informed consent to participate in the study upon receiving the study protocol. The participants remained relaxed with their eyes closed when collecting EEG signals, and the sampling rate was 250Hz for 15 minutes. Data was collected via the typical International 10-20 System to obtain 19 channels. The electrodes used were Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2.

3.3.2 Dataset 2

To demonstrate that the proposed method has good generality on both ample and small datasets, it is insufficient to validate using only one dataset. So, the second dataset selected should ideally encompass a larger number of individuals [40].

The dataset comprising 45 SZ individuals and 39 healthy individuals was collected and established by Moscow University [41]. The Mental Health Research Center (MHRC) confirmed the diagnoses of all patients, which involved 45 boys with schizophrenic disorders (infant schizophrenia and schizotypal and schizoaffective disorders (F20, F21, and F25 according to the ICD-10)) with similar symptoms. During the examination at the MHRC, none of the enrolled patients received chemotherapy. The patients' ages varied from 10 years and 8 months to 14 years. The control group consisted of 39 healthy schoolboys aged from 11 years to 13 years and 9 months, with a mean age of 12 years and 3 months in both groups.

EEG recordings were obtained from 16 electrodes placed according to the international 10-20 system at O1, O2, P3, P4, Pz, T5, T6, C3, C4, Cz, T3, T4, F3, F4, F7, and F8, and monopolarly referenced to coupled ear electrodes, in wakeful relaxed adolescents with closed eyes. The sampling rate is 128 Hz, and the time of one trial is 1 minute.

3.4 Denoising paradigms

A total of 66 paradigms, categorized into five denoising approach categories, which can be considered as 66 grids to deal with the training problem, are implemented on the raw data. This section will provide a detailed specification of these paradigms.

The first category which is also the initial step in the EEG processing pipeline is the bad channels process. In this step, three approaches are available: bad channel retention, bad channel interpolation, and bad channel removal. This approach category is founded on a bad channel detection algorithm, specifically, an iterative detection method proposed by [13]. The removal involves replacing the columns of observed samples identified as bad channels with zero vectors, while the interpolation refers to spherical interpolation. Note that the three choices reflect considerations regarding the resolution ratio mentioned in the previous chapter.

The second procedure is the re-reference process, which is still based on the methodology outlined in [13]. Two options are presented here: performing robust re-referencing or retaining the original reference. Note that performing robust re-reference is available only when bad channel retention hasn't been carried out, because the re-reference is dependent on the preceding bad channel interpolation step.

Next, we move on to the third procedure—the filtering process. By default, line noise removal (LNR) is carried out using the EEGLAB plugin CleanLine in MATLAB [42, Chapter 7.3.4]. This choice is made because CleanLine is a robust plugin that minimizes the risk of losing genuine EEG signals. Moreover, there are two options left in this

procedure—LNR+ 1 HZ high pass filter, and LNR+1-50 HZ band pass filter. Those filters are FIR filters constructed by EEGLAB tool-basic FIR filter, with the default parameters inputted (but unchecked removal bad channels).

The following step is the bad epoch process. Depending on the choice made, certain blocks of observed samples may be replaced by zero blocks (if bad epoch removal is selected), or these blocks will remain unchanged (if bad epoch retention is selected). The bad epoch (block) detection is implemented using the EEGLAB plugin-clean raw data, which is based on Artifact Subspace Reconstruction [43]. The plugin is used with its default settings.

In the final category, we address the decomposition process, presenting three distinct options. If ICA with artifactual components removal is selected, fast ICA will be performed on the dataset which has previously undergone processing through the four procedures [44, 45]. Then, with the assistance of the EEGLAB plugin-classify components, the components identified as non-brain will be systematically removed. Then, if MSPCA with Kaiser-rule-based principal components removal (level 5, wavelet “sym4”) is selected, the dataset will undergo Discrete Wavelet Transform (DWT) and principal components analysis (PCA) for decomposition, and the principal components with eigenvalues lower than the average eigenvalue will be removed [22]. Note that in order to reduce the computational complexity, this option is only available when the reference retention and the bad channel retention are selected in advance. Finally, if retention is selected, no data will be changed in this procedure.

After the specification, let's provide a straightforward rationale for the selection and arrangement of the preprocessing steps.

Generally, the preprocess of raw EEG can be broken down into several essential procedures, including filtering, re-referencing, resampling, processing bad channels, removing bad data epochs, performing ICA with bad artifactual components removal, segmentation, baseline correction, etc. The order of these procedures can vary depending on the specific analysis.

Because used to analyze the event related potential (ERP) especially, the baseline correction is not involved in our study. Meanwhile, our segmentation is unrelated to ERPs; instead, it serves as a means to prepare EEG signals for input into the neural network. Besides, to avoid sample size reduction, the resampling is not concerned.

Concerning the remaining five procedures, our choices are well-grounded and have been recommended or introduced by the EEGLAB documentation. Our re-reference and bad channel process are both cited from [13] which emphasizes that their pipeline represents an early-stage operation, and where the bad channel interpolation is demanded to be performed

initially to ensure the robustness of the average re-reference. Then the filter we implement is a standard FIR filter (the LNR has been introduced clearly). Here, the high-pass filter is configured with a cut-off frequency of 1Hz to effectively remove linear trends, while the 50Hz cut-off frequency is set to eliminate high-frequency noise, such as Muscle Artifacts, while retaining the primary EEG frequency bands (α , β , θ , and δ waves). The choice to place the filtering operation third in the sequence is intended to reduce computational demands for subsequent steps. Next, it is desirable to perform the bad epoch process if there are considerable data that suffer from high amplitude noises, and the ASR is used since it is a mature plugin. At last, just costing a little calculation power, the fast ICA is a quite common and outstanding method to address various types of EEG noise. Regarding MSPCA, compared to fast ICA, it offers faster signal processing, and denoising based on MSPCA does not require the signal source classification performed by the neural network. However, this method's operational process may be less straightforward compared to the method based on ICA. The two methods' generality makes us place them as the final step so that we can leave the special noises to the corresponding professional tools.

3.5 Model inside EIEF-1D-CNN

Although *Corollary 1* indicates that every end-to-end classifier is qualified to seek out the optimal denoising paradigm. There are still many classifiers that aren't reliable. Some classifiers may lack sensitivity to noise variations, making the task of distinguishing the optimal paradigm challenging. Others might require excessive training time, potentially compromising the practicality of the EIEF. To avoid these obstructions, The 1D-CNN proposed by [18] is selected to be the classifier inside EIEF. This selection is based on several

key advantages, including the network's compact size and minimal trainable parameters. These characteristics endow the 1D-CNN with heightened sensitivity while keeping training times relatively short, ensuring both the reliability and practicality of the EIEF. Table 2 details the layers of the Net. For dataset 1, single trials recorded EEG signals for 15 minutes at a rate of 250 Hz, while for dataset 2, single trials recorded EEG signals for 1 minute at a rate of 128 Hz. To balance the sample sizes, all 67 datasets corresponding to dataset 1 will be divided into irrelevant epochs with the size of 19×6250 each. Conversely, datasets corresponding to dataset 2 will be divided into irrelevant epochs with the size of 19×512 each. Since dataset 2 has three fewer channels than dataset 1, the data of 3×512 in each dataset 2 epoch are padded with zeros. Consequently, the processed datasets for dataset 1 roughly comprise 1131 samples each, while those for dataset 2 roughly comprise 1260 samples each (the quantity may fluctuate due to the bad epochs removal). This approach ensures that dataset 2's segmented samples are not disproportionately fewer.

3.6 Model for diagnosis system-3D-CNN

The properties of the 1D-CNN make it reasonable that being selected by EIEF, but will cause low stability and generalization. So when pursuing the first objective outlined in the introduction-constructing an automatic SZ diagnosis system, if intending to apply the framework's result to the construction, we shall select a DNN with a bigger scale to make the system perform better. Meanwhile, *Corollary 1* tells us that the optimal dataset determined using the 1D-CNN also holds true for other end-to-end classifiers.

Here, that end-to-end classifier is the 3D-CNN proposed by [19], which operates on three-dimensional input data. The input data of that is in the form of the combination of

Table 2 Parameter details of each layer of 1D-CNN used

Layers	Type of Layer (DP)	Activation	Output Size(Dataset 1/2)	Kernel Size	Stride
1	Convolution	Relu	5 @ 6248 / 5 @ 510	3	1
2	Max pooling	—	5 @ 3124 / 5 @ 255	2	2
3	Convolution(0.5)	Relu	5 @ 3122 / 5 @ 253	3	1
4	Max pooling	—	5 @ 1561 / 5 @ 126	2	2
5	Convolution(0.5)	Relu	5 @ 1559 / 5 @ 124	3	1
6	Average pooling	—	5 @ 779 / 5 @ 62	2	2
7	Convolution	Relu	5 @ 777 / 5 @ 60	3	1
8	Average pooling	—	5 @ 388 / 5 @ 30	2	2
9	Convolution	Relu	5 @ 386 / 5 @ 28	3	1
10	Global Average pooling	—	5 / 5	—	—
11	Fully connected	Relu	2 / 2	—	—
12	Softmax	—	2 / 2	—	—

DP Dropout rate if exists.

The numbers before all @ are channel numbers for convolution layers

Table 3 Parameter details of each layer of 3D-CNN used

Layers	Type of Layer (DP)	Activation	Output Size	Kernel Size	Stride
1	3D convolution	Relu	20@6 × 4 × 503	2 × 2 × 10	1 × 1 × 1
2	3D convolution(0.5)	Relu	30@3 × 2 × 494	2 × 2 × 10	2 × 2 × 1
3	3D convolution	Relu	40@1 × 494	3 × 2 × 1	1 × 1 × 1
4	2D convolution(0.5)	Relu	50@1 × 485	1 × 10	1 × 1
5	2D convolution(0.5)	Relu	60@1 × 475	1 × 10	1 × 1
6	Max pooling	–	1@60 × 238	1 × 2	1 × 2
7	2D convolution	Relu	2@1 × 238	60 × 1	1 × 1
8	Average pooling	–	2	1 × 238	1 × 238
9	Softmax	–	2	–	–

DP Dropout rate if exists.

Each convolution layer in the Net is followed by a batch normalization layer which isn't mentioned in the table.

The numbers before all @ are channel numbers for convolution layers

two-dimensional spatial topological structure and temporal dimension. In the initial few layers, three-dimensional convolution is used to simultaneously extract Spatio-temporal features. Subsequent layers involve spatial fusion to amalgamate high-level spatial features, with only temporal features being output. The following convolutional layers primarily focus on central time feature extraction, followed by intensive prediction embedding. Finally, a softmax layer is employed to produce classification results.

Table 3 details the layers of the Net. Notably, slight adjustments to the Net structure is necessary due to differences in channel numbers between our dataset and the dataset referenced in [19], as well as disparities in the classification tasks. The changed input format of the Net is illustrated in Fig. 3. Besides, we alter the channel numbers of the Net to (20,30,40,50,60), as a hyper-parameters adjustment, and keep the time segmentation unchanged, which indicates the data in the 2 (origin datasets) × 67 datasets will be segmented into irrelevant epochs with the size of 7 × 5 × 512 before inputting. Note that the raw EEG of any subject corresponds to a X , and any data segment after the segmentation corresponds to a \bar{Z} . Besides, especially for dataset 2, the non-existent channels in Fig. 3 will be padded with zeros.



Fig. 3 Input Format of 3D-CNN

4 Experiment and verification

4.1 Details of experiments and hyper-parameters adjustment

To begin this chapter, it is essential to provide clarity on the implementation procedures and the specifics of the upcoming experiments.

First, a unit experiment is defined as follows: Based on K-fold Cross-validation, a series of Nets are trained by K trainsets from a dataset processed by a particular paradigm, then tested by K testsets from the same dataset. If a unit experiment is based on a subject-independent dataset split, it's referred to as a SI unit experiment (SIUE) otherwise a subject-dependent unit experiment (SDUE). For datasets originating from dataset 1 which comprises 14 HCs and 14 SZs, each SIUE implements 14-fold cross-validation, among which the k-th fold takes the samples from the k-th healthy control and the k-th schizophrenia as the testset, and the samples from the rest of the subjects as the trainset. For datasets originating from dataset 2, each SIUE implements 10-fold cross-validation, and the split groups are described by Table 4. Or else in one SDUE, no matter which dataset it is based on, 10-fold random Cross-validation will be adopted.

