Biomedical Signal Processing and Control 68 (2021) 102663

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Deep learning via ECG and PPG signals for prediction of depth of anesthesia

Meghna Roy Chowdhury [a], Ravichandra Madanu [b], Maysam F. Abbod [c], Shou-Zen Fan [d], Jiann-Shing Shieh [b,*]

[a] Department of Electronics and Communication Engineering, Vellore Institute of Technology, Chennai, 600127, India
[b] Department of Mechanical Engineering, Yuan Ze University, Taoyuan, 32003, Taiwan
[c] Department of Electronics and Computer Engineering, Brunel University London, Uxbridge, UB8 3PH, UK
[d] Department of Anesthesiology, College of Medicine, National Taiwan University, Taipei, 100, Taiwan

A R T I C L E I N F O

Keywords:
Electrocardiogram (ECG)
Photoplethysmography (PPG)
Bispectral index (BIS)
Convolutional neural network (CNN)
Depth of anesthesia (DoA)

A B S T R A C T

During surgeries, the amount of used anesthetic depends on the physical conditions of the patient and is immensely critical. The conventionally used BIS Quantro machine which measures the Bispectral Index (BIS) level in order to help doctors administer anesthesia, is quite expensive. In this paper, an economic, accurate and state-of-the-art technique is presented to predict the depth of anesthesia (DoA) via advanced deep learning models using 512 Hz Electrocardiogram (ECG) and 128 Hz Photoplethysmography (PPG). The study is conducted based on signal collected from 50 patients acquired during surgery at National Taiwan University Hospital (NTUH). First, heatmaps of the ECG and PPG signals (individual and combined subplots) are generated using MATLAB by filtering 5 s windows to match the frequency of the BIS Quantro Machine which is 0.2 Hz. Then, various deep learning models comprising 5, 6, 8, 10 and 19 layered CNNs are trained using data of 40 patients and tested using the remaining 10 patients. The heatmap images of ECG and PPG are fed as inputs to the CNN models separately and using two input channels. The best accuracy achieved is 86 % which is attained using 10 layered CNN with Tensorflow backend, with combined ECG and PPG heatmaps as inputs. This study uses inexpensive signals, minimum data reconstruction, minimum memory and timing constrains to achieve a decent accuracy, and so it can be used by even small hospitals.

1. Introduction

Anesthesiologists play a very crucial role in the medical world. Anesthesia is one of the most important ingredients of any surgery. It enables doctors to perform surgery on patients with unconsciousness and painlessness. According to the definition of general anesthesia, current practices consist of four main components: hypnosis, analgesia, amnesia, and muscle relaxation [1]. A state of general anesthesia is produced by anesthetics that act on the spinal cord, and the brain (the stem, cortex and thalamus). The precision and accuracy of administering the appropriate depth of anesthesia are vital. If the dose is too light, the patient may become aware of the surgical stimulus and too deep patients are at risk of other complications. However, anesthesiologists have multiple inconsistent definitions of the anesthetic state and have no standard measurement to assess it. So far, valuation indices like Bispectral Index (BIS), entropy, auditory evoked potential (AEP), surgical stress index (SSI) can help give objective reports of general anesthesia [2]. Indices like analgesia nociception index (ANI) and surgical pleth index (SPI) are measures of pain (nociception) [3]. A heart rate variability parameter called similarity index (SI) was derived as a consciousness level parameter [4].

Earlier, anesthesia used to be administered manually based on the patient's physiological response. However, the quantity of anesthesia to be given does not avoid external interference [5]. This inaccuracy may be fatal. Lan et al. has said that direct measurements like heart rate, blood pressure (BP), respiratory parameters, temperature, blood oxygen saturation, cannot provide sufficient information of the autonomic nervous system (ANS) and central nervous system (CNS), which are related to the depth of anesthesia (DoA), level of surgical stress, and nociceptive changes. Although these direct measurements of vital signs are available easily, their response is delayed, and precise modeling and anesthesia control are very challenging for anesthesiologists. In order to reach the proper anesthetic state for surgery, hypnotic and analgesic effects are considered as major fundamental pharmacologic

* Corresponding author.
E-mail addresses: meghnaroy.chowdhury2016@vitstudent.ac.in (M. Roy Chowdhury), s1078707@mail.yzu.edu.tw (R. Madanu), Maysam.Abbod@brunel.ac.uk (M.F. Abbod), shouzen@ntuh.gov.tw (S.-Z. Fan), jsshieh@saturn.yzu.edu.tw (J.-S. Shieh).

https://doi.org/10.1016/j.bspc.2021.102663
Received 27 October 2020; Received in revised form 4 April 2021; Accepted 16 April 2021
Available online 4 May 2021
1746-8094/

components. Under general anesthesia, memory and awareness are critical components of DoA. Moreover, the autonomic nervous system (ANS) and central nervous system (CNS) are related to the DoA, level of surgical stress, and nociceptive changes [2]. In early times, the concept of depth of anesthesia was considered illusory, and due to the various components of anesthesia it was difficult to expect that a single index can be used successfully for measuring the depth of anesthesia in general [6,7]. But over the years, a lot of researchers have laid the importance of DoA being a reliable and significant index to monitor the change of anesthetic drug effects and being a safe and good quality index to determine DoA [8–10]. Therefore, predicting DoA with accuracy has attracted growing attention since it provides patients a safe surgical environment in case of secondary damage caused by intraoperative awareness or brain injury [5].

There are accepted commercial indices like the Bispectral index (Aspect Medical Systems, Newton, MA, USA), entropy (GE Healthcare, Helsinki, Finland), auditory evoked potential that are used to find the level of unconsciousness (depth of anesthesia) [10]. Furthermore, devices that implement BIS are the only ones currently approved by the US Food and Drug Administration (FDA) for marketing as monitors of anesthetic effect on the brain. The BIS, most popular used in the hospital, is a complex parameter derived from analysis in time domain, frequency domain and high order statistics [11]. It has been a common measure to assess the depth of sedation while administering anesthesia and is recommended as a guide for the administration of hypnotic drugs during anesthesia. BIS is presented as a numerical index ranging from 100 (awake) to 0 (isoelectric EEG). Values below 60 imply that the patient is almost certainly unconscious. The BIS index offers considerable advantages, most notably extensive clinical validation [12]. Kissin has stated that it is possible to conclude that the BIS is most promising as a monitor of unconsciousness [7]. The BIS monitor gives us the level of DoA, which is a measure of the level of unconsciousness (i.e., hypnosis component of anesthesia). The BIS monitor, derived from electroencephalogram (EEG) data, has been used as a statistical predictor of the level of hypnosis and has been proposed as a tool to reduce the risk of intraoperative awareness [8]. The BIS index is based on the power distribution of the Fourier transform of the EEG signal and quantifies the phase coupling between different EEG frequencies [13]. It is derived using a composite of measures from EEG signal processing techniques, including bispectral analysis, power spectral analysis, and time-domain analysis. These measures were combined via an algorithm to optimize the correlation between the EEG states and the clinical effects of anesthesia, and to quantify these effects using the BIS monitoring value [14]. BIS values between 40 and 60 are recommended for surgery under general anesthesia [15]. In fact, there is a research paper by Kreuer S et al. where a new device (The Narcotrend™) to measure depth of anesthesia was investigated based on 14 substages and compared to BIS [16]. The FDA cleared BIS monitoring in 1996 for assessing the hypnotic effects of general anesthetics and sedatives. The FDA further stated in 2003 that "A reduction in awareness provides a public health benefit, in that BIS technology can now provide anesthesiologists with a way to reduce this often debilitating, yet preventable medical error" [14].

