Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **9**, Number **2**, May **2024**

HRTS: Hierarchical Rauch-Tung-Striebel Smoother With Online Learning Priors for EEG Denoising

Chuan-Sheng Wang†

Department of Automatic Control Technical
Polytechnic University of Catalonia
Autonomous Region of Catalonia, Barcelona, Spain
wangcs95@163.com

Ling Zhang†

School of Computer and Data Science
Minjiang University
No.200, Xiyuangong Road, Fuzhou University Town, Fuzhou City, Fujian Province
zhling30@163.com

Tian-Lin Fu

School of Computer and Data Science
Minjiang University
No.200, Xiyuangong Road, Fuzhou University Town, Fuzhou City, Fujian Province
1003785211@qq.com

Zhao-Qi Chen

College of Computer and Big Data
Fuzhou University
No.2 Wulongjiang North Road, Fuzhou University Town, Fuzhou City, Fujian Province
zhaoq_chen@163.com

Fu-Quan Zhang*

School of Computer and Data Science
Minjiang University
No.200 Xiyuangong Road, Fuzhou University Town, Fuzhou City, Fujian Province
Digital Media Art
Key Laboratory of Sichuan Province
Sichuan Conservatory of Music
Fuzhou Technology Innovation Center of Intelligent Manufacturing Information System
Minjiang University
No.200 Xiyuangong Road, Fuzhou University Town,Fuzhou City, Fujian Province
Engineering Research Center for ICH Digitalization and Multi-source Information Fusion(Fujian Polytechnic Normal University)
Fujian Province University
No.1 Campus New Village, Longjiang Street, Fuqing City, Fujian Province
zfq@mju.edu.cn

† The first two authors contributed equally to this work.
*Corresponding author: Fuquan Zhang

Received June 14, 2023, revised September 17, 2023, accepted December 27, 2023.

Abstract. *Electroencephalogram (EEG) is a nonlinear signal that reflects the physiological state of the brain at different times, containing rich information. However, the possible interference during the collection and transmission process often leads to a large amount of unnecessary noise in the EEG signal. Traditional denoising methods may face difficulties in high-dimensional data processing and the inability to effectively handle different types of Gaussian white noise. To better eliminate the interference of Gaussian white noise, this article specifically adopts the Hierarchical Rauch-Tung-Striebel Smoother (HRTS) method. This method can effectively integrate the learned structural prior EEG signals into the state space model, thereby describing the relationship between EEG signals and noise. Capture the spatiotemporal characteristics of EEG signals through hierarchical modeling, and optimize the components of EEG signals through estimation and prediction of hidden variables. Finally, the mean square error (MSE) is used as an evaluation indicator to compare and evaluate the denoising results. Empirical research has shown that the HRTS algorithm can not only effectively reduce runtime and better process high-dimensional data, but also significantly improve the quality of EEG signals, effectively suppress noise interference, and more accurately reflect the characteristics of brain activity. Compared to denoising algorithms such as Kalman filtering and Kalman wavelet filtering, the HRTS algorithm has more advantages in denoising EEG signals.*
Keywords: EEG Denoising, Rauch–Tung–Striebel smoother, Hierarchical State Space Model, Expectation-Maximum algorithm.

1. **Introduction.** EEG is a graphical representation of the electrical activity of the human brain. It is made by a sophisticated instrument that measures the electrical signals generated by the electrical activity of neurons in the brain from the surface of the scalp, amplifies them, and records them. Thus, EEG signals can provide most of the required information about brain activity and reflect the functional state of the brain [1]. It is widely used in brain-computer interfaces, clinical diagnostics, computational neuroscience, etc [2, 3, 4]. Artifacts arising from muscle activity, blinking, and work-frequency interference in EEG signals are an important issue in EEG signal processing studies. These artifacts must be corrected before further analysis [5]. Generally speaking, the EEG signal has strong background noise and weak signal, and eliminating unnecessary noise in the EEG signal can make the EEG signal clearer and more accurate.

To correct the artifacts in EEG signals, many techniques have been proposed for EEG denoising. Among them, Mashhadi et al. [6] proposed a deep learning model U-NET, which learns to purify contaminated EEG signals by converting each EEG signal into an image and inputting it into the U-NET model. Although deep learning-based models have good performance in removing EOG artifacts from EEG signals, deep learning methods typically require a large amount of labeled data for training to achieve better performance, which increases time and money costs. In addition, Tahir Akhtar et al. [7] proposed a method that combines independent component analysis (ICA) and wavelet denoising to automatically remove artifacts from multi-channel EEG data. The author combines spatial constrained ICA with wavelet denoising to effectively remove noise caused by artifacts while preserving the activity of EEG signals. However, the process of selecting and setting parameters in ICA and wavelet denoising methods in this method is not specified, which may affect the processing effectiveness of the method. The most commonly used method at present is the ICA method based on blind source separation. This method can separate neuronal signals from other signals, thereby removing noise. Although ICA can effectively separate signals, each component in the results may represent multiple source signals or components containing noise, and the resulting results may lack some interpretability. ICA requires a large amount of computing resources to process a large amount of data, which may require a long running time.

HKF is designed based on a hierarchical state space(SS) Model Inspired by Locher et al. [8], we refer to the online learning-based a priori hierarchical Kalman Filter (HKF) denoising method and combine it with EEG signal characteristics to design the hierarchical Kalman Smoother Hierarchical Rauch-Tung- for EEG denoising. Specifically, HRTS consists of a structured evolution learned online that utilizes this prior knowledge for EEG denoising. It does not require pre-training in a supervised manner and preserves the interpretability of the KF due to its online covariance estimation and pre-learned structured evolution while accelerating the denoising process. These properties enable HKF to remove artifacts well on multi-level, high-dimensional, multi-source EEG data and obtain relatively complete signal information.

We adopted the method by Locher et al. [8]. and applied it to EEG data denoising, and found the following advantages.

Good Retention of Information at the Peaks. Since there are many different scales of noise in the EEG signal, such as physiological noise, environmental noise, etc. These noises may cause interference at the peak of the signal and affect the signal quality. The HRTS algorithm uses a multi-level filter structure, which can filter and smooth the signal several times while estimating the state and noise variance of the signal at each level, to estimate the true value of the signal more accurately, and retain the information at the peak better and improve the quality of the signal. One of the multi-level filtering structures can eliminate low-frequency noise by using a high time scale while using a low time scale to eliminate high-frequency noise. This allows better retention of information at the peaks and improves the quality of the signal.

More Adaptive and Shorter Running Time. HRTS is a state-space model-based method that models the dynamic model of a signal and updates the model parameters according to the actual signal. Thus, it can adaptively adjust the parameters by adapting to different types of signals. Not only that, HRTS uses a recursive algorithm to estimate the variance of the noise while estimating the signal, which not only shortens the running time but also removes artifacts better.

Stronger for Emergent Noise and Outliers. Traditional denoising methods usually can only handle some specific noise types and are weaker for emergent noise and outliers. The recursive algorithm used by HRTS, on the other hand, allows it to better handle these problems by estimating the dynamic model of the signal.