Next, it is crucial to select the hyperparameters for the CNN structures and training strategies for the two CNNs. Notably, the selection of hyper parameters in this paper differs from usual papers where the goal is to achieve the best classification performance. The purpose of the selection is to highlight the influences of denoising approaches on the classification results. So that if the hyper parameters lead to low fitting capability, low stabilities in each training or a slow learning speed, on any of the datasets, it will significantly constrain the aforementioned influences, indicating an incorrect selection. There are three hyper parameters requiring adjusting in the adjustment- the CNN's channel number,

Table 4 Split groups for datasets originating from dataset 2

Group	SZs Included	HCs Included
1	022W, 32W, 33W, 088W, 103W	S10W, S12W, S18W, S20W
2	113W, 155W, 156W, 192W, 219W	S26W, S27W, S31W, S42W
3	221W, 249W, 276W, 307W, 312W	S43W, S47W, S50W, S53W
4	314W, 342MW, 382W, 387-2W, 387-3W	S55W, S59W, S60W, S72W
5	401W, 423W, 429W, 454-1W, 485W	S78W, S85W, S94W, S152W
6	508W, 509W, 510-1W, 515W	S153W, S154W, S155W, S157W
7	517W, 540W, 548W, 573W	S158W, S163W, S164W, S165W
8	575W, 585W, 576W, 642W	S167W, S169W, S170W, S173W
9	683W, 719W, R229W, R416W	S174W, S176W, S177W, S178W
10	S083W, S084-1W, S351W, S435W	S179W, S182W, S196W

the learning rate, and the learning rate scheduler (others are specified directly). We only roughly describe the strategy for the adjustment which isn't the key point of the paper, that is: In order to carry on a grid search, some alternative values of the three parameters were listed first. Then, utilizing these values, a series of SIUEs are executed. Following an assessment of the CNNs trained in the SIUEs based on the aforementioned three aspects, the optimal values for the parameters can be determined.

Here the result of adjustment will be listed. The hyper parameters for the 1D-CNN are the same among all unit experiments. The LR is 0.001, the batch size 128, the max training epoch 150, the loss cross-entropy, adam optimizer is applied, the LR scheduler is ReduceLRonPlateau for all epochs, plus the channel numbers of the Net stay original.

The hyper parameters for the 3D-CNN are also the same, and are listed as follows: the LR is 0.006, the batch size 128, the max training epoch 150, the loss cross-entropy, adam optimizer is applied, the LR scheduler is warm-up in the first 20 epochs, and ReduceLRonPlateau in the rest epochs, plus the channel numbers of the Net is (20,30,40,50,60).

4.2 Platforms and softwares

The following Table 5 lists all the software packages used, along with their versions and purposes (excluding common ones like NumPy). Note that fast ICA can be directly executed by EEGLAB, and MSPCA with signal denoising can be executed by MATLAB function "wmspca".

About the hardwares, all codes written with MATLAB are run on the PC with i7-6800k (3.4GHz), and 16GB (3000MHz) Ram. And all codes written with Python are run on the server with E5-2698 (2.2GHz), 503GB Ram, and Tesla V100-SXM2-32GB \times 8.

4.3 Implementation of EIEF

This section focus on SIUEs taking the 1D-CNN as the classifier. To ensure consistency and minimize unforeseen variations, each dataset is subjected to 8 parallel SIUEs, resulting in a total of $2 \times 67 \times 8$ SIUEs within this section.

To effectively visualize the impact of different denoising paradigms, it is necessary to represent the 66 paradigms with

Table 5 Software list

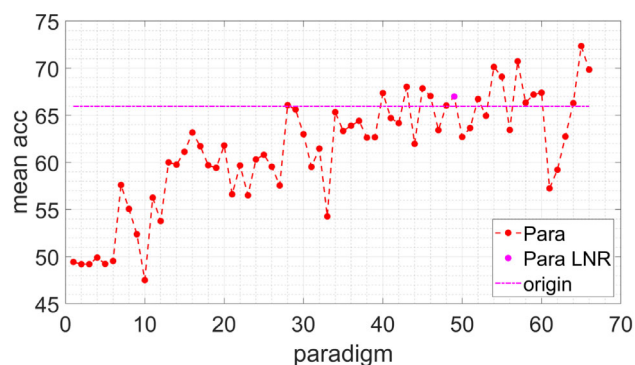
Environment (Version)	Packages and Plugins (Version)	Purpose
Python (3.7.9)	PyTorch (1.13.1)	Training and testing of the networks
	Neurokit2 (0.2.1)	Feature extraction
	Scipy (1.12.0)	T-test
	Pandas(1.4.0)	Data packing
MATLAB (2020a) andEEGLAB (v2021.1)	cleanline-master (v2.00)	Line noise removal
	clean-rawdata (v2.4)	Bad epoch removal
	PrepPipeline (v0.56.0)	Bad channel removal, interpolation. Robust re-reference
	ICLabel (v1.3)	ICA component classification

unique integer codes ranging from 1 to 66. The rules for encoding are as follows. The code 1 signifies the paradigm that sequentially employs robust re-referencing, bad channels removal, line noise removal, bad epochs retention, and decomposition retention, and then per time the code changes upward by 1, an approach belonging to a certain denoising approach categories will change. The change priority of approach categories is ranked as the filter first, the bad epoch second, and then the decomposition, the re-reference, the bad channel. The change direction of the bad channel is from removal to interpolation to retention. The re-reference's is from using the robust reference to retention. The filter's is from line noise removal (LNR) to LNR+1 HZ high pass filter to LNR+1-50 HZ band pass filter. The bad epoch's and the decomposition's shift from retention to ICA to MSPCA. Following this encoding scheme, the impacts of different paradigms are visually represented using line graphs. Unless otherwise specified, in the following sections, all mentioned metrics for evaluating each paradigm are calculated on the respective testsets and averaged across relevant parallel SIUEs.

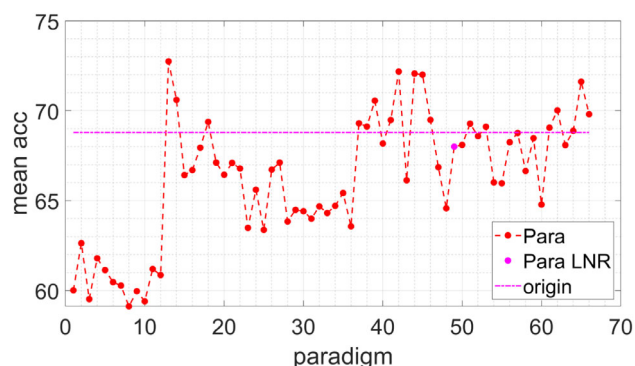
As illustrated in Fig. 4, the influence of denoising approaches on diagnostic mean-accuracies (mean accuracy across parallel SIUEs, cross-validation and across 101-150 epochs.) is shown by the two lines marked "Para". For dataset 1, the maximum accuracy difference of paradigms is up to 24.83% and the STD of 66 accuracies is 5.99%. For dataset 2, the two metrics are 13.63% and 3.55%. This clearly demonstrates that denoising approaches exert a substantial influence on classification outcomes, affirming the experimental feasibility of EIEF.

Compared with the processed datasets, the two baselines (marked "origin") shows a relatively high level of performance. Only 17 paradigms (dataset 1) and 18 paradigms (dataset 2) yield higher accuracies than the baselines, with the highest achieving 6.39% (dataset 1) and 3.95% (dataset 2) increases, while the lowest record an 18.44% (dataset 1) and 9.67% (dataset 2) decreases. Admittedly, such a result has been influenced by the selection of the paradigm pool. Yet, suppose the selection has a generality to some extent. In that case, we can infer that an inadequate denoising paradigm will sorely harm the EEG effective information of SZ classification, which easily happens. Besides, the harm exerts a more pronounced negative influence on the diagnosis performance than the positive influence caused by an excellent paradigm.

Beside mean-accuracy, the influences of denoising approaches on classification's mean-loss, mean-recall, mean-precision, and mean-specificity are illustrated in Fig. 5. Similar inferences like the above can be delivered if you check the 8 figures thoroughly. Except for the mean-recalls



(a) Dataset 1



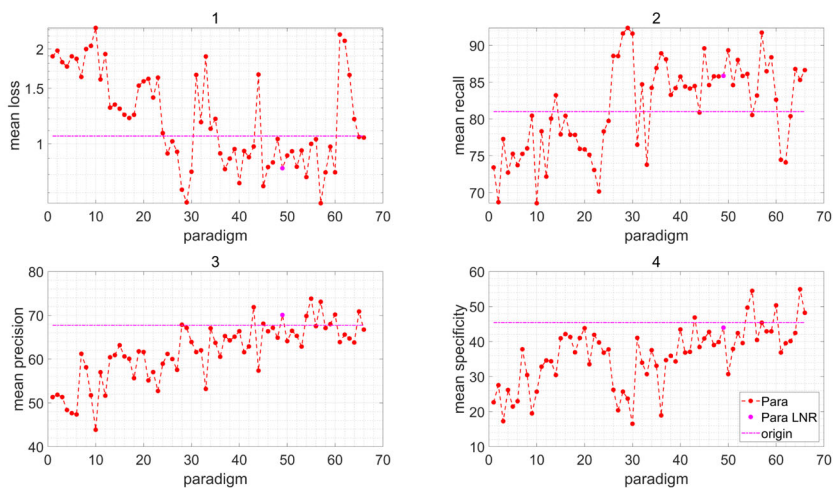
(b) Dataset 2.

Fig. 4 Max accuracy versus denoising paradigm. **Para LNR** The result of the SIUE with the line noise removal performed only; **origin** The result of the SIUE without denoising

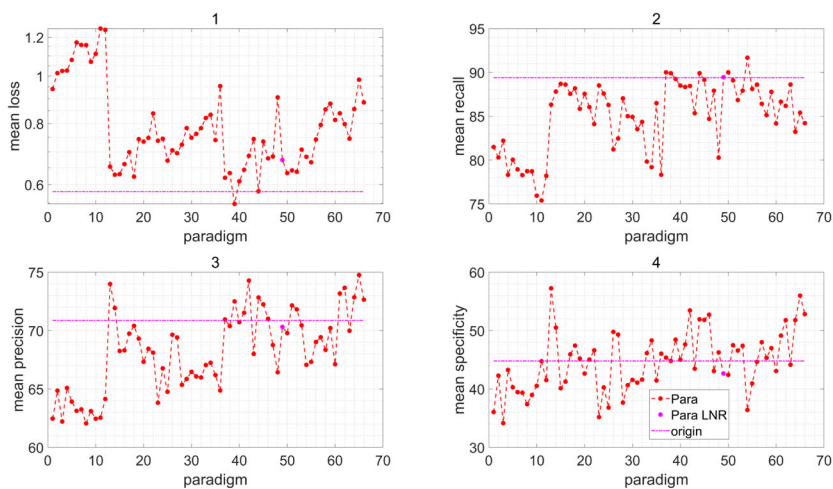
of dataset 1 and the mean-specificities of dataset 2, the harms from poor paradigms are more significant than the gains from excellent paradigms. Utilizing the Pearson correlation coefficient (CC) calculation between the two mean-loss lines and any of the other metric lines separately, we can assert that the paradigms' influences on these five metrics are fundamentally similar, as evidenced by CC values of -0.89, -0.87, -0.40, and -0.85 (dataset 1) and of -0.73, -0.86, -0.65, -0.29 (dataset 2), respectively.

After the discussion, let's return to the core of EIEF's functionality. According to (9), the optimal estimation is up to the selection of metric function. Although five function candidates are plausible, loss is not the most intuitive, and recall, precision, and specificity are not unequivocally guided by loss. Therefore, we have opted for accuracy as the metric function. Consequently, for dataset 1 the paradigm that yields the highest accuracy is considered the optimal estimation for the *restoration function* of effective information, denoted as paradigm 65. And for dataset 1, the optimal estimation is paradigm 13. The following Table 6 details the denoising approaches inside the two optimal paradigms.

Fig. 5 Mean-loss versus denoising paradigm 1; mean-recall versus denoising paradigm 2; mean-precision versus denoising paradigm 3; mean-specificity versus denoising paradigm 4; (The meaning of the legend is the same as that in Fig. 4)



(a) Dataset 1



(b) Dataset 2

4.4 Construction of SZ diagnosis system

This section primarily focuses on the development of an SZ diagnosis system characterized by high stability and generalization. The key objectives here involve inputting the two datasets processed by the two optimal paradigms into the 3D-CNN for training, resulting in the establishment of an SZ Diagnosis system based on these training iterations. Furthermore, a comparative analysis is conducted between this system and the system trained using the original datasets. Additionally, our results on two datasets are compared with state-of-the-art (SOTA) results obtained by other methods

that also utilized the same raw datasets. Both SIUEs and SDUEs will be implemented.

