Heart Rate Variability-Derived Features Based on Deep Neural Networks for Monitoring Depth of Anaesthesia

Jian Zhan¹,², Zuo-xi Wu¹, Zhen-xin Duan¹, Gui-ying Yang¹, Zhi-yong Du¹, Xiao-hang Bao and Hong Li¹, *

* Correspondence: lh78553@163.com

1 Department of Anaesthesiology, Second Affiliated Hospital of Army Medical University, Chongqing 400037, China
2 Department of Anaesthesiology, Affiliated Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China

Abstract

Background: Estimating the depth of anaesthesia (DoA) is critical in clinical anaesthesiology. Electroencephalograms (EEGs) have been widely used for monitoring the DoA; however, they may be inaccurate under certain conditions.

Methods: In this study, we propose a novel method to evaluate the DoA based on multiple heart rate variability (HRV)-derived features combined with a discrete wavelet transform and deep neural networks (DNNs). Four features were extracted from an electrocardiogram, including the HRV high-frequency power, low-frequency power, high-to-low-frequency power ratio, and sample entropy. Next, these features were used as inputs for the DNN, which used the expert assessment of consciousness level as the
reference output. Finally, the DNN was compared with the logistic regression (LR), support vector machine (SVM), and decision tree (DT) models. The data of 23 anaesthesia patients were used to assess the proposed method.

Results: The results demonstrated that the accuracies of the four models, in distinguishing the anaesthesia states, were 86.2% (LR), 87.5% (SVM), 87.2% (DT), and 90.1% (DNN). Our method outperformed the LR, SVM, and DT methods.

Conclusions: The proposed method could accurately distinguish between different anaesthesia states, thus, providing an alternative or supplementary method to EEG monitoring for the assessment of DoA.

Keywords: depth of anaesthesia; heart rate variability; deep neural network; discrete wavelet transform
Background

General anaesthesia is a reversible state of anaesthetic drug-induced loss of consciousness during surgery [1]. Current research indicates that both the central nervous and autonomic systems are related to the depth of anaesthesia (DoA) of a patient [2]. Excessively deep or shallow DoA can result in harmful complications for patients; a DoA that is too shallow increases the risk of intraoperative awareness [3], and a DoA that is too deep can cause delayed recovery [4], cognitive dysfunction, and even death [5]. Therefore, accurate DoA monitoring is crucial to reduce the complications associated with overdose or insufficiency of anaesthetics and guarantee the safety and quality of anaesthesia.

However, the mechanisms of action of general anaesthetics are still not completely understood [6, 7], and there is currently no ‘gold standard’ for evaluating DoA [1]. In recent years, a variety of DoA monitors based on electroencephalograms (EEG) signals, such as bispectral index (BIS), Narcotrend, and entropy have been widely used during surgery [8-10]. However, EEG signals only show the functions of the central nervous system and the indices based on these signals are not sufficiently accurate to indicate a precise DoA [11, 12]. Moreover, there are various shortcomings related to this approach. First, these indices were developed using adult patients, and are unsuitable for infants and children. Second, these indices are not recommended for use with some general anaesthetics, such as ketamine and nitrogen dioxide [13]. In addition, EEG signals are subject to interference originating from the noise of the medical equipment in the operating room. Therefore, it is essential to seek new methods of DoA monitoring to overcome the drawbacks of mainstream methods based on EEG signals [14] and improve DoA monitoring accuracy.

Electrocardiograms (ECGs) are internationally used in standard monitoring during general anaesthesia [15] and provide important physiological signals, such as heart rate.
(HR), blood pressure (BP), and peripheral oxygen saturation (SpO₂), which are used to assist anaesthesiologists in evaluating the DoA. In addition, there are certain advantages in the use of ECGs. On the one hand, ECG signals are more stable and less susceptible to noise than EEG signals. On the other hand, in comparison with the EEG, the electrode sensors used for ECG signal acquisition are cheaper, rendering ECG a more cost-effective method. More importantly, ECG signals are closely related to DoA. During general anaesthesia, different anaesthetic drugs affect ECG signals[16, 17]. Previous studies have found that heart rate variability (HRV) derived from an ECG is regulated by the central nervous and autonomic systems, and closely related to the DoA during surgery[18-20]. Therefore, HRV may be used as an alternative or important supplementary method of EEG monitoring in terms of DoA evaluation [13, 21].

Owing to the strong nonlinear characteristics of the EEG and ECG, nonlinear analysis methods may be used in studies of anaesthesia[22, 23]. Sample entropy (SampEn) is a typical nonlinear analysis method that was developed to study HRV [24, 25] and provide an improved assessment of DoA during surgery[26, 27]. In addition, three features of HRV, including the high-frequency power(HF), low-frequency power(LF), and ratio of high-to-low-frequency power(HF/LF), are related to the autonomic nervous system and have been implemented in anaesthesia research[28, 29].

