DoA assessment based on EEG DFA and Entropy features

Xing Chen[1[0000-0001-7627-9632]], Bo Song[1[0000-0002-3450-7489]], Peng Wen[1[0000-0003-0939-9145]] and Yan Li[2[0000-0002-4694-4926]]

[1] School of Engineering, University of Southern Queensland, QLD 4350, Australia
u1112086@umail.usq.edu.au
[2] School of Mathematics, Physics and Computing, University of Southern Queensland, QLD 4350, Australia

Abstract. In this study, a novel method is proposed to combine modified detrended fluctuation analysis (DFA) and entropy to extract features of electroencephalogram (EEG), which are then processed using a random forest algorithm to generate a new DoA index. The bispectral index (BIS) was used as the reference standard. The proposed DoA index achieved Pearson and Spearman correlation coefficients of 0.97 ($p<0.01$) and 0.95 ($p<0.01$) with the BIS index, respectively. Additionally, the mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) were 20.45, 4.52, and 2.85, respectively. These results indicate that the proposed DoA index is more accurate in patients' consciousness level assessment.

Keywords: Anesthesia, Depth of Anesthesia, Electroencephalogram.

1 Introduction

General anesthesia is often given to surgical patients to induce a state of loss of consciousness, disappearance of external stimuli, decrease of motor function and memory forgetfulness by anesthetics inhibiting the central nervous system. The purpose is to help intraoperative patients tolerate medical procedures and ensure comfort and safety. However, it is significant to maintain appropriate anesthesia depth as both overdose and underdose can pose significant risks to patients. Overdose would prolong postoperative recovery, cause serious brain injury and even increase the postoperative mortality of patients [1], [2]. While underdose leads to intraoperative awareness, which can induce severe physical and psychological trauma [3], and symptoms of posttraumatic stress disorder (PTSD) in patients [4].

In clinical operations, anesthetists often judge the Depth of Anesthesia (DoA) through the vital signs and physiological parameters of patients, such as blood pressure, heart rate, sweating, and tears applied in the PRST score [5]. However, the qualitative judgment of DoA based on descriptive behavioral assessments of patients' response to the irritant reaction would inevitably lead to the assessment error of anesthesia status. And the inconsistent conclusions on DoA may be drawn from different anesthetists.

Since the above method cannot be recorded in a real-time manner, several methods have been proposed, such as heart rate [6], evoked potential [7], ECG [8], and other biomedical signals [9], and some of them have already been applied clinically, especially electroencephalogram (EEG). As the representation of brain activity, EEG can directly track the variation of the central nervous system induced by the anesthetics [10]. Take a widely applied EEG-based anesthesia monitor, BIS monitor, as an example, the EEG was recorded from the 2 or 4 sensors placed on the forehead, and then processed with the proposed especially methods to present a 0-100 dimensionless number to quantify the depth of anesthesia ranging from very deep anesthesia to awake state. However, the BIS monitor suffers from the variability of EEG patterns during different stages and anesthetics applied [11], which disrupt the EEG representation.

Therefore, many methods have been proposed in the past decade for the exploration like entropy, fractal analysis and deep learning methods. Shalbaf et al. implemented sample entropy (SampEn) and permutation entropy (PermEn) with Aritificial Neural Network (ANN) model to quantify the DoA over the 2 data groups of sevoflurane and propofol. The multivariate empirical mode decomposition with multi-scale PermEn was applied on DoA monitoring by [13]. Liang et al. compared various entropy methods in classifying 3 consciousness states [14]. Jospin et al. and Gifani et al. both applied Detrended Fluctuation Analysis (DFA) and focused on finding the optimal scale range to better represent the relationship between the EEG sequence and anesthesia state [15], [16]. Nguyen-Ky et al. modified DFA and built a model thus improving the discrimination between the awake state and light anesthesia state [17]. Liu et al. applied Short-time Fourier Transform (STFT) and transformed the time series into spectrum, thus classifying the awareness state via Convolutional Neural Network (CNN) [18]. Asfar et al. further composited Long-short term memory (LSTM) and attention layers on discriminating awareness levels.

Despite the fact that several investigations have been made, the results were not as satisfactory as invented. Most efforts have been made to search the optimal scaling range for the sequences under anesthesia states individually or overall. The self-similarity variations calculated from the linear trend reduction under different awareness levels cannot present linear persistently. Furthermore, repeated calculations on increasing scales suffer from computation efficiency, which prevents DFA application in real-time monitoring.

In this study, we proposed a novel method that integrated DFA and entropy-based feature extraction with an optimized random forest model to predict the depth of anesthesia (DoA) as shown in Fig. 1. The dataset utilized in this study comprises EEG recordings from 73 patients, which were split into training, validation, and test sets in a 60%, 20%, and 20% ratio, respectively. Specifically, EEG recordings from 45, 14, and 14 patients were used for feature selection, model evaluation, and performance assessment. To ensure reproducibility, a random number generator was set to the default seed and algorithm.

For each EEG recording, a sliding window technique was applied. While longer windows enhance effectiveness, they reduce computational efficiency. Given the need for real-time application, we set the window length to 10 seconds with a 1-second step

size, ensuring that the DoA index updates every second. The BIS index, corresponding to the final second of each window, serves as the reference for the DoA prediction.

During preprocessing, a zero-phase bandpass finite impulse response (FIR) filter was applied to the EEG signals to ensure phase consistency and minimize distortion. The cutoff frequencies of 0.5 Hz and 47 Hz were chosen based on literature recommendations to capture relevant brain activity efficiently.

For feature extraction, DFA was employed to measure self-affinity. The EEG time series was repeatedly partitioned into sub-sequences, and the fluctuation was computed as the mean square difference from the linear fitting of each sub-sequence. This fluctuation was then analyzed using linear regression on a log-log plot of fluctuations versus scales. However, this method can be sensitive to outliers introduced by temporal distortions in the EEG, potentially compromising the accuracy of DoA assessments. To mitigate this, we explored using the quartile-based threshold to reject these outliers. Furthermore, the necessity of the flipping procedure as suggested by previous studies, where the time series is reversed and processed to capture additional information, is evaluated and compared with standard methods.

