Brain Informatics

RESEARCH

Open Access

Accurate depth of anesthesia monitoring based on EEG signal complexity and frequency features

Tianning Li[1], Yi Huang[1*], Peng Wen[2] and Yan Li[1]

Abstract

Accurate monitoring of the depth of anesthesia (DoA) is essential for ensuring patient safety and effective anesthesia management. Existing methods, such as the Bispectral Index (BIS), are limited in real-time accuracy and robustness. Current methods have problems in generalizability across diverse patient datasets and are sensitive to artifacts, making it difficult to provide reliable DoA assessments in real time. This study proposes a novel method for DoA monitoring using EEG signals, focusing on accuracy, robustness, and real-time application. EEG signals were pre-processed using wavelet denoising and discrete wavelet transform (DWT). Features such as Permutation Lempel–Ziv Complexity (PLZC) and Power Spectral Density (PSD) were extracted. A random forest regression model was employed to estimate anesthetic states, and an unsupervised learning method using the Hurst exponent algorithm and hierarchical clustering was introduced to detect transitions between anesthesia states. The method was tested on two independent datasets (UniSQ and VitalDB), achieving an average Pearson correlation coefficient of 0.86 and 0.82, respectively. For the combined dataset, the model demonstrated an R-squared value of 0.70, a RMSE of 6.31, a MAE of 8.38, and a Pearson correlation of 0.84, showcasing its robustness and generalizability. This approach offers a more accurate and reliable real-time DoA monitoring tool that could significantly improve patient safety and anesthesia management, especially in diverse clinical environments.

Keywords Depth of anesthesia (DoA), Electroencephalogram (EEG), Permutation Lempel–Ziv Complexity (PLZC), Power spectral density (PSD), Hurst exponent algorithm, Hierarchical clustering, Random forest regression

1 Introduction

The assessment of depth of anesthesia (DoA) is a critical aspect of patient care during surgical procedures and it is necessary to prevent intraoperative awareness and excessive anesthetic dosing [1]. While heart rate variability, processed electroencephalogram (pEEG) indices, and auditory evoked potentials assist specialists in measuring DoA, they are subject to individual variations. As a direct reflection of the brain's response to anesthetic agents, EEG signals offer valuable information for determining consciousness levels during anesthesia.

Despite the widespread use of tools like the Bispectral Index (BIS) for DoA assessment, limitations persist in real-time surgery monitoring. These tools are sensitive to individual variations, computationally complex, and often fail to generalize across diverse patient populations. Advances in machine learning and deep learning have proposed solutions, leveraging features such as wavelet coefficients, entropy measures, and frequency analysis techniques for enhanced decision-making. However, the real-time reliability and generalizability of these solutions remain significant challenges [2, 3].

*Correspondence:
Yi Huang
Yi.Huang@unisq.edu.au
[1] School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, Australia
[2] School of Engineering, University of Southern Queensland, Toowoomba, Australia

In parallel, Artificial Intelligence's role in EEG analysis for medical diagnostics has seen promising developments. For instance, the automatic classification of sleep stages has been refined through graph convolutional networks, achieving impressive accuracy and execution efficiency [4]. Schizophrenia identification now utilizes dynamic functional connectivity analysis with deep learning, offering high accuracy and insight into neural network variations [5]. Real-time epilepsy seizure detection methods have harnessed discrete wavelet transform alongside machine learning for accurate classification and onset detection [6]. Moreover, the identification of Autism Spectrum Disorders in children has advanced through multimodal diagnostic frameworks, combining EEG with eye-tracking data for a comprehensive assessment [7].

However, current DoA monitoring systems have yet to fully exploit the potential of machine learning and EEG analysis for real-time, reliable, and generalizable monitoring across diverse clinical environments. To address this gap, this study proposes an EEG-based method for DoA estimation, integrating Permutation Lempel–Ziv Complexity (PLZC) and Power Spectral Density (PSD) to capture the complexity and spectral features of EEG signals. The proposed approach is validated across two independent datasets, demonstrating robust and generalizable performance. Moreover, the method is designed for real-time application in clinical settings, ensuring computational efficiency while maintaining high accuracy.

This paper is organized to first introduce the study's aims, followed by an exploration of diverse datasets and methodologies for DoA assessment, and concludes with empirical validation, analysis of DoA state transitions, and discussions on the potential of this method to enhance patient care during surgery.

2 Related work

Considerable progress has been made in easing the study and construction of highly accurate models for the analysis of EEG signals in the assessment of DoA, thanks to recent advancements in machine learning (ML) algorithms and statistical metrics. The majority of the time, training is based on the labels supplied by classification models, clinical assessment of the DoA state (CAD) for index design, and industrial models (like the BIS). Regression approaches, such as the Gaussian process, support vector machine (SVM) regression, artificial neural network, and Random Forest classifier, are the traditional machine learning methods that have been used recently. The most recent approaches are categorised in Tables 1, 2 with details on the feature extraction technique used, along with the model-building methodology used and evaluation criteria.

Recent literature also contains a variety of deep learning-based model-building techniques. Based on EEG signal analysis, these techniques have been used to describe the intricate and nonlinear interactions between EEG signals and the DoA. A combination of a 1×1 convolution network and a deep residual shrinkage network (DRSN) demonstrated a high positive connection, as evidenced by the Spearman's rank correlation coefficient of 0.9344 (PSI) on a wide variety of data extracted after wavelet treatment [15]. Utilising a CNN model that analysed 60-channel EEG signals utilising network and graph features, a correlation of 0.872 with the perturbational

Table 1 Latest DoA assessment algorithms for index design/regression since 2022

Author	Feature extraction method	Model building Method	Outcome
Huang et al. [8]	Range of entropy features including fuzzy entropy	Gaussian process regression	Correlation = 0.9491 (BIS)
Schmierer et al. [9]	Empirical wavelet transformation (EWT) with Second Order Differential Plot (SODP) and Spectral Entropy (SE)	SVM regression	Correlation = 0.834, Choen's Kappa of 0.809 (BIS)
Alsafy and Diykh [10]	Hierarchical dispersion entropy (HDE) with wavelet transforms (WT)	Artificial neural network	Average coefficient of determination = 0.965 (BIS)
Chen et al. [11]	EEG variability and EEG analysis with a range of spectral-domain and entropy-domain features	Long short-term memory	Correlation = 0.70, AUC = 0.93 (BIS). Correlation = 0.80, AUC = 0.93 (CAD)
Lee et al. [12]	60-channel EEG with spatiotemporal dynamics	Convolutional neural network (CNN)	Correlation = 0.872 (perturbational complexity index)
Shi et al. [13]	WT with 14 features extracted including SEF and Sample Entropy	Deep residual shrinkage network (DRSN) & 1×1 CNN	Spearman's rank correlation coefficient = 0.9344 (PSI)
Shahbakhti et al. [14]	Parameter-free features based on entropy, power and frequency, fractal, and variation	Random forest regressor	Correlation = 0.80 and 0.79 (BIS) for Databases I and II, Mean absolute error (MAE) of 7.1 and 9.0 for Databases I and II

Table 2 Latest DoA assessment algorithms for classification Since 2022

Author	Feature extraction method	Model building Method	Outcome
Casey et al. [15]	256-channel EEG with occipital delta power and power spectrum density (PSD)	SVM	AUC = 0.622 (CAD)
Xiao et al. [16]	19-channel EEG with phasal relationship	Random Forrest classifier	Accuracy = 93.88% (based on 2 states, CAD)
Anand et al. [17]	Range of time series features (63) extracted		Accuracy = 83% (based on 2 states BIS)
Wang et al. [18]	No Feature extraction	Long short-term memory (LSTM)	Accuracy = 81.8% (CAD)
Dutt and Saadeh [19]	Stationary wavelet transform (SWT) with fractal, non-linear and spectral features	Multilayer perceptron regressor network	Accuracy = 96.8% (CAD)
Dutt and Saadeh [20]			Accuracy = 97.1%, $R^2 = 0.9$, MAE = 1.5
Zhang et al. [21]	Frequency components sampling and Frequency domain self-attention	Frequency Enhanced Hybrid Attention Network	Accuracy = 89.63% (CAD)

complexity index (PCI) was discovered [12] that employed network and graph properties to analyse 60-channel EEG signals. However, when working with data that has a grid-like structure, like image data, convolutional neural networks (CNN) are excellent in extracting hierarchical features in an automatic and adaptive manner. But their computation intensity is high, which makes them difficult to use in some situations.

Permutation Lempel–Ziv Complexity (PLZC) is a signal complexity analysis technique that is a variant of Lempel–Ziv Complexity (LZC). It utilizes permutation vectors to quantify the neighboring values of the time series and employs them to assess the Depth of Anaesthesia (DoA) [22]. As a more accurate representation of the mutual relationship between neighboring signal points compared to classical LZC, PLZC exhibits high sensitivity in assessing low-amplitude (high-frequency) EEG signals during changes in brain state. Due to its foundation in the relative loudness of the signal, PLZC is more robust to noise than the conventional Lempel–Ziv Complexity. Because of this property, PLZC is especially helpful in differentiating between different states of consciousness in EEG readings [23, 24]. A common tool in EEG analysis for DoA assessment is Power Spectrum Density (PSD). It offers a detailed view of the power distribution across the whole frequency range. In comparison to awake patients, anaesthetized patients' EEG signals often exhibit more power in the lower frequency bands (delta and theta) and less power in the higher frequency bands (alpha and beta). The decomposition of the electroencephalogram signal into its frequency components' power reveals that under general anesthesia, the electroencephalogram is organized into distinct oscillations at specific frequencies [25]. The information derived from EEG power spectrum analysis can be enhanced by incorporating newer EEG indices, such as the bispectral index and approximate entropy. Additionally, other neurophysiological monitors, including auditory evoked potentials or somatosensory evoked potentials, can further supplement the gathered data [26]. These approaches have shown potential in the DoA assessment, but there are still issues to be resolved, like the sensitivity to individual differences, the computational cost of feature extraction, and the generalizability of methods across different patient populations because most anaesthesia datasets are private [27–29]. Thus, creating a reliable and effective EEG-based DoA estimation technique is essential.