HRTS also has certain limitations. Specifically, this method uses prior knowledge from online learning to denoise EEG signals. The performance may not be stable enough during the initial learning stage, resulting in poor denoising performance. In response to the above limitations, we believe that improving the performance and stability of the algorithm is the primary task. This may include designing more efficient online learning strategies, developing more stable initialization methods, and automating parameter selection. In addition, efforts should be made to improve the adaptability of algorithms to handle EEG data of different quality levels and types.

2. **Related Works.** Brain-computer interface is a new type of human-computer interaction system used to realize the communication between humans and the external environment. Feature extraction and pattern classification of EEG signals have become a hot issue in Brain-Machine Interface research, and the removal of EEG artifacts is also an important part of the research. Therefore, it is important to denoise EEG data. EEG denoising can be divided into two categories: traditional EEG denoising and deep learning-based EEG denoising. In this chapter, the advantages and disadvantages of the above two methods will be described in detail.

2.1. **Traditional EEG Denoising.** Traditional EEG denoising methods include wavelet denoising, independent component analysis, etc. Traditional EEG denoising methods usually use simple and intuitive techniques and algorithms that are easy to understand and implement. For example, a wavelet-based EEG eye artifact denoising technique was proposed by Zikov et al. [9]. This technique uses generic thresholding and statistical thresholding functions to efficiently process wavelet-based EEG signals to remove noise. However, they do not threshold all wavelet coefficients because the frequency of ocular artifacts is very low, so the thresholding is performed in the lower frequency band. The inaccuracy of threshold estimation may result in the removal of useful information from the signal. Traditional EEG denoising methods can also be used in combination with other denoising methods to improve the denoising effect. McMenamin et al. [10] proposed a method for the validation of ICA-based scalp and source localization EEG myogenic artifact correction. The method applies independent component analysis to the EEG signal, decomposing it into source signals. By identifying and excluding muscle artifacts in the ICA component, the source of the EEG signal can be located more accurately. However, the method also requires some prior assumptions and parameter settings, which may also affect the accuracy and reliability of the results.

Traditional EEG denoising methods are easy to use, interpretable, and have some denoising effects. However, they suffer from dependence on the amount of data and vulnerability to parameter selection. Parameter selection is a very important part of the EEG denoising process. However, traditional methods have certain drawbacks in the estimation and selection of parameters. In contrast, the method in this paper is able to achieve the best noise separation according to the characteristics of the input EEG signal by adaptively adjusting the parameters and thresholds within the algorithm. By automatic adjustment, our method can effectively optimize the denoising results without manually adjusting the parameters. In addition, our method has a low computational complexity and short processing latency capability.

In summary, Traditional EEG denoising methods include wavelet denoising, independent component analysis, etc. Although traditional methods are easy to understand and implement, there is also a dependence on data volume and vulnerability to parameter selection. Parameter selection is a very important part of the EEG denoising process. In contrast, the method proposed in this article can adaptively adjust parameters and achieve optimal noise separation based on the characteristics of the input EEG signal.

2.2. **EEG Denoising Based on Deep Learning.** With the continuous development of science and technology, researchers have conducted in-depth analyses of EEG signal characteristics and noise sources and used deep learning methods to work on improving signal quality and removing noise. Some of the commonly used methods in this regard include combining convolutional neural networks (CNN), recurrent neural networks (RNN), and generative adversarial networks (GAN). The application of these methods aims to automatically learn the features and noise distribution in EEG signals through the learning and optimization capabilities of deep learning models to achieve denoising and signal enhancement.

A new end-to-end 1D ResCNN model proposed by Sun et al. [11] for automatic removal of artifacts in contaminated EEG signals. However, this method removes artifacts and noise while removing high-frequency noise in the spectrum, this may affect the quality of the original EEG signal. The HRTS algorithm consists of two main steps, namely filtering and smoothing. Filtering is used for real-time estimation of the system state, while smoothing is used for more accurate estimation of the state over the entire time series. This enables the HRTS algorithm to handle artifacts and noise while preserving

the quality of the original signal. Zhang et al. [12] proposed a new convolutional neural network model for removing EEG muscle artifacts, and compared with other types of convolutional neural networks, the model solves the problem of overfitting to a large extent. However, the insufficient amount of electromyogram (EMG) data contained in the dataset during the training of the model may lead to the inability of the neural network to learn the complex features of magnetoencephalography (MEG). Both methods mentioned above involve signal processing and estimation issues, but HRTS is mainly used for system state estimation, while convolutional neural network models are mainly used for signal denoising and feature learning. HRTS can effectively solve this problem when processing large amounts of data [13, 14, 15].

Compared with deep learning, the HRTS algorithm relatively does not require large-scale data annotation because it focuses on subspace analysis and dimensionality reduction analysis, and does not involve complex annotation and deep learning. When analyzing EEG signals, the HRTS algorithm usually focuses on finding correlations between different frequencies and channels in high-dimensional space, and then reducing them to low-dimensional subspaces. This helps to capture spatial features in EEG signals, such as electrode interactions and connections in brain regions. Moreover, it can adaptively adjust the parameters of the filter based on the time-varying characteristics of the signal, thereby better-removing noise. Usually, the HRTS algorithm can provide good interpretability because its output is based on linear or nonlinear subspace representations.

In summary, these deep learning methods show potential in EEG signal denoising and improve the signal quality to some extent. However, deep learning methods also need to consider practical issues such as data volume, model complexity, and availability of labeled data when applied. The HRTS algorithm proposed in this paper has high denoising capability and simple and easy-to-understand implementation steps through adaptive parameter adjustment and the application of the hierarchical Kalman filter. It provides a more effective and reliable solution for EEG signal denoising.

3. **Methodology.** In this paper, we propose a hierarchical Kalman smoother for EEG denoising. Specifically, the method in this paper is divided into the following parts: first, we describe the EEG denoising task and the SS model (Sec. 3.1). To learn the parameters required in the denoising process, we fit the prior evolution function $\mathbf{f}_t(\cdot)$ using Taylor's least squares (LS) and the expectation-maximization(EM) algorithm to estimate the process noise covariance matrix $\mathbf{Q}_t$ and the observation noise covariance matrix $\mathbf{R}_t$ for the prior data (Sec. 3.2). Finally, we describe the process of forward and backward propagation of the Kalman smoother using the learned parameters (Sec. 3.3). Figure 1 is an overview of our method.

3.1. **Task Modeling. Denoising task:** The EEG denoising task is the process of extracting clean EEG signals from noisy EEG signals. We record the noise-free multichannel EEG signal at moment t as $\mathbf{x}_\mathrm{t} \in \mathbb{R}^\mathrm{n}$ and the noise-added EEG signal as $\mathbf{y}_\mathrm{t} \in \mathbb{R}^\mathrm{n}$:

$$\Psi : \{\mathbf{y}_i\}_{i=1}^{t} \mapsto \hat{\mathbf{x}}_{t|t}, \quad \Psi^* = \arg\min \mathbb{E} \left\| \hat{\mathbf{x}}_{t|t} - \mathbf{x}_t \right\|^2 \tag{1}$$

SS model: Dynamical systems are used to describe the evolution of an object or system over time. In our denoising task, the use of SS models is a canonical approach for modeling discrete-time EEG dynamical systems [16].