Table 7 lists the best performances obtained through multiple training runs of the network on two original datasets and their respective optimal datasets, categorized into SIUE and SDUE. First, let us delineate the two primary applications of the diagnostic system. Suppose clinicians aim to classify individuals visiting the hospital for the first time, who are potential patients. In this scenario, they should input pre-collected EEG data from both patients and healthy subjects into the optimal paradigm determined by EIEF. Subsequently, this processed and segmented data should be

Table 6 Approach details of the optimal paradigms

Dataset	Paradigm Code	Bad channels	Re-reference	Filter	Bad epoch	Decomposition
1	65	Retention	Retention	LNR+ 1	removal	MSPCA
2	13	removal	Retention	LNR	Retention	Retention

Table 7 Performance of four classifiers

Dataset	Training path	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)
1	Or + SIUE	88.91	95.49	88.39	77.65
	Op + SIUE	89.85	92.29	92.32	84.26
	Or + SDUE	97.96	98.86	97.44	96.86
	Op + SDUE	99.94	99.92	99.97	99.97
2	Or + SIUE	82.33	70.03	94.22	96.17
	Op + SIUE	87.40	83.63	91.41	91.67
	Or + SDUE	92.86	87.30	99.57	99.45
	Op + SDUE	98.02	97.30	98.98	98.79

Or The classifier is trained on the original dataset. **Op** The classifier is trained on the optimal dataset

employed to pre-train a 3D-CNN. After collecting EEG data from a new potential patient and applying the same denoising paradigm, the pre-trained CNN can provide clinicians with an estimate of the likelihood that the individual has SZ, as long as they divide the segment number with positive outputs by the total number. According to the 2 lines of “Op + SIUE”, the diagnosis system based on dataset 1 recognizes an SZ patient with the possibility of 92.29%, and a healthy man with the possibility of 86.37%. And for the system based on dataset 2, the 2 indicators are 83.63% and 91.67%. Otherwise, Suppose clinicians expect to monitor one patient’s or potential patient’s progress, they simply need to ensure that the pre-collected EEG data of that person exists in the dataset inputted into the EIEF. And the subsequent procedures remain consistent with the prior usage. With the 99.94% accuracy on dataset 1 and the 98.02% accuracy on dataset 2, patient’s progress could be traced precisely. Note that since the dataset accumulates EEG data from more subjects over time, all subjects’ EEG data must be collected under the same conditions, which avoids the situation where the optimal paradigm may not be suitable for certain subjects.

After that, let’s focus on the performance changes before and after implementing the optimal paradigm. Whether trained in SIUEs or SDUEs, the classifiers trained on the optimal datasets both achieve better accuracies. Let’s focus on dataset 1 first. In the SIUEs, it can be observed that the optimal denoising paradigm led to an overall accuracy improvement of 0.94%. Additionally, precision and specificity increase by 3.93% and 6.61%, respectively, while recall decrease by 3.20%. This indicates that the system, while slightly sacrificing the possibility of correctly identifying patients, better reduces the likelihood of misclassifying healthy individuals as patients. In the SDUEs, the diagnostic system’s capability is significantly enhanced to a level that is almost impossible to surpass. Now turning to Dataset 2, it is noteworthy that the gains achieved by EIEF on Dataset 2 are more remarkable, as reflected in the 5.07% (SI) and 5.16% (SD) accuracy improvements. Regardless of the Split way, the gains obtained by the system are mainly manifested in a

significant increase in correctly identifying patients (recall), albeit at the cost of slightly increasing the probability of misclassifying healthy individuals. These findings collectively underscore the positive impact of EIEF on classifier performance.

Last, the comparison between our work and the other popular methods comes. The methods and results are in Tables 8 and 9. In the comparison with works conducted on dataset 1, our results are competitive in all SD researches. In SI researches, since the experimental results of Lillo et al. were obtained by excluding the subject HC-12, our results still stand out compared to all other works. In the comparison to works conducted on dataset 2, our results can be considered excellent in all SD researches, as we are only slightly behind the best-performing research by 1%. In SI researches, the work of Shen et al. stands out as exceptionally outstanding, but our results are not too far behind.

4.5 Verification for EIEF mechanism

Two distinct approaches will be undertaken to validate the mechanism of EIEF. They are both based on replacing the 1D-CNN with 3D-CNN to implement another grid search. The first approach involves SI dataset splitting before training, confirming that the optimal denoising paradigms identified through different DNNs exhibit similarity. The second approach entails SD splitting, aimed at rationalizing the condition associated with SI splitting within EIEF (Condition 6).

To avoid accidents in the same light, for one dataset and one split method, 4 relevant unit experiments need to be implemented, indicating there are $67 \times 4 \times 2$ SIUEs and $67 \times 4 \times 2$ SDUEs involved in this section. Besides, note that the 5 following metrics are both measured on testsets, and averaged across parallel unit experiments, cross-validation and epochs 101-150, akin to Section 4.3.

As illustrated in Fig. 6, the influence of denoising approaches on diagnostic performances is shown by the lines marked “SI”. From them, the denoising’s tremendous

Table 8 Comparison with S.O.T.A works used the same dataset 1

Split way	Related work	Analysis technique	Classification	Result (%)
SD	Oh et al. [18]	Multi-channel EEG epochs are applied as input	1D-CNN	●ACC: 98.07 ●REC: 97.32 ●SPE: 98.17
	Jahmunah et al. [27]	●2nd order Butterworth filter ●157 nonlinear features	SVM-RBF	●ACC: 92.91 ●REC: 93.45 ●SPE: 92.25
	Aslan et al. [49]	2D- Spectrograms	2D-CNN	●ACC: 97.00 ●REC: 96.22 ●SPE: 98.31
	Krishnan et al. [31]	●Multivariate Empirical Mode Decomposition (MEMD) ●Recursive Feature Elimination	SVM	●ACC: 93.00 ●F-1: 93.00
	Prabhakar et al. [28]	Black Hole (BH) optimization	SVM-RBF	●ACC: 92.17
	Baygin [32]	●Tunable Q-Factor Wavelet Transform ●25 Statistical Moments ●ReliefF	Ensemble Subspace kNN	●ACC: 99.12
	Aslan et al. [50]	2D-Scalo-grams by Continuous Wavelet Transform(CWT)	CNN-VGG16	●ACC: 99.50 ●REC: 99.79 ●SPE: 99.79
	Ilakiyaselvan et al. [51]	Reconstructed phase space images	CNN + Voting Classifier +Inception-v4	●ACC: 99.37 ●REC: 99.68 ●SPE: 99.06 ●F-1: 99.37 ●AUC: 99.92
	Bagherzadeh et al. [52]	2D transfer entropy connectivity matrix images from EEG	pretrained CNN-LSTM	●ACC: 99.90 ●REC: 99.54 ●SPE: 100.00
	Sharma et al. [34]	1-D EEG time series data with ICA denoising	SzHNN hybrid CNN + LSTM	●ACC: 99.90 ●REC: 100.00 ●SPE: 99.80 ●PRE: 99.80
	Zülfikar et al. [37]	●Decomposition of EEG signals using the EMD method ●HS images for the first four IMF components	CNN-VGG16	●ACC: 98.2 ●F-1: 98.00 ●AUC: 98.80
	Khare et al. [36]	Margenau-Hill time-frequency distribution	CNN-SchizoNET	●ACC: 99.95 ●REC: 99.96 ●SPE: 99.96 ●PRE: 99.96 ●F-1: 99.96 ●Kappa: 99.90
		Our work	EIEF for denoising paradigm selection	3D-CNN
SI	Buettner et al. [47]	99-point spectrum from 99 frequency bands	Random Forest	●ACC: 71.43 ●REC: 100.00 ●SPE: 60.00 ●PRE: 50.00
	Oh et al. [18]	Multi-channel EEG epochs are applied as input	1D-CNN	●ACC: 81.26 ●REC: 75.42 ●SPE: 87.59
	Wu et al. [53]	1-D EEG time series data	Recurrent AutoEncoder	●ACC: 81.81 ●REC: 80.30 ●SPE: 83.37
	Lillo et al. [54]	●Processing of brain micro-states ●Random walk built	1D-CNN	●ACC: 93 ●REC: 86 ●SPE: 100 ●PRE: 100
	Sharma et al. [34]	1-D EEG time series data with ICA denoising	SzHNN hybrid CNN + LSTM	●ACC: 90.11 ●REC: 88.46 ●SPE: 91.66 ●PRE: 92.03
		Our work	EIEF for denoising paradigm selection	3D-CNN

ACC: Accuracy. REC: Recall. SPE: Specificity. PRE: Precision. F-1: F1-score. AUC: Area under Curve

Table 9 Comparison with S.O.T.A works used the same dataset 2

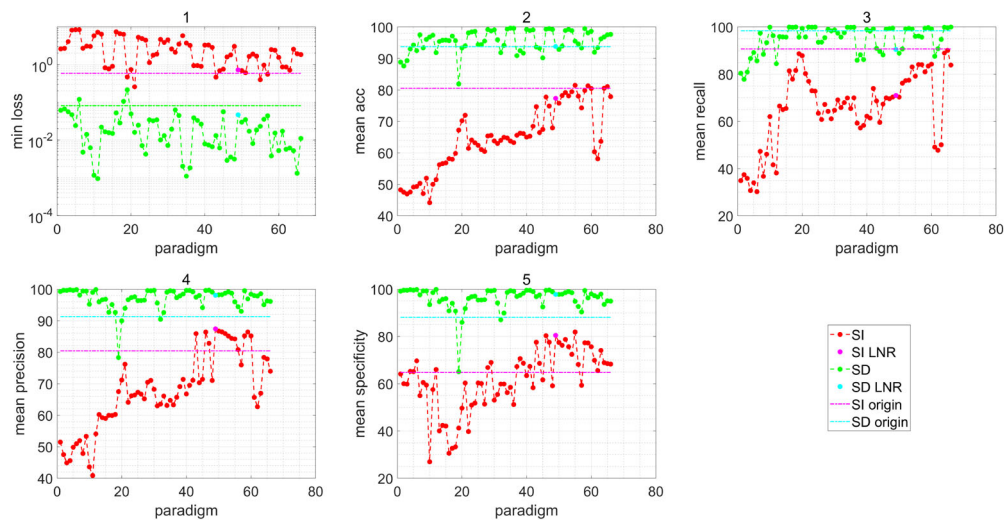
Split way	Related work	Analysis technique	Classification	Result (%)
SD	Naira et al. [55]	Convert EEG into Correlation Coefficient heat map	2D-CNN	●ACC: 90 ●REC: 90 ●SPE: 90
	Phang et al. [56]	●vector-autoregressionbased directed connectivity ●graph-theoretical complex network (CN) measures	DNN-DBN	●ACC: 95
	Aslan et al. [49]	2D- Spectrograms	2D-CNN	●ACC: 95.00 ●REC: 95.35 ●SPE: 94.72
	Calhas et al. [57]	FFT and STFT	Siamese neural network(SNN)	●ACC: 95.00 ●REC: 98.00 ●SPE: 92.00
	Supakar et al. [58]	1-D EEG time series data	RNN- LSTM	●ACC: 98.00 ●REC: 98.00 ●SPE: 98.00
	Sharma et al. [34]	1-D EEG time series data with ICA denoising	SzHNN hybrid CNN + LSTM	●ACC: 99.50 ●REC: 99.40 ●SPE: 99.59 ●PRE: 99.60
	Aslan et al. [50]	2D-Scalo-grams by Continuous Wavelet Transform(CWT)	CNN-VGG16	●ACC: 98.00 ●REC: 98.00 ●SPE: 98.00
	Zülfikar et al. [37]	●Decomposition of EEG signals using the EMD method ●HS images for the first four IMF components	CNN-VGG16	●ACC: 96.02 ●F-1: 96.00 ●AUC: 96.00
	Khare et al. [36]	Margenau-Hill time-frequency distribution	CNN-SchizoNET	●ACC: 98.14 ●REC: 97.93 ●SPE: 98.39 ●PRE: 98.61 ●F-1: 98.27
	Our work	EIEF for denoising paradigm selection	3D-CNN	●ACC: 98.02 ●REC: 97.30 ●SPE: 98.98 ●PRE: 98.79
SI	Phang et al. [59]	connectivity feature extraction for classification	Multi-domain Connectome CNN	●ACC: 91.69 ●REC: 97.78 ●SPE: 92.5
	Sharma et al. [34]	1-D EEG time series data with ICA denoising	SzHNN hybrid CNN + LSTM	●ACC: 89.60 ●REC: 88.43 ●SPE: 91.27 ●PRE: 91.80
	Shen et al. [60]	●continuous wavelet transform ●cross mutual information	Customized 3D-CNN	●ACC: 97.74 ●REC: 96.91 ●SPE: 98.53
	Shen et al. [35]	●multivariate auto-regressive mode ●coherence functional brain network	Customized 3D-CNN	●ACC: 98.47 ●REC: 99.26 ●SPE: 97.23
	Our work	EIEF for denoising paradigm selection	3D-CNN	●ACC: 87.40 ●REC: 83.63 ●SPE: 91.41 ●PRE: 91.67

ACC: Accuracy. REC: Recall. SPE: Specificity. PRE: Precision. F-1: F1-score. AUC: Area under Curve