As a result of the complicated changes in patient vital signals during different anaesthesia states, it is necessary to use multiple physiological features to evaluate DoA. Recently, several studies based on multiple EEG features have been conducted to assess DoA. Ferreira et al. used multiple features of the blink reflex frequency domain to perform a multi-class logistic regression (LR) to predict DoA in eleven patients who were subjected to propofol anaesthesia[30]. In the work of Shalbaf et al., a support vector machine (SVM)
with Shannon permutation entropy and frequency features was used to estimate the DoA in seventeen patients who were subject to sevoflurane anaesthesia[31]. Lee et al. used four EEG parameters, including the burst suppression ratio, power of electromyogram, 95% spectral edge frequency, and relative beta ratio to construct a deep decision tree (DT) to evaluate DoA[32]. Gu et al. implemented an artificial neural network (ANN) to integrate four EEG features to assess DoA[33]. These results indicate that it is necessary to combine multiple time and frequency features to improve DoA assessment methods. Moreover, Liu et al. evaluated DoA based on only the similarity and the distribution index of HRV combined with an ANN and obtained higher accuracy for DoA estimation[13]. Deep neural networks (DNNs) developed from ANNs have also been used to monitor DoA[34, 35]. Thus, we propose the hypothesis that multiple time and frequency features of HRV based on DNNs could improve DoA monitoring accuracy and provide an alternative or key supplementary method for EEG monitoring in the assessment of DoA.

**Methods**

This study protocol was approved by the Institutional Ethics Committee of the Second Affiliated Hospital of the Army Medical University (2018–029). Written informed consent was obtained from each patient. Twenty-three ASA physical status I or II adult patients, aged from 20 to 70 years old, scheduled to undergo elective abdominal surgery were recruited. Exclusion criteria included patients with neurological and cardiovascular diseases or a known allergy history of anaesthetics.

All patients underwent preoperative fasting for at least 8 h. After the electrodes were placed on the patient chest wall, anaesthesia was usually induced by intravenous midazolam, propofol, sufentanil, and cisatracurium. Sevoflurane together with propofol
and remifentanil were used to maintain anaesthesia. Additional drugs (such as sufentanil and atropine) were administrated when approaching the end of surgery. Table 1 summarises this information in detail. Physiological signals (such as ECG, BP, HR, and SpO$_2$) were measured to guarantee the safety of the patients under different anaesthesia states. The attending anaesthetist adjusted the DoA accordingly, using these observed signals and their own experience.

In this study, ECG signals were recorded from twenty-three adult patients under general anaesthesia. The signals were recorded using a Philips MP60 monitor (Intellivue; Philips, Foster City, CA, USA). The operation time was 1—3 h. Raw ECG data were sampled at a 500-Hz sampling frequency.

Expert Assessment of Consciousness Level

The EACL is the average value of the DoA assessment score determined by five experienced anaesthesiologists (i.e., attending physicians) based on clinical recordings and their own experience[26]. The states of general anaesthesia are classified as: anaesthesia induction, anaesthesia maintenance, and anaesthesia recovery. Anaesthesia induction means that the anaesthesia depth gradually increases, anaesthesia maintenance means that the anaesthesia depth is relatively stable, and anaesthesia recovery means that the anaesthesia depth gradually decreases. These three states represent the different states of anaesthesia depth. The obtained EACL value is a single number from 0–100, similar to the BIS (with 100 denoting ‘fully awake’ and 0 denoting ‘isoelectricity’). During surgery, the clinical information recorded included: (1) vital signs (e.g., HR, BP, SpO$_2$), (2) anaesthetic events, including induction, intubation and extubation of anaesthesia, addition of muscle relaxant drugs, and airway management, (3) surgical events, including the start and end of
the surgical procedure and the occurrence of noxious stimulus, (4) other clinical signs, including unusual responses, movement, and arousability under induction and recovery, and (5) any other related events, such as patient demography. In addition, as experienced anaesthesiologists have undergone a significant amount of training, their evaluation scores could be defined as a highly accurate examples of estimation for evaluating the DoA.

**ECG Preprocessing**

Body movements and medical device frequency noise are the main artifacts in ECG recordings. These artifacts seriously affect the analysis results of the ECG signals. Therefore, data preprocessing is essential for evaluating DoA, and can normalize and facilitate subsequent analysis. Preprocessing usually includes data format conversion, noise cancellation, and data rearrangement. In this study, the outliers beyond the threshold (85 ms) were first removed by comparing the current RR interval, the interval between R peaks in two adjacent heartbeats of the ECG, in the ECG signal with the average of the previous ten sampling points. The RR interval is shown in Fig.1. Second, a low-pass(16 Hz) filter and a high-pass(8 Hz)filter were used consecutively to remove the baseline drift and frequency noise from the medical devices. The filters do not interfere in the frequency information of the ECG signals. The frequency information (3-45 Hz) was retained for resampling. Third, the average method was used to remove the electromyogram (EMG) artifacts and transient high-amplitude artifacts (THA)[36]. The average method compares whether the average value of the current RR interval and the past 10 sampling points is less than the threshold and removes the point if it exceeds the threshold. Fourth, the ECG data was resampled to 1.67 Hz[37]. In addition, some sampling points were far beyond the normal biological voltage, meaning only the sampling points from -5 to 5 mV were retained. Finally, 153-s
epochs including 256 RR intervals were extracted from the artifact-free ECG. In this study, four kinds of features were used as the inputs of the DNN to evaluate the DoA. These include three frequency features and one SampEn feature. The HF, LF, and HF/LF ratio were calculated as the frequency domain features and SampEn of the RR interval as the entropy feature. In addition, four algorithms were applied to the ECG data processing. The processing flow of the ECG is depicted in Fig. 2.