To determine the depth of anesthesia in recent times, quantitative modeling approaches are being replaced by qualitative techniques, especially in the field of artificial intelligence. Hybrid models (quantitative/qualitative) and hybrid intelligent algorithms (neural networks, fuzzy logic, evolutionary computing) have been applied to anesthesia [2]. Convolutional Neural Network (CNN) is a well-known deep learning architecture inspired by the natural visual perception mechanism of the living creatures. Among different types of deep neural networks, convolutional neural networks have been most extensively studied [17]. Deep neural networks are computational models consisting of multiple processing layers, with each layer being able to learn increasingly abstract, higher-level representations of the input data relevant to perform specific tasks. They have dramatically improved the state of the art in speech recognition, image recognition, strategy games, and in medical applications [18]. CNN models have been deployed to detect diseases like arrhythmias from short segments of ECG recordings, REM Behavior Disorder (RBD) diagnosis from the EEG, brain seizures & postanoxic coma, etc [19–24]. However, the application of deep learning is seldom studied on anesthesia. This is because it is not easy for researchers to have the patient's total anesthetic state, apart from the extremely strict requirements of CNN on the operating environment [5]. There has been prior research on predicting DoA using various vital signs like electroencephalography (EEG), electromyography (EMG), heart rate (HR), pulse, systolic blood pressure (SBP), diastolic blood pressure (DBP), and signal quality index (SQI) via artificial neural network (ANN) and has produced a less mean absolute error of of 6.54 with 6.69 of standard deviation in comparison with BIS which has mean absolute error of 12.31 and 13.06 of standard deviation, in predicting the DoA [25]. There has also been a study of prediction of DoA using EEG Spectrum using Convolutional Neural Network which has given the accuracy of 93 %, interpreted as CNN's deep learning to approximate the DOA by senior anaesthesiologists, which highlights the potential of deep CNN combined with advanced visualization techniques for EEG-based brain mapping [5]. Horiguchi and Nishikawa evaluated studies of anesthesia based on monitoring the heart rate with the drug propofol [26,27].

A state of general anesthesia is produced by anesthetics that act on the spinal cord and the stem and cortex of the brain; monitoring of EEG patterns is therefore useful [28–30]. Indices derived from EEG have been widely used to quantify DoA. The BIS is EEG-based and EEG has a reducing anesthetic drug consumption and resulting in faster wake-up and recovery from anesthesia, hence it gives accurate measures in many cases [13]. However, EEG signals are known to be inaccurate under certain conditions. Although EEG-based spectral indices have been applied commercially for nearly 20 years, they are still not part of standard anaesthesiology practice, and the reasons for this are complex [10]. EEG signals developed from adult patient cohorts, and are not strictly relevant to infants or younger patients, thereby providing lower accuracy [31]. EEG cannot generate precise DoA measurements for certain drugs [32,33], especially when ketamine and nitrous dioxide are used. EEG signals are sensitive to noise, and therefore more complex algorithms and resources for noise filtering are required. Disposable EEG electrodes is much more expensive than using other physiological signal sensors [10]. Alternatively, heart rate variability has been widely accepted as a potentially good predictor of anesthetic depth. An HRV related unconsciousness SI was derived by Huang et al. and its value was observed at both conscious and anesthetized state during isoflurane anesthesia. Electrocardiography (ECG) proved to be good alternative for measuring the HRV, instead of EEG [4].

ECG and PPG signals are widely used to measure the cardiac cycle and monitor heart rate are ECG and Photoplethysmography (PPG) [34–36]. Byeon et.al conducted a comparative analysis of deep models in biometrics using scalograms of an electrocardiogram and has got good results [37]. The ECG provides important clinical physiological signals and is highly recommended for continuous monitoring and ensuring international standards for the safe practice of anesthesia [38]. Different anesthetics affect the Q wave and T wave interval of an ECG during anesthetic induction, and rhythmic-to-non-rhythmic observations from the ECG can provide anesthetic information [39,40]. The PPG is a non-invasive and inexpensive waveform that contains information related to the balance (or analgesia level) between nociception and anti-nociception undergoing general anesthesia, which has been ascertained to result in physiological reactions of the autonomic nervous system (ANS) [41–43]. A study conducted by Zhang et al. shows that photoplethysmography-derived parameters appear to be more suitable in monitoring the nociceptive component of balanced general anesthesia [44]. Therefore, with ECG and PPG monitoring, we can get a balanced state of general anesthesia. In addition, several studies have shown that ECG and PPG help in calculating various standard pain indices like ANI, SPI and SSI. ANI derived from the heart rate variability is based on ECG data derived from two single-use ANI electrodes. Promising results of an

SPI index based on heartbeat interval and pulse wave amplitude of the finger-PPG signal were reported [3]. Huiku et al. proposed the SSI [45], which is a simple numerical measure of the surgical stress level under anesthesia. Two continuous variables, the interval between successive hearts beats and PPG are used to calculate SSI. Also, ECG and PPG are easy and cheap to acquire. All hospitals can afford to get these signals from patients. On the other hand, the process of EEG signals is complex and expensive. Although the accuracy using ECG and PPG signals may be a bit lower than the accuracy of EEG signals, the former signals can be used in smaller hospitals where EEG detection is not feasible due to the cost. The relation between BIS values and ECG and PPG signals has not been studied much via deep learning, and thereby the research of predicting the depth of anesthesia using ECG and PPG signals is quite new.

In this paper, we present research on predicting DoA using the vital signs ECG and PPG with deep learning models. Based on the BIS values attained every 5 s, the 512 Hz ECG signal and 128 Hz PPG signal are filtered and heatmaps are generated for every 5 s of the surgery duration. CNN models are designed and trained to predict the DOA levels from ECG and PPG heatmaps, which presents an intuitive mapping process with decent efficiency and reliability. Through our research, we have proposed a cost-effective solution with a decent accuracy to help predict the depth of anesthesia.