3 Methods

The study introduces an advanced methodology for the real-time monitoring of the depth of anesthesia via EEG signals. This approach integrates signal processing techniques with machine learning algorithms to facilitate precise DoA estimation. Initially, EEG signal acquisition is undertaken, followed by preprocessing, which includes wavelet-based denoising and discrete wavelet transform for frequency filtering. Subsequently, significant statistical features are extracted utilizing Permutation Lempel–Ziv Complexity and Power Spectral Density analysis. These features are then employed to train a Random Forest regression model that classifies the DoA levels. The culmination of this methodology is the generation of a new index for DoA monitoring, indicative of transitions in anesthetic states. Figure 1 delineates the comprehensive workflow, illustrating the process from the initial EEG data preprocessing to the final prediction of DoA states and transitions.

3.1 Pre-processing

Accurate discrimination of anesthetic response from raw EEG data can be challenging due to the presence of noise. These include, but are not limited to, muscle activity, eye movement, and various other artifacts which are especially prevalent in the awake state. As a result, it

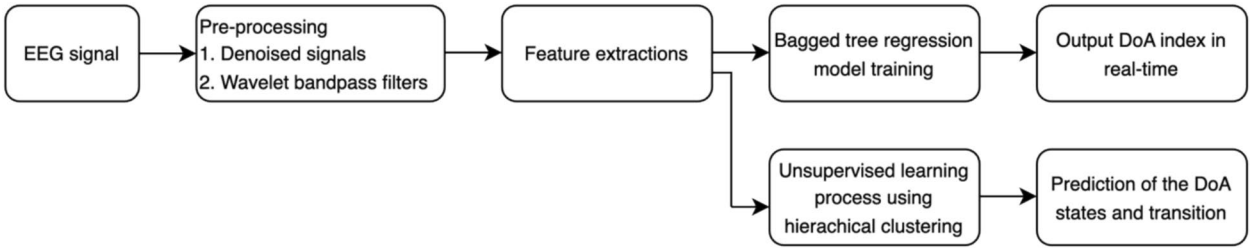

Fig. 1 The framework of the new DoA index development and the states prediction

was necessary to put all raw EEG signals through a preprocessing phase to filter those noise. A wavelet threshold method grounded in entropy was applied to remove low-amplitude noise and spike noise, and the details of de-noising EEG signals are in [8, 30–32].

One of the most challenging aspects in depth of anesthesia assessment using EEG signals is the accurate identification and extraction of relevant features amidst the varying frequency components. However, not all features across all frequency bands show significant differences among different anesthesia states. Thus, discrete wavelet transform (DWT) is proposed to decompose the EEG signals into various frequency sub-bands, allowing for the examination of frequency components with higher relevance to the aesthetic depth [33]. The formula for DWT is as follows [34, 35]:

$$C(a,b) = \frac{1}{\sqrt{a}} \int \bar{\psi} \left(\frac{t-b}{a} \right) x(t) dt \tag{1}$$

where ψ is the analyzing wavelet method, 'a' and 'b' are the parameters of time dilation and time translation, respectively.

After DWT decomposition, two style coefficients the detail coefficients and approximation coefficients of each sub-bands are calculated in Eqs. 2 and 3.

$$A_j(n) = \sum_{l=-\infty}^{+\infty} g(l-2n) A_{(j-1)}(l), j = 1, 2, ..., J \tag{2}$$

$$D_j(n) = \sum_{l=-\infty}^{+\infty} h(l-2n) A_{(j-1)}(l), j = 1, 2, ..., J \tag{3}$$

where $A_j(n)$ and $D_j(n)$ are the approximation coefficients and detail coefficients at level j respectively.

A 5-level DWT employing the 'db12' wavelet technique is utilized to decompose the data from UniSQ database, which has a sample rate of 128 Hz, and 'db16' was applied to the data from VitalDB database. The resulting cD1, cD2, cD3, cD4, cD5 and cA5 detail coefficients correspond to the EEG sub-bands outlined in Table 3.

Table 3 Decomposed levels and corresponding frequency bands

Sub-band j	Decomposed level signal	Frequency Bands (Hz)
1	cD1	64–128
2	cD2	32–64
3	cD3	16–32
4	cD4	8–16
5	cD5	4–8
5	cA5	0–4

3.2 Feature extraction

In this study, two distinct algorithms were employed for feature extraction from the processed EEG signals: a multiple signal classification (MUSIC) method and a permutation Lempel–Ziv complexity algorithm. The MUSIC method was chosen for its high resolution in frequency domain analysis, while the PLZC algorithm was selected for its robustness against noise and ability to capture dynamic changes in brain activity. These complementary methods allow for a comprehensive assessment of the depth of anesthesia based on the extracted features from EEG data.

In the PLZC algorithm, the time series of EEG data is first transformed into a sequence of ordinal patterns or permutations using the permutation entropy (PE) method. Each ordinal pattern represents a specific arrangement of the elements in a sliding window of a certain length along the data sequence.

$$PE = -\sum p(\pi) * \log(p(\pi)) \tag{4}$$

where $p(\pi)$ represents the probability of a given permutation π occurring in the time series data and embedding dimension m and time delay τ are set empirically at 4 and 1, respectively.

Then, the Lempel–Ziv complexity is applied to the ordinal input data pattern sequence to compute the complexity of the sequence. The complexity of the ordinal pattern sequence is calculated using the following equation:

$$PLZC = \frac{c(n)[log_{m!}c(n) + 1]}{n} \tag{5}$$

where $c(n)$ denotes the total length of the symbol sequence.

In applications of the MUSIC method for extracting features from the power spectral density (PSD) of wavelet coefficients, the procedures are referred to as outlined in previous research [36, 37] and are also summarised as below:

1. The coefficients A_j and D_j are utilized as input signals for the eigenvector method:

$$D : \begin{cases} D_j \rightarrow P(D_j) = ED_j \\ A_j \rightarrow P(A_j) = EA_j \end{cases} \tag{6}$$

where j represents a six-level wavelet decomposition, and the principal eigenvector corresponds to 6.

2. The average values of EA_j and ED_j are computed as follows:

$$Mean : \begin{cases} D_j \rightarrow M(ED_j) = mean\,(ED_j) \\ A_j \rightarrow M(AD_j) = mean\,(AD_j) \end{cases} \tag{7}$$

3. Standard deviation (STD) of EA_j and ED_j are determined as follows:

$$STD : \begin{cases} D_j \rightarrow S(ED_j) = std(ED_j) \\ A_j \rightarrow S(AD_j) = std(AD_j) \end{cases} \tag{8}$$

4. Derived from steps 2 and 3:

$$D : \begin{cases} M_j \rightarrow \frac{1}{2}\{\log(mean(M(ED_j))) + \\ \quad log(mean(M(AD_j)))\} \\ S_j \rightarrow \frac{1}{2}\{\log(mean(S(ED_j))) + \\ \quad log(mean(S(AD_j)))\} \end{cases} \tag{9}$$

5. Deriving the feature:

$$PSD = \frac{k1 \times M_j + k2 \times S_j}{k_3} \tag{10}$$

The constant parameters *k1*, *k2*, and *k3* are chosen empirically as 28, 90, and 3, respectively.

3.3 Development and assessment of the regression model

In this study, a bagged tree regression (BTR) algorithm was employed to estimate the depth of DoA based on the features extracted from EEG data. The BTR algorithm, an ensemble learning approach that merges predictions from multiple decision trees to boost model accuracy and robustness, was utilized in our investigation for DoA estimation using EEG-derived features.

The dataset was divided into training and testing subsets to facilitate the training and evaluation of the BTR model. The training set constituted 80% of the data, while the remaining 20% was designated as the testing set. This partitioning ensured a reliable evaluation of the model's performance on previously unseen data, thus providing an accurate representation of its real-world applicability.

The hyperparameters of the BTR model, such as maximum tree depth, minimum samples for split, and minimum samples per leaf, were optimized to maximize prediction accuracy. A grid search, coupled with cross-validation, was conducted to systematically explore the hyperparameter search space and evaluate each combination's performance, ensuring that the model was well-tuned for the dataset characteristics.

Upon determining the optimal hyperparameters, the BTR model was trained using the training dataset and the selected hyperparameters. The trained model was subsequently utilized to estimate the depth of anesthesia on the testing dataset, and its performance was subsequently evaluated on the testing dataset using R-squared (R^2), Root Mean Square Error (RMSE), and Bland–Altman analysis. The R^2 value quantified the proportion of variance in the observed BIS values explained by the proposed DoA index, with higher values indicating a better model fit. The RMSE measured the average prediction error, with lower values reflecting greater predictive accuracy. In addition, the Bland–Altman method was employed to assess agreement between the proposed index and the BIS index by plotting the differences against their means, offering insights into consistency. To further evaluate the relationship between the estimated DoA and BIS values, the Pearson correlation coefficient was calculated, providing a measure of the linear correlation between the two indices. Values closer to one suggest a stronger positive relationship.

3.4 Prediction of the DoA states and transition

In addition to regression analysis, the overall effectiveness of the new index in predicting the DoA states was evaluated by a novel method that combines the Hurst exponent algorithm, unsupervised learning, and the anesthesia agent usage information [38, 39].

First, the Hurst exponent algorithm was used to extract features from the EEG signals, which provide a representation of the underlying DoA states. For each EEG signal, we calculate the Hurst exponent values (H) using a rescaled range method. The rescaled range is calculated as:

$$(R/S)_n = \frac{1}{\left(\frac{N}{m}\right)} \sum_{n=1}^{\left\lceil \frac{N}{m} \right\rceil} (R_n/S_n) \tag{11}$$

The Hurst exponent is then obtained by finding the slope of the log–log plot of the rescaled range as a function of the time scale:

$$R_r = \min_{n=1,2,\ldots} \{mean(R_n)\} \tag{12}$$

These features (H values) are then utilized in the subsequent unsupervised learning process using hierarchical clustering. Next, an unsupervised learning method, hierarchical clustering, was employed to classify the extracted features into distinct DoA states. The resulting clusters indicated state transition changes, allowing for the identification of the transitions between different levels of anesthesia.

To further validate the effectiveness of the proposed method, the information of anesthesia agent usage provided by anesthetists was analyzed. The loss of consciousness (LOC) and recovery of consciousness (ROC) transition points, as observed by anesthetists, and were compared with the state change points detected by the proposed method. Additionally, these transition points were compared with the BIS values to investigate the agreement between the new index and the BIS index.

4 Results

4.1 Subjects and data collection

The first datasets utilized in this study were gathered from 28 adult participants, and we called the UniSQ database. The approvals for the data collection and conducting the research were obtained from the Human Research Ethics Committee of the University of Southern Queensland (Approval No: H09REA029) and the Human Research Ethics Committee of the Toowoomba and Darling Downs Health Service District (Approval No: TDDHSD HREC 2009/016).