Sample Entropy

Entropy, as a nonlinear dynamic parameter that can measure the incidence of new information in a time series, can be described as a regularity or degree of randomness indicator. If the number of sequences in the time series is more complicated, the entropy will have a higher value and vice versa.

SampEn is an improved algorithm proposed by Richman and Moorman [38] based on approximate entropy, which reduces the deviation caused by self-matching. The SampEn function is a negative logarithm, indicating whether two similar sequences of m consecutive data points remain similar at the next point (m + 1).

\[ SE(m, r, N) = - \ln \left( \frac{C_{m+1}(r)}{C_{m}(r)} \right) \]  

(1)

The definition of \( C_m \)[26] is as follows:

\[ C_m(r) = \frac{\text{number of all probable pairs}(i,j) \text{ with } |x_i^m - x_j^m| < r, i \neq j}{\text{number of all probable pairs}, \text{ i.e. } (N-m+1)(N-m)} \]  

(2)

where \(|x_i^m - x_j^m|\) is the distance between points \(x_i^m\) and \(x_j^m\) in the dimension space, \(m\) and \(r\) denote the tolerable standard deviation of the time series, and \(N\) represents the length of the time series. Thus, the SampEn of the RR interval was calculated, and the analysis window was found to be consistent with that obtained using frequency domain analysis. In
addition, 256 consecutive RR intervals were used as the time-domain window. We set the
parameter to $m=5$, and $r$ is 0.3 times the standard deviation of the original signal in the
current window.

**Frequency Domain Algorithm**

Wavelet transform (WT) is a typical nonlinear analysis technique and one of the most
useful methods for biological signal analysis, especially in cases of continuous signals with
various frequency features [39, 40]. It is usually categorized into continuous WT and
discrete WT (DWT). Due to the discontinuity of the RR interval sampling, DWT is used for
time-frequency analysis. In addition, DWT is an effective wavelet transform analysis
method and can be used for assessing DoA [41]. Therefore, in this study, DWT was used for
the frequency domain analysis of the RR interval. First, RR interval detection was
performed according to Hamilton's method [42]. Second, 256 consecutive RR intervals were
set as the time window of the DWT, with a translation step of 5 RR intervals. Third, the
db4, a discrete wavelet of the Daubechies wavelets was set as the mother wavelet. The
length of the db4 wavelet was set to eight and the vanishing moment to four. The wavelet
decomposition level was set to seven. In addition, DWT was used for the time frequency
analysis in this study, following the work of Shensa et al. [43]. The LF and HF were
0.04-0.15 Hz and 0.15-0.4 Hz, respectively. The frequency-domain features of interest were
the HF, LF, and HF/LF ratio. The absolute values of the original HF and LF were retained in
the frequency domain feature to classify different anaesthesia states.

The HRV power was defined as the sum of squares of the time domain coefficients at a
certain frequency after the DWT. The calculation formula for the HRV power is as follows:
The discrete time signal $x(n)$ is defined as:

$$x(n) = \{x_n\},$$

where $x(n)$ is the $n^{th}$ digit in the sequence $x$, and $n$ is an integer.

**Logistic Regression**

LR is a classification algorithm used to predict the probability of classifying dependent variables. In this study, this is a combination algorithm of a linear and sigmoid function.

The linear function is defined as:

$$f(x) = \omega^T x,$$

where $\omega$ is the weight of each feature, and $T$ stands for the transpose of the vector.

The sigmoid function is given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}},$$

where $e$ is the natural logarithm. The $\sigma(x)$ maps the linear prediction result to the range of 0-1, indicating the probability of the prediction result. The mean squared error of the sigmoid function was used to measure the difference between the actual and predicted values.

The optimal regression function is as follows:

$$y = \sigma(f(x)) = \sigma(\omega^T x) = \frac{1}{1 + e^{-\omega^T x}}.$$

Therefore, the parameter $y$ can be described as a linear combination of the prediction observations.

**Support Vector Machine**
A support vector machine is a supervised learning algorithm that can be applied to classification problems. It is a linear classifier with the largest geometric interval in the feature space. In this study, we deployed classification modelling using the Lagrange multiplier method, while the specific calculation steps follow those in previous research [33].

**Decision Tree**

A decision tree is a multi-classification supervised learning algorithm. It can be used not only for classification problems but also regression problems, and is represented by a tree structure in which each internal node represents a judgment condition of an attribute, each branch represents the output of the judgment result, and each leaf node represents a classification result. The construction of a DT aims to classify the conditions by judging the size of the information gain under a certain feature. The formula of the information gain is determined by the difference of the entropy minus the conditional entropy. The formula of the conditional entropy is as follows:

\[
H(Y|X) = \sum_{i=1}^{n} p_i H(Y|X = x_i). \tag{8}
\]

Conditional entropy represents the uncertainty of the random variable \(Y\) under the condition that the random variable \(X\) is known.

Therefore, the formula of the information gain is as follows:

\[
g(Y, X) = H(Y) - H(Y|X), \tag{9}
\]

**Deep Neural Network**

An ANN is a nonparametric parallel computing model, which is similar to the neural structure of the human brain [44]. It usually consists of an input layer, a hidden layer, an
output layer, and numerous interconnected nodes in multiple layers. The nodes in the input layer receive external information, while the output layer outputs the result. Between the input and output layers, there are usually one or more hidden layers used to identify complex features in the data [45]. Thus, ANNs support continuous self-learning and error correction, and can analyse new problems and obtain optimal results. Supervised learning and unsupervised learning are the two learning rules in ANNs.