2. Materials and methods

This study has been approved by the Research Ethics Committee, National Taiwan University Hospital (NTUH) in Taiwan. Furthermore, written informed consent was received for permission by the patients. In total, a dataset of 50 patients during surgical operation was used for the evaluation. The research has been divided into three major parts, namely data collection, data preprocessing and finally deep learning. The data (several vital signs like ECG, PPG, EEG, heart rate, blood pressure, etc) was collected from the surgical operation room at NTUH in Taipei, Taiwan using the techniques demonstrated in [46]. To evaluate the model and compare it to the BIS signal, whose sampling rate is also 0.2 Hz, the raw 128 Hz PPG signal and 512 Hz ECG signal are filtered and analyzed every 5 s, 640 and 2560 points respectively. Heatmaps are then generated and segregated as per the BIS and SQI values. These heatmaps are used as inputs for the CNN model. Various CNN models were implemented to maximize prediction accuracy. The deep learning part of DoA was divided into three categories as shown in Fig. 1.

The signals were processed and the heatmaps were generated using MATLAB R2020a. Deep learning using various CNN models were codes using Python3 and implemented on Google Colab. A total of 50 patients' data was used.

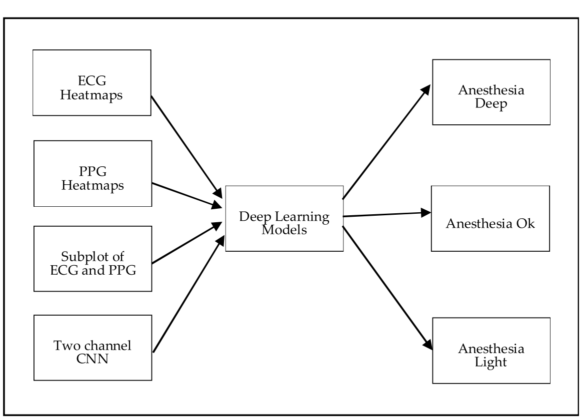

Fig. 1. Deep Learning Block Diagram.

Table 1
BIS Value Categorization.

BIS Range	Consciousness
0–40	Anesthesia Deep
40–60	Anesthesia OK
60–100	Anesthesia Light

2.1. Data collection process

As per [46], the equipment that were used to collect data are the physiological monitor Phillips IntelliVue MP60, BIS Quatro Sensor module, and a laptop for data-logging. First, the data measurements of vital signs like heart rate, blood pressure, SPO2 are collected through an MP60 monitor. ECG and PPG signals are also measured. The monitor is connected to NPort through UART serial communication port and it uses TCP/IP protocol for transmission. The NPORT transmits the data received wirelessly to the repeater. The data received from the repeater is then transmitted to the PC. The connection is verified using ping and handshake; when the connection is released, the transmission stops.

The logged data consists of continuous as well as discrete data. The continuous signal data that are used for this study are 512 Hz ECG signal and 128 Hz PPG signal recorded for the whole duration of the surgery of each patient. The discrete data used are the BIS value and SQI. These are recorded with a 0.2 Hz frequency. The BIS values attained from the BIS Quantro sensor module are categorized as Anesthesia Deep, Anesthesia Light, and Anesthesia Ok. Anesthesia Deep implies that the patient is in a state of unconsciousness and too deep. Anesthesia ok implies that the patient is suitable for surgery. While anesthesia light implies that the patient is awake, and surgery cannot proceed. Table 1 shows the range of BIS values for each category. The signal quality index (SQI) is also significant to determine whether a signal is contaminated with noise or not. The SQI of a Good Quality signal must be at least 40. An SQI below 40 signifies Low Quality. Based on the SQI of each BIS value attained in a 5-second interval, the signal is categorized it into low signal quality or good signal quality. Thereby, only the ECG and PPG signals are considered which fall under good signal quality for analysis.

Most of the medical datasets are unbalanced, that is, the number of conventional samples is much larger than the number of unconventional instances. In this case, it was observed that the percentage of samples in the category of anesthesia deep was relatively higher than that of the other two categories. Although it is peculiar that in our dataset as shown in Table 2, over 50 % of the data lies in Anesthesia Deep category, we have calculated the average BIS values of each category and it was seen that the average BIS values of the 'Anesthesia Deep' category is approximately 36.8 ± 4.6, which is close to the BIS values of Anesthesia OK (BIS range 40–60). It is fully coincidental that the Anesthesia Deep Category has more data than Anesthesia OK in the samples that we considered. Moreover, we have balanced the dataset while training our model, i.e. we used data augmentation techniques to ensure that we have equal number of data from each category to predict the depth of anesthesia.

To generate input images for Deep Learning, the raw signals of ECG and PPG are converted into images format (2D matrix) in jpg format and stored in specific folders. This was done using MATLAB.

2.2. Data preprocessing

The first cardinal step for analysis is data preprocessing. Here, the data accumulated needs to be converted to an image like format so that we can apply deep learning models. This is done by creating matrices. To do this, we first filter 5 s of raw ECG signal (512 Hz) and raw PPG signal (128 Hz) which gives us 2560 samples of ECG and 640 samples of PPG

Table 2
The distribution of the images in each category along with their average value and standard deviation.

BIS Category	Percentage of images (%)	Mean ± SD of BIS Value
Anesthesia Deep	56	36.8 ± 4.6
Anesthesia OK	35	45.9 ± 6.3
Anesthesia Light	9	67.0 ± 11.4

signal. Two different matrices are created with 2650 and 640 rows respectively. Each column of the matrix corresponds to one BIS value which is attained for the 5 s. Plots of ECG and PPG signals of each corresponding BIS value for each 5-second span of the surgery duration are generated. It is observed that each in all the BIS categories and the ECG signal follows a PQRST wave pattern. It is pretty hard to differentiate the plots of the ECG signals among the 3 categories. On the other hand, it is observed that the PPG signal follows its conventional two-peak plot. The plots are not enough to feed as input to our deep learning models. Next, heatmaps of ECG and PPG signals for each 5-second span of the surgery are generated and saved as jpg images. Based on the frequency of the two signals, the heatmap size of each ECG signal is 51 × 51 and the heatmap size of the PPG signal is 26 × 26. The heatmaps of the 50 patients were collected and we go a total of around 14,000 images. These images were categorized into Training and testing data in the ratio of 8:2. Fig. 2 shows a sample of ECG plots and heatmaps obtained for each BIS category (Anesthesia Deep, Anesthesia Light, Anesthesia Ok) and Fig. 3 show the same for PPG Plots. After balancing the dataset, we have 4715 images in each category (Anesthesia Deep, Anesthesia Light, and Anesthesia Ok). In total, there are 14,145 images for analysis.