The EEG and BIS datasets were initially collected from 37 patients at Toowoomba St Vincent's Hospital using a BIS VISTA™ monitoring system, version 3.22. From the original data pool, only 28 cases were included in this study to ensure a high quality of EEG data. Cases with low signal quality index (SQI), as well as those with incomplete recordings, were excluded. This selection ensured that the study was based on comprehensive and accurate data, thereby increasing the reliability of the results. Anesthesiologists on site documented all events occurring intraoperatively as well as intravenous dosing in dedicated data log fields. The compiled data logs incorporated a multitude of elements, including but not limited to, the BIS index, raw EEG data, the SQI, impedance, electromyography (EMG) data, along with a log of monitor errors. Demographic information for all participants is presented in Table 4. Typical drug administration for the subjects involved the use of preoperative medications, intravenous midazolam at 0.05 mg/kg, fentanyl at 1.5–3 μg/kg, or alfentanil at 15–30 μg/kg.

The second database employed in this study, designated as the VitalDB database, was acquired from the Department of Anesthesiology and Pain Medicine at Seoul National University Hospital, South Korea [40]. This publicly available database encompasses a broad spectrum of physiological signals collected from patients who underwent surgical procedures under general anesthesia. Despite the VitalDB database containing 6,388 cases, 28 cases with satisfactory SQI from the EEG signals were selected at random. The selection ensured that the number of cases was consistent with that of the UniSQ database, thereby maintaining a balance in the influence of each database on the final model design. This approach ensures the comparability and coherence of the two datasets. Therefore, only the EEG signals from these selected cases were utilized in the analysis and subsequently compared with the UniSQ database. The incorporation of the VitalDB database facilitated a comprehensive assessment of the proposed methodology across varied patient demographics and clinical settings. In-depth demographic information and anesthetic administration data were provided in Table 4, and can be accessed via the VitalDB website (https://vitaldb.net/dataset/).

4.2 Pre-processing

The wavelet-based preprocessing method successfully removed low-amplitude noise and spike interference from the raw EEG signals, improving signal quality and facilitating feature extraction. Table 5 and Table 6 summarize the selected features from the UniSQ and VitalDB databases after the preprocessing stage.

The EEG signals were decomposed using discrete wavelet transform (DWT) into five levels, corresponding

Table 4 Patient demographics and intraoperative drug usage

Variable/Description	UniSQ Database Details	VitalDB Database Details
Age (year)	2–83	16–78
Weight (kg)	55–130	45.1–88
Height (cm)	154–194	147.9–180.5
Gender (F/M)	15/22	10/18
Midazolam (mg)	2–5	30–190
Alfentanil (μg)	500, 750, 1000	500, 750, 1000
Propofol (mg)	90–200	80–150
Parecoxib (mg)	40	5–30
Fentanyl (ug)	100, 150	50, 100
EEG Channels	2 (Channel 2 used for analysis)	2 (Channel 2 used for analysis)
Sampling Rate	128 Hz	128 Hz
BIS Recording	1 value/sec	1 value/sec
Window Size	56 s	56 s
Overlap	55 s	55 s
Preprocessing	None (signal not denoised prior to analysis)	None (signal not denoised prior to analysis)

Table 5 Selected features from UniSQ database

Decomposed level signal	Features
CD1	N/A
CD2	PLZC
CD3	PLZC, PSD
CD4	PSD
CD5	PSD
CA5	N/A

Table 6 Selected features from VitalDB database

Decomposed level signal	Features
CD1	N/A
CD2	PSD
CD3	PLZC, PSD
CD4	PLZC
CD5	PLZC, PSD
CA5	N/A

to specific EEG frequency sub-bands. From these sub-bands, key features—Permutation Lempel–Ziv Complexity (PLZC) and Power Spectral Density (PSD)—were extracted for further analysis. As shown in Tables 5 and 6, different feature combinations were selected from each sub-band, with no features extracted from CD1 and CA5 for either database. The preprocessing techniques improved the clarity of the signals, ensuring that relevant features for DoA estimation could be efficiently extracted. This enhancement laid a strong foundation for the accurate and reliable estimation of depth of anesthesia using the proposed index.

Tables 5 and 6 illustrate the selected features from different decomposed levels for the UniSQ and VitalDB datasets. The discrepancy in the selected features is a result of our dataset-specific feature selection process. Each dataset underwent independent feature selection to determine the most informative features for its unique characteristics, ensuring that the model for each dataset was optimized for performance. This approach allowed for a meaningful comparison between models trained on individual datasets and a generalized model trained on both datasets, highlighting the trade-offs between dataset-specific optimization and generalizability.

4.3 Feature selection

In the UniSQ database, the application of a 5-level discrete wavelet transform (DWT) using the 'db12' wavelet resulted in significant improvements in feature discrimination. PLZC values extracted from decomposition levels D_2 and D_3 demonstrated the highest sensitivity to transitions between anesthesia states, including the loss of consciousness (LOC), the maintenance of anesthesia, and the recovery of consciousness (ROC). This distinction is clearly illustrated in Fig. 2, which depicts PLZC values for a representative case across three key time intervals: LOC (0–500 s), stable anesthesia (1100–1800s), and ROC (3800–4150 s).

For the VitalDB database, optimal performance was achieved using a 6-level decomposition with the 'db16' wavelet. Feature extraction from levels D_3 and D_4 provided the most reliable discrimination between anesthetic states, with PLZC values showing strong

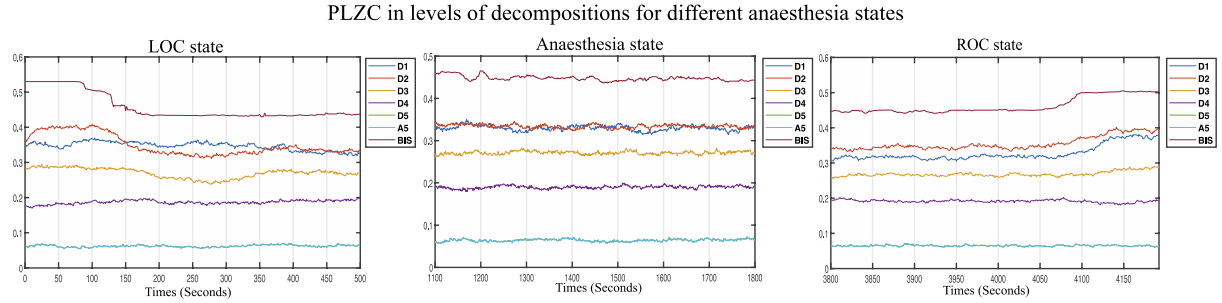

Fig. 2 PLZC with the different levels of decompositions for UniSQ database in case 16, and the comparison in three different anaesthetic states, including LOC state (0–500 s), anaesthetic state (1100–1800s), and ROC state (3800–4150 s)

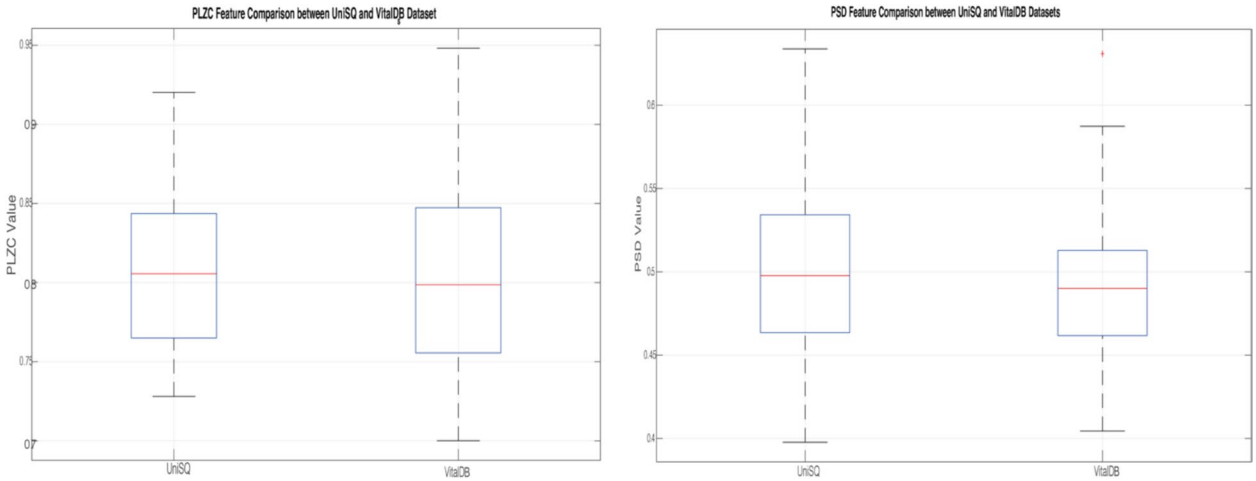

Fig. 3 Boxplots Comparing PLZC and PSD Feature Distributions Between UniSQ and VitalDB Datasets

correlations with the transitions between these states. The decomposition levels D_3 and D_4 were particularly effective in capturing variations in brain activity associated with changes in anesthetic depth.

The differences in the selected features between the UniSQ and VitalDB datasets (as shown in Tables 5 and 6) can be attributed to several factors. Firstly, the UniSQ and VitalDB databases differ in terms of patient demographics, clinical environments, and the types of anesthetic agents used, which likely influenced the EEG signals. In addition, the use of different wavelet techniques ('db12' for UniSQ and 'db16' for VitalDB), due to variations in sampling rates, also contributed to distinct spectral characteristics. These factors necessitated the selection of different features from different decomposition levels for each dataset to optimize performance. Adapting the feature selection to the specific characteristics of each dataset resulted in improved DoA estimation accuracy, as reflected in the evaluation metrics.

Through these analysis, we found the optimal level of decompositions and the mother wavelet for each database.

A Mann–Whitney U test was conducted to determine whether there were significant differences between the PLZC and PSD feature distributions for the UniSQ and VitalDB datasets (Fig. 3). The results indicated no significant difference between the PLZC distributions ($p=0.6760$) and no significant difference between the PSD distributions ($p=0.5718$). This suggests that both features exhibit similar behavior across the two datasets, reflecting comparable EEG signal properties in terms of complexity (PLZC) and spectral characteristics (PSD).