The DNN used in this study contains one input layer, two hidden layers, and one output layer. There are ten neural nodes in the first hidden layer and seventeen neural nodes in the second. Furthermore, the back-propagation algorithm, which is the most commonly used learning algorithm [33], was implemented in the proposed DNN model. For the construction of the DNN model, the four features of the HRV were set as input, and a highly accurate example of DoA estimation (i.e., mean EACL value determined by five experienced anaesthesiologists) was used as the output. The flowchart of the DNN construction is shown in Fig. 3.

**Performance Analysis**

The 23 datasets in this study were divided into training and test datasets. Eighty percent of the data were used to train the model, and 20% of the data were used to test the model. Training and testing were performed simultaneously to reduce model over-fitting. Due to the limited number of samples, a 5-fold cross-validation strategy was used to evaluate the generalization ability of the predictors. The performance of the predictors was quantified based on the results of cross-validation using the precision, recall, and classification accuracy. The three parameters are defined as follows:
Precision is defined as the ratio of the number of correct classifications of an anaesthesia state $N_{\text{detected}}$ to the total number of classifications of the same type of anaesthesia state $N_{\text{total}}$.

$$P_i = \frac{N_{\text{detected}}}{N_{\text{total}}},$$  \hspace{1cm} (10)

Where $i$ represents the three anaesthesia states (anaesthesia induction, anaesthesia maintenance, anaesthesia recovery).

Recall is defined as the ratio of the number of correct classifications of an anaesthesia state to the number of actual occurrences of this anaesthesia state $N_{m,\text{total}}$.

$$R_i = \frac{N_{\text{detected}}}{N_{m,\text{total}}},$$  \hspace{1cm} (11)

where $m$ is the actual number of one anaesthesia state.

Classification accuracy is defined as the ratio of the total number of correctly identified anaesthesia states $N_{\text{detected}}$ to the sum of all anaesthesia states $N_{\text{total}}$.

$$ACC = \frac{N_{\text{detected}}}{N_{\text{total}}}.\hspace{1cm} (12)$$

**Statistical analysis**

Statistical analyses were performed in SPSS 22.0 (SPSS Inc., Chicago, IL). Data were expressed as mean (SD) or percentage, where appropriate. Ternary classification outcome parameters were expressed as events (percentages). Data are presented as tables, box-and-whisker diagrams, and correlation graphs. In addition, we calculated the distribution of the four features in the three anaesthesia states. The Pearson’s correlation coefficient between the EACL and the four features of the DNN model was also calculated to estimate the efficacy of the proposed method. The performances of four classification methods were compared: LR, the SVM, the DT, and the proposed DNN. Owing to the
small sample size in this study, the sample does not satisfy a normal distribution. The nonparametric Wilcoxon signed-rank test may be applied to non-normal distribution data. Therefore, the four classification methods were compared using the Wilcoxon signed-rank test. $p < 0.05$ was considered statistically significant.

Results

Fifty-two adult patients were enrolled in this study; twenty-three were analysed and twenty-nine were excluded. Among the excluded patients, thirteen were excluded because they declined to participate in the study, eight were excluded as they did not meet the inclusion criteria, and eight were excluded due to technical failure. The details of the selection procedure are shown in Fig. 4. Patient demographics and clinical characteristics are shown in Table 1.

In this study, RR interval resampling, discrete wavelet transformation of the ECG signals, and frequency domain analysis were conducted. The DNN model was then constructed with two hidden layers. In addition, we analysed the distribution characteristics of the four features under three different anaesthesia states and the correlation of four features with the EACL. Finally, we analysed the precision, recall, and classification accuracy of four methods to distinguish between the three anaesthesia states of the patient during surgery.

The ECG signals collected from the 23 patients were resampled to remove the EMG artifacts and the THA. The target of the filtering was intended to increase the reliability of the data analysis. The RR interval resampling was performed using the interference-free ECG signal. Because the time window (153 s) of the DWT that corresponds to 256 RR
intervals, continuous wavelet decomposition was performed at seven different frequencies.

Three HF and four LF wavelets were then obtained from the wavelet decomposition of the
time course in the frequency domain.

During the anaesthesia induction state, the HF/LF ratio was below one, while the
absolute value was relatively high. Under the anaesthesia maintenance state, the ratio
exceeded that of the induction state without interference from unexpected events. The
absolute values of the HF and LF were below those of the induction state. In the
anaesthesia recovery state, the HF/LF ratio significantly exceeded one. The trends of these
ratios are shown in Fig. 5(e). In addition, Fig. 5 shows the original ECG signal, filtered ECG
signal, filtered RR interval, HF, LF, HF/LF ratio, and EACL in the same time period. The
voltage of the filtered ECG signal was mainly between 0-2.5 mV. During the sampling
period, the voltage of the ECG was relatively stable with no significant changes. The
filtered RR interval, HF, LF, and HF/LF ratio were significantly reduced before reaching a
relatively stable level. The trend of change in the three frequency features was similar to
that of the EACL.