2.3. Deep learning

The images of 50 patients were used for analysis. Several models

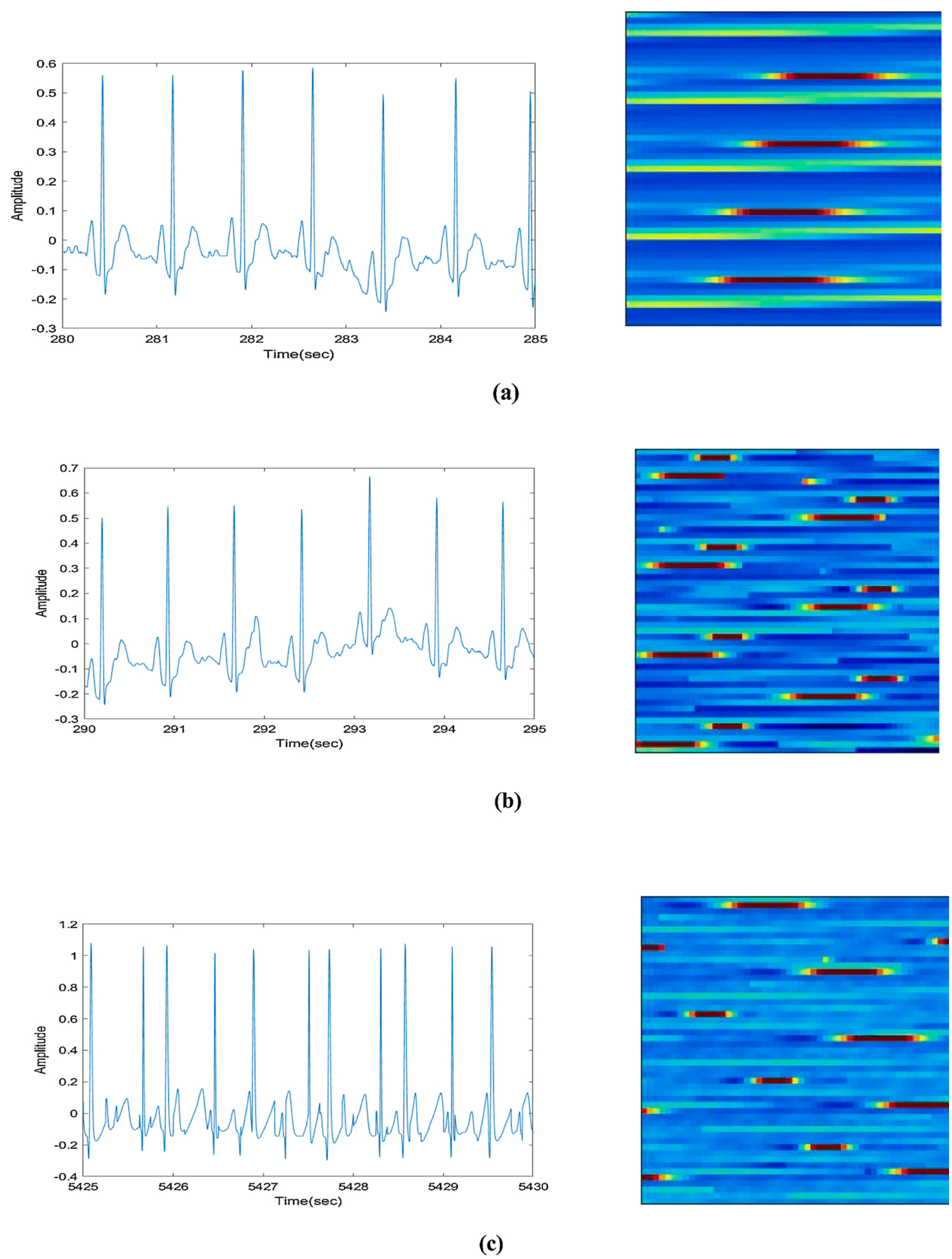

Fig. 2. ECG plots and heatmaps of (**a**) Anesthesia Ok; (**b**) Anesthesia Deep; (**c**) Anesthesia Light.