4.4 Ablation studies

An ablation study was undertaken on combined datasets to assess the relative impact of model features and parameter settings on the performance of the DoA estimation model. This study extends beyond feature contribution to include the examination of the sliding

Table 7 Ablation Study Outcomes including Sliding Window Analysis

Variation	Description	Regression Method	R^2	RMSE	MAE	Pearson Coefficient
Baseline	Full model with all features and original parameters	BTR	0.70	6.49	8.36	0.84
Without PLZC	Model without PLZC feature	BTR	0.67	6.93	9.05	0.81
Without PSD	Model without PSD feature	BTR	0.65	7.12	8.96	0.79
Linear Model	Full model using Linear Regression	Linear	0.50	8.68	9.16	0.62
SVM Model	Full model using SVM	SVM	0.63	7.31	9.08	0.67

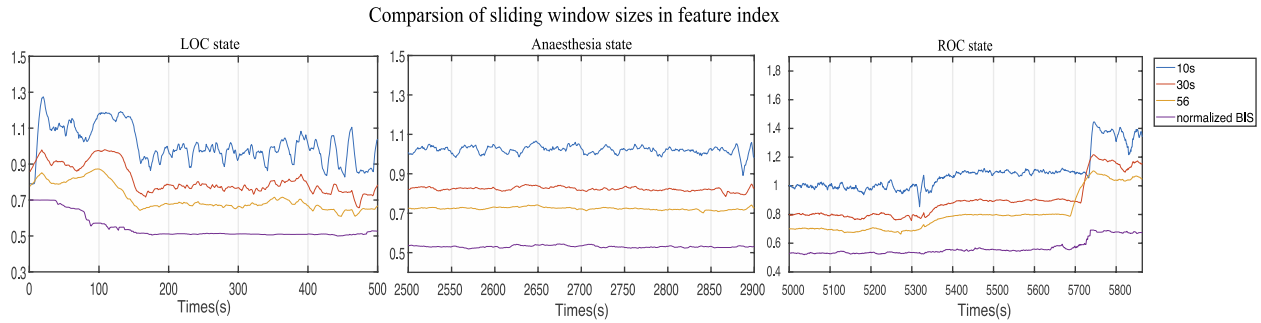

Fig. 4 PSD-D_3 in three different sliding window sizes for case 16 VitalDB database and the comparison in three different anaesthetic states, including LOC state (0–500 s), anaesthesia state (2500–2900 s), and ROC state (5500–5800 s)

window length, a key parameter influencing feature extraction and real-time monitoring capabilities.

The 'Baseline' model in Table 7 represents the full feature set optimized for a 56-s sliding window, serving as the standard for comparison. The removal of PLZC and PSD features from the model underscores their individual significance, as evidenced by the decrease in R^2 and increase in RMSE and MAE values. The ablation of the regression method further clarifies the superior performance of the Random Forest algorithm over Linear Regression and SVM. These findings are pivotal for understanding the robustness of the predictive model and guiding future refinements for DoA monitoring.

In this study, we tested several different sliding window sizes, including 10 s, 30 s, and 56 s, and evaluated their performance in terms of feature extraction (using the PSD-D_3 feature) and anesthetic state discrimination. Figure 4 illustrates the comparison of feature extraction results using these sliding window sizes. Also, we tested the performance of the model on combined databases in terms of various metrics such as R^2, RMSE, MAE, and Pearson Coefficient and the results are summarized in Table 8. Based on our analysis, a fixed-length sliding window of 56 s was selected as it provided a good balance between real-time performance and accurate feature extraction.

Table 8 Performance metrics of estimation model for different sliding window sizes

	10 s	30 s	56 s
R^2	0.57	0.63	0.70
RSME	8.62	8.01	6.31
MAE	11.67	8.89	8.38
Pearson Coefficient	0.76	0.81	0.84

4.5 Results and evaluations

The performance of the proposed DoA index estimation method was evaluated using a random forest regression model. In this study, the Random Forest model was configured with the following hyperparameters: 100 trees to balance computational efficiency and predictive accuracy, a maximum tree depth of 25 to prevent overfitting, a minimum of 2 samples per split, a minimum of 1 sample per leaf, and the use of sqrt as the maximum number of features considered for each split. Bootstrap sampling was enabled to ensure ensemble diversity. Several key metrics were used to assess the model's accuracy and reliability, including R-squared (R^2), root mean square error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient. Together, these metrics provide a comprehensive understanding of how well the model predicts the depth of anesthesia in comparison to the BIS index.

Table 9 Results of the testing cases evaluation from UniSQ database and VitalDB database individually

UniSQ database	Patient number							VitalDB database	Patient number						
Evaluation methods	01	09	10	16	26	34	Average		03	45	46	48	74	106	Average
R^2	0.70	0.89	0.76	0.76	0.83	0.57	0.75		0.55	0.64	0.63	0.65	0.81	0.82	0.68
RSME	6.91	6.76	8.84	6.02	6.87	6.75	7.03		8.50	6.62	4.01	4.36	4.85	6.62	5.83
MAE	5.73	7.38	9.65	4.69	6.47	6.80	6.79		10.13	11.37	8.89	6.65	6.09	6.81	8.32
Pearson coefficient	0.84	0.94	0.87	0.87	0.91	0.75	0.86		0.74	0.80	0.80	0.80	0.90	0.90	0.82

R^2 reflects how much of the variance in the BIS index is explained by the proposed DoA index, with higher values indicating a better fit. The RMSE and MAE represent the average prediction errors, with lower values indicating more accurate predictions. Lastly, the Pearson correlation coefficient measures the linear relationship between the predicted DoA index and the BIS index, where values closer to 1 indicate stronger agreement.

As shown in Table 9, the model performed well for the UniSQ database, achieving an average R^2 of 0.75, an RMSE of 7.03, an MAE of 6.79, and a Pearson correlation coefficient of 0.86. These values suggest that the proposed DoA index closely matches the BIS index, with the model explaining 75% of the variation in the BIS values and showing a relatively low prediction error. Similarly, for the VitalDB database, the model yielded an average R^2 of 0.68, an RMSE of 5.83, an MAE of 8.32, and a Pearson correlation coefficient of 0.82, further demonstrating the method's effectiveness across different patient datasets.

When combining the UniSQ and VitalDB datasets (Table 10), the model maintained strong performance, with an average R^2 of 0.70, an RMSE of 6.31, an MAE of 8.38, and a Pearson correlation coefficient of 0.84, A Pearson correlation coefficient of 0.84 for the combined dataset indicates that there is a high level of agreement between the predicted DoA values and the actual BIS values, which suggests that the model generalizes effectively across different clinical conditions and patient populations.

To further assess the agreement between the proposed DoA index and the BIS index, a Bland–Altman analysis was conducted (Fig. 5). The Bland–Altman plot shows that 94.51% of the differences between the two indices lie within the limits of agreement ($diff \pm 2sd$), confirming a high level of agreement. The mean bias was − 0.50, with limits of agreement at 19.95 and -20.94, supporting the robustness of the proposed method in estimating the depth of anesthesia.

Finally, scatter plots with 95% confidence intervals (Fig. 6) illustrate the relationship between the proposed DoA index and the BIS index across the UniSQ, VitalDB, and combined datasets. The R^2 values for the datasets ranged from 0.65 to 0.66, indicating that the proposed DoA index accounts for a substantial portion of the variance in the BIS index, particularly within the 30–60 range (anesthetic states), where precision was highest.

4.6 DoA states and transition prediction

For each database (UniSQ and VitalDB), the classification results revealed clear patterns corresponding to different anesthesia states. The transitions between these states were further validated by comparing the identified states with anesthesia agent usage information and BIS index values, as demonstrated in Figs. 7, 8 and 9.

In Figs. 7, 8 and 9, the vertical lines indicate the administration times of anesthetic agents such as alfentanil and propofol, marking the onset and offset of sevoflurane. These markers provide a temporal reference, showing the correlation between the administration of anesthetics and the observed DoA state transitions. Notably, the transition from awake to anesthetic states aligns with the timing of propofol and sevoflurane administration, as indicated in the figures.

Figure 9 highlights a specific transition from deep anesthesia to moderate anesthesia around the 500-s mark. This change shows a stronger correlation between the detected states and the proposed DoA index compared to the BIS index, suggesting that the proposed method may provide a more accurate reflection of anesthesia state transitions.

Overall, these results demonstrate the effectiveness of the proposed DoA index in identifying distinct anesthesia states and capturing state transitions in real-time, with high correspondence to clinical anesthesia events.

4.7 Comparison with state-of-the-art

The Table 11 below compares the proposed method with several state-of-the-art approaches for Depth of Anesthesia (DoA) estimation based on EEG data primarily from the frontal cortex, using two accessible databases. The three state-of-the-art approaches were applied to the same training data and testing data as the proposed method.

Compared to Shi et al. [13], our proposed method achieves a comparable correlation coefficient (CC) using a random sampling strategy, but with only one EEG

Table 10 Results of the testing evaluations based on the training model of combined UniSQ and VitalDB databases

Patient number														
Evaluation methods	**01**	**09**	**10**	**16**	**26**	**34**	**Average**	**03**	**45**	**46**	**48**	**74**	**106**	**Average**
R^2	0.69	0.87	0.58	0.77	0.68	0.53	0.69	0.56	0.72	0.67	0.60	0.83	0.83	0.70
RSME	7.03	6.26	8.43	5.05	6.01	7.16	6.66	9.00	6.20	5.00	6.32	4.71	6.62	6.31
MAE	5.93	10.12	12.39	4.87	7.59	6.46	7.89	10.66	10.51	9.08	7.37	5.82	6.84	8.38
Pearson Coefficient	0.83	0.93	0.76	0.88	0.82	0.73	0.83	0.75	0.85	0.82	0.78	0.91	0.91	0.84