In addition to DFA, three other features—sample entropy (SampEn), permutation entropy (PermEn, and fuzzy entropy (FuzzEn)—were extracted from each EEG epoch, with parameter settings based on [12], [14]. The features were normalized by setting the mean to 0 and the standard deviation to 1 to improve convergence during training. The random forest model was optimized by adjusting hyperparameters such as the applied method, number of learning cycles, learning rate, minimum leaf size and maximum number of splits.

Model performance was evaluated using the mean square error (MSE) between the predicted DoA index and the target BIS index on the validation set. The model with the lowest MSE was selected for further investigation. In the test group, the proposed DoA index was compared with other methods using metrics such as Spearman correlation, MSE, root mean square error (RMSE), and mean absolute error (MAE). Spearman correlation is preferred over Pearson correlation due to its ability to capture nonlinear relationships without assumptions about data distribution. MAE represents the absolute prediction error, while MSE and RMSE magnify the influence of larger errors, making them more sensitive to outliers. Unlike MSE, which squares the error, RMSE and MAE preserve the original scale, making them more intuitive. While RMSE tends to be larger than MAE due to the squared errors, this difference highlights discrepancies in error distribution. Finally, we introduce the R^2 score, which normalizes the error and provides a consistent evaluation standard across different methods.

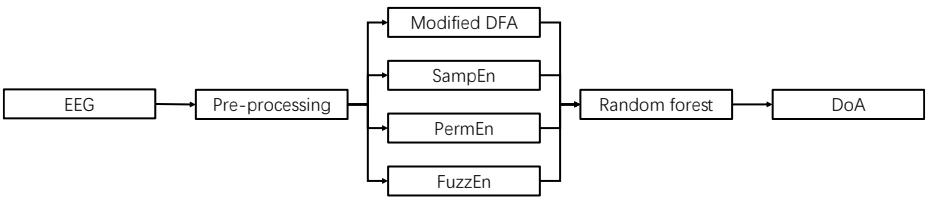

Fig. 1. Flowchart

2 Method

2.1 Data Collection

We collected EEG signals from 73 adult patients undergoing general anesthesia using the ASPECT A-1050 monitor (Aspect Medical Systems, Natick, MA, USA). The electrodes were placed on the forehead by an experienced anesthesiologist, ensuring proper skin cleaning and impedance levels. The EEG data was recorded at 128 Hz, and the BIS index was recorded simultaneously. All procedures were approved by the USQ ethics committee, and written informed consent was obtained from each patient.

The dataset was randomly divided into training, validation, and test groups, comprising 70%, 15%, and 15% of the data, respectively, as shown in Table 1. For convenience, the level of consciousness was categorized into five states—A, B, C, D, and E—based on BIS ranges of 0-20, 21-40, 41-60, 61-80, and 81-100, respectively.

Table 1. Data separation

Group	Patients	State Percentage		
		A	533	0.19%
		B	41615	14.51%
Train	45	C	159315	55.55%
		D	68315	23.82%
		E	17039	5.94%
		A	1	0.00%
		B	16521	15.56%
Validate	14	C	60462	56.96%
		D	15980	15.06%
		E	13178	12.42%
		A	15	0.02%
		B	9456	10.65%
Test	14	C	49898	56.19%
		D	23609	26.58%
		E	5831	6.57%

2.2 Preprocessing

Raw EEG signals are often contaminated by various types of noise, such as ocular artifacts, muscle artifacts, cardiac artifacts, and extrinsic artifacts [20]. These unavoidable components can provide inaccurate information and hinder subsequent analysis [21]. To build a reliable DoA index, most studies preprocess EEG data, commonly applying a bandpass filter. We filtered the raw EEG segments using a bandpass filter with a frequency range of 0.5 to 47 Hz and a minimum order, as suggested by [22].

Additionally, we applied a zero-phase filter to eliminate phase distortion. Since the EEG was processed in real-time, with data being added and the BIS index updated every second, we adopted a sliding window method with a 10-second EEG segment at a 1-second step size as suggested by [19]. The sliding window procedure on different EEG epochs and the relevant BIS index are shown in Fig. 2.

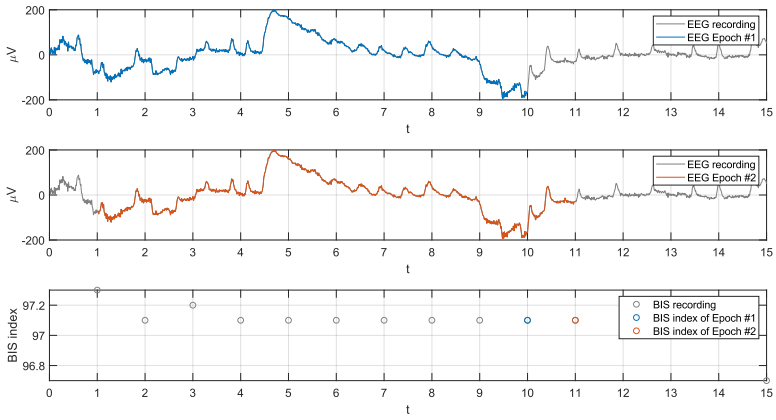

Fig. 2. The subplot of the last second EEG epoch (top, blue line), the current second EEG epoch (middle, red line), and the relevant BIS index (bottom) for the last second (blue circle) and the current second (red circle).