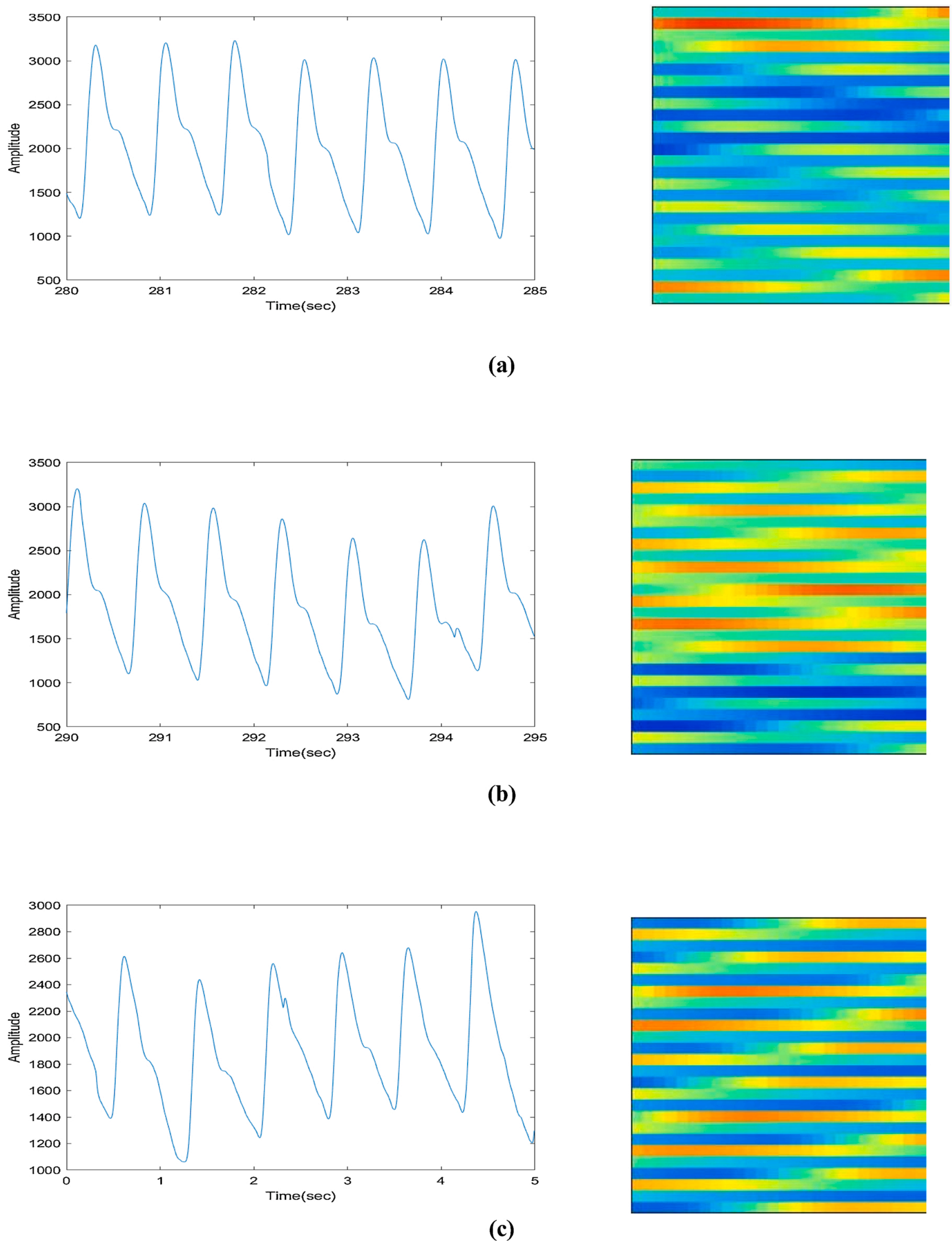

Fig. 3. PPG plots and heatmaps of (**a**) Anesthesia Ok; (**b**) Anesthesia Deep; (**c**) Anesthesia Light.

with a different number of layers were implemented. AlexNet a five convolutional layered model was used with some fine-tuning. Next, models with 6, 8, and 10 convolutional layers were used. Finally, VGG 19 was used with pre-trained weights of the ImageNet dataset. The first experiment was CNN performed with different models using the following ECG heatmaps and PPG heatmaps separately. Second, a combination of ECG&PPG heatmaps was used as inputs to our deep learning models. Finally, a two- Stream Multi-channel CNN was performed with ECG and PPG heatmap images an input. The output has 3 classes based on the DoA prediction - Anesthesia OK, Anesthesia Deep, Anesthesia Light. Deep learning models were coded in Python and the tool Google Colab was used for implementation. High computational time and GPU capacity are required for training complex models with many layers. Google Colab provides users with free access to Google GPU. To reduce the computational cost, filters were incorporated in the convolutional layers. Dropouts and Batch Normalization were added in

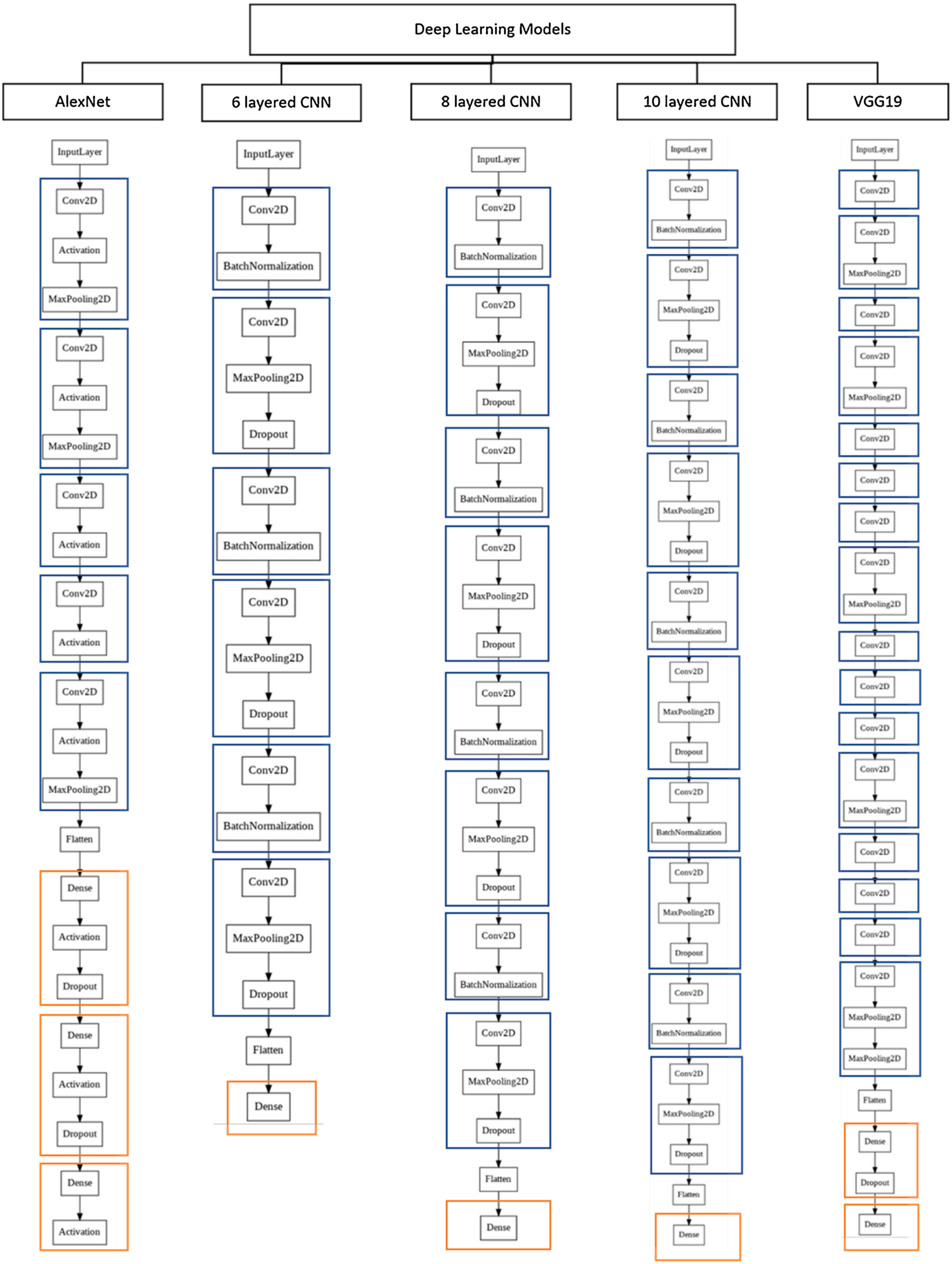

Fig. 4. Structure of Deep Learning models.