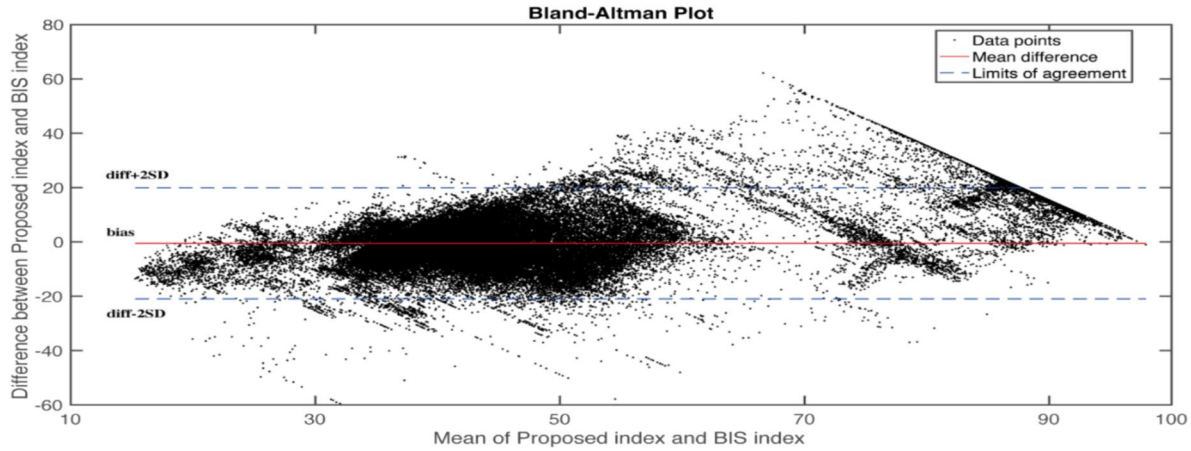

Fig. 5 Bland–Altman plot between the proposed DoA index against the BIS index with 95% limits of agreement

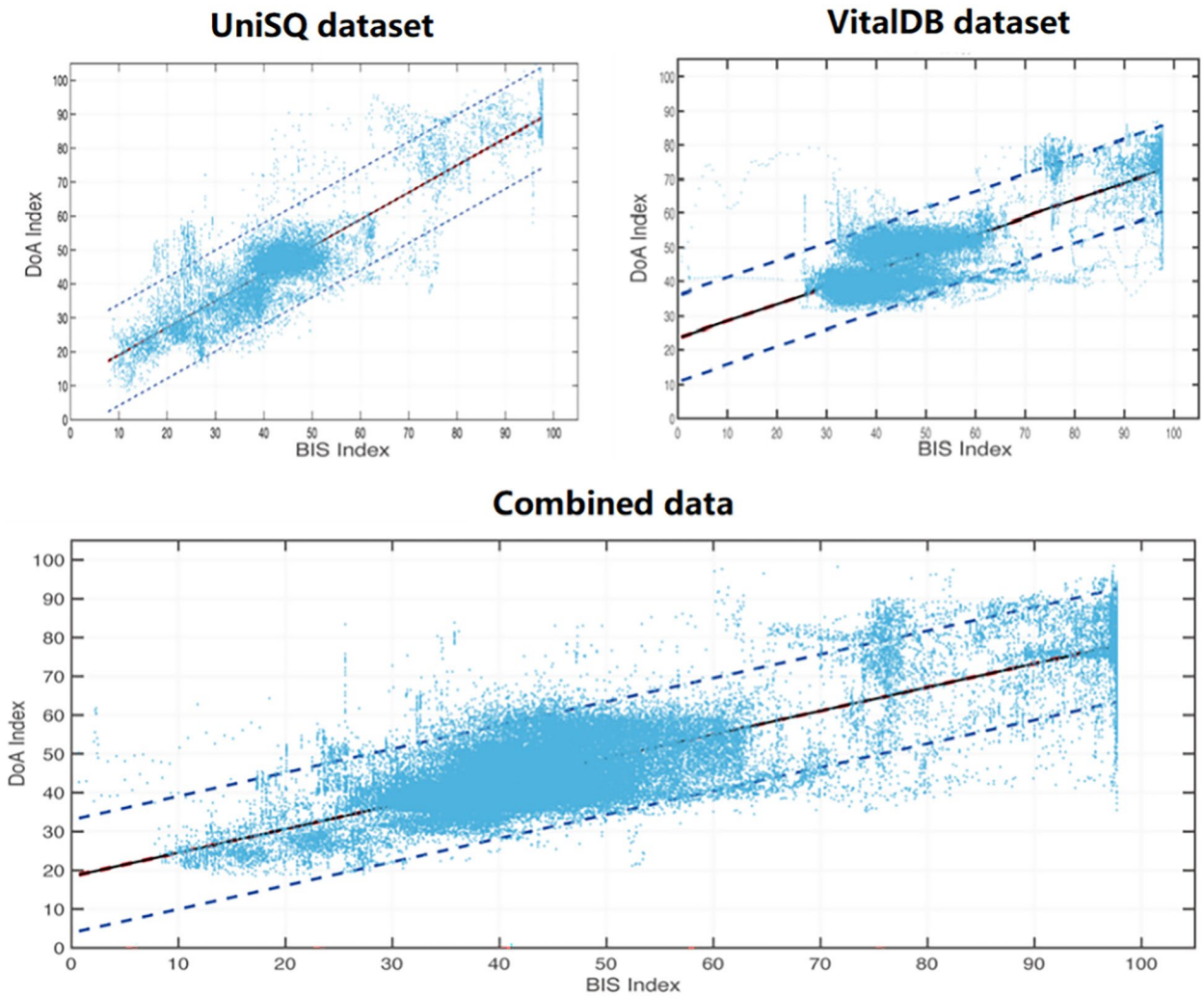

Fig. 6 The scatter plot with 95% confidence interval demonstrated the proposed DoA index against the BIS index

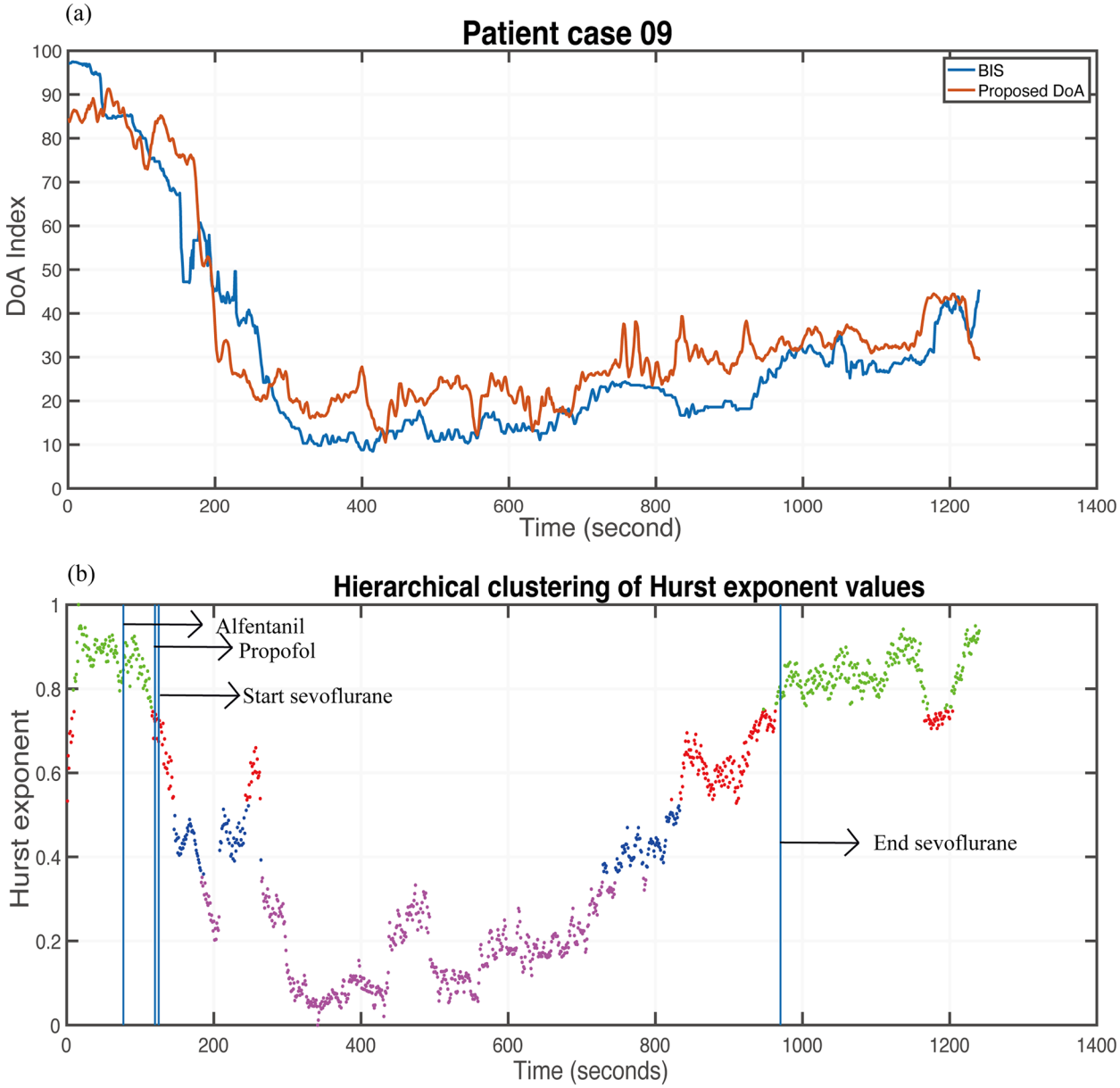

Fig. 7 **a** Comparison between the BIS index and the proposed DoA index for the patient case number 9. **b** Anaesthetic states classification using a hierarchical clustering method for patient case number 9

channel instead of four. This significantly reduces wearable complexity while maintaining accuracy. Although Chen et al. [11] achieved a slightly lower CC, they employed a more complex LSTM model on a single-channel dataset. Our simpler Random Forest model demonstrates competitive performance with fewer computational demands.

Our method also yields a higher mean CC compared to Shahbakhti et al. [14], despite both using random sampling as the evaluation strategy. Additionally, our approach leverages Wavelet decomposition, which is computationally simpler compared to techniques like SWT and shows improved accuracy with a lower MAE.

While some approaches, such as Shi et al. [13], employed more complex EEG decomposition techniques like WT-CEEMDAN-ICA, our method maintains a balance between performance and computational simplicity by using DWT with PLZC and PSD features, achieving competitive accuracy across multiple datasets.