The DNN structure used in this study consisted of four layers: an input layer with four
nodes, a hidden layer with ten nodes, a second hidden layer with seventeen nodes, and an
output layer with one node (as shown in Fig. 6). The data of 23 patients were processed
simultaneously for training and testing to avoid over-fitting and improve classification
accuracy.

In this study, four features of the HRV were selected as the input of the DNN model.
Specifically, these were the HF, LF, HF/LF ratio, and the SampEn of the RR interval; the
EACL was used as the reference output. Figure 7 depicts the distribution characteristics of
the four features under three different anaesthesia states. The HF during the anaesthesia
The HF during the induction state was significantly higher than that of the anaesthesia maintenance state (p < 0.001). The HF during the recovery state were significantly higher than that of the anaesthesia maintenance state (p < 0.001). These results can be reasonably explained in terms of the clinical practice, and may be related to the emotional tension of the patient during anaesthesia induction and recovery resulting in a higher physiological voltage. Moreover, the LF gradually decreased during the three anaesthesia states. The HF/LF ratio during the anaesthesia recovery state was significantly higher than those of the anaesthesia induction and maintenance states. There was a gradual increase in the ratio, which reflected the changes during the different anaesthesia states. Finally, the SampEn of the RR interval gradually increased under the three anaesthesia states. These results indicate that four frequency domain features can be used for effective classification and as features for evaluating the DoA.

The correlations between the four features and the EACL are depicted in Fig. 8. The results indicate a clear correlation between the HF, LF, HF/LF ratio, and RR interval SampEn, and the EACL. There is a positive correlation between the HF, LF, HF/LF ratio, and EACL. There is a negative correlation between the RR interval SampEn, and the EACL. These parameters can be used for the construction of the DNN model. Therefore, our method is expected to provide a reliable reference for anaesthesiologists to accurately assess the DoA. The most striking finding in Figure 8 is the low correlation between the four features and the target EACL. In addition, the four parameters are mainly distributed in the EACL value range of 40-80, which is consistent with the actual clinical conditions.

To compare the performance of the four models and EACL, the precision, recall, and classification accuracy of the models were used to determine the different anaesthesia states. The precision and recall values of 23 test datasets of the anaesthesia induction,
maintenance, and recovery states are listed in Table 1. In addition, the classification
accuracies of the three different anaesthesia states were obtained through the calculation of
the recall and precision. The precision and recall of the four models during the anaesthesia
induction and recovery states were lower than those during the maintenance state, which
could be attributed to the shorter time and smaller sample size of the former states. In
future research, we intend to increase the sample size to further validate the proposed
method. In addition, The DNN model yielded a classification accuracy of 90.1% (p < 0.05),
which was higher than that of other three models, i.e., 86.2% (LR), 87.5% (SVM), and 87.2%
(DT). A comparison of LR, the SVM, the DT, and the DNN is presented in Table 2.

Discussion

This study proposed a new method for DoA assessment based on multiple HRV features,
including three frequency domain features and one time domain SampEn feature,
combined with a DNN. In addition, this study compared the proposed DNN model with
LR, an SVM, and a DT in terms of DoA estimation. The data of 23 patients under general
anaesthesia were used for assessing the proposed method. Each of the four models
provided high accuracy in classifying the anaesthesia maintenance state, although the
proposed method exhibited better performance in detecting the three different anaesthesia
states than the three conventional methods. However, all models exhibited poor
performance in identifying the anaesthesia induction and recovery states.

Accurate monitoring of the DoA is crucial to guarantee the safety of surgery patients.
Anaesthesiologists use physiological vital signs and their own experience to evaluate
levels of consciousness during operations. The key parameters of interest are generally BP,
HR, and SpO₂[13], although these parameters cannot accurately reflect the actual DoA.
New methods such as BIS, Narcotrend, and Entropy were developed to evaluate the DoA[8-10], and are effective to some extent but bring certain drawbacks. Furthermore, HRV based on ECG signals is correlated with autonomic nervous system function and can be used with general anaesthesia[18-20]. These facts are widely accepted in the field of anaesthesia, and ECG monitoring has been used in the present DoA study.

To improve the accuracy of DoA estimation based on ECG signals, we proposed a practical HRV-derived method designed to correspond with the EACL, the highly accurate example of DoA estimation that anaesthesiologists insist on when developing methods of monitoring the DoA. In this study, although EACL was scored by experienced anaesthesiologists with subjective opinions, the average value of the five doctors' scores was used as the EACL. In addition, the results in this study indicated that the change of four parameters could reflect the change of anaesthesia states as a result of anaesthetic drugs acting on the autonomic nervous system[2, 28]. We observed the correlations of four features with the EACL. The parameters were mainly distributed in the EACL value range of 40-80, which is an interesting phenomenon and consistent with clinical reality. However, the correlation between a single index and the EACL was not strong, and the synergy between the four indices can be improved to classify the different anaesthesia states. Thus, the proposed method still requires future evaluation and development regarding the mathematical algorithms employed. Moreover, feature selection details are also a necessary component of further research.