Meanwhile, inspired by the cyclic smooth signal representation and the periodic ECG signal denoising task [8, 17, 18], using the dynamic evolution process of EEG time series, we model the brain electrical signals using a hierarchical SS model of dynamic system modeling. It is decomposed into multiple periodic segments and each segment is dynamically modeled and features are extracted. The dataset is divided into an a priori dataset

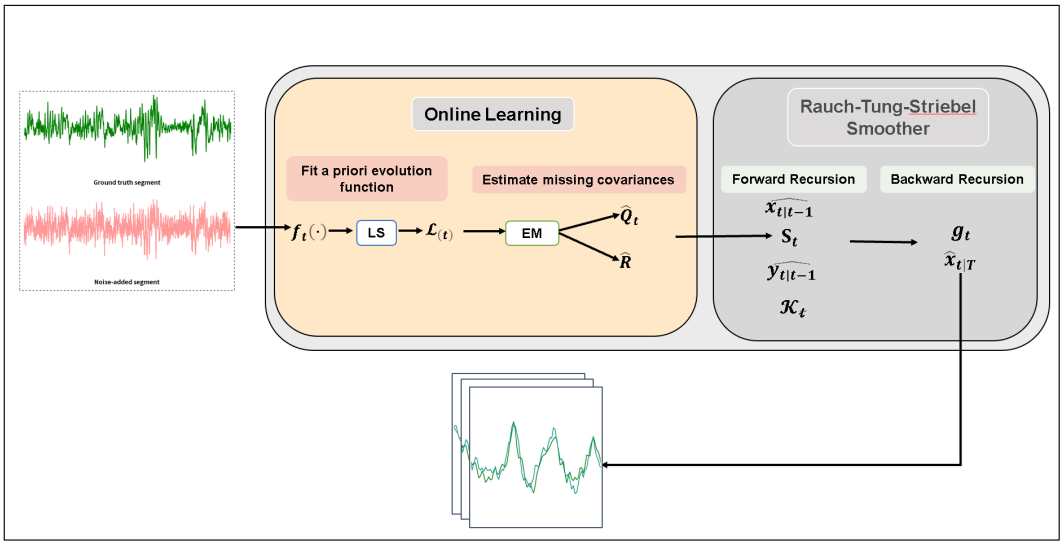


FIGURE 1. HRTS. shows the denoising method in this study. First, the dataset $\mathbf{y}_t$ with Gaussian noise added is divided into a priori and test datasets and fed into a hierarchical SS model for modeling. The state evolution function $\mathbf{f_t}(\cdot)$ is obtained by point-to-point mapping. Taylor least squares (LS) is used to learn the model parameters. Next, the noise covariance matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$ of the prior data are estimated using the EM algorithm. The learned prior knowledge is fed into the Kalman smoother for the forward and backward propagation process. Finally, the denoised signal $\hat{\mathbf{x}}_{t|\mathrm{T}}$ is obtained by this method.

and a test dataset, and for each time step $t \in \{1, ..., T\}$ we denote it as:

$$\mathbf{x}_{\tau,t} = \mathbf{f}_t\left(\mathbf{x}_{\tau,t-1}\right) + \mathbf{e}_{\tau,t}, \quad \mathbf{e}_{\tau,t} \sim \mathcal{N}\left(0, \mathbf{Q}_t\right). \tag{2a}$$

$$\mathbf{y}_{\tau,t} = \mathbf{x}_{\tau,t} + \mathbf{v}_{\tau,t}, \quad \mathbf{v}_{\tau,t} \sim \mathcal{N}\left(0, \mathbf{R}_t\right). \tag{2b}$$

The equations describe the state changes of the EEG data at different time points and the resulting observations. In particular, the state evolution function $\mathbf{f_t}(\cdot)$ describes the process of transferring the system state from moment t-1 to a point-by-point vector mapping at moment t. We will describe the state evolution function $\mathbf{f}(\cdot)$ in detail in (Sec. 3.2). In addition, the evolutionary noise $\mathbf{e}_{\tau,t}$ represents the random perturbation of the system state during the transfer process, while the observation noise $\mathbf{v}_{\tau,t}$ represents the random error of the observation results. The covariance matrices of both noises obey a Gaussian distribution $\mathcal{N}(0, \mathbf{N}_t)$.

In EEG signals, the state evolution noise is usually used to describe the evolution error of the signal in time, and the sources may include natural variations of EEG signals, small changes in electrode positions, etc. Observation noise is mainly used to describe the random error of the EEG signal during the measurement process, and the sources include poor contact of electrodes, industrial frequency interference, environmental electromagnetic interference, etc.

For the test data set, we denote it as:

$$\mathbf{x}_{\tau,t} = \mathcal{F}\left(\mathbf{x}_{\tau-1,t}\right) + \boldsymbol{\epsilon}_{\tau,t}, \quad \boldsymbol{\epsilon}_{\tau.t} \sim \mathcal{N}\left(0, \mathcal{Q}_{\tau,t}\right) \tag{3a}$$

$$\mathbf{y}_{\tau,t} = \mathbf{x}_{\tau,t} + \boldsymbol{\nu}_{\tau,t}, \quad \boldsymbol{\nu}_{\tau,t} \sim \mathcal{N}\left(0, \mathcal{R}_{\tau,t}\right) \tag{3b}$$

To complete the hierarchical modeling, we note that the state vector and observation of the EEG signal and its corresponding observations for the first cycle τ are:

$$\mathcal{X}_\tau = [\mathbf{x}_{\tau,1}, \ldots, \mathbf{x}_{\tau,\mathrm{T}}] \in \mathbb{R}^{m\times \mathrm{T}} \tag{4a}$$

$$\mathcal{Y}_\tau = [\mathbf{y}_{\tau,1}, \ldots, \mathbf{y}_{\tau,\mathrm{T}}] \in \mathbb{R}^{m\times \mathrm{T}} \tag{4b}$$

where m is the dimension of the state vector and T is the length of the time series.

3.2. **Online Learning.** To fit the missing prior evolution function $\mathbf{f}_t(\cdot)$ in the hierarchical SS model and learn the evolutionary noise covariance $\mathbf{Q}_t$ and the measurement noise covariance $\mathbf{R}_t$. In this section, we use the Taylor approximation algorithm and the EM algorithm with less computational cost by unsupervised learning Update the process parameters.

3.2.1. *Fitting A Priori Evolution Function.* The role of the state evolution function $f(\cdot)$ in the Kalman filter is to describe the process of state change of a dynamical system, i.e., the process of transferring from the state of $\tau - 1$ to the state of the next cycle τ. In the SS model of the Kalman filter, the state variable $\mathbf{x}_t$ is a time-varying hidden variable, which is usually not directly observed but can only be inferred indirectly from the observed variable $\mathbf{y}_t$ and the state transfer function $f(\cdot)$.