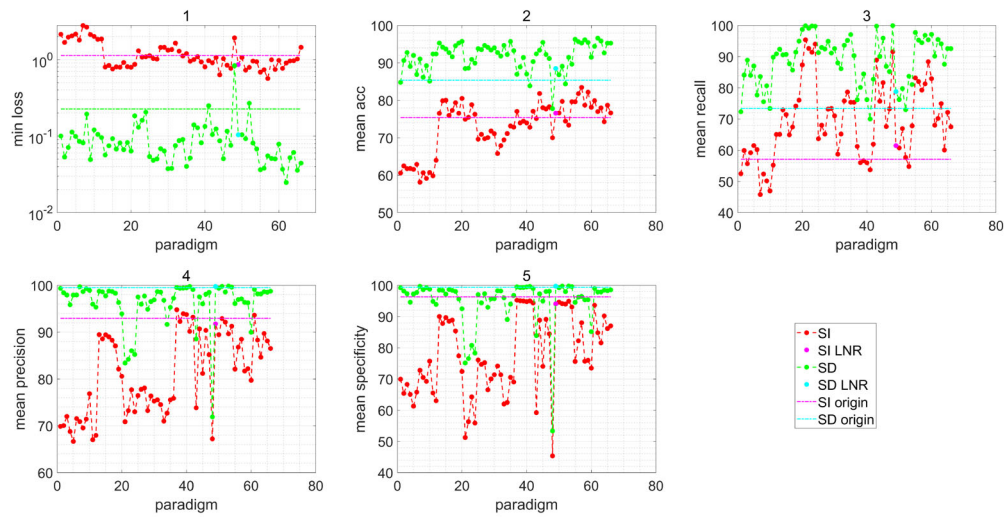
influence still holds. As a matter of fact, the optimums searched by the two Nets are actually different, because they are just the optimal estimations of the effective information rather than the information itself. Whereupon, the difference is caused by the classifiers' difference and shouldn't be blamed on hypothesis 1. Analyzing Table 10, we observe that both 1D-CNN and 3D-CNN, when subjected to grid search, yielded optimal denoising paradigms that improved classifier performance. Except for Dataset 2, where the optimal paradigm obtained using 3D-CNN did not enhance the

performance of the 1D-CNN classifier, but it did not lead to performance degradation. This suggests that EIEF demonstrates good robustness, consistently identifying denoising paradigms that enhance the proportion of effective information, and this enhancement is not dependent on changes in the network used by the diagnostic system.

As for the second verification way, a statistical analysis of the disparity between the "SD" and "SI" lines yields Table 11, which demonstrates that the SD partitioning method results in higher average performance with minimal performance



(a) Dataset 1



(b) Dataset 2.

Fig. 6 Mean-accuracy versus denoising paradigm: 1; Mean-loss versus denoising paradigm: 2; mean-recall versus denoising paradigm: 3; mean-precision versus denoising paradigm: 4; mean-specificity versus

denoising paradigm: 5 **SI** performance change line based on SIUEs; **SD** performance change line based on SDUEs; The rest legends are the same as those in Fig. 4

fluctuations. All the indicators push us conclude that one advanced CNN (like our 3D-CNN) can remember every subject’s EEG pattern and reach a high test performance in the SD evaluation no matter what denoising approach is applied, and such a characteristic runs counter to the core objective of EIEF, which is to identify the optimal metric value. In that case, the dataset partitioning approach required for EIEF should ideally be the SI split.

So far, the grid searches based on 1D-CNN with SI split, 3D-CNN with SD, and 3D-CNN with SI have all been completed. Next, a more detailed robustness analysis will be

conducted. We will utilize T-tests to analyze the impact of changes in approach selection within each denoising approach category on network performance, for the three grid searches. Additionally, we will compare the analysis results between 1D-CNN and 3D-CNN, as well as between 3D-CNN with SI and SD, to assess whether changes in network and data split way alter the impact of denoising approaches on performance. For the two SI-based grid searches, the data split is determined and non-random, thus paired T-tests are used. For the SD-based grid search, since the data are randomly shuffled and grouped, two-sample T-tests are

Table 10 Comparison of the effects of optimal paradigms generated from different networks on the performances of the classifiers

Dataset	Net for Diagnosis	Net inside EIEF	Code of Optimum	ACC (%)	REC (%)	SPE (%)	PRE (%)
Dataset 1	1D-CNN	1D-CNN	65	72.34	85.31	54.93	70.86
		3D-CNN	55	69.09	80.54	54.48	73.80
		–	Raw Data	65.95	81.00	45.44	67.70
	3D-CNN	1D-CNN	65	81.09	90.09	68.54	77.88
		3D-CNN	55	81.47	79.19	81.88	84.22
		–	Raw Data	80.52	90.69	67.76	80.43
Dataset 2	1D-CNN	1D-CNN	13	72.75	86.32	57.23	73.99
		3D-CNN	57	68.77	86.42	48.00	69.43
		–	Raw Data	68.80	89.39	44.79	70.87
	3D-CNN	1D-CNN	13	76.53	65.09	90.02	89.47
		3D-CNN	57	83.46	79.31	88.03	88.50
		–	Raw Data	75.39	57.14	96.32	92.95

ACC: Accuracy. REC: Recall. SPE: Specificity. PRE: Precision.
All included indicators represent averages rather than the optimal values

employed. In tests, one metric value (such as the accuracy corresponding to paradigm 43) is averaged across multiple parallel unit experiments.

Looking at Table 12. For Dataset 1, there are 9 pairs of means (20.45%) showing differences in sign. Considering only the 13 pairs of means with p-values below 5% for both networks, only 1 pair of means (7%) exhibits differences in sign. When considering only accuracies, no pairs of means exhibit differences in sign. For Dataset 2, there are 16 pairs of means (36.36%) showing differences in sign. Considering p-values, 5 out of 15 pairs (33.33%) exhibit differences, while considering only accuracies, 3 out of 11 pairs (27.27%) exhibit differences. These data suggest that out of the 44 total effects resulting from 11 types of change of the denoising

approaches, the direction of most effects remains consistent regardless of the internal network used by EIEF. Turning to Table 13. For dataset 1, there are 15 pairs of mean values (34%) that differ in sign. When considering the p-values, 4 out of 21 pairs (19.04%) show differences. Looking solely at accuracies, there are 3 out of 11 pairs (27.27%) with differences. For dataset 2, 5 pairs of mean values (11.36%) differ in sign. When considering the p-values, 2 out of 23 pairs of mean values (8.70%) show differences. Considering only accuracies, 2 out of 11 pairs of mean values (18.18%) exhibit differences. The analysis suggests that the direction of most effects does not change with the variation in data split strategy during network training.

Table 11 Statistical indicators of performance metrics

Dataset	Statistical indicator	Split way	Acc. (%)	Rec. (%)	Pre. (%)	Spe. (%)	Loss
Dataset 1	MMD	SI	37.29	59.87	46.57	54.93	11.81
		SD	17.74	22.15	21.59	34.83	1.51
	STD	SI	10.49	15.89	12.39	13.01	2.95
		SD	3.49	5.70	3.35	5.03	0.24
	Mean	SI	64.86	65.48	67.66	61.55	6.06
		SD	95.43	94.82	97.11	96.17	0.24
Dataset 2	MMD	SI	25.33	49.58	28.13	49.87	2.94
		SD	18.82	29.94	27.93	46.42	1.85
	STD	SI	6.78	12.19	8.73	12.84	0.66
		SD	3.70	7.91	4.85	7.60	0.28
	Mean	SI	73.15	69.17	81.21	77.78	2.06
		SD	91.67	88.71	96.39	95.06	0.40

Each indicator is calculated across the 66 metric values corresponding to the 66 paradigms.
MMD Maximum minimum difference

Table 12 Filter’s, decomposition’s and bad epochs’ influences via T-test

Dataset	Factor	Value1	Value2		Accuracy 1D(%) / 3D(%)	Precision 1D(%) / 3D(%)	Recall 1D(%) / 3D(%)	Specificity 1D(%) / 3D(%)
dataset 1	Filter	LNR	LNR+1	Mean diff	0.72/0.38	1.09/0.90	-0.99/2.47	2.81/ - 1.61
				P-value	26.96/63.40	15.44/38.21	29.47/ 2.72	2.56/18.59
		LNR	LNR+1-50	Mean diff	0.52/0.12	1.54/0.64	-1.59/0.46	4.18/0.72
				P-value	44.13/87.85	5.78/54.79	9.10/66.21	0.20/62.94
		LNR+1	LNR+1-50	Mean diff	-0.21/ - 0.26	0.46/-0.26	-0.59/ - 2.00	1.37/2.33
				P-value	73.58/67.48	56.11/75.17	52.26/ 1.59	25.96/4.76
	Bad epoch	Removal	Retention	Mean diff	2.13/2.40	-0.62/1.21	1.67/5.58	1.23/-1.08
				P-value	0.18/0.77	49.02/22.91	8.04/ 0.00	27.97/39.71
	Decompose	ICA	Retention	Mean diff	0.92/3.45	1.13/2.18	-2.07/7.04	5.21/0.08
				P-value	17.90/ 0.01	18.01/ 4.99	4.49/0	0/95.79
		ICA	MSPCA	Mean diff	2.75/8.96	4.01/12.05	4.22/14.35	2.40/4.05
				P-value	12.14/ 0.39	8.37/ 0.19	9.84/ 0.09	43.45/28.13
		Retention	MSPCA	Mean diff	1.24/7.74	0.51/15.09	5.35/7.55	-2.97/7.63
				P-value	40.68/ 0.24	79.14/ 0	1.14/3.79	35.67/0.92
	Bad channel	Removal	interpolation	Mean diff	-7.58/ - 10.86	-8.23/ - 13.42	-9.63/ - 7.30	-1.41/ - 12.43
				P-value	0/0	0/0	0/0.04	29.13/ 0
		Removal	Retention	Mean diff	-6.63/ - 16.64	-9.19/ - 20.31	-8.92/ - 2.82	-4.82/ - 32.16
				P-value	0/0	0/0	0/16.14	1.29/0
		interpolation	Retention	Mean diff	-1.59/ - 9.79	-3.09/ - 11.04	-0.98/ - 14.46	-4.22/ - 6.67
				P-value	8.07/ 0	0.82/0	44.71/ 0	2.44/0.03
	Ref	Robust	Retention	Mean diff	-5.86/ - 9.31	-5.08/ - 11.55	-1.15/ - 17.93	-11.19/3.16
				P-value	0/0	0/0	39.01/ 0	0/20.63
	Filter	LNR	LNR+1	Mean diff	-0.73/ - 1.44	-0.20/0.30	-1.00/ - 2.47	-0.58/ - 0.37
				P-value	11.85/ 0.46	71.24/65.60	10.81/ 1.02	66.89/70.99
		LNR	LNR+1-50	Mean diff	-0.17/ - 1.23	0.13/1.79	-0.11/ - 5.54	-0.25/3.84
				P-value	71.83/ 1.40	81.56/ 0.81	85.38/ 0	85.11/ 0.15
		LNR+1	LNR+1-50	Mean diff	0.56/0.22	0.33/1.50	0.89/-3.07	0.33/4.21
				P-value	20.81/69.21	51.96/ 2.27	17.61/ 0.10	79.66/ 0
	Bad epoch	Removal	Retention	Mean diff	-0.64/ - 0.23	-0.59/ - 1.40	-1.20/2.07	0.18/-2.88
				P-value	9.53/60.51	17.34/ 0.71	2.00/1.33	86.74/ 0.09
Decompose	ICA	Retention	Mean diff	-1.31/0.57	-1.61/ - 6.58	-2.36/12.11	-0.03/ - 12.61	
			P-value	0.21/35.66	0.10/0	0/0	97.61/ 0	
	ICA	MSPCA	Mean diff	-2.43/3.24	-4.27/ - 4.99	0.99/14.04	-6.11/ - 8.73	
			P-value	1.01/7.59	0.05/0.82	40.98/ 0	1.94/0.27	
	Retention	MSPCA	Mean diff	-1.40/ - 0.68	-2.58/2.71	3.45/-7.24	-7.12/6.90	
			P-value	13.72/61.35	2.19/5.88	0.43/0.29	1.03/0.07	
Bad channel	Removal	interpolation	Mean diff	-2.91/ - 3.69	-2.62/ - 4.88	-2.38/ - 2.10	-3.29/ - 5.57	
			P-value	0/0	0/0	0.02/13.90	1.14/0	
	Removal	Retention	Mean diff	-0.14/ - 1.05	-0.55/ - 4.91	-0.70/7.19	0.46/-10.68	
			P-value	86.22/24.56	55.38/ 0	38.94/ 0	83.51/ 0	

Table 12 continued

Dataset	Factor	Value1	Value2		Accuracy 1D(%) / 3D(%)	Precision 1D(%) / 3D(%)	Recall 1D(%) / 3D(%)	Specificity 1D(%) / 3D(%)
		interpolation	Retention	Mean diff	1.50/-2.84	1.39/-0.13	-0.29/ - 3.52	3.50/-2.11
				P-value	2.22/0.12	7.91/89.77	71.26/ 1.79	6.32/20.36
	Ref	Robust	Retention	Mean diff	-5.72/ - 11.04	-4.91/ - 12.12	-6.39/ - 11.21	-4.71/ - 10.93
				P-value	0/0	0/0	0/0	0.03/0

Mean diff: Metrics' marginal mean of value 1 subtracts metrics' marginal mean of value 2.