The aim of this study was to propose the use of the four features of HRV based on DWT and a DNN to monitor the DoA. Previous studies have found that HRV is an effective and powerful index for this task. A short-term parameter of HRV could distinguish between the waking and isoflurane anaesthesia states [46]. Liu et al. used the similarity and
distribution index of HRV based on an ANN to assess the DoA [13]. However, HRV based on time domain or frequency domain signals could miss crucial information that is necessary for evaluating the DoA, thereby affecting the performance of the ANN. Thus, the performance of an ANN based solely on HRV time domain or frequency domain features as input is not more accurate than our strategy. In addition, DWT analysis can accurately estimate the transient changes in HRV during the general anaesthesia state[41, 47]. Therefore, the method proposed in this study was used to improve the classification accuracy of the DoA, differing from previous studies in terms of HRV-based DoA evaluation.

Although our method has a high classification accuracy for DoA evaluation, it still presents some shortcomings and a need for further improvement. First, the sample size in this study is small, resulting in poor performance in distinguishing between anaesthesia induction and recovery states. These changes could be attributed to the limited amount and quality of the datasets used during induction and recovery, along with undetected changes during these states. Additionally, we did not consider the impact of the high variability between individuals in the human ECG on the performance of the DNN model because of the small sample size. Individualized prediction of the DoA could be achieved by increasing the sample size and improving the precision of the DNN model. Second, the DNN could be improved by increasing the number of input features and meeting the individual conditions of different patients, thereby assessing the DoA more precisely. In this study, cross-validation was used to train and test the model to avoid over-fitting, ensure model generalization, and improve the performance of the DNN. To minimize personal error, the mean values of the EACL were used as the output of the ANN model [26]. Although we used an HRV-based DWT combined with a DNN in this study, the
combination of different time frequency analyses based on HRV-derived features could also be implemented to improve the classification accuracy of the DoA in the future. Nevertheless, the use of other wavelets is more in line with HRV changes for time frequency analysis, provides outstanding performance, and is worthy of subsequent research. Third, the DoA in this study was classified into the three anaesthesia states in the DNN model. It is necessary to explore new methods of DoA evaluation with higher precision, better performance, and more classifications (e.g., four or more states) in subsequent work.

**Conclusions**

This study combined multiple HRV-derived features, including three frequency-domain features and one time-domain feature, with a DNN to evaluate the DoA. The results demonstrated that the proposed method could accurately identify the three different anaesthesia states. The classification accuracy of the DNN model was similar to that of the EACL, and better than that of the three traditional machine-learning algorithms with which it was compared. In addition, this study indicates that DNN models based on HRV analysis can serve as effective methods for DoA assessment. Our method is expected to assist clinical anaesthesiologists in the accurate evaluation of the DoA and in avoiding anaesthesia crisis events. However, there is currently no ‘gold standard’ that may be employed in comparing DoA evaluation methods. Thus, the proposed method can be combined with other general anaesthesia-related physiological signals, such as EEG, to be more clinical, meaningful, and helpful in further improving the accuracy of DoA estimation.
Abbreviations

DoA: depth of anaesthesia; EEG: electroencephalogram; HRV: heart rate variability; DWT: discrete wavelet transform; DNN: deep neural network; HF: high-frequency power; LF: low-frequency power; HF/LF: high-to-low-frequency power ratio; RR interval: The interval between R peaks in two adjacent heartbeats of the ECG; EACL: expert assessment of consciousness level; LR: logistic regression; SVM: support vector machine; DT: decision tree; BIS: Bispectral index; ECG: electrocardiogram; HR: heart rate; BP: blood pressure; SpO2: peripheral oxygen saturation; SampEn: Sample entropy; ANN: artificial neural network; ASA: American Society of Anaesthesiology; Hz: hertz; EMG: electromyogram; THA: transient high-amplitude artifacts; WT: wavelet transform; DWT: discrete wavelet transform; SD: Standard Deviation; BMI: body mass index

Declarations

Ethics approval and consent to participate

Ethical approval for Institutional Ethics Committee of the Second Affiliated Hospital of Army Medical University prior to patient enrolment. Written informed consent was obtained from the patients.

Consent for publication

Not applicable.

Availability of data and materials

The datasets are not publicly available, but available from the corresponding author on reasonable request.

Competing interests

The authors declare no conflict of interest.
Funding

This study was supported by National Key Research and Development Project (2018YFC0117200) and Clinical Research Project of Army Medical University (No.CX2019LC114 and 2018JSLC0015).

Authors’ contributions

ZJ: study design, data analysis, writing paper. WZX: data analysis, writing paper. DZX: data collection. YGY: data collection, data analysis, manuscript revision. DZY and BXH: study design, manuscript revision. LH: study design, data analysis, writing paper, manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Editage (www.editage.cn) for English language editing and senior engineer Qin-yuan Yu and engineer Yi-wei Chen for providing guidance and help in machine learning algorithms, Chongqing Abacus Software Co., Ltd.

References

1. Anesthesiology ASoATFoIAJ: Practice advisory for intraoperative awareness and brain function monitoring: a report by the american society of anesthesiologists task force on intraoperative awareness. 2006, 104(4):847.

2. Lan JY, Abbod MF, Yeh RG, Fan SZ, Shieh JS: Review: Intelligent Modeling and Control in Anesthesia. Journal of Medical and Biological Engineering 2012, 32(5):293-307.

3. Avidan MS, Jacobsohn E, Glick D, Burnside BA, Zhang L, Villafranca A, Karl L, Kamal S, Torres B, O’Connor M et al: Prevention of intraoperative awareness in a high-risk surgical population. The New England journal of medicine 2011, 365(7):591-600.