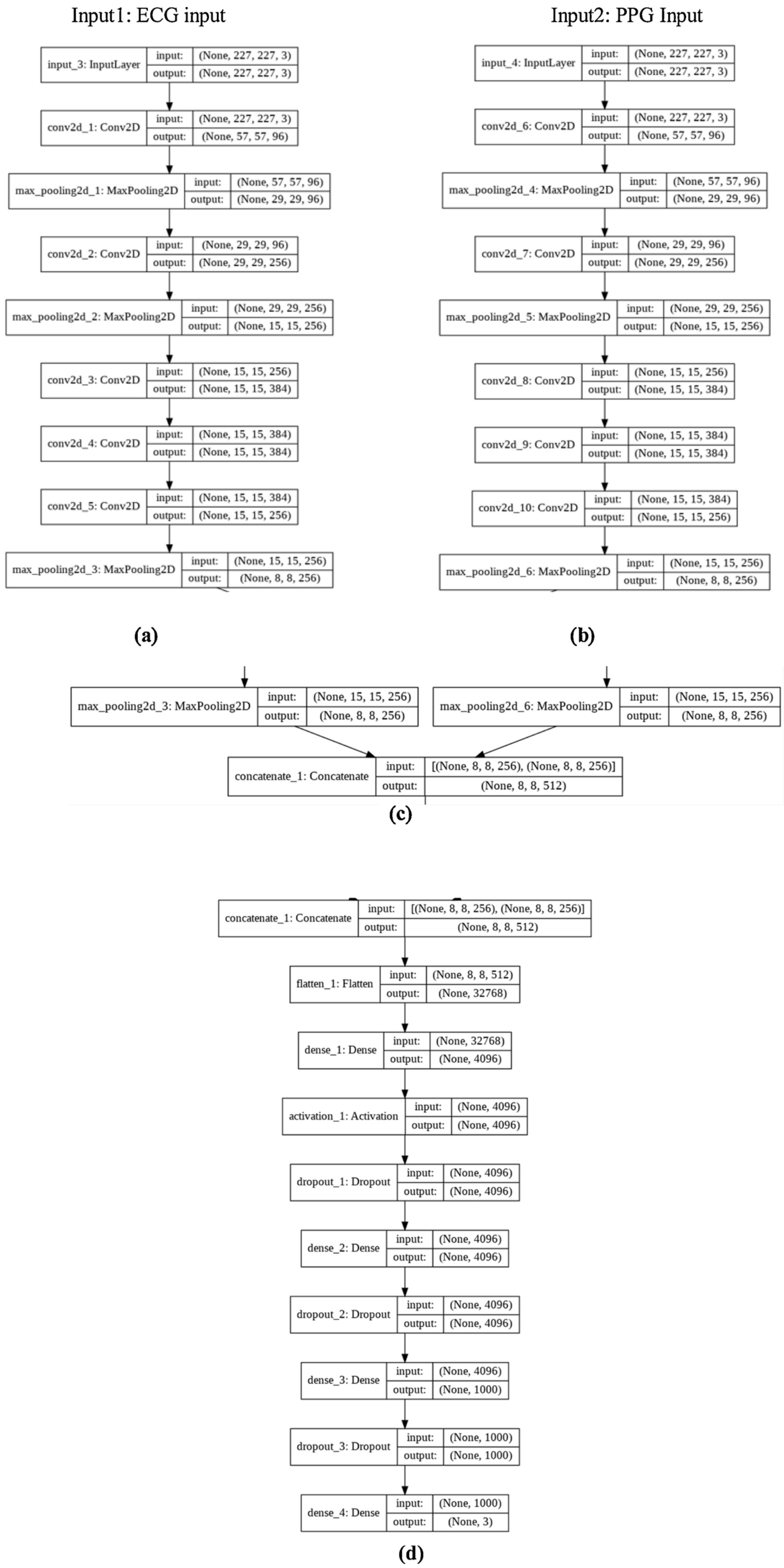

Fig. 5. Layers of Two-channel CNN (**a**) Convolution layer with ECG input; (**b**) Convolution layer with PPG input; (**c**) Concatenation Layer; (**d**) Fully Connected Layer. ('None' implies that the batch size is flexible; it is determined during training).

Table 3
ECG and PPG accuracy.

Model Structure	Accuracy (%)	
	ECG	PPG
AlexNet	72	62
Six layered model	72	63
Eight layered model	75	64
Ten layered model	77	65
VGG	70	62

the models to prevent overfitting. The models were compiled using categorical cross entropy since the output has three categories (Anesthesia Ok, Anesthesia Deep and Anesthesia Light). The activation functions used were ReLU and tanh and the optimizer used is Adam. The learning rate was varied from 0.001 to 0.0005. For faster computation and better accuracy, a learning rate of 0.005 was used. The batch size of 64, 128, and 256 were experimented. Finally, the best accuracy was achieved using a batch size of 64. Each model was trained for 50 epochs. Model checkpoint was used to save the weight every epoch for testing. Further, the validation loss was checked using ReduceLROnPlateau, that is, the learning rate was decreased by a factor of 0.1 whenever the loss function remained around the same value for more than 2 epochs. The minimum learning rate was kept to 0.00001. Tensorflow was used as the backend of the training process. The inputs to the models are the images of heatmaps in the RGB format. The VGG and AlexNet models used a 224 $\times$ 224 size of the input images and the rest of the models used 128 $\times$ 128 size. The input and output arrays are a 4D array which consists of the batch size, input width, input eight and input depth. The batch size is flexible and is fed during the training process. The image width and height are 224, 224 or 128, 128 depending on the model. The depth of the image is 3 since we have used an RGB image. Finally, Softmax as the activation for the fully connected layers since the categorization was for three categories.

Fig. 4 shows the structure of each model used. 'InputLayer' represents the heatmap images that are fed in each model; 'Conv2d' represents the convolution layer; 'Activation' represents the activation function used in the convolution layer; 'MaxPooling2D', 'Batch Normalization' and 'Dropout' are used in each respective Convolution layer to prevent overfitting and reduce the computational cost; 'Flatten' represents the conversion of 2D convolution layers to 1 dimension; 'Dense' represents the fully connected layers. Each blue box represents one convolution block, and the orange box represents one Fully connected block. In AlexNet, 5 convolutional layers have been used according to the predefined structure with 96, 256, 384, 384 and 256 kernels respectively. We have fine-tuned the fully connected layers using three fully connected layers with 4096, 4096 and 1000 neurons respectively. Dropout was used to decrease the number of samples. The activation function used was tanh. In the 6 layered CNN, 6 convolutional layers have been used with 32, 32, 64, 64, 128, 128 kernels respectively. There is one dense layer with 1000 neurons. In the 8 layered CNN, 8 convolutional layers have been used with 64,64, 96, 96, 256,256, 384 and 384 kernels respectively. There is one fully connected layer consisting of 1000 neurons. In the 10 layered CNN model, 10 convolutional layers have been used with 32, 32, 64, 64, 96, 96, 128, 128, 256, and 256 filters respectively. There is one dense layer of size 1000. In the 6, 8 and 10 layered CNN, ReLU activation function is used in the convolution blocks. VGG19 is a CNN model with 19 layers - 16 convolutional layers and 3 fully connected layers. The convolutional layers were pre-trained