By using a finite Taylor expansion over time, we express the evolution of the state variables $\mathrm{d}\mathbf{x}_t$ is equivalent to the direction of change of the state vector over the time interval Δ_t as follows:

$$\mathrm{d}\mathbf{x}_t \triangleq \mathbf{x}_{t+\Delta t} - \mathbf{x}_t = \sum_{k=1}^{K} \frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t^k} \cdot \frac{\Delta t^k}{k!}, \quad K \to \infty \tag{5}$$

We use Equation (5) to represent the system state differences at time step t as a matrix of $\Phi^\top$ and use the first K terms of the Taylor expansion to approximate the amount of change in the state quantities. where Φ denotes the time interval vector,$\mathbf{\Theta}_t$ denotes the K-order derivative of the system state vector x_t.

$$\mathrm{d}\mathbf{x}_t \approx \Phi^\top \cdot \mathbf{\Theta}_t, \quad \Phi^\top = \left(\Delta t, \frac{\Delta t^2}{2}, \ldots, \frac{\Delta t^K}{K!}\right) \tag{6}$$

We use the Taylor approximation for $f_t(\cdot)$, and bringing in the parameters in Equation (6), we can structure the state evolution function simply as.

$$\mathbf{f}_t\left(\mathbf{x}_{t-1}\right) = \mathbf{x}_{t-1} + \Phi^\top \cdot \mathbf{\Theta}_t \tag{7}$$

We use LS linear regression to learn the model parameters $\mathbf{\Theta}t$. For multiple EEG observations $t \in \mathcal{T}$, $\tau = 1, \ldots, N$. The LS is defined by means of a time window, and the weights of the window can be chosen arbitrarily, but need to satisfy the constraint that the weighted sum is 1. The window length is $\tilde{L} = L_1 + L_2 + 1$, where w_ℓ is the sliding window weight. $\hat{\mathbf{y}}_{t+\ell}$ denotes the observed quantity obtained using the current parameter estimation,$\mathbf{y}_{t+\ell}$ denotes the true observed quantity:

$$\mathcal{L}(t) = \sum_{\tau=1}^{N} \sum_{\ell=-L_1}^{L_2} w_\ell \cdot \left|\mathrm{d}\hat{\mathbf{y}}_{\tau,t+\ell} - \mathrm{d}\mathbf{y}_{\tau,t+\ell}\right|^2 \tag{8}$$

By exploiting the temporal correlation within the EEG signal, smoother a priori information can be obtained, resulting in better denoising effects.

3.2.2. *Learning The Missing Noise Covariance.* While traditional methods for estimating noise covariance of SS models require estimation operations on the whole data set, which may get into the dilemma of local minima and high computational effort, the EM algorithm can update the parameters step by step at each observation of the data. This method only needs to store the state information of the previous time step observations instead of storing the whole data set and has higher accuracy and faster convergence [19, 20, 21].

In this paper, we use an iterative EM algorithm with alternating expectation steps(E-steps) and maximization steps (M-steps) for maximum likelihood estimation to obtain the expected covariance of the observed and implied state vectors. By iteratively updating the covariance, we can gain a deeper understanding of the correlation between the implied states and the observable variables. Therefore, it is possible to balance the significance of the implied state and the observable variables on the system more effectively, and thus improve the prediction accuracy of the states and observations.

E-step is a conditional expectation, where the expected covariance is calculated by RTS smoother, and $\mathbf{X}$, $\mathbf{Y}$ denote the expected covariance of the hidden states and observations, respectively:

$$\begin{aligned} \mathbf{X}_t^{\mathrm{II}} = \left\|\hat{\mathbf{x}}_{t|\mathrm{T}}\right\|^2 + \mathbf{P}_{t|T}, \quad \mathbf{Y}_t^{\mathrm{II}} = \|\mathbf{y}_t\|^2 \\ \mathbf{XY}_{t,t} = \hat{\mathbf{x}}_{t|\mathrm{T}}\mathbf{y}_t^{\top}, \\ \mathbf{XX}_{t,t-1} = \hat{\mathbf{x}}_{t|\mathrm{T}} \cdot \hat{\mathbf{x}}_{t-1|\mathrm{T}} + \mathbf{P}_{t|T} \cdot \mathcal{G}_{t-1}^{\top}. \end{aligned} \tag{9}$$

Specifically, $\mathbf{X}_t^{\mathrm{II}}$ describes the correlation between different components of the implied state vector within the same moment, and $\mathbf{Y}_t^{\mathrm{II}}$ describes the correlation of the observable variables themselves. The expected covariance matrix $\mathbf{XY}_{t,t}$ reflects the degree of correlation between the implied state vector and the observable variables, while $\mathbf{XX}_{t,t-1}$ reflects the correlation and uncertainty transfer between the implied state vector at adjacent time steps, describing the degree of correlation of the state vector at adjacent time points.

M-step is performed based on the noise covariance given by E-step by maximizing the log-likelihood function and updating the noise covariance matrices $\hat{\mathbf{Q}}_t$ and $\hat{\mathbf{R}}_{\mathbf{t}}$:

$$\begin{aligned} \tilde{\mathbf{Q}}_t = \mathbf{X}_t^{\mathrm{II}} - 2 \cdot \mathbf{XX}_{t,t-1} + \mathbf{X}_{t-1}^{\mathrm{II}}.\hat{\mathbf{Q}}_t = \frac{1}{\tilde{L}} \cdot \sum_{\ell=-L_1}^{L_2} \tilde{\mathbf{Q}}_{t+\ell}. \\ \tilde{\mathbf{R}}_t = \mathbf{Y}_t^{\mathrm{II}} - 2 \cdot \mathbf{YX}_{t,t} + \mathbf{X}_t^{\mathrm{II}}.\hat{\mathbf{R}} = \frac{1}{\mathrm{T}} \cdot \sum_{\tilde{t}=1}^{\mathrm{T}} \tilde{\mathbf{R}}_{\tilde{t}}. \end{aligned} \tag{10}$$

For $\hat{\mathbf{Q}}_t$, we first compute the intermediate variable $\tilde{\mathbf{Q}}_t$, which contains the state estimate $\mathbf{X}_t^{\mathrm{II}}$ at the current moment t and the estimation error covariance matrix $\mathbf{XX}_{t,t-1}$, and its state estimation error covariance matrix $\mathbf{X}_{t-1}^{\mathrm{II}}$ for the previous moment $t-1$. For a certain range of time steps ℓ, we weight the summation of $\tilde{\mathbf{Q}}_{t+\ell}$ to obtain the maximum likelihood estimate of $\mathbf{Q}_t$ at the current moment.

The M-Step calculation produces a new set of noise covariance parameter estimates, which will be used in subsequent E-step calculations to iteratively update the state estimates and the estimation error covariance matrix to obtain more accurate model parameter estimates, and thus better denoising results.

3.3. **Rauch-Tung-Striebel.** Rauch-Tung-Striebel is an extension of KF to smooth EEG data y_t by the learned prior function $\mathbf{f}_t$ and evolutionary noise covariance $\hat{\mathbf{Q}}_t$ and observation noise covariance $\hat{\mathbf{R}}_{\mathbf{t}}$ to obtain more accurate estimation results. The smoothing task is performed in the context of the SS model, i.e., observation recovery and hidden state

estimation, which differs from Kalman filtering in that Kalman smoothing uses both past and future observations to estimate the system state, resulting in smoother results.