4. Misal US, Joshi SA, Shaikh MM: Delayed recovery from anesthesia: A postgraduate...
5. Lobato EB, Gravenstein N, Kirby RR: Complications in anesthesiology: Lippincott Williams & Wilkins; 2008.

6. Hemmings Jr HC, Akabas MH, Goldstein PA, Trudell JR, Orser BA, Harrison NL: Tips: Emerging molecular mechanisms of general anesthetic action. 2005, 26(10):503-510.

7. Chau PL: New insights into the molecular mechanisms of general anaesthetics. British journal of pharmacology 2010, 161(2):288-307.

8. Bruhn J, Ropcke H, Hoeft AJA: Approximate Entropy as an Electroencephalographic Measure of Anesthetic Drug Effect during Desflurane Anesthesia. 2000, 92(3):715-726.

9. Chan MTV, Cheng BCP, Lee TMC, Gin TJJ: BIS-guided Anesthesia Decreases Postoperative Delirium and Cognitive Decline. 2013, 25(1):33-42.

10. Kreuer S, Biedler A, Larsen R, Altmann S, Wilhelm WJA: Narcotrend Monitoring Allows Faster Emergence and a Reduction of Drug Consumption in Propofol–Remifentanil Anesthesia. 2003, 99(1):34-41.

11. Bruhn J, Myles PS, Sneyd R, Struys MM: Depth of anaesthesia monitoring: what's available, what's validated and what's next? British journal of anaesthesia 2006, 97(1):85-94.

12. Ye S-y, Park J-m, Kim J-h, Jung J-h, Jeon A-y, Kim I-c, Son J-m, Nam K-g, Baik S-w, Ro J-h: Development for the evaluation index of an anaesthesia depth using the bispectrum analysis. 2009, 4:67-70.

13. Liu Q, Ma L, Chiu RC, Fan SZ, Abbod MF, Shieh JS: HRV-derived data similarity and distribution index based on ensemble neural network for measuring depth of
Ahmed MU, Li L, Cao J, Mandic DP: Multivariate multiscale entropy for brain consciousness analysis. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society: 2011: IEEE; 2011: 810-813.

Merry AF, Cooper JB, Soyannwo O, Wilson IH, Eichhorn JH: International standards for a safe practice of anesthesia 2010. 2010, 57(11):1027-1034.

Huh IY, Kim DY, Sung M, Lee M, Park SE: Change of QT variability index during general anesthesia. Korean J Anesthesiol 2016, 69(3):250-254.

Oji M, Terao Y, Toyoda T, Kuriyama T, Miura K, Fukusaki M, Sumikawa K: Differential effects of propofol and sevoflurane on QT interval during anesthetic induction. Journal of clinical monitoring and computing 2013, 27(3):243-248.

Hsu C-Y, Tsai M-Y, Huang G-S, Lin T-C, Chen K-P, Ho S-T, Shyu L-Y, Li C-Y: Poincaré plot indexes of heart rate variability detect dynamic autonomic modulation during general anesthesia induction. 2012, 50(1):12-18.

Huhle R, Burghardt M, Zaunseder S, Wessel N, Koch T, Malberg H, Heller AR: Effects of awareness and nociception on heart rate variability during general anaesthesia. Physiological measurement 2012, 33(2):207-217.

Kanaya N, Hirata N, Kurosawa S, Nakayama M, Namiki A: Differential effects of propofol and sevoflurane on heart rate variability. Anesthesiology 2003, 98(1):34-40.

Lee B-R, Won D-O, Seo K-S, Kim HJ, Lee S-W: Classification of wakefulness and anesthetic sedation using combination feature of EEG and ECG. In: 2017 5th International Winter Conference on Brain-Computer Interface (BCI): 2017: IEEE; 2017: 88-90.
22. Akbarian B, Erfanian A: Automatic Seizure Detection Based on Nonlinear Dynamical Analysis of EEG Signals and Mutual Information. Basic Clin Neurosci 2018, 9(4):227-240.

23. Wei Q, Liu Q, Fan S, Lu C, Lin T, Abbod MF, Shieh JJE: Analysis of EEG via Multivariate Empirical Mode Decomposition for Depth of Anesthesia Based on Sample Entropy. 2013, 15(9):3458-3470.

24. Udhayakumar RK, Karmakar C, Palaniswami MJIToBE: Understanding Irregularity Characteristics of Short-Term HRV Signals Using Sample Entropy Profile. 2018, 65(11):2569-2579.

25. Alangari HM, Sahakian AVJIToBE: Use of Sample Entropy Approach to Study Heart Rate Variability in Obstructive Sleep Apnea Syndrome. 2007, 54(10):1900-1904.

26. Jiang GJ, Fan S-Z, Abbod MF, Huang H-H, Lan J-Y, Tsai F-F, Chang H-C, Yang Y-W, Chuang F-L, Chiu Y-FJBi: Sample entropy analysis of EEG signals via artificial neural networks to model patients’ consciousness level based on anesthesiologists experience. 2015, 2015:343478.

27. Huang J, Fan S, Abbod MF, Jen K, Wu J, Shieh JJE: Application of Multivariate Empirical Mode Decomposition and Sample Entropy in EEG Signals via Artificial Neural Networks for Interpreting Depth of Anesthesia. 2013, 15(9):3325-3339.