A black bold entry indicates that the effect of the denoising approach varies in direction between the two networks.

A red entry indicates that the variation in the denoising approach leads to a significant change in the corresponding metric (at the 95% significance level).

A red bold entry indicates that the variation in the denoising approach leads to a significant change in the corresponding metric for both networks (at the 95% significance level)

Table 13 Filter's, decomposition's and bad epochs' influences via T-test

Dataset	Factor	Value1	Value2		Accuracy SD(%) / SI(%)	Precision SD(%) / SI(%)	Recall SD(%) / SI(%)	Specificity SD(%) / SI(%)	
dataset 1	Filter	LNR	LNR+1	Mean diff	-0.28/0.38	-0.67/0.90	0.41/2.47	-1.08/ - 1.61	
				P-value	46.48/63.40	7.82/38.21	49.01/ 2.72	6.19/18.59	
			LNR	LNR+1-50	Mean diff	-0.24/0.12	-0.72/0.64	0.60/0.46	-1.18/0.72
					P-value	53.51/87.85	5.35/54.79	32.1/66.21	3.85/62.94
			LNR+1	LNR+1-50	Mean diff	0.04/-0.26	-0.05/ - 0.26	0.20/-2.01	-0.10/2.33
					P-value	90.27/67.48	86.24/75.17	75.27/ 1.59	80.31/ 4.76
		Bad epoch	Removal	Retention	Mean diff	4.62/2.40	2.03/1.21	5.97/5.58	2.95/-1.08
					P-value	0/0.77	0/22.91	0/0	0/39.71
		Decompose	ICA	Retention	Mean diff	0.68/3.45	-1.71/2.18	3.43/7.04	-2.63/0.08
	P-value				3.46/0.01	0/4.99	0/0	0/95.79	
			ICA	MSPCA	Mean diff	1.01/8.96	-0.60/12.05	2.29/14.35	-0.64/4.05
					P-value	4.49/0.39	21.11/ 0.19	0.35/0.09	34.57/28.13
			Retention	MSPCA	Mean diff	0.66/7.74	1.56/15.09	-0.80/7.55	2.33/7.63
					P-value	23.94/ 0.24	0/0	41.64/ 3.79	0/0.92
		Bad channel	Removal	interpolation	Mean diff	-2.11/ - 10.86	-1.34/ - 13.42	-2.25/ - 7.30	-2.01/ - 12.43
					P-value	0/0	0.03/0	0.03/0.04	0.03/0
				Retention	Mean diff	-1.46/ - 16.64	-3.81/ - 20.31	1.99/-2.82	-5.54/ - 32.16
					P-value	0.27/0	0/0	0/16.14	0/0
		interpolation	Retention	Mean diff	-0.84/ - 9.79	0.89/-11.04	-2.55/ - 14.46	1.21/-6.67	
				P-value	6.12/ 0	0.06/0	0.03/0	0.06/0.03	
	Ref	Robust	Retention	Mean diff	-0.28/ - 9.31	2.07/-11.55	-2.99/ - 17.93	2.84/3.16	
				P-value	46.40/ 0	0/0	0/0	0/20.63	
dataset 2	Filter	LNR	LNR+1	Mean diff	-0.95/ - 1.44	-0.46/0.30	-1.04/ - 2.47	-0.79/ - 0.37	
				P-value	5.33/ 0.46	23.52/65.60	31.36/ 1.02	15.54/70.99	
			LNR	LNR+1-50	Mean diff	-0.47/ - 1.23	1.59/1.79	-3.23/ - 5.54	2.65/3.84
					P-value	36.89/ 1.40	0.65/0.81	0.08/0	0.44/0.15
			LNR+1	LNR+1-50	Mean diff	0.48/0.22	2.05/1.50	-2.20/ - 3.07	3.44/4.21
					P-value	32.97/69.21	0.02/2.27	1.47/0.10	0.01/0.01

Table 13 continued

Dataset	Factor	Value1	Value2		Accuracy SD(%) / SI(%)	Precision SD(%) / SI(%)	Recall SD(%) / SI(%)	Specificity SD(%) / SI(%)
Bad epoch	Removal	Retention	Mean diff		-1.29 / - 0.23	-1.92 / - 1.40	0.09 / 2.07	-2.96 / - 2.88
			P-value		0.16 / 60.51	0/0.71	91.00 / 1.33	0/0.09
Decompose	ICA	Retention	Mean diff		1.28 / 0.57	-4.42 / - 6.58	8.04 / 12.11	-6.46 / - 12.61
			P-value		0.32 / 35.66	0/0	0/0	0/0
	ICA	MSPCA	Mean diff		0.05 / 3.24	-3.10 / - 4.99	4.07 / 14.04	-4.48 / - 8.73
			P-value		92.89 / 7.59	0/0.82	0/0	0/0.27
	Retention	MSPCA	Mean diff		-6.69 / - 0.68	0.93 / 2.71	-13.32 / - 7.24	1.00 / 6.90
			P-value		0 / 61.35	0.01 / 5.88	0/0.29	0.20 / 0.07
Bad channel	Removal	interpolation	Mean diff		-0.47 / - 3.69	-0.41 / - 4.88	-0.61 / - 2.10	-0.41 / - 5.57
			P-value		32.34 / 0	46.06 / 0	51.55 / 13.90	63.43 / 0
	Removal	Retention	Mean diff		0.90 / - 1.05	-4.14 / - 4.91	6.78 / 7.19	-5.93 / - 10.68
			P-value		14.61 / 24.56	0/0	0/0	0/0
	interpolation	Retention	Mean diff		-2.45 / - 2.84	-1.87 / - 0.13	-1.83 / - 3.52	-3.11 / - 2.11
			P-value		0.25 / 0.12	2.70 / 89.77	22.94 / 1.79	2.33 / 20.36
Ref	Robust	Retention	Mean diff		0.61 / - 11.04	2.74 / - 12.12	-2.70 / - 11.21	4.36 / - 10.93
			P-value		20.66 / 0	0/0	0.37 / 0	0/0

Mean diff: Metrics' marginal mean of value 1 subtracts metrics' marginal mean of value 2.

A **black bold entry** indicates that the effect of the denoising approach varies in direction between the two split ways.

A **red entry** indicates that the variation in the denoising approach leads to a significant change in the corresponding metric (at the 95% significance level).

A **red bold entry** indicates that the variation in the denoising approach leads to a significant change in the corresponding metric for both split ways (at the 95% significance level)

In summary, while variations in the networks used by EIEF, the data split way, and other unmentioned hyperparameters may somewhat hinder our ability to assess the proportion of effective information in the dataset using network performance metrics, overall, EIEF demonstrates sufficient robustness to remain practically meaningful.

4.6 Comparison between two classifiers

From Figs. 4, 5, 6, it seems like, whether the 1D-CNN or the 3D-CNN is selected to perform a grid search, both models consistently yield a reasonable optimal paradigm. To clarify why the 1D-CNN is included within EIEF while the 3D-CNN is not, let's delve into a comparative analysis of the searches

conducted by these two classifiers. The $60 \times 4 \times 2$ SIUEs performed by the 3D-CNN and the $60 \times 8 \times 2$ SIUEs performed by 1D-CNN will be analyzed again (unit experiments corresponding to the original dataset aren't involved), from the aspects of the mean and the max accuracy, the transverse accuracy STD, and the training time.

As listed in Table 14, 3D-CNN-based classifiers exhibit higher mean accuracies but lower extreme accuracies compared to 1D-CNN-based classifiers. This suggests that the 1D-CNN might be more suitable for constructing the diagnosis system. However, in the case that the system requires re-training frequently because of an increasing number of patient visits, the higher STD of multi-training makes the scalability of a 1D-CNN-based system questionable.

Table 14 Classifiers' average search performances

Classifiers	Dataset 1			Dataset 2		
	Acc.(max/mean%)	STD(%)	Cons.(s)	Acc.(max/mean%)	STD(%)	Cons.(s)
1D-CNN	83.78/61.31	15.89	250	83.96/66.29	11.54	290
3D-CNN	77.14/64.86	6.87	1150	80.91/73.18	6.56	335

All the performances are averaged across the cross-validation first and the 60 paradigms second.

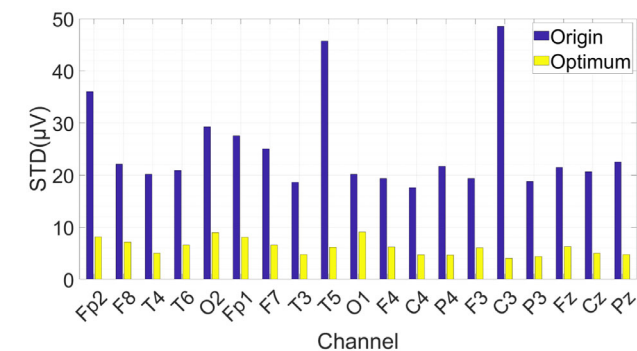
For one paradigm and one fold, the **Max Acc.** is found among the 101-150 epochs among the parallel SIUEs, the **Mean Acc.** is averaged across the same epochs and SIUEs, the **STD** is the STD of the mean accuracies for the parallel SIUEs (STD of the 4 mean accuracies for the 3D-CNN-based, 8 for the 1D-CNN-based), the **Cons.** is the average time consumption across all the trainings with 150 epochs in the parallel SIUEs)

Moreover, the 3D-CNN, with its larger size and deeper convolutional depth, holds the potential for improved generalization capabilities in deep learning. Collectively, this evidence indicates that the 3D-CNN offers greater stability and suggests it as the preferred choice for system construction. Conversely, EIEF assesses paradigms with the metric averaged across several epoch-fixed SIUEs, which dilutes the influence caused by the instability of the 1D-CNN. Moreover, 1D-CNN's training time is significantly less than 3D-CNN. And it can have EIEF implemented faster. Thus, we incorporate it into EIEF.

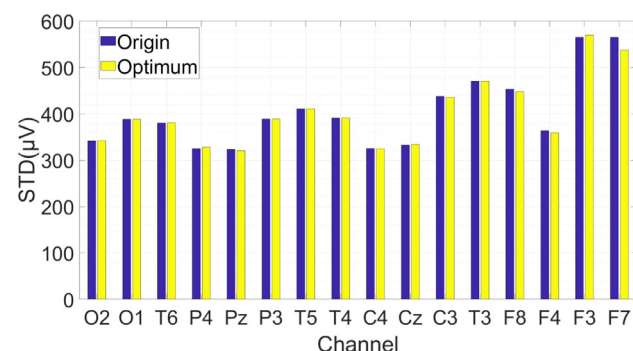
4.7 Confirmation of improved electrobiological markers

In this section, We will focus on experimentally elucidating that the dataset processed through our framework is more suitable for the pathological analysis of schizophrenia, making the analysis on this dataset more reliable.

As an overview, the data STDs of all the channels of the original and optimal datasets are shown in Fig. 7. Figure. 7(a) shows, for dataset 1, the STD of each channel in the origin is higher than the optimum, and indicates the optimal paradigm



(a) Dataset 1



(b) Dataset 2.

Fig. 7 STDs of every channel of the original and optimal datasets (All subjects' EEGs are considered)

found by EIEF does subtract a lot of EEG components from the original data. Considering that the performance metrics prove that such a subtraction hasn't caused a loss of the SZ classification effective Information, so obviously, it becomes evident that EIEF enhances the proportion of that information within the EEG dataset. This, in turn, reinforces the extraction of SZ electrobiological markers, making it more compelling. However, for dataset 2, the STDs of the origin and the optimum are very close in value. This is because the optimum for dataset 2 consists of linear noise removal and bad channel removal. Actually, dataset 2 does not exhibit significant power-line interference. And channel removal only leads to the loss of several channels in the EEG data of some subjects within the 84 subjects, without removing any components of these channels as a whole. Therefore, even if the proportion of effective Information actually increases, this improvement will not be reflected in the standard deviations. So, the next job is to determine what kind of influence EIEF has on electrobiological markers.

4.7.1 Features extraction and analysis method

Here, we take the difference between the two time-domain feature distributions in the 14 SZ patients and the 14 healthy subjects as an electrobiological marker. And the following Table 15 lists the 38 features selected.