In Kalman smoothing, we first use the Kalman filter to process the observations to obtain optimal estimates of the system state. Then, we use these optimal estimates and the observed data to obtain a smoothed estimate of the system state by recursively iterating over them. This process can be implemented using the "forward-backward" algorithm.

3.3.1. *Forward Recursion.* Forward propagation is an important step in the Kalman smoother. In fact, the forward propagation process is closely related to Kalman filtering, as they both involve the estimation and prediction of the system state. Therefore, it can be said that the process of forward propagation is the process of Kalman filtering. Specifically, forward propagation is divided into two steps:

(i) Predict Step: In the prediction step, the state of the next moment is predicted using the dynamic SS equations of the system, i.e., a priori estimates are computed. This step is usually done using the prediction equation of the Kalman filter, where the prediction equation predicts the state of the next moment based on the state estimate of the previous moment and the control input:

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{f}_t\left(\hat{\mathbf{x}}_{t-1|t-1}\right), \mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \mathbf{Q}_t \tag{11a}$$

$$\hat{\mathbf{y}}_{t|t-1} = \hat{\mathbf{x}}_{t|t-1}, \mathbf{S}_t = \mathbf{P}_{t|t-1} + \mathbf{R}_t \tag{11b}$$

The state estimate of the previous moment $\hat{\mathbf{x}}_{t-1|t-1}$ is predicted by the state transfer function $\mathbf{f}_t(\cdot)$ to obtain the state prediction of the current moment $\hat{\mathbf{x}}_{t|t-1}$.
$hat\mathbf{y}_{t|t-1}$ means that in the absence of new observations, the predicted value of the system state is equal to the current state estimate.

$\mathbf{S}_t$ is the covariance matrix used to determine the difference between the state predicted and observed values in the state update. The smaller $\mathbf{S}_t$ indicates that the difference between the predicted and observed values is smaller. It is obtained by summing the process noise covariance matrix $\mathbf{P}_{t|t-1}$ and the observation noise covariance matrix $\mathbf{R}_t$. The initial value of $\mathbf{P}_{t|t-1}$, which reflects the uncertainty of state prediction, is determined by the covariance matrix $\mathbf{Q}_t$ of the process noise learned in the great likelihood estimation stage and then iteratively computed by recursive updating.

(ii) Update Step: In the update step, the observations at the current moment are used to update the prior estimates, i.e., the posterior estimates are computed. This step is usually done using the update equation of the Kalman filter, where the update equation updates the state estimates based on the predicted and observed values.

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathcal{K}_t \cdot \Delta\mathbf{y}_t, \quad \Delta\mathbf{y}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}, \tag{12a}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathcal{K}_t \cdot \mathbf{S}_t \cdot \mathcal{K}_t^{\top}, \mathcal{K}_t = \mathbf{P}_{t|t-1} \cdot \mathbf{S}_t^{-1}. \tag{12b}$$

The Kalman gain $\mathcal{K}_t$ is the most important parameter in the update phase, which is used to weigh the uncertainty between the predicted and observed values and combine them to update the state estimates. The $\mathcal{K}_t$ can be regarded as a weight coefficient whose magnitude determines the proportion of weights the Kalman filter algorithm assigns between the prediction model and the observation model.

When $\mathcal{K}_t$ is large, the Kalman filter algorithm trusts the observations more and therefore more weight is placed on the observed model. Conversely, when $\mathcal{K}_t$ is small, the Kalman filter algorithm trusts the predicted values more, and therefore more weight is placed on the prediction model. By adjusting the size of $\mathcal{K}_t$, the Kalman filter algorithm can be better adapted to different systems and environments, thus improving the accuracy and stability of state estimation.

3.3.2. *Backward Recursion.* The process of RTS backward propagation is similar to the update step mentioned earlier, the difference lies in the different direction of information propagation. The Kalman filter propagates from the current moment t to the next moment $t+1$, while the RTS algorithm back-propagates from the future moment T to the current moment t.

$$\hat{\mathbf{x}}_{t|\mathrm{T}} = \hat{\mathbf{x}}_{t|t} + \mathcal{G}_t \cdot \Delta\mathbf{x}_t, \quad \Delta\mathbf{x}_t = \overleftarrow{\hat{x}}_{t+1|T} - \overleftarrow{\hat{x}}_{t+1|t} \tag{13a}$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} - \mathcal{G}_t \cdot \Delta\mathbf{P}_{t+1} \cdot \mathcal{G}_t \tag{13b}$$

$$\Delta\mathbf{P}_{t+1} = \mathbf{P}_{t+1|\mathrm{T}} - \mathbf{P}_{t+1|t}, \mathcal{G}_t = \mathbf{P}_{t|t}\mathbf{P}^{-1}_{t+1|t} \tag{13c}$$

The Kalman gain $\mathcal{K}_t$ is calculated using the observed and estimated error covariance matrix at the current moment, while the RTS algorithm calculates the Kalman gain $\mathcal{G}_t$ using the error covariance matrix at the current moment t and the future moment $t+1$. The denoised EEG signal $\hat{\mathbf{x}}_{t|\mathrm{T}}$ is obtained by correcting the state predictions with $\mathcal{G}_t$.

4. **Experiments.**

4.1. **Experimental Setup.**

4.1.1. *Data Bases.* We used a well-structured and standardized EEG dataset, EEGdenoiseNet, in this project, which records 64 channels of motor imagery and left- and right-handed movement tasks at a sampling frequency of 256 HZ and contains 52 subjects with 4514 clean EEG segments. The dataset segmented the data into 2-second periods. In addition, experts performed a visual inspection to ensure the cleanliness and usability of the data [22]. To better accommodate the Gaussian noise assumption and to improve the denoising effect and the stability of the smoother, we normalized the input noisy EEG signal $\mathbf{y}_\mathrm{t} \in \mathbb{R}^\mathrm{n}$ as follows:

$$y = \frac{2\left(y_t - \min\left(y_t\right)\right)}{\max\left(y_t\right) - \min\left(y_t\right)} - 1 \tag{14}$$

4.1.2. *Evaluation Index.* **Mean Square Error:** is a measure of the magnitude of the error between the predicted and actual values, and is used to characterize how well the smoothed denoising results match the noise-free EEG signal:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mathbf{x}}_{t|t} - x_t\right)^2 \tag{15}$$

where n denotes the number of sample points in the recording time period, $\hat{\mathbf{x}}_{t|t}$ denotes the smoothed mean of the noise-added signal $\mathbf{y}_\mathrm{t} \in \mathbb{R}^\mathrm{n}$ at moment t after denoising by the RTS, x_t is the actual observed value (Ground Truth).