28. Komatsu T, Kimura T, Sanchala V, Shibutani K, Shimada YJJoC, Anesthesia V: Effects of fentanyl-diazepam-pancuronium anesthesia on heart rate variability: A spectral analysis. 1992, 6(4):444-448.

29. Moak JP, Goldstein DS, Eldadah BA, Saleem A, Holmes C, Pechnik S, Sharabi YJJHrtOjotHRS: Supine low-frequency power of heart rate variability reflects baroreflex function, not
cardiac sympathetic innervation. 2007, 4(12):1523-1529.

30. Ferreira AL, Nunes CS, Mendes JG, Amorim P: Usefulness of the Blink Reflex to Assess the Effect of Propofol During Induction of Anesthesia in Surgical Patients. In: 2020; Cham: Springer International Publishing; 2020: 1057-1063.

31. Shalbaf A, Shalbaf R, Saffar M, Sleigh J: Monitoring the level of hypnosis using a hierarchical SVM system. Journal of clinical monitoring and computing 2020, 34(2):331-338.

32. Reports C-WJJS: Data Driven Investigation of Bispectral Index Algorithm. 2019, 9(1):1-8.

33. Gu Y, Liang ZH, Hagihira S: Use of Multiple EEG Features and Artificial Neural Network to Monitor the Depth of Anesthesia. Sensors 2019, 19(11):2499.

34. Liu Q, Cai J, Fan S-Z, Abbod MF, Shieh J-S, Kung Y, Lin LJIA: Spectrum Analysis of EEG Signals Using CNN to Model Patient’s Consciousness Level Based on Anesthesiologists’ Experience. 2019, 7:53731-53742.

35. Li R, Wu Q, Liu J, Wu Q, Li C, Zhao Q: Monitoring Depth of Anesthesia Based on Hybrid Features and Recurrent Neural Network. Front Neurosci 2020, 14:26.

36. Fraser GD, Chan ADC, Green JR, Macisaac D: Removal of electrocardiogram artifacts in surface electromyography using a moving average method. In: IEEE International Symposium on Medical Measurements & Applications: 2012; 2012.

37. Berger RD, Akselrod S, Gordon D, Cohen RJ: An efficient algorithm for spectral analysis of heart rate variability. IEEE Trans Biomed Eng 1986, 33(9):900-904.

38. Richman JS, Randall MJJAJoPH, Physiology C: Physiological time-series analysis using approximate entropy and sample entropy. 2000, 278(6):H2039-H2049.
Quiroga RQ, Sakowitz O, Basar E, Schürmann MJBRP: Wavelet transform in the analysis of the frequency composition of evoked potentials. 2001, 8(1):16-24.

Guler I, Ubeyli ED: Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. Journal of Neuroscience Methods 2005, 148(2):113-121.

Tai NK, Peng W, Yan L: Monitoring the Depth of Anesthesia Using Discrete Wavelet Transform and Power Spectral Density. In: Rough Sets and Knowledge Technology, 4th International Conference, RSKT 2009, Gold Coast, Australia, July 14-16, 2009 Proceedings: 2009; 2009.

Hamilton P: Open source ECG analysis. In: Computers in Cardiology: 2002; 2002.

Shensa, Processing MJJIToS: The discrete wavelet transform: wedding the a trous and Mallat algorithms. 1992, 40(10):2464-2482.

Bose NK, Liang P: Neural network fundamentals with graphs, algorithms, and applications: Mcgraw-Hill; 1996.

Zhang G, Hu MY, Patuwo BE, Research DCIJEJoO: Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis. 1999, 116(1):16-32.

Huang H-H, Lee Y-H, Chan H-L, Wang Y-P, Huang C-H, Fan S-ZJM, engineering b, computing: Using a short-term parameter of heart rate variability to distinguish awake from isoflurane anesthetic states. 2008, 46(10):977-984.

Pichot V, Buffiere S, Gaspoz J-M, Costes F, Molliex S, Duverney D, Roche F, Barthelemy J-CJCJoA: Wavelet transform of heart rate variability to assess autonomic nervous system activity does not predict arousal from general anesthesia. 2001, 48(9):859-863.
Figures

Fig. 1 RR interval in an ECG signal.

Fig. 2 Flow chart for ECG processing.

Fig. 3 Flowchart depicting the proposed DNN model.

Fig. 4 Study protocol.

Fig. 5 ECG data for the proposed method. (a)–(c) Raw ECG, filtered ECG, and filtered RR intervals. (d)–(e) HF, LF, and ratio of HF/LF. (f) EACL within the sampling period.

Fig. 6 Schematic representation of the proposed DNN structure. One input layer with four nodes, a hidden layer with ten nodes, a second hidden layer with 17 nodes, and an output layer with one node.

Fig. 7 Comparison between anaesthesia states. (A)–(D) Distributions of (A) HF, (B) LF, (C) the ratio of HF/LF, and (D) the RR interval SampEn values. I, II, and III represent anaesthesia induction, anaesthesia maintenance, and anaesthesia recovery, respectively. Vertical coordinates represent the four feature values.

Fig. 8 Correlations between the four features and EACL. (A)-(D) Correlations of HF, LF, ratio of HF/LF, and RR interval SampEn with the EACL, respectively. I, II, and III represent anaesthesia induction, anaesthesia maintenance, and anaesthesia recovery, respectively.