2.3 Detrended Fluctuation Analysis Algorithm

DFA was initially introduced by [23], and applied to measure long-range memory of the DNA sequence. The procedure is explained as follows:

For a given time-series $x(t)$ with a finite length N, DFA starts with centering and cumulating the original sequence to obtain the integrated sequence $Y(i)$,

$$Y(i) = \sum_{t=1}^{i}[x(t) - \langle x \rangle], i = 1, \dots, N \tag{1}$$

where $\langle x \rangle$ represents the mean of the sequence.

For a given box size s, $Y(i)$ is partitioned into m non-overlapping intervals of length s to build final $Y_v(i)$, $v = 1, \dots, m$. If N is indivisible by s, to avoid omitting samples at the end, the integrated sequence $Y(i)$ is flipped, partitioned similarly, and combined with $Y(i)$ to obtain the partitioned integrated sequence $Y_v(i)$, $v = 1, \dots, 2m$. Here we only considered the local trend $P_v(i)$ as linear in each interval, which is commonly used in EEG processing.

The mean square deviation between $P_v(i)$ and $Y_v(i)$ for each interval is calculated as:

$$F_s^2(v) = \langle \big(Y_v(i) - P_v(i)\big)^2 \rangle = \frac{1}{s}\sum_{i=1}^{s}(Y_v(i) - P_v(i))^2, v = 1, \dots, 2m. \tag{2}$$

The average over all segments is taken, and the square root is used to obtain the fluctuation $F(s)$ at a specific s,

$$F(s) = \left\{\frac{1}{2m}\sum_{v=1}^{2m}[F_s^2(v)]\right\}^{1/2} \tag{3}$$

The standard DFA focuses on how the fluctuation function $F(s)$ depends on s by repeating steps 2 to 4 for several time scales.

$$F(s) \sim s^{\alpha} \tag{4}$$

The self-affinity α was then calculated at the log-log scale of the fluctuation function. The whole procedure can be expressed in Fig. 3. According to the suggested scale range from 3 to 100 [15], [16], we searched the optimal scale range to extract the self-affinity. The self-affinity was calculated from the starting scale of 3, and the ending scale ranges from 4 to 100. The Pearson and Spearman correlation between the BIS index and the self-affinity extracted from the different ranges is plotted in Fig. 4. The maximum negative correlation can be found in the ending scale of 9 and 11 from the Pearson and Spearman correlation respectively. We selected the latter one since the results present a higher non-linear correlation.

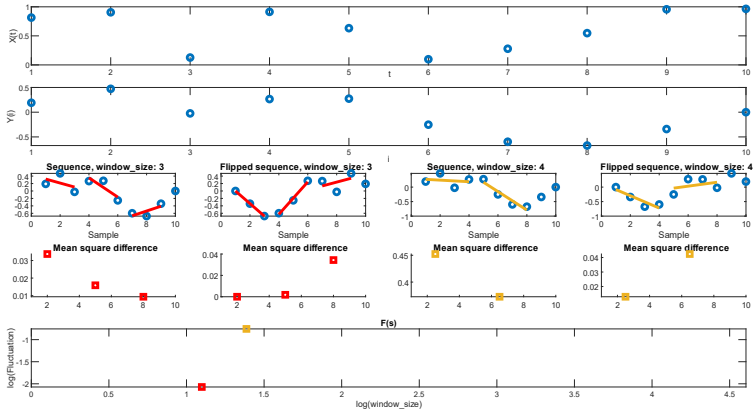

Fig. 3. Standard DFA procedure on a simulated time series, from the top to the end: the original time series; the integrated sequence; partition at different window size s; the mean square differences at each segment; and the loglog-plot on $F(s)$ versus s.

To avoid omitting samples at the end, the sequence was flipped and the procedure was repeated [17], [24]. However, this doubled sequence might not provide more information but lower the calculation efficiency. We investigated the correlation between the BIS index and the self-affinity from DFA both with and without the flipping step. As shown in Table 2, the Pearson and Spearman correlation of DFA on the doubled sequence within the flipping procedure is decreased from -0.2042 and -0.3822 to -0.2062 and -0.3838 compared with the single sequence without the procedure. Hence, we removed the procedure for the sake of efficiency and effectiveness.

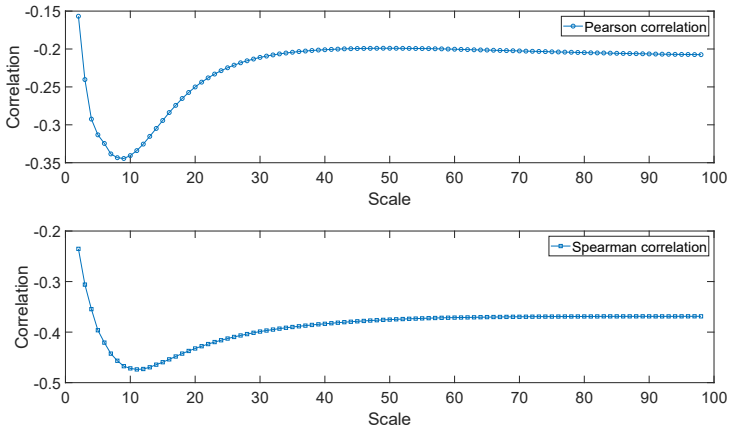

Fig. 4. Scatter plot on Pearson and Spearman correlation between the BIS index and the self-affinity from the different scales.

Table 2. Pearson and Spearman correlation between the feature and the BIS index.

Flipping procedure	Pearson correlation	Spearman correlation
DFA-Double	-0.2042 (p<0.01)	-0.3822 (p<0.01)
DFA-Single	-0.2062 (p<0.01)	-0.3838 (p<0.01)

Since the fluctuation $F(s)$ is calculated from the mean square error between the original sequence and local trends at each segment, the extracted feature self-affinity is easily influenced by the outliers of $F_s^2(v)$. Artifacts such as baseline drift will introduce abrupt EEG changes and cannot be easily removed by the bandpass filter. Therefore, we inspected the effect of the elimination of the outliers according to the threshold based on mean, median and quartile. As shown in Table 3, all the extracted self-affinity presents a negative relationship with the BIS index, which satisfies the assumption and matches the results proposed by [15], [16].