Due to the complexity of EEG signal composition and its difficulty in discerning periodicity, extracting structural features and normal dynamic characteristics of EEG is almost meaningless. But the complex nature of the brain's electrical activity and its non-linear dynamic characteristics result in diverse EEG patterns. Breaking down the signal into smaller subsystems can potentially modify the irregular patterns and dynamic attributes of the signal [61]. The complexity analysis of EEG signals mainly involves the following aspects of feature analysis: 1. Nonlinear features, which examine whether signals exhibit nonlinear dynamical characteristics, such as chaotic behavior, phase transitions, and fractal structures, etc. Nonlinear features reflect the complex dynamics and nonlinear interactions of signals. 2. Information features, which evaluate the amount of information and the information structure carried by signals, including measures such as information entropy, sample entropy, permutation entropy, SVD entropy etc. Information features can reveal the information organization and encoding methods of signals. 3. Stability features, which Investigate the stability and predictability of signals under different conditions, including stability analysis, long-term dependence, and self-similarity, etc. Stability features reflect the stability and predictability of signals. Most of the features we extracted can quantify one or more of the aforementioned aspects of complexity.

Table 15 Selected features

Feature name	Feature name	Feature name	Feature name
Sample En	Wiener En	Fisher Inf	SVD En
Differential En	Hjorth Exp	Kolmogorov En	Maximum En
Multiscale En	Fuzzy En	Permutation En	Attention En [63]
CWPermutation En [64]	Line Length [65]	Dispersion En [66]	Distribution En [67]
Increment En [68]	En of En [69]	Shannon En-Power [70]	Range En [71]
Slope EN [72]	Symbolic Dynamic En [73]	Tsallis En [74]	Fisher-Shannon Inf [75]
Bubble En [12]	Mean Inf Gain	MDFA-Width [76]	MDFA-Peak
MDFA-Mean	MDFA-Max	MDFA-Delta	MDFA-Asymmetry
MDFA-Fluctuation	MDFA-Increment	Poincaré-STD [62]	Poincaré-STA
Poincaré-SSHD	Poincaré-SDTC		

En Entropy; **Inf** Information; **Exp** Exponent

In addition to classical complexity analysis, we have also used a new set of features for analyzing potential patterns in EEG-graph features. These features are essentially geometric characteristics on the visualization graph, and their extraction primarily relies on the two-dimensional visualization of EEG signals. Based on Poincaré pattern of DWT coefficients of EEG signals, Akbari et al. proposed new geometrical features—standard descriptors of 2-D projection (STD), summation of triangle area using consecutive points (STA), as well as summation of shortest distance from each point relative to the 45-degree line (SSHD), and summation of distance from each point relative to the coordinate center (SDTC) [62]. And these features are used in our following analysis.

Specifically, For each dataset and each subject, the corresponding multi-channel data are split into irrelevant segments of 4 seconds and of the size of 19×1000 (dataset 1) and 16×512 (dataset 2) (the last segment that is less than 4 seconds will be thrown away). Then for each segment and each feature, we extract 19 (16) feature values for all 19 (16) channels (Note that any feature’s parameters are the default set by a python package—“NeuroKit2”). And for each feature from Poincaré plot, since the signals are decomposed to 5 components with the DWT level 4, we extract 5 feature values for one channel. So the dimension number of a feature vector will be 1026 ($19 \times 34 + 19 \times 5 \times 4$) for dataset 1 or 864 ($16 \times 34 + 16 \times 5 \times 4$) for dataset 2. Denote the original SZ feature matrix that consists of all feature vectors extracted from the SZ subjects from the original dataset as $M^{sz} \in \mathbb{R}^{n_{sz} \times L_d}$, where n_{sz} is the segment number from the SZ subjects, and L_d is a variable with optional values of 1027 for dataset 1 and 864 for dataset 2. And denote the i -th column of M^{sz} as m_i^{sz} which represents a feature’s values of a channel. Then, let $M^{hc} \in \mathbb{R}^{n_{hc} \times L_d}$ denote the original HC feature matrix and let m_i^{hc} denote the i -th column of it. Similar, for the

optimal dataset, we get M^{*sz} , m_i^{*sz} , M^{*hc} and m_i^{*hc} . After defining the symbols, we deliver the analysis method with Algorithm 1.

Algorithm 1 Discrimination of change of difference between two distributions.

```

Input:  $M^{sz}, M^{hc}, M^{*sz}, M^{*hc}$ 
Output: For each feature of each channel, figure out whether the difference between the two average feature values from the SZs and HCs is changed before and after EIEF.
1:  $C$  (Confidence level)  $\leftarrow 0.05$ 
2:  $T^P$  (P’s change threshold)  $\leftarrow 0.5$ 
3:  $D^e$  (diff enhanced)  $\leftarrow$  zero vector with the size of  $1 \times L_d$ .
4:  $D^w$  (diff weakened)  $\leftarrow$  the same zero vector as the above.
5:  $H^L$  (higher to lower)  $\leftarrow$  the same zero vector.
6:  $L^H$  (lower to higher)  $\leftarrow$  the same zero vector.
7: for all  $i \in \{1, 2, \dots, L_d\}$  do
8:    $P_{Le}^{or}$   $\leftarrow$  P value of the levene-test of  $m_i^{sz}$  and  $m_i^{hc}$ .
9:    $P_{Le}^{op}$   $\leftarrow$  P value of the levene-test of  $m_i^{*sz}$  and  $m_i^{*hc}$ .
10:  if  $P_{Le}^{or} > C$  then
11:     $P^{or}$   $\leftarrow$  P value of the standard t-test of  $m_i^{sz}$  and  $m_i^{hc}$ .
12:  else
13:     $P^{or}$   $\leftarrow$  P value of the Welch’s t-test of  $m_i^{sz}$  and  $m_i^{hc}$ .
14:  end if
15:   $P^{op}$   $\leftarrow$  P value got by the same method as 10 ~ 13 but by  $m_i^{*sz}$  and  $m_i^{*hc}$ .
16:   $D^{or}$   $\leftarrow$  Mean of  $m_i^{hc}$  – Mean of  $m_i^{sz}$ 
17:   $D^{op}$   $\leftarrow$  Mean of  $m_i^{*hc}$  – Mean of  $m_i^{*sz}$ 
18:  if  $P^{or} - P^{op} > T^P$  then
19:     $D^e(i) \leftarrow 1$ 
20:  else if  $P^{op} - P^{or} > T^P$  then
21:     $D^w(i) \leftarrow 1$ 
22:  else if ( $P^{op}, P^{or} < C$ ) & ( $D^{or} > 0$  &  $D^{op} < 0$ ) then
23:     $H^L(i) \leftarrow 1$ 
24:  else if ( $P^{op}, P^{or} < C$ ) & ( $D^{or} < 0$  &  $D^{op} > 0$ ) then
25:     $L^H(i) \leftarrow 1$ 
26:  end if
27: end for return  $D^e, D^w, H^L, L^H$ 
    
```

Table 16 Numbers of features that have mean difference changes after implementing the optimal denoising paradigm

Dataset	D^e	D^w	H^L	L^H	Total
Dataset 1	33	59	47	75	214
Dataset 2	7	26	0	0	33

Four types of changes are listed respectively

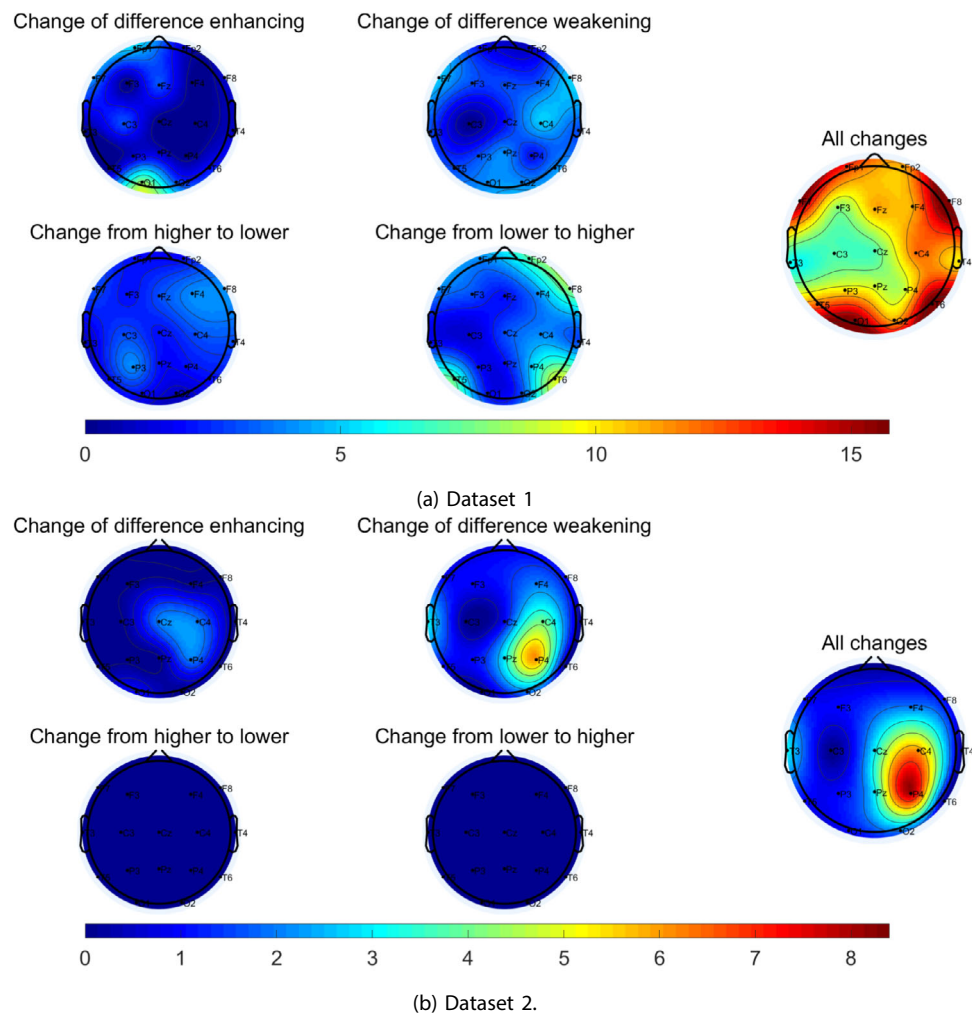
4.7.2 Analysis result

As has been detailed in Table 16, on dataset 1, for any of the 214/1026(20.86%) features, the difference between the two means calculated on SZ subjects and HC subjects show significant changes before and after the application of the optimal denoising paradigm to the raw dataset, which indicates at least 20.86% of SZ marker extractions lack reliability without the incorporation of EIEF. Besides, the numbers of features in the four categories of changes remains consistent in magnitude, indicating the widespread impact of the optimal paradigm. Next, considering dataset 2, even though from

Fig. 7 it appears that the standard deviation of data from each channel remains almost the same before and after denoising, there are still 33/864 (3.82%) features where the mean differences show significant changes. Additionally, given that each channel has been extracted with 38 types of features, for each dataset, we plot the changes' frequency of the paired means' differences for every channel in the 5 heatmaps in Fig. 8, to show the EIEF's effects on different brain regions visually. Intuitively, data from those channels with the high change frequencies shall contain more noises or more EEG components that are irrelevant to SZ recognition.

With all the evidence above, it is definite that, applying EIEF has discernible effects on EEG data's time-domain feature distributions, and then could help to make extracted SZ electrobiological markers more reliable since EIEF improves the proportion of the SZ classification effective Information information in the EEG dataset. However, as seen in the analysis results of dataset 2, EIEF may not necessarily induce substantial changes in the feature distribution; its effectiveness is constrained by the proportion of noise in the raw data and the selection of the denoising paradigm

Fig. 8 Brain heatmaps of changes in feature distribution differences before and after implementing



pool. Nonetheless, if the effect of EIEF is not significant, it suggests that the raw data itself has a high signal-to-noise ratio. In this case, we can consider that EIEF validates the reliability of extracting electrobiological markers from the raw data. As a practical recommendation for researchers, if expecting to uncover discriminable features qualified to be the valid biomarkers of SZ based on EIEF, they should put their dataset into EIEF to get the optimal denoising paradigm, use this paradigm to process their dataset, and do their marker extraction and analysis on the processed dataset. Two key considerations should be highlighted when employing EIEF. One is that the EEG data for all selected subjects must be collected in the same circumstance, which makes sure the noises in every subject's data are similar or completely identical. This helps avoid situations where the optimal paradigm is unsuitable for certain subjects. The other is the paradigm pool shall be selected carefully so that a pretty good paradigm for the dataset exists in the pool.