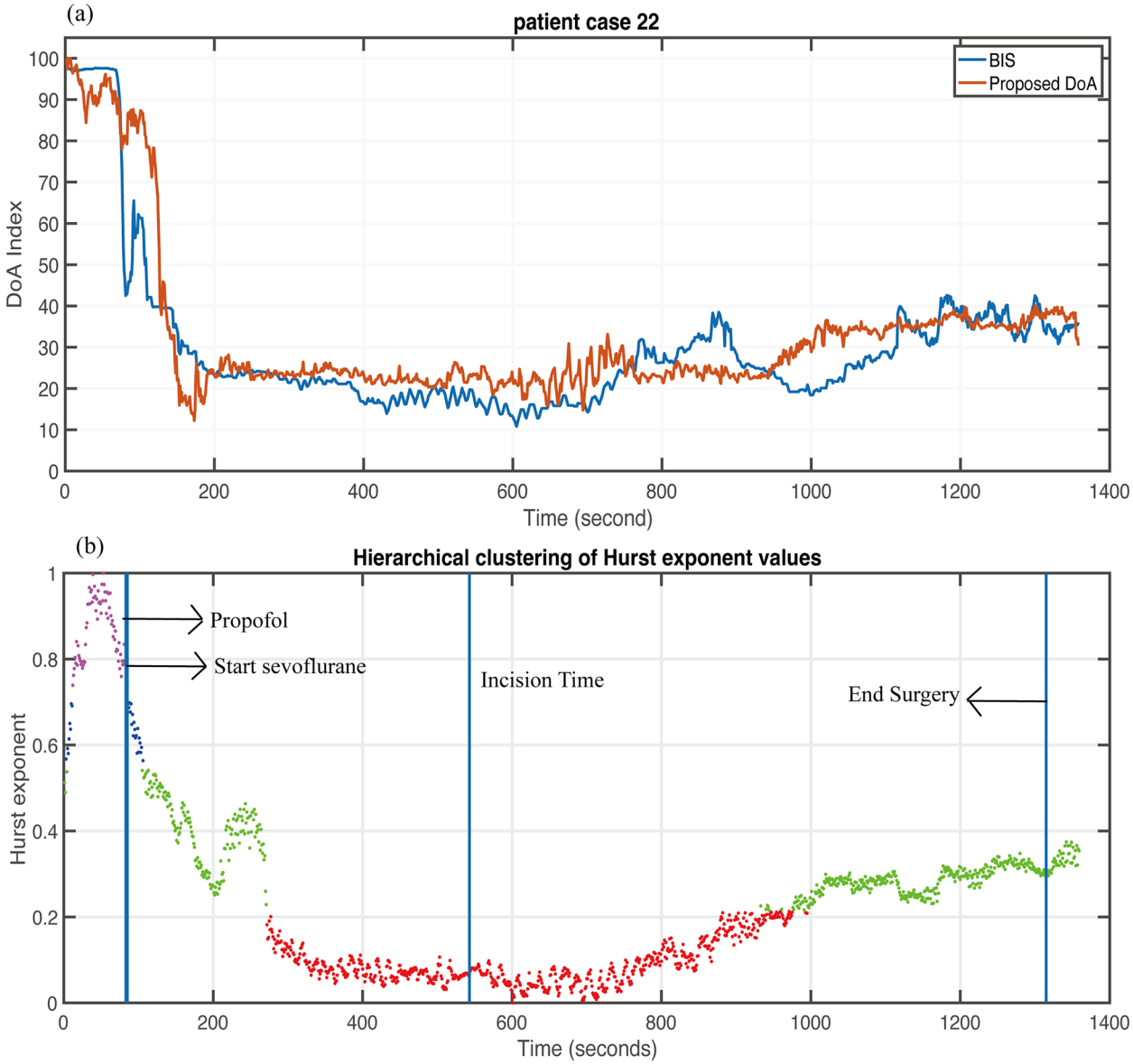

Fig. 8 **a** Comparison between the BIS index and the proposed DoA index for the patient case number 26. **b** Anaesthetic states classification using the hierarchical clustering method for patient case number 26

Furthermore, our approach has been evaluated using two independent databases, which enhances the generalizability of our findings. Unlike several other studies that were limited to a single dataset, our broader evaluation framework demonstrates consistent performance across varying conditions. This robustness underscores the applicability of our method to real-world scenarios, making it a reliable choice for practical deployment in diverse environments.

5 Discussion

In addressing the critical challenge of DoA estimation, this study delineates several key findings that collectively underscore its novelty and contribute substantively to the field. First, the dual-database strategy employed here, which utilizes both the UniSQ and VitalDB datasets, provides a model that has demonstrated robustness and generalizability across diverse patient demographics and anesthetic protocols—a notable advancement over studies limited to single, proprietary datasets [41–43]. The model's performance metrics, with an R^2 value of 0.70 and a Pearson correlation coefficient of 0.84, signify

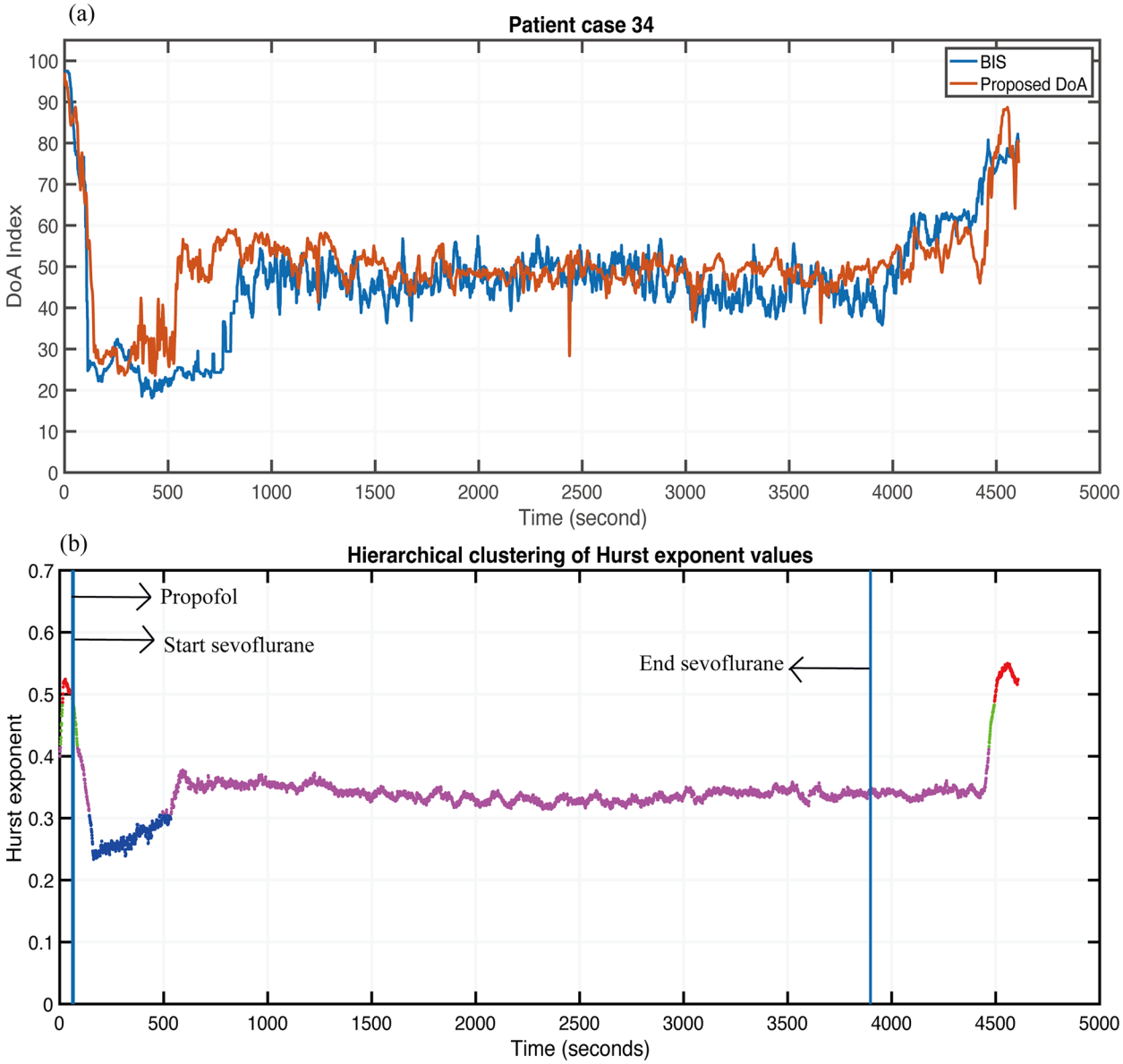

Fig. 9 **a** Comparison between the BIS index and the proposed DoA index for the patient case number 34. **b** Anaesthetic states classification using the hierarchical clustering method for patient case number 34

Table 11 Comparison between the proposed and state-of-the-art algorithms

Study	EEG Decomposition	Feature	Model	Training–testing strategy	Correlation (CC)	MAE	RMAE
Shi et al. [13] [Deep Residual Shrinkage Network]	WT-CEEMDAN-ICA	SampEn, SEF, PSI	Deep Residual Shrinkage	Cross-Subject Validation	0.84	10.51	6.15
Chen et al. [11] [Nonlinear EEGV Analysis]	None	SampEn, PeEn, DFA, PoP	LSTM	Train-Test Split	0.78	8.78	9.76
Shahbakhti et al. [14] [Nonparametric Feature Set]	SWT	Entropy, Band Power	Random Forest	Random Sampling	0.83	11.32	8.45
This study	Wavelet (DWT)	PLZC, PSD	Random Forest	Random Sampling	0.84	8.38	6.31

a substantial improvement over the long short-term memory (LSTM) networks that traditionally rely on BIS, which is frequently cited as the gold standard for DoA assessment [11]. Moreover, the integration of the Hurst exponent as a predictive variable into the DoA estimation model offers a novel perspective that enriches the multidimensional analysis of anesthesia dynamics [44]. These innovations not only bridge gaps in the current research landscape but also pave the way for future investigations geared toward enhancing patient safety and the management of anesthesia.

The substantial correlation with BIS scores validates the hypothesis that machine learning models when equipped with appropriate feature sets and trained on diverse datasets, can effectively estimate the depth of anesthesia. Acknowledging the inherent limitations of the BIS system, including its susceptibility to artifacts and inter-individual variability, the present study utilizes BIS as a comparative standard due to its widespread clinical adoption. The findings demonstrate that the proposed model maintains a high degree of correlation with BIS scores, bolstering its potential as a reliable DoA estimation tool within clinical settings. The robustness of the model's performance, coupled with its validation against a widely accepted benchmark, underscores its promise for future clinical deployment. To fortify these results, subsequent research will focus on corroborating the model's efficacy with clinician-labeled data, thereby reinforcing the model's practical relevance.