Table 3. Threshold fluctuation correlation with BIS index.

Outlier rejection	Pearson correlation	Spearman correlation
DFA	-0.3483 (p<0.01)	-0.4655 (p<0.01)
Median-based DFA	-0.4702 (p<0.01)	-0.5533 (p<0.01)
Mean-based DFA	-0.3792 (p<0.01)	-0.4887 (p<0.01)
Quartile-based DFA	-0.4951 (p<0.01)	-0.5576 (p<0.01)

2.4 Sample Entropy Algorithm

SampEn is a measure of regularity or predictability in time-series data and has been commonly used in the analysis of physiological signals, such as EEG and ECG [6], [22], to assess the complexity of these signals. It quantifies the likelihood that similar

patterns of data points (within a specified tolerance) will remain similar when compared over subsequent time points. For a time-series $x(t)$ with a finite length N, the procedure of SampEn can be explained as follows:

For a given embedding dimension m and tolerance r, the m-dimension template vector $Y_m(i)$ is defined as:

$$X_m(i) = \{X(i), X(i+1), \dots, X(i+m-1)\}, 1 \le i \le N-m+1 \quad (5)$$

The Chebyshev distance between two vectors $X_m(i)$ and $X_m(j)$ is calculated as:

$$Dist[X_m(i), X_m(j)] = max(|X_m(i+k)| - |X_m(j+k)|), 1 \le k \le m-1, i \ne j \quad (6)$$

Let $B_i(j)$ be the number of the distances between $X_m(i)$ and $X_m(j)$ that are less than a threshold computed from the tolerance r ($r * std$, the standard deviation of $x(t)$).

$$B_m^r(i) = \frac{1}{N-m+1} \sum_{j=1, j\ne i}^{N-m} B_i(j) \quad (7)$$

B_m^r stands for the whole vector satisfying the condition and is defined as:

$$B_m^r = \frac{1}{N-m} \sum_{i=1}^{N-m} B_m^r(i) \quad (8)$$

For $m+1$, we obtain B_{m+1}^r by repeating step 1 to 4 and the SampEn is then calculated as:

$$SampEn = -\ln\left(\frac{B_{m+1}^r}{B_m^r}\right) \quad (9)$$

Here we take m and r as 2 and 0.2 respectively as suggested by [12].

2.5 Permutation Entropy Algorithm

PermEn assesses the complexity of a time series by analyzing the ordinal patterns of its data points and is less sensitive to noise and data outliers compared to other entropy measures [25]. It involves assigning a symbolic representation to each data point based on its relative order within a sliding window. For a time-series $x(t)$ with a finite length N, the procedure of PermEn can be explained as follows:

For a given embedding dimension m and time delay τ, the vector $X_m^\tau(i)$ is constructed as:

$$X_m^\tau(i) = \{x(i), x(i+\tau), \dots, x(i+(m-1)\tau)\}, 1 \le i \le N-(m+1)\tau \quad (10)$$

The possible ordinal pattern is defined as $m!$ in number. Each vector $X_m^\tau(i)$ can be represented by one of them.

The probability P_j of each pattern occurring is defined as P_j, where $j = 1, \dots, m!$.

The normalized PermEn is calculated as:

$$PermEn = \frac{\sum_{j=1}^{m!} P_j \ln P_j}{\ln m!} \quad (11)$$

Here we take m and τ as 6 and 1 respectively as suggested by [12].

2.6 Fuzzy Entropy Algorithm

FuzzEn is a variation of sample entropy that incorporates fuzziness or uncertainty into the similarity calculation [26]. This method is particularly useful in situations where data uncertainty or imprecision is present, such as in medical diagnosis or financial forecasting. FuzzEn accounts for the degree of similarity between data points rather than treating them as strictly similar or dissimilar. Cao & Lin et al. combined FuzzEn with Empirical Mode Decomposition for measuring the complexity of EEG signals with eyes open and closed.

For a given embedding dimension m and time delay τ, the vector $X_m(i)$ is constructed as:

$$X_m(i) = \{X(i), X(i+1), \dots, X(i+m-1)\}, 1 \le i \le N-m+1 \tag{12}$$

For a given vector $X_m^\tau(i)$, determine the similarity degree $D_{ij}^{m,\tau}$ using the Chebyshev distance $Dist_{ij}^{m,\tau}$ between $X_m^\tau(i)$ and $X_m^\tau(j)$, and the fuzzy membership function fm:

$$D_{ij}^{m,\tau} = fm\left(Dist_{ij}^{m,\tau}\right) = \exp\left(-\frac{(Dist_{ij}^{m,\tau})^n}{\tau}\right) \tag{13}$$

Construct the function φ^m as:

$$\varphi^m = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{ij}^{m,\tau}\right) \tag{14}$$

Step 4: For $m+1$, repeat step 1 to 3 to obtain φ^{m+1}

$$FuzzEn = -\ln\frac{\varphi^{m+1}}{\varphi^m} \tag{15}$$

Here we take m and r as 2 and 0.1 respectively as suggested by [27].

2.7 Fitting Model

Random forest proposed by [28], has been applied and successfully involved in various practical EEG-based problems [29], including artifact detection and removal [30], EEG spike detection [31], motor imagery classification [32], seizure detection [33] and so on. Liu et al. applied it in DoA monitoring and outperformed than the regression model Support Vector Machines (SVM) and ANN [34].

Compared with decision tree, random forest introduces randomness by searching for the best feature among a random set of features rather than the given set. This randomness enhances the stability and accuracy of predictions. For regression tasks, the criteria are set by mean square error, and iterations continue until satisfying the condition. Here, the hyperparameters including methods, number of leaning cycles, learning rate, minimum leaf size and maximum number of splits, are varied in the range: “Bootstrap

aggregation" or "Least-squares boosting (LSBoost)", $[10-500]$, $[1^{-3}, 1^{-2}, \ldots, 1]$, $[1-50]$ and $[1-99]$.