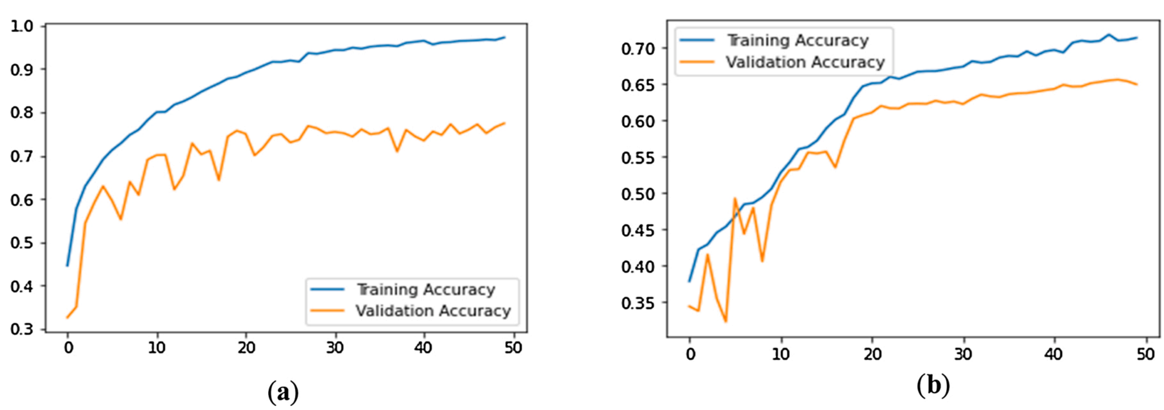

Fig. 6. Accuracy plot of the best model (10 layered CNN) of (**a**) ECG; (**b**) PPG.

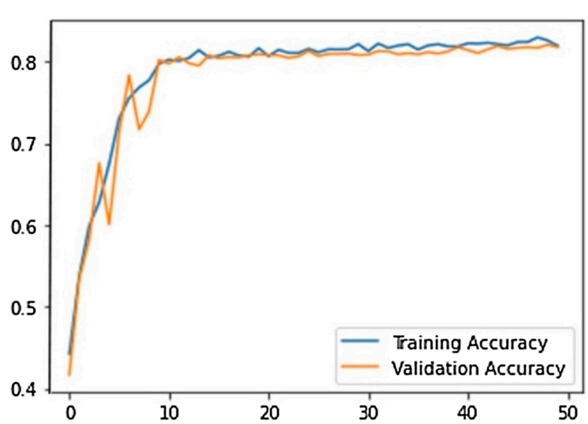

Fig. 7. Accuracy plot of the best model (10 layered CNN) – ECG and PPG subplot.

Table 4
ECG and PPG subplot accuracy.

Model Structure	Accuracy (%)
AlexNet	69
Six layered model	75
Eight layered model	78
Ten layered model	86
VGG	67

using the ImageNet dataset weights, and the fully connected layers are fine-tuned according to the model. In this architecture, two fully connected layers of size 512 and 128 neurons are used respectively.

The last type of structure used is Two Channel CNN. It is a type of deep learning structure where we feed in two types of data parallelly. The demonstration can be referred to in [47] A similar structure was used in our analysis. ECG and PPG heatmaps are used as inputs to our model as shown in Figs. 2 and 3. The convolutional layers take place individually and the output layers of each input are finally concatenated as shown in Fig. 5(a), (b) and (c). After this, the fully connected layers are performed as one as shown in Fig. 5(d). Similarly, the rest of the models that are described above were performed using two-channel CNN, where the convolution layers were performed individually. The outputs of the two blocks were concatenated and the fully connected layers were performed together.

3. Results

In this study, the depth of anesthesia was predicted based on BIS values acquired. The heatmaps generated from the signals of PPG and ECG were used as inputs data to our model. Various deep learning models like AlexNet and VGG19 were used. Besides, a 6 layered, 8 layered, and 10 layered model were also used. VGG19 model was pre-trained using the ImageNet weights and the rest were trained from scratch. A total number of 14,145 images were used, and the training and testing sets were split in 80 % and 20 %. Each model was trained for 50 epochs with a batch size of 64. AlexNet and the 6 layered model took around 1.5 h for training, the 8 and 10 layered models took around 2 h to train. The VGG19 model took 1.5 h for training. First, a smaller dataset of 25 patients was used for training, followed by evaluating the complete dataset of 50 patients. A significant difference in accuracy was observed, probably due to the variance.

In all the 3 experiments, we observe that the 10 layered model gives us the maximum accuracy compared to the other models. In the first experiment with ECG and PPG inputs, we get an accuracy of 77 % and 65 % respectively with this model. With the subplots of ECG and PPG as inputs, the accuracy achieved with this model is 86 % and with Two-channel CNN we get an accuracy of 75 %.

VGG19, despite having the maximum number of layers, has given us a relatively lower accuracy in all the three experiments conducted. With ECG inputs the accuracy is 67 %, with PPG 62 %. With ECG & PPG subplots and two-channel CNN, an accuracy of 70 % and 65 % are achieved respectively while using VGG19. This was because of the pre-trained weights. The GPU constraint restricted us from training such a complex model from scratch.

Overall, in this research focus is on the accuracy when ECG and PPG images are fed together, as this gives a good insight of hypnosis components of anesthesia. We get an accuracy of 86 % when ECG and PPG data are fed as subplots in a 10 layered model. With two-channel CNN, the maximum accuracy achieved is 75 %. Again, this was attained while using a 10 layered model.

3.1. ECG and PPG analysis result

The maximum accuracy was achieved when we used a 10 layered model. Table 3 shows the accuracy achieved using different models. It was observed that the ECG signals more accurately predict DoA compared to the PPG signals. This result matches with previous studies conducted. Fig. 6 shows the accuracy plots over 50 epochs during the training of the 10 layered model of ECG and PPG signal.

3.2. ECG and PPG subplot analysis result

This experiment gives us maximum accuracy. ECG and PPG signal heatmaps produced in MATLAB were combined into a single image as a subplot and have been used to predict DoA. The heatmap on top represents the ECG signal, and the plots on the bottom of each image represent the PPG signals. The maximum accuracy of 86 % was achieved when we used 50 patients' data with a 10 layered model. The accuracy plot of this model over 50 epochs is shown in Fig. 7. Table 4 shows the accuracy table with each model implemented.

3.3. Two-channel CNN analysis result

In this experiment, we have used ECG and PPG images as described

Table 5
Two-channel CNN accuracy.