5.1 Significance of DWT levels, pre-processing, feature determination, and mother wavelet selection

The decision to examine various levels of the DWT in this investigation was driven by the necessity to ascertain the most appropriate level of decompositions for extracting features from the EEG signals. The DWT decomposes a signal into distinct frequency sub-bands, with each level of the decomposition focusing on specific frequency ranges. Identifying the optimal level of a decomposition ensures that the features extracted are relevant to the anesthesia depth, thereby contributing to the development of a novel and accurate DoA index.

A comprehensive analysis in this study determined the optimal level of decompositions and the most suitable mother wavelet for each dataset, which significantly impacted the accuracy and reliability of the DoA estimation model. This process ensured that the selected features, PLZC and PSD, were accurately differentiated between various anesthesia states. Such meticulous consideration provided a strong foundation for the development of a dependable and precise DoA estimation model using the proposed index.

Pre-processing and feature determination play a pivotal role in enhancing the quality of the raw EEG signals and selecting the most relevant features for the DoA estimation. The wavelet threshold method and DWT were utilized to eliminate noise and facilitate feature extraction. The pre-processing techniques improved the clarity and quality of the EEG signals, enabling more an accurate feature extraction, which is essential for a reliable DoA estimation model.

5.2 The size of the sliding window

The sliding window technique was implemented to ensure an optimal real-time response during the depth of anesthesia monitoring. After experimented with various window sizes, a window size of 56 s and an overlap of 55 s between two adjacent windows were selected. With this approach, the DoA index could be updated every second, significantly reducing the time delay for monitoring the DoA.

Smaller window sizes, such as 10 s and 30 s, may provide faster updates but may also compromise the quality of the feature extraction, particularly during transitions between anesthesia states. This is due to the limited amount of data available for smaller windows, which might not capture the dynamics of the EEG signals sufficiently. The chosen sliding window size of 56 s achieved an optimal balance between the real-time performance and accurate feature extraction. It allows for continuous updates of the DoA index while maintaining a high level of accuracy in differentiating between various anesthesia states. This choice contributes to the overall effectiveness and reliability of the proposed DoA estimation method.

5.3 Significance of using UniSQ and VitalDB databases and the combined datasets

The employment of two independent databases, UniSQ and VitalDB, allowed for a comprehensive evaluation of the proposed method. The results obtained from the datasets provided insights into the robustness and generalizability of the method across diverse patient populations and anesthesia protocols. By evaluating the performance of the proposed method on individual datasets and then on a combined dataset, the study could demonstrate its effectiveness in a broader context.

The experimental results indicated that the proposed method performed well on both individual databases and the combined datasets. When comparing the evaluation metrics between the individual databases, the UniSQ database had slightly higher R^2 and Pearson correlation coefficient values, while the VitalDB database showed a lower RMSE value. These differences could be attributed to variations in patient populations and anesthesia protocols. The combined datasets can be considered a valuable

approach to further validate the method's performance and ensure its applicability across different patient populations and clinical scenarios. The overall performance remains strong, supporting the robustness and generalizability of the proposed method.

5.4 Insights on anaesthesia states transitions

The examination of anaesthesia state transitions offered valuable insights into the behaviour of the proposed method in response to changing DoA conditions, as illustrated in Figs. 7, 8 and 9. The Hurst exponent values calculated for each database served as the input features for the hierarchical clustering algorithm, with the resultant clusters representing different anesthesia states such as awake, light anesthesia, moderate anesthesia, and deep anesthesia.

The vertical lines in Figs. 7, 8 and 9, representing the administration times of anesthetic agents, provided a visual guide linking the administration of specific anesthetics with the observed DoA state transitions. Notably, the start time of the anesthetic agents aligned with the anaesthesia states transition time from awake to anaesthesia state, indicating the capability of the proposed method to accurately track these transitions in real time.

Furthermore, in Fig. 9, the transition from deep anaesthesia to middle anaesthesia around the 500th-second mark demonstrated a stronger correlation with the detected states and the proposed DoA index than the BIS index recorded during the surgery. This finding supports the effectiveness of the proposed method in accurately identifying different anesthesia states and responding to transition events, lending further credence to the robustness and utility of the proposed DoA estimation method in real-world clinical scenarios.

Although the ground truth data from attending anaesthetists was not available for direct comparison, the agreement between the identified states, anesthesia agent usage information, and BIS index values indicate the reliability of the proposed method. The observed patterns in the classification results contribute to a better understanding of the depth of anesthesia and the transitions between different states during surgical procedures. These findings demonstrate the potential of the proposed method in providing a more comprehensive assessment of the depth of anesthesia using EEG signals.

5.5 Methodological efficacy

This research has examined the efficacy of the chosen analytical methods through a comprehensive ablation analysis. Such an examination is critical for affirming the methodological underpinnings of the proposed DoA estimation model.

Permutation Entropy (PE) and Power Spectral Density (PSD) were selected for feature extraction after a careful evaluation of their capacity to encapsulate the intricate attributes of EEG signals under anesthetic influence. These methods demonstrated superior capability in distilling pertinent characteristics from the EEG signals, which are inherently non-linear and exhibit non-stationary behavior. While Fourier-based approaches were considered, their performance in capturing the transient dynamics of EEG during anesthesia was less compelling when contrasted with the nuanced fidelity afforded by PE and PSD.

The robustness of the random forest regression model was underscored by its performance amidst the multifaceted challenges posed by EEG data analysis, notably its inherent noise and dimensionality. The results from the ablation study have illuminated the robust performance of the random forest approach, particularly in contrast to linear regression and SVM, which were found to be less adept in this context.

The model's parameters were optimized through a grid search strategy, complemented by cross-validation to mitigate overfitting. This approach was pivotal in navigating the intricate balance between model complexity and generalizability, culminating in a model finely attuned to the complexities embedded within the EEG data.

The insights gleaned from the ablation study, alongside the methodical optimization of model parameters, have collectively validated the chosen features and regression methodology. The corroboration of these methodological choices bolsters the proposed model's potential for clinical deployment. It also acts as a harbinger for subsequent research endeavors aimed at enhancing the precision and reliability of DoA monitoring tools.

In this investigation, the validation of machine learning models was rigorously addressed, with an 80:20 train-test partitioning ratio chosen to ensure robust training and meaningful validation, a standard practice for datasets of this scale and complexity. To ascertain model stability and generalizability, a k-fold cross-validation approach, ranging from 5 to 10 folds, was adopted in Table 12. The resulting R^2 and RMSE metrics across varying folds demonstrated the model's consistent performance, affirming its resilience against data partitioning discrepancies and aligning with validation methodologies documented in the existing scientific literature.

The clinical relevance of our study is highlighted by the ability to offer a more adaptive and precise tool for DoA monitoring. By leveraging a novel combination of PLZC and PSD for feature extraction, our method demonstrates a high correlation with standard clinical assessments, addressing individual patient variability with notable accuracy. The validation against two diverse

Table 12 K-fold cross-validation results

Training models	k = 5		k = 6		k = 7		k = 8		k = 9		k = 10	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Linear	0.52	11.08	0.51	11.10	0.52	11.08	0.52	11.08	0.53	11.06	0.52	11.08
SVM	0.73	6.61	0.72	6.69	0.74	6.52	0.72	6.69	0.70	7.14	0.70	7.14
Gaussian Process Regression Models	0.79	5.66	0.79	5.66	0.79	5.66	0.80	5.62	0.79	5.66	0.79	5.66
Bagged Trees	0.82	5.25	0.82	5.25	0.81	5.31	0.82	5.25	0.81	5.31	0.82	5.25
Wide Neural Network	0.78	5.71	0.78	5.71	0.78	5.71	0.78	5.71	0.78	5.71	0.78	5.71

datasets—UniSQ and VitalDB—not only solidifies the method's reliability but also its applicability to various clinical environments. These results represent a critical advancement toward personalized anesthetic management, potentially reducing the risk of anesthesia-related adverse events and aligning with the evolving paradigm of precision medicine.

5.6 Limitation of study

While the proposed approach for DoA estimation using PLZC and PSD along with random forest regression has shown promise, it is important to acknowledge certain limitations. The sensitivity of these methods to individual patient differences, such as age and gender, may affect the accuracy of the DoA estimation. Computational complexity is another concern, especially when processing large EEG datasets, which may pose challenges for real-time applications. Additionally, the generalizability of the proposed method across diverse patient populations remains to be rigorously tested, particularly given the proprietary nature of most anesthesia datasets which limits access to a broader range of data for validation. Moreover, the exploration of advanced deep learning techniques and the integration of clinician-labeled data are identified as future research directions to enhance model validation and address current limitations. Future research should focus on addressing these limitations by refining the computational efficiency of feature extraction techniques and enhancing the method's adaptability to different patient demographics.

6 Conclusion

The present study introduces a methodological framework for estimating the Depth of Anesthesia via electroencephalographic signals, thereby addressing existing gaps in literature. Utilizing a series of computational stages—including data pre-processing, Discrete Wavelet Transform decomposition, feature extraction, and Random Forest Regression—the research delineates a comprehensive methodology substantiated through rigorous empirical validation across two divergent databases: UniSQ and VitalDB. Such an amalgamation of data sources enhances the robustness and generalizability of the model across disparate patient populations and anesthetic protocols.

In the conclusion of the study, the results indicate that the developed model demonstrates strong performance, marked by a high degree of correlation with BIS and robustness across varied datasets. The model achieves an R^2 value of 0.70 and a Pearson correlation coefficient of 0.84, indicating substantial improvements in depth of anesthesia assessment compared to established LSTM networks [18]. In reference to Tables 1 and 2, although the correlation figures do not surpass all previously reported results [12], the model's lower computational intensity presents a practical advantage, particularly in clinical applications where efficiency and accuracy are both critical.

In line with advanced developments in deep learning, our future research will pivot towards employing methodologies such as transformers and attention mechanisms in EEG anesthesia monitoring. Transformers, which have revolutionized sequence modeling due to their unique self-attention capabilities [45], present a promising avenue for enhancing DoA estimation. Their application in interpreting EEG patterns can provide unparalleled precision, as evidenced in recent studies [46]. Similarly, attention learning, known for isolating pivotal features within complex data sequences [47], can significantly refine EEG signal interpretation. Incorporating these advanced models aligns with the evolving landscape of AI in medical diagnostics [48], particularly for personalized anesthetic care where patient-specific factors are paramount.