In this study, we compare the performance of random forest with other models commonly used in DoA prediction, including linear models, SVM, ANN and decision trees. The performance is evaluated based on correlation coefficient, MSE, RMSE and MAE.

3 Result

3.1 Model Selection

As shown in Table 4, the random forest model achieves the highest Pearson correlation 0.6430 (p-value<0.01). It also attains a low MSE of 176.7423, RMSE of 13.2944, and MAE of 8.3865. These results demonstrate the random forest model's superior performance in capturing the underlying trend and providing accurate predictions for DoA.

Table 4. Performance evaluation on the regression models.

Model	Pearson correlation	MSE	RMSE	MAE
Linear	0.6274 (p<0.01)	188.2222	13.7194	10.6388
SVM	0.6274 (p<0.01)	173.6409	13.1773	9.9380
NN	0.6274 (p<0.01)	188.2222	13.7194	10.6388
Decision tree	0.4271 (p<0.01)	260.1368	16.1288	11.9830
Random forest	0.6430 (p<0.01)	176.7423	13.2944	10.0644

3.2 Feature Combination DFA

To verify the effectiveness of the proposed modified DFA and feature combination, we measured the performance of the DoA index from a random forest model using one to four feature inputs. The quartile-based self-affinity feature alone does not present the highest correlation with the BIS index compared with the entropy methods, its combination with entropy-based features significantly lowers the MSE, RMSE, and Mean Absolute Error (MAE). Specifically, these metrics decreased from 133.0488, 11.5347, and 8.0047 to 72.7818, 8.5312, and 5.7753, respectively. These results demonstrate the superior performance of the proposed feature combination in awareness identification.

Table 5. Performance evaluation on the regression models.

Model	MSE	RMSE	MAE
Modified DFA	176.7423	13.2944	10.0644
SampEn	262.7302	16.2090	11.1085
PermEn	185.4590	13.6183	10.4100
FuzzEn	287.3668	16.9519	12.4655

Alpha_SampEn	177.5950	13.3265	9.1611
Alpha_PermEn	166.0570	12.8863	9.7160
Alpha_FuzzEn	168.6683	12.9872	9.5670
SampEn_PermEn	179.0892	13.3824	9.3173
PermEn_FuzzEn	193.7137	13.9181	10.2417
SampEn_FuzzEn	229.5838	15.1520	10.3540
Alpha_SampEn_PermEn	159.9403	12.6468	8.7352
Alpha_SampEn_FuzzEn	135.5484	11.6425	8.0325
Alpha_PermEn_FuzzEn	323.7834	17.9940	13.3155
SampEn_PermEn_FuzzEn	133.0488	11.5347	8.0047
Alpha_SampEn_PermEn_FuzzEn	122.3104	11.0594	7.5957

3.3 Model Optimization

The optimization process for the random forest model used in real-time DoA monitoring involved searching over various parameters, including methods, the number of learning cycles, learning rate, and minimum leaf size, to minimize the estimated cross-validation loss (MSE). The final results selected the “Bag” aggregation method and 496 trained trees.

The comparison of real-time DoA monitoring between the unoptimized and optimized random forest models is shown in Fig. 5. The distortions are mitigated by the model optimization and the R-square is increased from -0.2063 to 0.8778, respectively. A negative R-square indicates that the unoptimized model fits the data worse than simply returning the mean of the DoA index. The significant increase in the R-square value after optimization demonstrates a substantial reduction in estimation error.

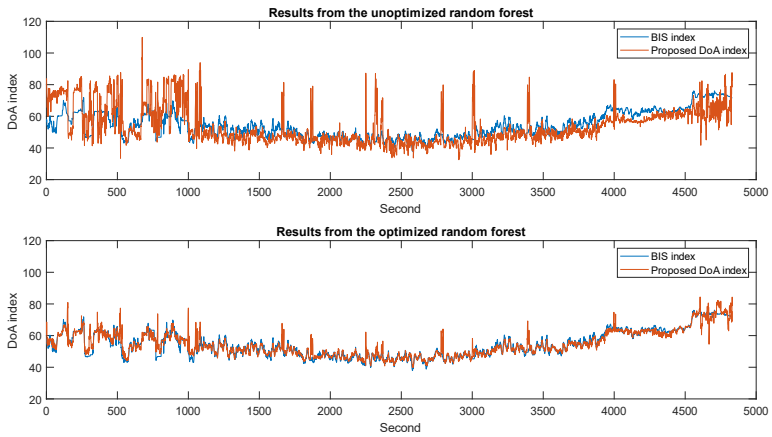

Fig. 5. random forest model output comparison.

3.4 DoA

Since Jospin et al. and Gifani et al. focus solely on feature extraction without a fitting model, we predicted the DoA using the same random forest model as the proposed method. Although they both investigated optimal DFA parameter settings, Figure 8 shows no significant difference in DoA estimation between these methods and the proposed model. Both methods also suffer from under-fitting. For the entropy-based DoA monitoring, Shalbaf et al. can represent peri-operative trends but it is limited in identifying awake and deep anesthesia states. The LSTM model focuses only on the lower DoA range and cannot accurately represent real awareness variations. While the method proposed by (Dutt and Saadeh 2023) performs well, it occasionally produces abnormal indices larger than 100.

Additionally, the scatter plot shown in Fig. 7 (A) displays the relationship and R-square as 0.8778 ($p<0.01$). Fig. 7 (B) presents the results of Bland-Altman analysis within bias as -0.0018 and limit agreement as 8.8627 and -8.8663.