Further, we convert the mean square error into decibel units, which are used to characterize the denoising effect of our method:

$$MSE[db] = 10\log_{10}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mathbf{x}}_{t|t} - x_t\right)^2\right) \tag{16}$$

Time: In applications in the field of brain-computer interfaces, it is crucial to acquire and process EEG signals in a timely manner, as this will directly affect the user's interaction experience. Therefore, EEG signal denoising is processed in real-time or near real-time speed in order to detect any abnormal or unexpected situations in a timely manner.

4.2. **Statistical Analysis.** Our work involves utilizing the Kalman filter(KF) and its variants [23], the extended Kalman filter (EKF) [24], the Kalman smoothing algorithm (Rauch-Tung-Striebel (RTS) smoother) [25] and the wavelet transform-based nonlinear wavelet transform thresholding (NWT) [26], the Kalman wavelet filtering algorithm (KW) and the particle smoother particle filtering (PF) [27, 28] and other methods for comparison. We also report a deep learning CNN-based Kalman filtering (KF-CNN) [29] for reference.

We have compared the above methods and listed the comparison results in Table 1. First, from the perspective of MSE, the HRTS proposed in this paper performs the best, indicating that its denoising effect is better than other methods. This is because HRTS can make full use of the dynamic information of the signal and effectively denoise the signal by building a hierarchical SS model. However, we also notice that the wavelet transform-based denoising method has a faster processing speed in the time dimension. This is due to the fact that wavelet denoising methods can decompose the signal by fast wavelet transform or discrete wavelet transform with relatively low computational complexity [26, 30, 31, 32, 33, 34, 35, 36, 37, 38]. However, the effect of wavelet denoising is constrained by the adaptability of the selected wavelet basis and the accuracy of the threshold selection. To obtain the desired denoising effect, the wavelet basis function should match the frequency characteristics of the EEG signal. If a wavelet basis that does not match the signal characteristics is selected, it may filter out important information in the signal, or introduce unwanted artifacts that can interfere with the interpretation of the EEG signal and produce misleading results. At the same time, if the threshold function is set too strictly, important components of the EEG signal may be mistakenly deleted during the denoising process, leading to the loss of key information about the EEG activity and making subsequent analysis and interpretation difficult.

Although in the time dimension, our method is only slightly inferior to the NWT method. However, this choice is reasonable considering the denoising effect. In order to realize efficient brain-computer interface applications, we need high-quality EEG signals and also have strict requirements on the response speed of denoising. Therefore, it is crucial to find a method that can achieve an optimal balance between speed and denoising effect. The complexity and noise interference of EEG signals make sophisticated denoising important because it helps us to accurately extract useful physiological information from the signals. Reducing noise interference is decisive for recognizing and understanding specific patterns, events, or abnormalities in EEG signals. Therefore, when evaluating algorithms, we should not only aim for fast response but should emphasize maintaining the clarity and accuracy of the EEG signal.

From the data in Table 1, we find that the particle filtering denoising effect is not satisfactory. This is because the denoising effect of particle filtering is influenced by the number of particles. Particle filtering requires a set of particles to represent the probability distribution of the system state, and the choice of the number of particles affects the accuracy and speed of the filtering. In general, the higher the number of particles, the better the accuracy and denoising effect of particle filtering, but it also increases the computational effort and computational time. In order to obtain an accurate estimate, a sufficient number of particles is required to cover the entire state space. If the number of particles is smaller than the number of samples of the EEG signal, it may result in insufficient smoothness during denoising. Our HRTS method uses an additional smoothing step after the filter denoising to improve the smoothness of the signal. This is important for tasks such as signal classification, feature extraction, and pattern recognition in brain-computer interfaces. The application of particle filter in EEG signal denoising needs to be combined with specific application scenarios and signal characteristics to select the

TABLE 1. Empirical evaluation: MSE [dB], Time [sec]

Denoising Method	MSE	Time
KF	-21.67	61.89
EKF	-21.22	108.64
RTS	-19.38	60.19
*PF	-1.94	718.26
NWT	-17.38	**0.72**
KW	-17.73	81.64
KF-CNN	-20.73	724.47
Ours	**-22.15**	13.83

The highlighted numbers are closest to the value of ground truth (SOTA).
*The PF method was applied with a particle count of 500

appropriate number of particles. Meanwhile, using an adaptive algorithm to select the number of particles is also a feasible method [39].

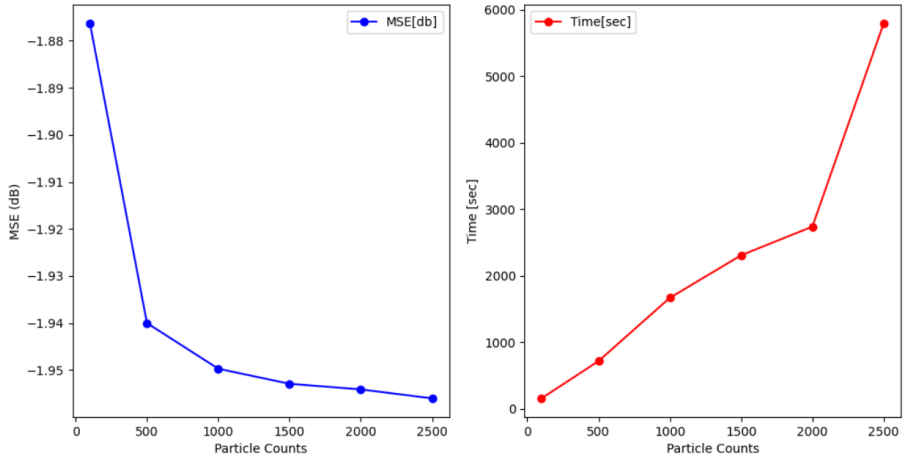
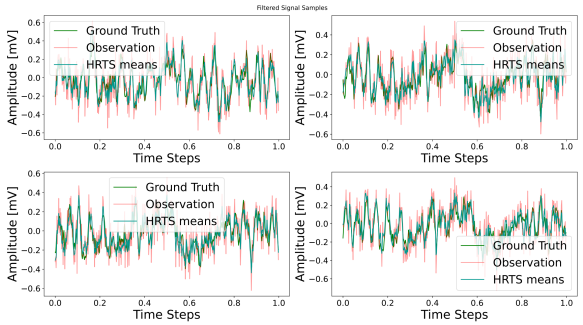


FIGURE 2. Relationship between MSE, Particle Count, and Time.

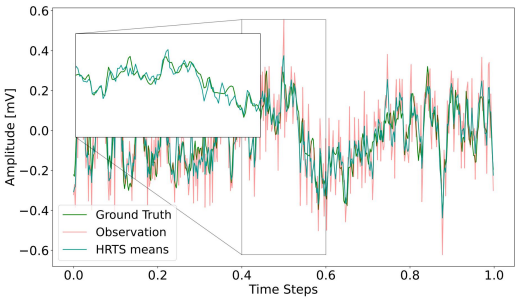
In order to better evaluate the effect of particle number on the denoising effect of EEG, we conducted an extended experiment of particle filtering. The experimental results are presented in Figure 2, which show that the denoising effect of EEG signals gradually increases with the increase of the number of particles and basically reaches saturation when the number of particles reaches 2000. Although the denoising effect is not much different from that when the number of particles is 1000, the computation time increases dramatically. This indicates that increasing the number of particles in a certain range can improve the accuracy and denoising effect of particle filtering. However, when the number of particles is too large, the computational complexity and computational memory will increase rapidly while the denoising effect is improved, and the computational efficiency will decrease significantly. In addition, particle filtering is a filtering algorithm based on the Monte Carlo method, and the repeatability of its results is poor and there is a certain uncertainty. In contrast, the hierarchical Kalman smoothing algorithm has better controllability and repeatability.