Acknowledgements
The authors would like to acknowledge the use of the datasets collected at Toowoomba St Vincent's Hospital in this study.

Author contributions
Conceptualization, TL, YH; methodology, YH; software, TL,YH; validation, TL, YH; formal analysis, TL and YH; investigation, YH; resources, TL, YH; data curation, TL, YH; writing—original draft preparation, YH; writing—review and editing, TL,

YH, PW and YL; visualization, YH; supervision, PW and YL. All authors have read and agreed to the published version of the manuscript.

Funding
Not applicable.

Availability of data and materials
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 17 June 2024 Accepted: 31 October 2024
Published online: 21 November 2024

References

1. Khan FH, Ashraf U, Altaf MAB, Saadeh W (2018) A patient-specific machine learning based EEG processor for accurate estimation of depth of anesthesia 2018 IEEE biomedical circuits and systems conference (BioCAS). IEEE, pp. 1–4
2. Sebel PS, Bowdle TA, Ghoneim MM, Rampil IJ, Padilla RE, Gan TJ et al (2004) The incidence of awareness during anesthesia: a multicenter United States study. Anesth Analg 99(3):833–839
3. Samuelsson P, Brudin L, Sandin RH (2007) Late psychological symptoms after awareness among consecutively included surgical patients. J Am Soc Anesthesiol 106(1):26–32
4. Ji X, Li Y, Wen P (2022) Jumping knowledge based spatial-temporal graph convolutional networks for automatic sleep stage classification. IEEE Trans Neural Syst Rehabil Eng 30:1464–1472
5. Supakar R, Satvaya P, Chakrabarti P (2022) A deep learning based model using RNN-LSTM for the detection of schizophrenia from EEG data. Comput Biol Med 151:106225
6. Shen M, Wen P, Song B, Li Y (2023) Real-time epilepsy seizure detection based on EEG using tunable-Q wavelet transform and convolutional neural network. Biomed Signal Process Control 82:104566
7. Torres JMM, Clarkson T, Hauschild KM, Luhmann CC, Lerner MD, Riccardi G (2022) Facial emotions are accurately encoded in the neural signal of those with autism spectrum disorder: a deep learning approach. Biol Psychiatry 7(7):688–695
8. Huang Y, Wen P, Song B, Li Y (2022) Real-time depth of anaesthesia assessment based on hybrid statistical features of EEG. Sensors 22(16):6099
9. Schmierer T, Li T, Li Y (2022) A novel empirical wavelet SODP and spectral entropy based index for assessing the depth of anaesthesia. Health Inf Sci Syst 10(1):10
10. Alsafy I, Diykh M (2022) Developing a robust model to predict depth of anesthesia from single channel EEG signal. Phys Eng Sci Med 45(3):793–808
11. Chen Y-F, Fan S-Z, Abbod MF, Shieh J-S, Zhang M (2022) Nonlinear analysis of electroencephalogram variability as a measure of the depth of anesthesia. IEEE Trans Instrum Meas 71:1–13
12. Lee M, Sanz LR, Barra A, Wolff A, Nieminen JO, Boly M et al (2022) Quantifying arousal and awareness in altered states of consciousness using interpretable deep learning. Nat Commun 13(1):1064
13. Shi M, Huang Z, Xiao G, Xu B, Ren Q, Zhao H (2023) Estimating the depth of anesthesia from EEG signals based on a deep residual shrinkage network. Sensors 23(2):1008
14. Shahbakhti M, Krycinska R, Beiramvand M, Hakimi N, Lipping T et al (2024) Wearable EEG-based depth of anesthesia monitoring: a non-parametric feature set. IEEE Sens J 24(11):18098–18107
15. Casey CP, Tanabe S, Farahbakhsh Z, Parker M, Bo A, White M et al (2022) Distinct EEG signatures differentiate unconsciousness and disconnection during anaesthesia and sleep. Br J Anaesth 128(6):1006–1018
16. Xiao Z, Xu Z, Ma L (2022) Construction of complex brain network based on EEG signals and evaluation of general anesthesia status. In: Xiao Z (ed) International conferencee on intelligent robotics and applications. Springer, Cham
17. Anand RV, Abbod MF, Fan S-Z, Shieh J-S (2023) Depth analysis of anesthesia using eeg signals via time series feature extraction and machine learning. Sci 5(2):19
18. Wang Q, Liu F, Wan G, Chen Y (2022) Inference of brain states under anesthesia with meta learning based deep learning models. IEEE Trans Neural Syst Rehabil Eng 30:1081–1091
19. Dutt MI, Saadeh W (2022) An accurate EEG-based deep learning classifier for monitoring depth of anesthesia. 2022 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, pp. 1768–1772.
20. Dutt MI, Saadeh W (2023) Monitoring level of hypnosis using stationary wavelet transform and singular value decomposition entropy with feedforward neural network. IEEE Trans Neural Syst Rehabil Eng. https://doi.org/10.1109/TNSRE.2023.3264797
21. Zhang H, Wu H, Chen Q, Xia Y (2024) FEHANet: A frequency enhanced Hybrid attention network for bispectral index score estimation. Biomed Signal Process Control 95:106431
22. Bai Y, Liang Z, Li X, Voss LJ, Sleigh JW (2015) Permutation Lempel-Ziv complexity measure of electroencephalogram in GABAergic anaesthetics. Physiol Meas 36(12):2483
23. Tosun PD, Abásolo D, Stenson G, Winsky-Sommerer R (2017) Characterisation of the effects of sleep deprivation on the electroencephalogram using permutation Lempel-Ziv complexity, a non-linear analysis tool. Entropy 19(12):673
24. Liang Z, Shao S, Lv Z, Li D, Sleigh JW, Li X et al (2020) Constructing a consciousness meter based on the combination of non-linear measurements and genetic algorithm-based support vector machine. IEEE Trans Neural Syst Rehabil Eng 28(2):399–408
25. Purdon PL, Sampson A, Pavone KJ, Brown EN (2015) Clinical electroencephalography for anesthesiologists: part I: background and basic signatures. Anesthesiology 123(4):937–960
26. Otto KA (2008) EEG power spectrum analysis for monitoring depth of anaesthesia during experimental surgery. Lab Anim 42(1):45–61
27. Liu Q, Cai J, Fan S-Z, Abbod MF, Shieh J-S, Kung Y et al (2019) Spectrum analysis of EEG signals using CNN to model patient's consciousness level based on anesthesiologists' experience. IEEE Access 7:53731–53742
28. Eagleman SL, Vaughn DA, Drover DR, Drover CM, Cohen MS, Ouellette NT et al (2018) Do complexity measures of frontal EEG distinguish loss of consciousness in geriatric patients under anesthesia? Front Neurosci 12:645
29. Obert DP, Schweizer C, Zinn S, Kratzer S, Hight D, Sleigh J et al (2021) The influence of age on EEG-based anaesthesia indices. J Clin Anesth 73:110325
30. Diykh M, Li Y, Abdulla S (2020) EEG sleep stages identification based on weighted undirected complex networks. Comput Methods Programs Biomed 184:105116
31. Siuly, Li Y, Wen P (2013) Identification of motor imagery tasks through CC–LR algorithm in brain computer interface. Int J Bioinf Res Appl 9(2):156–172
32. Siuly S, Li Y (2015) Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach. Neural Comput Appl 26:799–811
33. Saadeh W, Khan FH, Altaf MAB (2019) Design and implementation of a machine learning based EEG processor for accurate estimation of depth of anesthesia. IEEE Trans Biomed Circuits Syst 13(4):658–669
34. Vetterli M, Herley C (1992) Wavelets and filter banks: theory and design. IEEE Trans Signal Process 40:2207–2232
35. Mallat S (1999) A wavelet tour of signal processing. Elsevier, Amsterdam
36. Nguyen-Ky T, Wen P, Li Y, Gray R (2011) Measuring and reflecting depth of anesthesia using wavelet and power spectral density. IEEE Trans Inf Technol Biomed 15(4):630–639

37. Eriksson A, Stoica P, Soderstrom T (1994) Markov-based eigenanalysis method for frequency estimation. IEEE Trans Signal Process 42(3):586–594
38. Nguyen-Ky T, Wen P, Li Y (2014) Monitoring the depth of anaesthesia using Hurst exponent and Bayesian methods. IET Signal Proc 8(9):907–917
39. Lee H-C, Ryu H-G, Park Y, Yoon SB, Yang SM, Oh H-W et al (2019) Data driven investigation of bispectral index algorithm. Sci Rep 9(1):13769
40. Lee H-C, Park Y, Yoon SB, Yang SM, Park D, Jung C-W (2022) VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. Scientific Data 9(1):279
41. Zanner R, Schneider G, Meyer A, Kochs E, Kreuzer M (2021) Time delay of the qCON monitor and its performance during state transitions. J Clin Monit Comput 35:379–386
42. Wang G, Liu Z, Feng Y, Li J, Dong H, Wang D et al (2019) Monitoring the depth of anesthesia through the use of cerebral hemodynamic measurements based on sample entropy algorithm. IEEE Trans Biomed Eng 67(3):807–816
43. Zakaria L, Desowska A, Berde CB, Cornelissen L (2023) Electroencephalographic delta and alpha oscillations reveal phase-amplitude coupling in paediatric patients undergoing sevoflurane-based general anaesthesia. Br J Anaesth 130(5):595–602
44. Lee H, Noh GJ, Joo P, Choi BM, Silverstein BH, Kim M et al (2017) Diversity of functional connectivity patterns is reduced in propofol-induced unconsciousness. Hum Brain Mapp 38(10):4980–4995
45. Zhang H, Goodfellow I, Metaxas D, Odena A (2019) 'Self-attention generative adversarial networks. Int Conf Mach Learning PMLR 97:7354–7363
46. Bahador N, Jokelainen J, Mustola S, Kortelainen J (2021) Multimodal spatio-temporal-spectral fusion for deep learning applications in physiological time series processing: a case study in monitoring the depth of anesthesia. Information Fusion 73:125–143
47. Zhang D, Yao L, Chen K, Monaghan J (2019) A convolutional recurrent attention model for subject-independent EEG signal analysis. IEEE Signal Process Lett 26(5):715–719
48. Tveit J, Aurlien H, Plis S, Calhoun VD, Tatum WO, Schomer DL et al (2023) Automated interpretation of clinical electroencephalograms using artificial intelligence. JAMA Neurol. https://doi.org/10.1001/jamaneurol.2023.1645

Publisher's Note