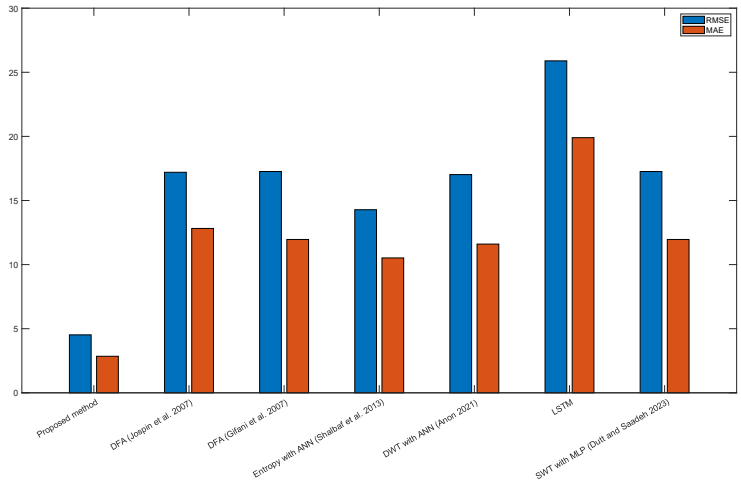

Fig. 6. DoA performance comparison on RMSE and MAE.

4 Discussion

DoA estimation is one of the most important indices in intra-operative monitoring and post-operative recovery. While previous studies have focused on EEG-based spectrum features, they often overlook temporal features. We believe that appropriate modified time-domain analysis within an ensemble learning method can perform better than commonly used spectrum features within machine learning.

In this study, we modified the time-domain analysis method, DFA, and combined it with SampEn, PermEn, and FuzzEn as inputs to a random forest model for identifying awareness levels. The proposed method was investigated using intra-operative EEG recordings and BIS index data from 73 patients. Metrics such as MSE, RMSE, MAE, and R-square indicated the effectiveness of the proposed method in discriminating

anesthesia states. However, the method still struggles with detecting very deep anesthesia states and abrupt distortions.

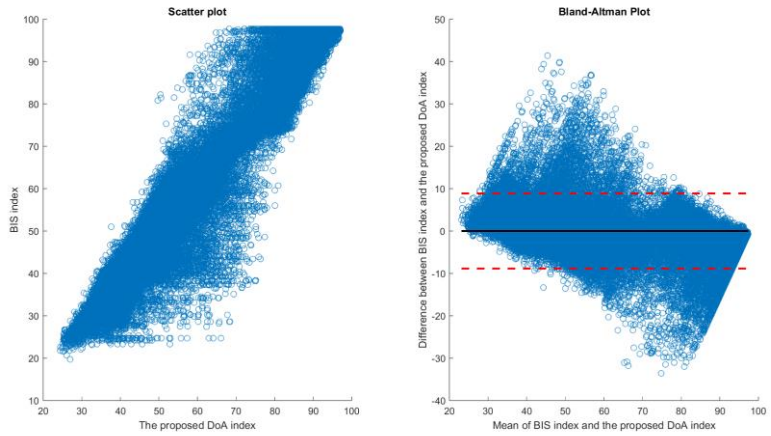

Fig. 7. (A) The scatter plot of the proposed DoA index. (B) The Bland-Altman plot of the BIS index and the proposed DoA index.

Previous studies [35] have shown that the SFS and SEF95 metrics dominate in deep anesthesia states due to their ability to capture frequency variations and phase coupling. For very deep anesthesia states, characterized by burst suppression patterns, the conscious level corresponds to the quasi-periodic alternation between high-voltage slow waves (bursts) and low voltage or isoelectric periods (suppression) [36]. According to the American Clinical Neurophysiology Society (ACNS), burst suppression is defined as a background with more than 50% attenuation or suppression, with bursts lasting a minimum of 0.5 seconds [37].

The proposed feature extraction methods convert the input sequence into several templates and investigate their relationships, rather than detecting these alternations. Additionally, the lack of very deep anesthesia state recordings impedes performance. To assess the pros and cons of the proposed method, we further examined the best and worst performance cases. In Fig. 8, the proposed index presents the trend of intra-operative awareness variation with fewer disturbances and more stable monitoring results compared to the BIS index. However, abnormal outliers sometimes appear even in the best-performing recordings. In the worst performance case shown in Fig. 9, the proposed index produces larger and more frequent outliers.

Since only the modified DFA focuses on outliers in each second, the entropy methods applied here should be further inspected. All methods in this study process EEG epochs by breaking the sequence into multiple sub-sequences and investigating correlations among them, thus neglecting the connection between EEG epochs over time.