By analyzing Figure 3 (a) and (b), we can find a high correlation between the denoised signal and the original signal. The HRTS method we used can effectively reduce the interference of Gaussian noise in the EEG signal, while the quality of the denoised signal

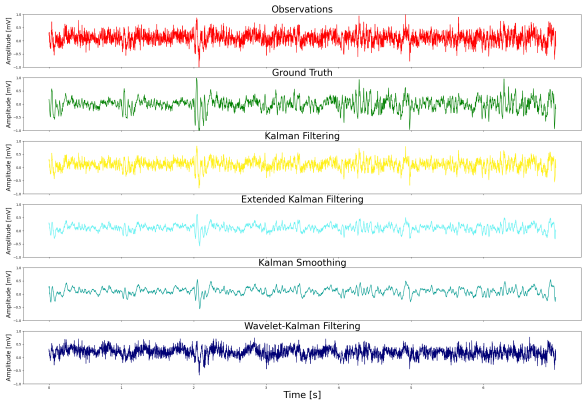
is close to that of the original signal. In particular, we can see from the illustration that the denoised signal almost overlaps with the ground truth, which indicates that the HRTS method successfully retains the original signal features and information without introducing significant distortion. Therefore, we can assume that the denoised signal is still representative of the original signal and can be used for relevant research and exploration.



(a) Superimposed View of Data.



(b) Superimposed View of Data.



(c) Separated View of Datas.

FIGURE 3. Denoised EEG signal.

Based on the comparison results in Figure 3 (c), we find that the comparison algorithms KF, EKF, KS, and KW used are able to remove Gaussian noise from the EEG signal to some extent. However, these algorithms, while removing the noise, also suppress the high-frequency part of the signal, leading to a reduction in the crest of the signal. This phenomenon may result in certain physiological features of the EEG signal not being effectively analyzed and identified, thus obscuring the information of the EEG physiological signal. However, considering the denoising effect and the degree of retention of signal features, the method in this paper has better performance in both the denoising effect and the degree of retention of signal features, which can better describe the physiological characteristics of EEG signals and provide a more reliable basis for further analysis and identification of EEG signals.

4.3. **Interpretability Analysis. Hierarchical SS Model:** Deep learning and Kalman filtering use different strategies for feature learning. Deep learning automatically extracts features mainly by analyzing a large amount of training data, while Kalman filtering is a state-space model-based filter that accurately estimates the state by modeling the dynamical processes of the system. Considering the strong spatio-temporal correlation of EEG signals, we use a hierarchical SS model for a more accurate description. The model

decomposes the EEG signal into multiple levels in the form of a sliding window, and each level includes multiple variables. By utilizing this a priori knowledge and performing state estimation based on the dynamics model of the system, the time-varying characteristics and dynamics of EEG signals can be captured more accurately, and more valuable features can be extracted to improve the denoising effect. Based on more accurate EEG signal analysis and modeling, our method can better extract meaningful feature information and thus has a broader application prospect in the field of EEG denoising.

EM Algorithm: The selection of the Kalman smoothing initial state estimate has a significant impact on the performance of the Kalman filter [40], and the selection of the initial process noise covariance matrix $\mathbf{P}_{t|t-1}$ and the observation noise covariance matrix $\mathbf{R}_t$ determines whether the iterations converge toward convergence. If the selection is improper, it may lead to poor denoising performance of the smoother, slow convergence, poor filter stability, and large estimation errors. To address this problem, this paper adopts the method of EM algorithm to optimize the selection of initial state estimates in Kalman smoothing to improve the denoising performance and generalization ability of the smoother. Specifically, suitable initial values are calculated by iterating over the prior data set, and then these initial values are applied to the subsequent Kalman smoothing process. The experimental results show that the method proposed in this paper can effectively improve the denoising performance and generalization ability of the smoother.

In the field of feature learning, deep learning models have demonstrated good performance in many areas. Typically, these models use optimization algorithms such as gradient descent to learn model parameters, but this requires a large amount of data and computing resources. Additionally, the performance and effectiveness of deep learning models often need to be adjusted and optimized according to specific application scenarios, resulting in poor generalization ability. When dealing with denoising EEG, the accuracy and robustness of deep learning models still face challenges. In contrast, our approach uses the EM algorithm to iteratively learn signal parameters in an unsupervised manner, with strong generalization ability and denoising speed, while also ensuring denoising effectiveness without the need for additional adjustments [41].

5. Conclusions.

In this paper, we use an HRTS method based on a hierarchical state-space model, which is able to extract structural prior information from EEG signals and apply it to remove Gaussian white noise removal in EEG. By integrating the structural prior information with the state space model, we can effectively filter EEG signals, thereby enhancing the quality of the EEG data. When compared with conventional filtering algorithms and deep learning denoising methods, our benchmark tests reveal that the HRTS method can reduce time costs while maintaining effective denoising. The focus of this research is Gaussian noise in EEG signals. In future studies, we plan to extend this method to process EEG data with artifacts from electrooculography (EOG) and electromyography (EMG) and use the denoised EEG signals in brain-computer interface applications. Furthermore, we intend to explore more optimized computational methods to further decrease the time consumption of our method.

Funding Statement. This work is supported by The first batch of industry-university cooperation collaborative education project funded by the Ministry of Education of the People's Republic of China, 2021, Project No.:202101071001; Supported by Minjiang University school-level scientific research project funding, Projects No.:MYK17021, MYK18033, MYK21011; Supported by Minjiang University Introduced Talents Scientific Research Start-up Fund, Projects No.:MJY21030; Supported by Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Project No.:21DMAKL01;

Supported by Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Project No.:23DMAKL02; Supported by The Ministry of Education's first batch of industry-university cooperation collaborative education projects in 2024 (project number: 231104497282106); Supported by Fujian Province Software Industry Technology Innovation Key Research and Industrialization Project (Second Batch) AI Intelligent Modeling Platform Based on Deep Learning Technology Project.

REFERENCES

[1] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: control signals review," *Neurocomputing*, vol. 223, pp. 26–44, 2017.

[2] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.

[3] T.-Y. Wu, Q. Meng, L. Yang, S. Kumari, and M. Pirouz, "Amassing the security: An enhanced authentication and key agreement protocol for remote surgery in healthcare environment." *CMES-Computer Modeling in Engineering & Sciences*, vol. 134, no. 1, pp. 317–341, 2023.

[4] T.-Y. Wu, L. Wang, and C.-M. Chen, "Enhancing the security: A lightweight authentication and key agreement protocol for smart medical services in the ioht," *Mathematics*, vol. 11, no. 17, p. 3701, 2023.