5 Discussion of advantages and limitations

5.1 Advantages

1. In general, research on EEG-based schizophrenia identification often innovates on one or two nodes in the open-loop technical chain, such as proposing new denoising methods, novel classification features, new feature selection methods, or specialized classification network architectures, etc. In contrast, our study centered around EIEF focuses on closing the loop of the entire technical chain. We use classification metrics to provide feedback for the denoising method selection, thereby improving the performance of the diagnosis system and the reliability of biomarker analysis. This represents a structured innovation across the entire technical chain. Following our direction, future researchers can continuously optimize the details of each node on this structure to achieve better results.
2. In recent years, most researchers in the field have often conducted experiments that are either SI or SD, without clearly specifying the type of experiment they are conducting. Furthermore, they compare SI experimental indicators with SD ones. Our study conducted two types of experiments separately, and in the comparison section, we summarized and clarified the types of experiments conducted by other researchers.
3. The structured innovation inherently possess robustness. In other words, as long as there is a denoising paradigm in the pool suitable for the dataset, the selection of paradigms by EIEF can definitely improve the perfor-

mance of the diagnosis system and the reliability of biomarker analysis. Our experiments on two datasets have also demonstrated this point.

4. We have incorporated dynamic performance metrics during the training process of the classifier into the criteria for selecting the classification model for the SZ diagnosis system, rather than solely relying on the highest accuracy of the model. This is because the optimal accuracy is always achieved on a certain dataset, and overly focusing on accuracy will make the system lack practicality. After all, we hope that the system can be continuously retrained as the hospital database grows, to improve its diagnostic capabilities.
5. In terms of our experiments themselves, the effectiveness of a large number of denoising approaches on the two public datasets has been tested and described. This can provide a basis for selecting denoising approaches for researchers who wish to conduct research on the two datasets.

5.2 Limitations and future solutions

To be frank, our work is significant but not perfect. And here are the limitations, and the possible solutions we adopt in the future.

1. In EIEF, our grid search strategy that requires training a 1D-CNN for each dataset processed by each paradigm in the pool makes the computation for the search very time-consuming, which brings about a restriction that the pool couldn't contain too many denoising paradigms. Hence, applying advanced search algorithms to reduce time consumption has been in our future blueprint. A feasible approach is: for the five categories of denoising approaches, each category is described by a vector θ_i , where small variations in the values of θ_i represent changes in the parameters of a method within that category, while large variations can represent changes of the methods within the category. Combining the 5 θ_i , along with another ϕ vector to control the execution order of the 5 categories, we form a θ . Clearly, a determined θ value can produce a corresponding processed dataset, and then obtain a corresponding test accuracy through training. In this way, the discrete denoising paradigm pool in this study becomes a continuous pool containing an infinite paradigms. Therefore, as long as we define the domain of θ , we can use intelligent optimization algorithms to more efficiently and accurately find the best estimate of the restoration function $f_i T$. As for the selection of the optimization algorithm, since our optimization objective lacks a function analytical

expression, heuristic algorithms have become the best solutions to tackle this problem. For examples: Firefly Algorithm (FA), Grey Wolf Optimizer (GWO), Cuckoo Search Algorithm (CSA), and etc are all within our consideration.

2. In the Section 3.2, we demonstrate that, under ideal conditions, the optimal denoising paradigm found through all DNNs will remain consistent. However, the ideal conditions can never be fully satisfied under experimental environment. It has been confirmed that the 1D-CNN has a different optimum from the 3D-CNN. So, Several questions remain unanswered in the current research. We have found that, with changes of the network structure inside EIEF, paradigms that were originally thought to damage effective information, lead to an improvement in the test accuracy of the network on denoised datasets compared to the original datasets. What is the underlying mechanism behind this phenomenon? Besides, is it merely a coincidental result that the 3D-CNN exhibits better test accuracy on the optimal dataset obtained using 1D-CNN? If another type of DNN is selected, could the so-called optimal dataset potentially lead to a decrease in test accuracy? Furthermore, what conditions must be met by two classifiers to ensure that the two optimal paradigms obtained are completely consistent or similar?

To answer these questions, it is necessary to conduct two types of experiments. Firstly, ablation experiments should be conducted by making subtle changes to the structure of 1D-CNN, and then EIEF will be performed to observe whether the obtained optimal paradigm exhibits robustness. Secondly, the type of network within EIEF should be changed, such as using different CNNs or vision transformers (ViTs) like Convit, Convnext, Darknet, EfficientNet, etc. One intuitive speculation is that the more complex the network used in EIEF, the stronger its generalization ability. Therefore, the network itself can perform denoising to some extent, weakening the impact of denoising paradigms on test accuracy. Our main focus will be to examine whether changes in the network alter the direction of this impact (positive impact turning into negative impact). Additionally, we need to carefully compare the various results produced by different networks and analyze how the commonalities and differences in the networks manifest in their static and dynamic performance indicators. The aforementioned experiments are extremely resource-intensive, but we will make every effort to obtain better hardware conditions to acquire greater computational power for further optimization of EIEF.

6 Conclusion

Allow me to reiterate the three objectives that drive researchers to develop an automated SZ diagnosis system: Get the system itself; Identify and validate interpretable, discriminative features that qualify as SZ markers; construct the SZ electrophysiological markers. As previously discussed, researchers working on creating this system face significant challenges due to the inherent characteristics of EEG data, characterized by its low SNR. This presents two primary issues. One is that the noises dilute the information in EEG, leading to a decline in the diagnosis system's performance and frustrating the first objective. The other is the lack of enormous EEG datasets, causing the noise components exhibit varying distributions within the two sub-datasets representing distinct groups, ultimately making the extracted SZ markers from a dataset unreliable (frustrate the second objective) and some neurons in a trained DNN contributed from noises (frustrate the third objective). To address these challenges, denoising is necessary. However, while denoising is effective in reducing noise, it may also result in the unintended loss of valuable information required for specific classification tasks.

In such a condition, to eliminate various noises from EEG data with requisite information reserved, we have proposed our EIEF based on three key ideas. The first key is task-centric, emphasizing that our framework must selectively retain EEG components containing the effective information relevant to a specified classification task. It's important to note that this task is defined with respect to a specific dataset. To ensure the preservation of effective information, it becomes imperative to quantify it. So, the second key is made that the test metrics of an end-to-end DNN could become the measurement. Given that the DNN is purposefully trained to address the defined task, these metrics provide a practical approach to measure the proportion of task-related effective information present in the EEG dataset. The ultimate objective is to devise a denoising paradigm that maximizes this proportion, which forms the primary focus of our third key principle. The third key is rooted in the concept of feedback. We have established a paradigm pool that generates a dataset pool. For every dataset in the pool, we have used DNN's test accuracy to measure the proportion, and fed the proportion back to the pool so as to find the dataset containing the most effective information and the optimal denoising paradigm. This approach capitalizes on the composability and flexibility of open-loop denoising paradigms while mitigating the risk of these paradigms unintentionally compromising EEG information by establishing a closed-loop system. All the mathematical derivations in Sections 3.1 and 3.2 align seamlessly with the three key ideas.

As mentioned above, EIEF is our first contribution in the paper. Next, the remaining two contributions will be concluded, along with their effects on the three objectives. One contribution is having constructed an SZ diagnosis system with outstanding performance. By preprocessing the dataset with the optimal paradigm identified by EIEF and subsequently training a 3D-CNN on it, we have successfully crafted an automatic SZ diagnosis system. This system exhibits commendable performance in the SD and SI evaluation when compared to S.O.T.A systems. Obviously, this contribution represents a significant stride toward fulfilling the first objective. As for the last contribution—assessing the changes of the EEG features' differences between the patients and the healthy before and after implementing the optimal denoising paradigm on raw dataset—we recognize its multifaceted significance. On one level, it directly bolsters our pursuit of the second objective since we have declared that the optimal paradigm makes extracting the SZ electrobiological markers more convincing in Section 4.7.1. Yet, delving deeper, the optimal paradigm is found based on a DNN (1D-CNN), or we can say it is found based on those uninterpretable automatically constructed features (also SZ markers), which actually sheds light on a new thought for the third objective, of leveraging the uninterpretable markers to enhance the reliability of the interpretable markers. Without these markers (without the DNN), the second key of EIEF shall be impracticable because, in that case, it can't be discriminated whether the low proportion of the effective information or the wrong feature selection causes the low performance of a traditional machine learning model. In contrast, the DNN's automated feature selection effectively eliminates the possibility of incorrect selection. Besides, amalgamating the uninterpretable markers with the feedback mechanism, the paper transforms the conventional end-to-end DNN into a raw-to-processed-to-end model. At this point, the extraction of SZ markers is carried out on the processed dataset, which mitigates the interpretability challenges of DNNs while preserving their inherent advantage in automatic feature extraction.

Author Contributions Tianyu Jing: Conceptualization, Investigation, Methodology, Formal analysis, Data curation, Coding, Writing-original draft, Visualization. Jiao Wang: Conceptualization, Supervision, Funding acquisition. Zhifen Guo: Conceptualization, Investigation, Writing-original draft, Formal analysis. Xindong Xu: Writing-original draft, Formal analysis, Coding. Fengbin Ma: Writing review & editing, Investigation. Longyue Fu: Investigation.

Funding This work was supported in part by the National Natural Science Foundation of China under Grant 61836011 and in part by the Fundamental Research Funds for the Central Universities of China under Grant N2104001.

Availability of data and materials All data generated or analyzed during this paper are available.

Code Availability Code can be available

Declarations

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication All authors read and approved the final manuscript.

Conflicts of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Green MF, Horan WP, Lee JJRN (2015) Social cognition in schizophrenia *16*(10):620–631
- Balasubramanian K, Ramya K, Gayathri Devi K (2022) Optimized adaptive neuro-fuzzy inference system based on hybrid grey wolf-bat algorithm for schizophrenia recognition from eeg signals. *Cogn Neurodyn* 19. <https://doi.org/10.1007/s11571-022-09817-y>
- Cohen MX (2014) *Analyzing Neural Time Series Data: Theory and Practice*. MIT press, ???
- Sharma M, Deb D, Acharya UR (2018) A novel three-band orthogonal wavelet filter bank method for an automated identification of alcoholic eeg signals. *Appl Int* 48:1368–1378
- Sadiq MT, Akbari H, Siuly S, Li Y, Wen P (2022) Alcoholic eeg signals recognition based on phase space dynamic and geometrical features. *Chaos, Soliton Fract* 158:112036
- Seal A, Bajpai R, Karnati M, Agnihotri J, Yazidi A, Herrera-Viedma E, Krejcar O (2023) Benchmarks for machine learning in depression discrimination using electroencephalography signals. *Appl Int* 53(10):12666–12683
- Zhong X, Gu Y, Luo Y, Zeng X, Liu G (2023) Bi-hemisphere asymmetric attention network: recognizing emotion from eeg signals based on the transformer. *Appl Intell* 53(12):15278–15294
- Barros C, Silva CA, Pinheiro AP (2021) Advanced eeg-based learning approaches to predict schizophrenia: Promises and pitfalls. *Artif Intell Med* 114:13. <https://doi.org/10.1016/j.artmed.2021.102039>
- Khare SK, March S, Barua PD, Gadre VM, Acharya UR (2023) Application of data fusion for automated detection of children with developmental and mental disorders: A systematic review of the last decade. *Inf Fusion*, 101898
- Ranjan R, Sahana BC, Bhandari AK (2024) Deep learning models for diagnosis of schizophrenia using eeg signals: Emerging trends, challenges, and prospects. *Arch Comput Methods Eng* 1–40
- Jiang X, Bian GB, Tian Z (2019) Removal of artifacts from eeg signals: A review. *Sensors (Basel)* 19(5):18. <https://doi.org/10.3390/s19050987>
- Manis G, Aktaruzzaman M, Sassi R (2017) Bubble entropy: An entropy almost free of parameters. *IEEE Trans Biomed Eng* 64(11):2711–2718
- Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA (2015) The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Front neuroinform* 9:16
- Di Flumeri G, Arico P, Borghini G, Colosimo A, Babiloni F (2016) A new regression-based method for the eye blinks artifacts correction in the eeg signal, without using any eeg