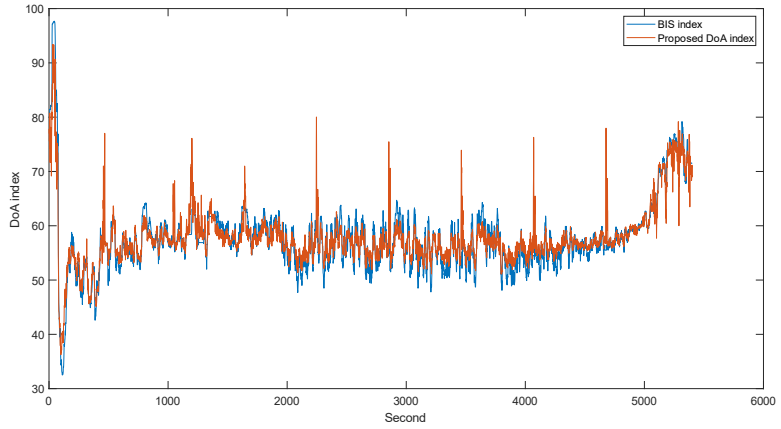

Fig. 8. Best performance Norm

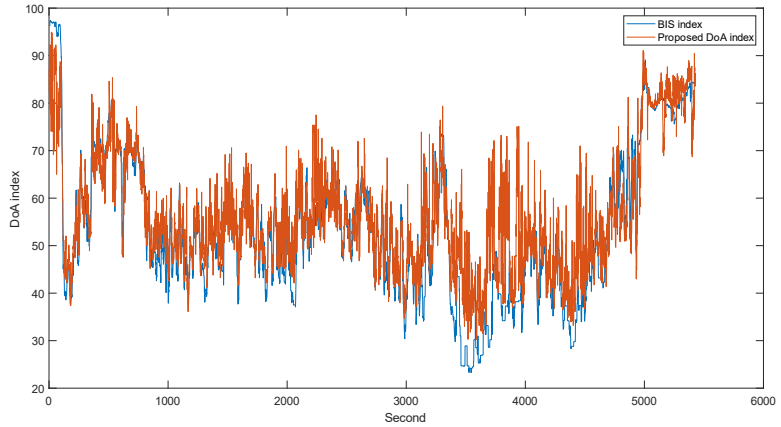

Fig. 9. Worst performance Norm

Although the proposed method can indicate the conscious level, it still limits the two considerations. The modified DFA improves the computation efficiency at the feature extraction stage, while model optimization requires further investigation and development, such as pruning in random forest. Moreover, the variability of human EEG and anesthetics should be considered in the future study.

5 Conclusion

In this study, we extracted self-affinity features and entropy features from EEG, and took them as inputs to a random forest model to assess the DoA level. The results demonstrated the effectiveness of the proposed method in discriminating levels of

awareness. The proposed method achieved relatively low MSE, RMSE, and MAE between the proposed DoA index and the BIS which is the most popular one. Additionally, the correlation and R-square values were generally high. These results highlight the computational efficiency of the proposed method and its feasibility for real-time intra-operative DoA monitoring systems, which is crucial for practical applications in clinical settings.

Future work will focus on incorporating more relevant features and applying advanced machine learning techniques to further improve the accuracy and robustness of DoA assessment.

References

[1] M.-L. Lindholm *et al.*, 'Mortality within 2 years after surgery in relation to low intraoperative bispectral index values and preexisting malignant disease', *Anesthesia & Analgesia*, vol. 108, no. 2, pp. 508–512, 2009.

[2] T. G. Monk, V. Saini, B. C. Weldon, and J. C. Sigl, 'Anesthetic management and one-year mortality after noncardiac surgery', *Anesthesia & Analgesia*, vol. 100, no. 1, pp. 4–10, 2005.

[3] M. S. Avidan *et al.*, 'Anesthesia Awareness and the Bispectral Index', *N Engl J Med*, vol. 358, no. 11, pp. 1097–1108, Mar. 2008, doi: 10.1056/NEJMoa0707361.

[4] K. Leslie, M. T. Chan, P. S. Myles, A. Forbes, and T. J. McCulloch, 'Posttraumatic stress disorder in aware patients from the B-aware trial', *Anesthesia & Analgesia*, vol. 110, no. 3, pp. 823–828, 2010.

[5] J. Smajic *et al.*, 'Assessment of depth of anesthesia: PRST score versus bispectral index', *Med Arh*, vol. 65, no. 4, pp. 216–20, 2011.

[6] J. Zhan *et al.*, 'Heart rate variability-derived features based on deep neural network for distinguishing different anaesthesia states', *BMC Anesthesiol*, vol. 21, no. 1, p. 66, Mar. 2021, doi: 10.1186/s12871-021-01285-x.

[7] R. J. Gajraj, M. Doi, H. Mantzaridis, and G. N. Kenny, 'Comparison of bispectral EEG analysis and auditory evoked potentials for monitoring depth of anaesthesia during propofol anaesthesia.', *British journal of anaesthesia*, vol. 82, no. 5, pp. 672–678, 1999.

[8] M. R. Chowdhury, R. Madanu, M. F. Abbod, S.-Z. Fan, and J.-S. Shieh, 'Deep learning via ECG and PPG signals for prediction of depth of anesthesia', *Biomedical Signal Processing and Control*, vol. 68, p. 102663, 2021.

[9] G. Wang *et al.*, 'Monitoring the Depth of Anesthesia Through the Use of Cerebral Hemodynamic Measurements Based on Sample Entropy Algorithm', *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 807–816, Mar. 2020, doi: 10.1109/TBME.2019.2921362.

[10] A. Shalbaf, M. Saffar, J. W. Sleigh, and R. Shalbaf, 'Monitoring the depth of anesthesia using a new adaptive neurofuzzy system', *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 671–677, 2017.

[11] M. Cascella, 'Mechanisms underlying brain monitoring during anesthesia: limitations, possible improvements, and perspectives', *Korean J Anesthesiol*, vol. 69, no. 2, pp. 113–120, Apr. 2016, doi: 10.4097/kjae.2016.69.2.113.

[12] R. Shalbaf, H. Behnam, J. W. Sleigh, A. Steyn-Ross, and L. J. Voss, 'Monitoring the depth of anesthesia using entropy features and an artificial neural network', *Journal of Neuroscience Methods*, vol. 218, no. 1, pp. 17–24, Aug. 2013, doi: 10.1016/j.jneumeth.2013.03.008.

[13] Q. Wei *et al.*, 'Analysis of EEG via Multivariate Empirical Mode Decomposition for Depth of Anesthesia Based on Sample Entropy', *Entropy*, vol. 15, no. 9, Art. no. 9, Sep. 2013, doi: 10.3390/e15093458.

[14] Z. Liang *et al.*, 'EEG entropy measures in anesthesia', *Front. Comput. Neurosci.*, vol. 9, Feb. 2015, doi: 10.3389/fncom.2015.00016.

[15] M. Jospin *et al.*, 'Detrended Fluctuation Analysis of EEG as a Measure of Depth of Anesthesia', *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 840–846, May 2007, doi: 10.1109/TBME.2007.893453.