[5] M. T. Akhtar, W. Mitsuhashi, and C. J. James, "Employing spatially constrained ica and wavelet denoising, for automatic removal of artifacts from multichannel eeg data," *Signal Processing*, vol. 92, no. 2, pp. 401–416, 2012.

[6] N. Mashhadi, A. Z. Khuzani, M. Heidari, and D. Khaledyan, "Deep learning denoising for eog artifacts removal from eeg signals," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2020, pp. 1–6.

[7] M. T. Akhtar, W. Mitsuhashi, and C. J. James, "Employing spatially constrained ica and wavelet denoising, for automatic removal of artifacts from multichannel eeg data," *Signal Processing*, vol. 92, no. 2, pp. 401–416, 2012.

[8] T. Locher, G. Revach, N. Shlezinger, R. J. van Sloun, and R. Vullings, "Hierarchical filtering with online learned priors for ecg denoising," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[9] T. Zikov, S. Bibian, G. A. Dumont, M. Huzmezan, and C. R. Ries, "A wavelet based de-noising technique for ocular artifact correction of the electroencephalogram," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*. IEEE, 2002, pp. 98–105.

[10] B. W. McMenamin, A. J. Shackman, J. S. Maxwell, D. R. Bachhuber, A. M. Koppenhaver, L. L. Greischar, and R. J. Davidson, "Validation of ica-based myogenic artifact correction for scalp and source-localized eeg," *Neuroimage*, vol. 49, no. 3, pp. 2416–2432, 2010.

[11] W. Sun, Y. Su, X. Wu, and X. Wu, "A novel end-to-end 1d-rescnn model to remove artifact from eeg signals," *Neurocomputing*, vol. 404, pp. 108–121, 2020.

[12] S. Zhang, Z. Zhang, Z. Chen, S. Lin, and Z. Xie, "A novel method of mental fatigue detection based on cnn and lstm," *International Journal of Computational Science and Engineering*, vol. 24, no. 3, pp. 290–300, 2021.

[13] K. K. Tseng, C. Wang, T. Xiao, C. M. Chen, M. M. Hassan, and V. H. C. D. Albuquerque, "Sliding large kernel of deep learning algorithm for mobile electrocardiogram diagnosis," *Computers & Electrical Engineering*, vol. 96, pp. 107 521–, 2021.

[14] E. K. Wang, C. M. Chen, M. M. Hassan, and A. Almogren, "A deep learning based medical image segmentation technique in internet-of-medical-things domain," *Future Generation Computer Systems*, vol. 108, pp. 135–144, 2020.

[15] K.-K. Tseng, R. Zhang, C.-M. Chen, and M. M. Hassan, "Dnetunet: a semi-supervised cnn of medical image segmentation for super-computing ai service," *The Journal of Supercomputing*, vol. 77, pp. 3594–3615, 2021.

[16] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2001.

[17] R. P. Brenner and N. Schaul, "Periodic eeg patterns: classification, clinical correlation, and pathophysiology," *Journal of Clinical Neurophysiology*, vol. 7, no. 2, pp. 249–268, 1990.

[18] G. B. Giannakis, V. Madisetti, and D. Williams, "Cyclostationary signal analysis," *Digital Signal Processing Handbook*, vol. 31, pp. 1–17, 1998.
[19] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," 1996.
[20] S. J. Orfanidis, "Applied optimum signal processing," *Author, Piscataway, NJ*, 2018.
[21] J. Dauwels, A. Eckford, S. Korl, and H.-A. Loeliger, "Expectation maximization as message passing-part i: Principles and gaussian messages," *arXiv preprint arXiv:0910.2832*, 2009.
[22] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, and Q. Liu, "Eegdenoisenet: A benchmark dataset for deep learning solutions of eeg denoising," *Journal of Neural Engineering*, vol. 18, p. 056057, 2021.
[23] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.
[24] G.-I. Jee, H. Kim, Y. Lee, and C. Park, "A gps c/a code tracking loop based on extended kalman filter with multipath mitigation," in *Proceedings of the 15th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 2002)*, 2002, pp. 446–451.
[25] S. Särkkä, "Unscented rauch–tung–striebel smoother," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 845–849, 2008.
[26] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: asymptopia?" *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 2, pp. 301–337, 1995.
[27] P. Del Moral, "Nonlinear filtering: Interacting particle resolution," *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, vol. 325, no. 6, pp. 653–658, 1997.
[28] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.
[29] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.
[30] A. Cohen, I. Daubechies, and P. Vial, "Wavelets on the interval and fast wavelet transforms," *Applied and Computational Harmonic Analysis*, 1993.
[31] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms i," *Communications on Pure and Applied Mathematics*, vol. 44, no. 2, pp. 141–183, 1991.
[32] N. Saravanan and K. Ramachandran, "Incipient gear box fault diagnosis using discrete wavelet transform (dwt) for feature extraction and classification using artificial neural network (ann)," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4168–4181, 2010.
[33] T. Fu, R. Deng, Y. Gao, and F. Zhang, "Representing scenes as compositional generative neural feature fields based on giraffe for 3d reconstruction of classroom scenes," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceeding of the 18th IIH-MSP 2022 Kitakyushu, Japan, Volume 2*. Springer, 2023, pp. 227–237.
[34] T. Fu, J. Shi, and H. Ke, "Research on fast encryption of electronic health record data based on privacy protection," in *IoT and Big Data Technologies for Health Care: Third EAI International Conference, IoTCare 2022, Virtual Event, December 12-13, 2022, Proceedings*. Springer, 2023, pp. 255–270.
[35] H. Ke, J. Shi, and T. Fu, "Research on secure storage of healthcare data in the environment of internet of things," in *IoT and Big Data Technologies for Health Care: Third EAI International Conference, IoTCare 2022, Virtual Event, December 12-13, 2022, Proceedings*. Springer, 2023, pp. 271–288.
[36] H. Xiong, C. Jin, M. Alazab, K.-H. Yeh, H. Wang, T. R. Gadekallu, W. Wang, and C. Su, "On the design of blockchain-based ecdsa with fault-tolerant batch verification protocol for blockchain-enabled iomt," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 1977–1986, 2021.
[37] H. Xiong, M. Yang, T. Yao, J. Chen, and S. Kumari, "Efficient unbounded fully attribute hiding inner product encryption in cloud-aided wbans," *IEEE Systems Journal*, vol. 16, no. 4, pp. 5424–5432, 2021.
[38] H. Xiong, Y. Hou, X. Huang, Y. Zhao, and C.-M. Chen, "Heterogeneous signcryption scheme from ibc to pki with equality test for wbans," *IEEE Systems Journal*, vol. 16, no. 2, pp. 2391–2400, 2021.
[39] O. Cappé and E. Moulines, "On-line expectation–maximization algorithm for latent data models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 3, pp. 593–613, 2009.
[40] I. Rusnak, "Optimal state estimation of nonlinear dynamic systems," in *Nonlinear Systems-Modeling, Estimation, and Stability*. IntechOpen, 2018.
[41] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, p. 051001, 2019.