- channel. In: 38th Annual International Conference of the IEEE Engineering-in-Medicine-and- Biology-Society (EMBC). IEEE Eng Med Biol Soc Conf Proceed, pp 3187–3190. Ieee, NEW YORK. <GotoISI>://WOS:000399823503135
15. Yang BH, Zhang T, Zhang YY, Liu WQ, Wang JG, Duan KW (2017) Removal of electrooculogram artifacts from electroencephalogram using canonical correlation analysis with ensemble empirical mode decomposition. *Cogn Comput* 9(5):626–633. <https://doi.org/10.1007/s12559-017-9478-0>
 16. Rabcan J, Levashenko V, Zaitseva E, Kvassay M (2020) Review of methods for eeg signal classification and development of new fuzzy classification-based approach. *Ieee Access* 8:189720–189734. <https://doi.org/10.1109/access.2020.3031447>
 17. Geiger BC, Kubin G (2018) Information Loss in Deterministic Signal Processing Systems. Springer, ???
 18. Oh SL, Vicnesh J, Ciaccio EJ, Yuvaraj R, Acharya UR (2019) Deep convolutional neural network model for automated diagnosis of schizophrenia using eeg signals. *Appl Sci Basel* 9(14):13. <https://doi.org/10.3390/app9142870>
 19. Wang Y, Huang Z, McCane B, Neo P (2018) Emotionet: A 3-d convolutional neural network for eeg-based emotion recognition. In: 2018 Int Joint Conf Neural Netw (IJCNN), pp 1–7. IEEE
 20. Radüntz T, Scouten J, Hochmuth O, Meffert B (2017) Automated eeg artifact elimination by applying machine learning algorithms to ica-based features. *J Neural Eng* 14(4):046004
 21. Sweeney KT, Ward TE, McLoone SF (2012) Artifact removal in physiological signals—practices and possibilities. *IEEE Trans Inf Technol Biomed* 16(3):488–500
 22. Sadiq MT, Yu X, Yuan Z, Aziz MZ (2020) Motor imagery bci classification based on novel two-dimensional modelling in empirical wavelet transform. *Electron Lett* 56(25):1367–1369
 23. Zangeneh Soroush M, Tahvilian P, Nasirpour MH, Maghooli K, Sadeghniaat Haghighi K, Vahid Harandi S, Abdollahi Z, Ghazizadeh A, Jafarnia Dabanloo N (2022) Eeg artifact removal using sub-space decomposition, nonlinear dynamics, stationary wavelet transform and machine learning algorithms. *Front Physiol*, 1572
 24. Sabeti M, Katebi S, Boostani R (2009) Entropy and complexity measures for eeg signal classification of schizophrenic and control participants. *Artif Intell Med* 47(3):263–274
 25. Parvinnia E, Sabeti M, Jahromi MZ, Boostani R (2014) Classification of eeg signals using adaptive weighted distance nearest neighbor algorithm. *J King Saud Univ-Comput Inf Sci* 26(1):1–6
 26. Murphy JR, Rawdon C, Kelleher I, Twomey D, Markey PS, Cannon M, Roche RA (2013) Reduced duration mismatch negativity in adolescents with psychotic symptoms: further evidence for mismatch negativity as a possible biomarker for vulnerability to psychosis. *BMC psychiatry* 13:1–7
 27. Jahmunah V, Oh SL, Rajinikanth V, Ciaccio EJ, Cheong KH, Arunkumar N, Acharya UR (2019) Automated detection of schizophrenia using nonlinear signal processing methods. *Artif Intell Med* 100:101698
 28. Prabhakar SK, Rajaguru H, Sun Hee K (2020) Schizophrenia eeg signal classification based on swarm intelligence computing. *Comput Intell Neurosci* : CIN 2020
 29. Sadiq MT, Yu X, Yuan Z, Zeming F, Rehman AU, Ullah I, Li G, Xiao G (2019) Motor imagery eeg signals decoding by multivariate empirical wavelet transform-based framework for robust brain-computer interfaces. *IEEE access* 7:171431–171451
 30. Sadiq MT, Yu X, Yuan Z, Aziz MZ, Rehman N, Ding W, Xiao G (2022) Motor imagery bci classification based on multivariate variational mode decomposition. *IEEE Trans Emerg Top Comput Intell* 6(5):1177–1189
 31. Krishnan PT, Raj ANJ, Balasubramanian P, Chen Y (2020) Schizophrenia detection using multivariate empirical mode decomposition and entropy measures from multichannel eeg signal. *Biocybern Biomed Eng* 40(3):1124–1139
 32. Baygin M (2021) An accurate automated schizophrenia detection using tqwt and statistical moment based feature extraction. *Biomed Signal Process Control* 68:102777
 33. Khare SK, Bajaj V (2021) A self-learned decomposition and classification model for schizophrenia diagnosis. *Comput Methods Prog Biomed* 211:106450
 34. Sharma G, Joshi AM (2022) Szhnn: a novel and scalable deep convolution hybrid neural network framework for schizophrenia detection using multichannel eeg. *IEEE Trans Instrum Meas* 71:1–9
 35. Shen M, Wen P, Song B, Li Y (2024) 3d convolutional neural network for schizophrenia detection using as eeg-based functional brain network. *Biomed Signal Process Control* 89:105815
 36. Khare SK, Bajaj V, Acharya UR (2023) Schizonet: a robust and accurate margenau– hill time-frequency distribution based deep neural network model for schizophrenia detection using eeg signals. *Physiol Meas* 44(3):035005
 37. Zülfikar A, Mehmet A (2022) Empirical mode decomposition and convolutional neural network-based approach for diagnosing psychotic disorders from eeg signals. *Appl Intell* 52(11):12103–12115
 38. Jakubovitz D, Giryas R, Rodrigues MRD (2019) In: Boche H, Caire G, Calderbank R, Kutyniok G, Mathar R, Petersen P (eds.) Generalization Error in Deep Learning, Springer, Cham, pp 153–193. https://doi.org/10.1007/978-3-319-73074-5_5. https://doi.org/10.1007/978-3-319-73074-5_5
 39. Olejarczyk E, Jernajczyk W (2017) Graph-based analysis of brain connectivity in schizophrenia. *PLoS one* 12(11):0188629
 40. Sadiq MT, Yu X, Yuan Z, Aziz MZ, Siuly S, Ding W (2021) Toward the development of versatile brain–computer interfaces. *IEEE Trans Artif Intell* 2(4):314–328. <https://doi.org/10.1109/TAI.2021.3097307>
 41. Borisov S, Kaplan AY, Gorbachevskaya N, Kozlova I (2005) Analysis of eeg structural synchrony in adolescents with schizophrenic disorders. *Hum Physiol* 31:255–261
 42. Mitra P (2007) Observed Brain Dynamics. Oxford University Press, ???
 43. Mullen T, Kothe C, Chi YM, Ojeda A, Kerth T, Makeig S, Cauwenberghs G, Jung TP (2013) : Ieee: Real-time modeling and 3d visualization of source dynamics and connectivity using wearable eeg. In: 35th Annual International Conference of the IEEE-Engineering-in-Medicine-and- Biology-Society (EMBC). IEEE Eng Med Biol Soc Conf Proceed, pp. 2184–2187. Ieee,NEW YORK. <GotoISI>://WOS:000341702102164
 44. Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13(4):411–430. [https://doi.org/10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5)
 45. Hyvarinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *Ieee Trans Neural Netw* 10(3):626–634. <https://doi.org/10.1109/72.761722>
 46. Buettner R, Beil D, Scholtz S, Djemai A (2020) Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute eeg recordings
 47. Buettner R, Hirschmiller M, Schlosser K, Rössle M, Fernandes M, Timm IJ (2019) Highperformance exclusion of schizophrenia using a novel machine learning method on eeg data. In: 2019 IEEE Int Conf E-Health Netw Appl Serv (HealthCom) pp 1–6. IEEE
 48. Singh K, Singh S, Malhotra J (2021) Spectral features based convolutional neural network for accurate and prompt identification of schizophrenic patients. *Proceedings of the Institution of Mechanical Engineers, Part H: J Eng Med* 235(2):167–184
 49. Aslan Z, Akin M (2020) Automatic detection of schizophrenia by applying deep learning over spectrogram images of eeg signals. *Trait Signal* 37(2)
 50. Aslan Z, Akin M (2022) A deep learning approach in automated detection of schizophrenia using scalogram images of eeg signals. *Phys Eng Sci Med* 45(1):83–96

51. Ilakiyaselvan N, Khan AN, Shahina A (2022) Reconstructed phase space portraits for detecting brain diseases using deep learning. *Biomed Signal Process Control* 71:103278
52. Bagherzadeh S, Shahabi MS, Shalbaf A (2022) Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal. *Comput Biol Med* 146:105570
53. Wu Y, Xia M, Wang X, Zhang Y (2022) Schizophrenia detection based on eeg using recurrent auto-encoder framework. In: *Int Conf Neural Inf Process*, pp 62–73. Springer
54. Lillo E, Mora M, Lucero B (2022) Automated diagnosis of schizophrenia using eeg microstates and deep convolutional neural network. *Expert Syst Appl* 209:118236
55. Naira T, Alberto C (2020) Classification of people who suffer schizophrenia and healthy people by eeg signals using deep learning. *schizophrenia*
56. Phang CR, Ting CM, Samdin SB, Ombao H (2019) Classification of eeg-based effective brain connectivity in schizophrenia using deep neural networks. In: *2019 9th Int IEEE/EMBS Conf Neural Eng (NER)*, pp. 401–406. IEEE
57. Calhas D, Romero E, Henriques R (2020) On the use of pairwise distance learning for brain signal classification with limited observations. *Artif intell med* 105:101852
58. Supakar R, Satvaya P, Chakrabarti P (2022) A deep learning based model using rnn-lstm for the detection of schizophrenia from eeg data. *Comput Biol Med* 151:106225
59. Phang CR, Noman F, Hussain H, Ting CM, Ombao H (2019) A multi-domain connectome convolutional neural network for identifying schizophrenia from eeg connectivity patterns. *IEEE J Biomed Health Inf* 24(5):1333–1343
60. Shen M, Wen P, Song B, Li Y (2023) Automatic identification of schizophrenia based on eeg signals using dynamic functional connectivity analysis and 3d convolutional neural network. *Comput Biol Med* 160:107022
61. Amer NS, Belhaouari SB (2023) Eeg signal processing for medical diagnosis, healthcare, and monitoring: A comprehensive review. *IEEE Access* 11:143116–143142. <https://doi.org/10.1109/ACCESS.2023.3341419>
62. Akbari H, Sadiq MT, Jafari N, Too J, Mikaelvand N, Cicone A, Serra Capizzano S (2023) Recognizing seizure using poincaré plot of eeg signals and graphical features in dwt domain. *Bratisl Med J*
63. Yang J, Choudhary GI, Rahardja S, Franti P (2020) Classification of interbeat interval time-series using attention entropy. *IEEE Trans Affect Comput*
64. Henry M, Judge G (2019) Permutation entropy and information recovery in nonlinear dynamic economic time series. *Econ* 7(1):10
65. Esteller R, Echaz J, Tcheng T, Litt B, Pless B (2001) Line length: an efficient feature for seizure onset detection. In: *2001 Conf Proceed 23rd Ann Int Conf IEEE Eng Med Biol Soc*, vol. 2, pp 1707–1710. IEEE
66. Rostaghi M, Azami H (2016) Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process Lett* 23(5):610–614
67. Li P, Liu C, Li K, Zheng D, Liu C, Hou Y (2015) Assessing the complexity of short term heartbeat interval series by distribution entropy. *Med Biol Eng Comput* 53(1):77–87
68. Liu X, Jiang A, Xu N, Xue J (2016) Increment entropy as a measure of complexity for time series. *Entropy* 18(1):22
69. Hsu CF, Wei SY, Huang HP, Hsu L, Chi S, Peng CK (2017) Entropy of entropy: Measurement of dynamical complexity for biological systems. *Entropy* 19(10):550
70. Guignard F, Laib M, Amato F, Kanevski M (2020) Advanced analysis of temporal data using fisher-shannon information: theoretical development and application in geosciences. *Front Earth Sci* 8:255
71. Omidvarnia A, Mesbah M, Pedersen M, Jackson G (2018) Range entropy: A bridge between signal complexity and self-similarity. *Entropy* 20(12):962
72. Cuesta-Frau D (2019) Slope entropy: A new time series complexity estimator based on both symbolic patterns and amplitude information. *Entropy* 21(12):1167
73. Li Y, Yang Y, Li G, Xu M, Huang W (2017) A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mrmr feature selection. *Mech Syst Signal Process* 91:295–312
74. Tsallis C (2009) Introduction to nonextensive statistical mechanics: approaching a complex world. *Springer* 1(1):2–1
75. Vignat C, Bercher JF (2003) Analysis of signals in the fisher-shannon information plane. *Phys Lett A* 312(1-2):27–33
76. Kantelhardt JW, Zschiegner SA, Koscielny Bunde E, Havlin S, Bunde A, Stanley HE (2002) Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Stat Mech Appl* 316(1-4):87–114

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Tianyu Jing¹ · Jiao Wang¹  · Zhifen Guo¹ · Fengbin Ma¹ · Xindong Xu¹ · Longyue Fu¹

✉ Jiao Wang
wangjiao@mail.neu.edu.cn

Tianyu Jing
2100760@stu.neu.edu.cn

Zhifen Guo
2210301@stu.neu.edu.cn

Fengbin Ma
2300927@stu.neu.edu.cn

Xindong Xu
2270958@stu.neu.edu.cn

Longyue Fu
2110329@stu.neu.edu.cn

¹ College of Information Science and Engineering, Northeastern University, Heping, Shenyang 110167, Liaoning, China