[16] P. Gifani, H. R. Rabiee, M. H. Hashemi, P. Taslimi, and M. Ghanbari, 'Optimal fractal-scaling analysis of human EEG dynamic for depth of anesthesia quantification', *Journal of the Franklin Institute*, vol. 344, no. 3, pp. 212–229, May 2007, doi: 10.1016/j.jfranklin.2006.08.004.

[17] T. Nguyen-Ky, P. Wen, and Y. Li, 'Improving the accuracy of depth of anaesthesia using modified detrended fluctuation analysis method', *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 59–65, Jan. 2010, doi: 10.1016/j.bspc.2009.03.001.

[18] Q. Liu *et al.*, 'Spectrum Analysis of EEG Signals Using CNN to Model Patient's Consciousness Level Based on Anesthesiologists' Experience', *IEEE Access*, vol. 7, pp. 53731–53742, 2019, doi: 10.1109/ACCESS.2019.2912273.

[19] S. Afshar, R. Boostani, and S. Sanei, 'A Combinatorial Deep Learning Structure for Precise Depth of Anesthesia Estimation From EEG Signals', *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3408–3415, Sep. 2021, doi: 10.1109/JBHI.2021.3068481.

[20] X. Jiang, G.-B. Bian, and Z. Tian, 'Removal of Artifacts from EEG Signals: A Review', *Sensors*, vol. 19, no. 5, Art. no. 5, Jan. 2019, doi: 10.3390/s19050987.

[21] C. Q. Lai, H. Ibrahim, M. Z. Abdullah, J. M. Abdullah, S. A. Suandi, and A. Azman, 'Artifacts and noise removal for electroencephalogram (EEG): A literature review', in *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Apr. 2018, pp. 326–332. doi: 10.1109/ISCAIE.2018.8405493.

[22] Y.-F. Chen, S.-Z. Fan, M. F. Abbod, J.-S. Shieh, and M. Zhang, 'Electroencephalogram variability analysis for monitoring depth of anesthesia', *J. Neural Eng.*, vol. 18, no. 6, p. 066015, Nov. 2021, doi: 10.1088/1741-2552/ac3316.

[23] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, 'Mosaic organization of DNA nucleotides', *Phys. Rev. E*, vol. 49, no. 2, pp. 1685–1689, Feb. 1994, doi: 10.1103/PhysRevE.49.1685.

[24] X. Li, F. Wang, and G. Wu, 'Monitoring Depth of Anesthesia Using Detrended Fluctuation Analysis Based on EEG Signals', *J. Med. Biol. Eng.*, vol. 37, no. 2, pp. 171–180, Apr. 2017, doi: 10.1007/s40846-016-0196-y.

[25] Z. Liang *et al.*, 'EEG entropy measures in anesthesia', *Frontiers in computational neuroscience*, vol. 9, p. 16, 2015.

[26] W. Chen, Z. Wang, H. Xie, and W. Yu, 'Characterization of surface EMG signal based on fuzzy entropy', *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 2, pp. 266–272, 2007.

[27] Z. Cao and C.-T. Lin, 'Inherent Fuzzy Entropy for the Improvement of EEG Complexity Evaluation', *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 1032–1035, Apr. 2018, doi: 10.1109/TFUZZ.2017.2666789.

[28] L. Breiman, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[29] M.-P. Hosseini, A. Hosseini, and K. Ahi, 'A Review on Machine Learning for EEG Signal Processing in Bioengineering', *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 204–218, 2021, doi: 10.1109/RBME.2020.2969915.

[30] M. N. Anastasiadou, M. Christodoulakis, E. S. Papathanasiou, S. S. Papacostas, and G. D. Mitsis, 'Unsupervised detection and removal of muscle artifacts from scalp EEG recordings using canonical correlation analysis, wavelets and random forests', *Clinical Neurophysiology*, vol. 128, no. 9, pp. 1755–1769, 2017.

[31] J.-E. Le Douget, A. Fouad, M. M. Filali, J. Pyrzowski, and M. Le Van Quyen, 'Surface and intracranial EEG spike detection based on discrete wavelet decomposition and random forest classification', in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2017, pp. 475–478. Accessed: Jun. 21, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8036865/

[32] M. Bentlemsan, E.-T. Zemouri, D. Bouchaffra, B. Yahya-Zoubir, and K. Ferroudji, 'Random forest and filter bank common spatial patterns for EEG-based motor imagery classification', in *2014 5th International conference on intelligent systems, modelling and simulation*, IEEE, 2014, pp. 235–238. Accessed: Jun. 21, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7280913/

[33] S. Bose, V. Rama, N. Warangal, and C. R. Rao, 'EEG signal analysis for Seizure detection using Discrete Wavelet Transform and Random Forest', in *2017 International Conference on Computer and Applications (ICCA)*, IEEE, 2017, pp. 369–378. Accessed: Jun. 21, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8079760/

[34] Q. Liu, Y.-F. Chen, S.-Z. Fan, M. F. Abbod, and J.-S. Shieh, 'EEG artifacts reduction by multivariate empirical mode decomposition and multiscale entropy for monitoring depth of anaesthesia during surgery', *Med Biol Eng Comput*, vol. 55, no. 8, pp. 1435–1450, Aug. 2017, doi: 10.1007/s11517-016-1598-2.

[35] I. J. Rampil, 'A primer for EEG signal processing in anesthesia', *The Journal of the American Society of Anesthesiologists*, vol. 89, no. 4, pp. 980–1002, 1998.

[36] F. Amzica, 'What does burst suppression really mean?', *Epilepsy & Behavior*, vol. 49, pp. 234–237, Aug. 2015, doi: 10.1016/j.yebeh.2015.06.012.

[37] L. J. Hirsch *et al.*, 'American Clinical Neurophysiology Society's Standardized Critical Care EEG Terminology: 2012 version', *J Clin Neurophysiol*, vol. 30, no. 1, pp. 1–27, Feb. 2013, doi: 10.1097/WNP.0b013e3182784729.