Model Structure	Accuracy (%)
Alexnet	69
Six layered model	67
Eight layered model	73
Ten layered model	75
VGG	65

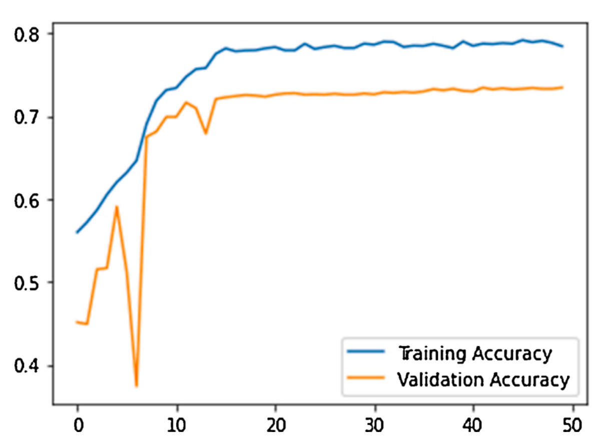

Fig. 8. Accuracy Plot of the best model (10 layered CNN) – two channel CNN.

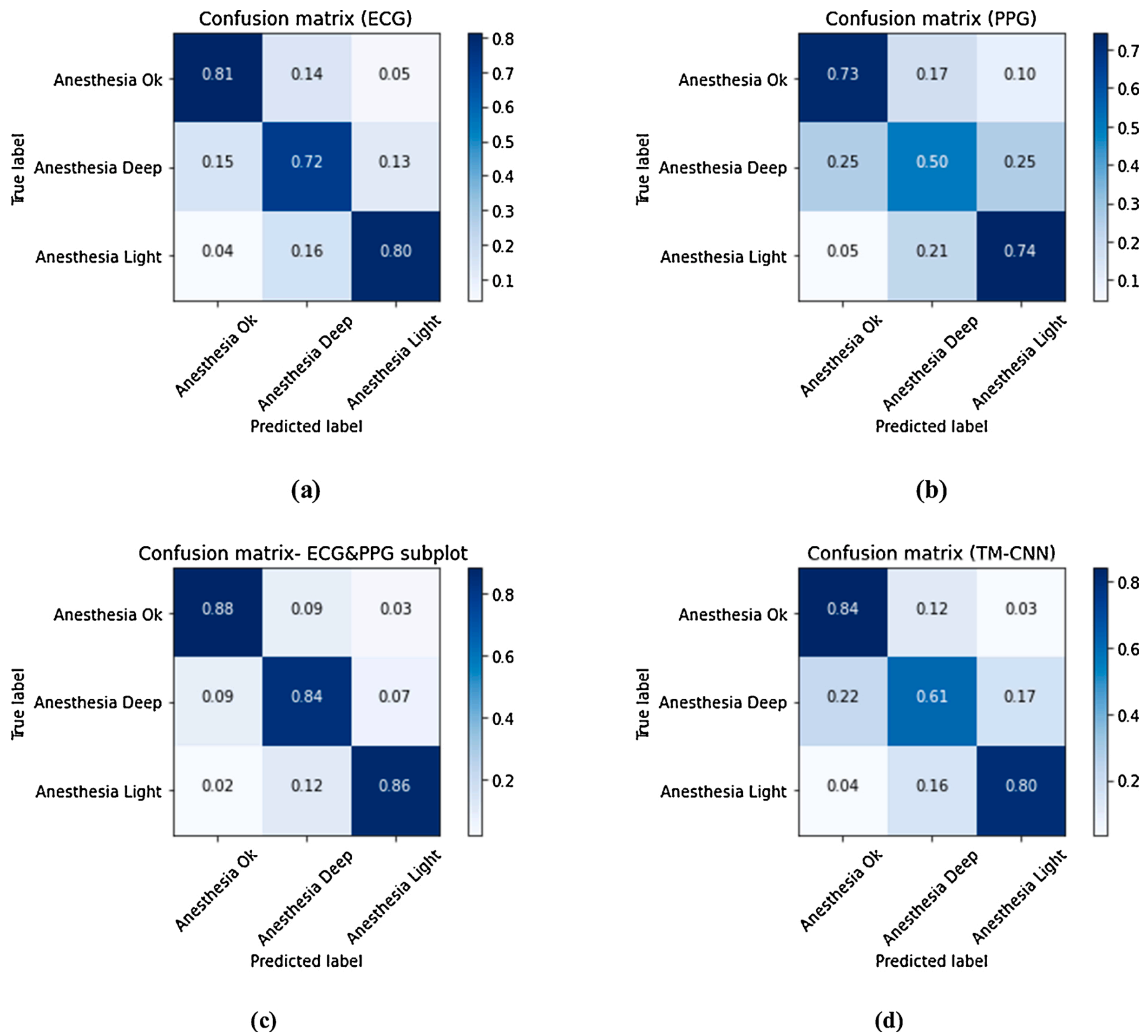

Fig. 9. Confusion Matrices of 10 layered CNN with normalized values of (**a**) ECG; (**b**)PPG; (**c**)ECG-PPG Subplot; (**d**) Two-channel CNN.

in Section 2.3. The maximum accuracy of 75 % was achieved when we used 50 patients' data with a 10 layered model. The accuracy plot of this model over 50 epochs is shown in Fig. 8. The accuracy of Two-channel CNN with all the models described in Section 2.3 is given in Table 5.

In order to verify the results, normalized confusion matrices of the best model, that is the 10 layered CNN, from each experiment are plotted. The confusion matrix shows the True Positives, True Negatives, False Positive and False Negatives of each category of classification. From Fig. 9, it can be concluded that the True Positives of the category Anesthesia Ok are the highest in all the cases, whereas Anesthesia Deep has the least number of True positive predictions among the 4 experiments conducted. This is because most of the BIS values 36.8 ± 4.6, which is very close to the BIS range of Anesthesia Ok. The images of Anesthesia Light, on the other hand had the values 67.0 ± 11.4, so the models could differentiate it from the other two categories.

Table 6
Summary of accuracy.

	AlexNet	6 layered CNN	8 layered CNN	10 layered CNN	VGG
ECG	72 %	72 %	75 %	77 %	70 %
PPG	62 %	63 %	64 %	65 %	62 %
ECG and PPG	69 %	75 %	78 %	86 %	67 %
Two-channel CNN	69 %	67 %	73 %	75 %	65 %

4. Discussion

In all, we have produced a good result in predicting DoA with ECG and PPG signal heatmaps used as a single image. In this research we have performed an analysis of ECG and PPG data separately and together using subplots and a two-channel CNN, to predict the depth of anesthesia (DoA). Table 6 shows a summary of the accuracy of all the experiments conducted. Various complex models were used consisting of a different number of layers (5, 6, 8, 10 and 19). The models with 5, 6, 8 and 10 layers were trained from scratch and 19 layered model VGG19 was trained using pre-trained weights of the ImageNet dataset due to constraints with GPU computational time and memory. Checkpoints have been used to store weights with the best validation accuracy after every epoch. The validation loss was monitored by reducing the learning rate during training to prevent overfitting. MATLAB is used to generate the heatmaps and the model was trained on Google Colab, which provides users with free access to Google GPU and TPU. Hence, smaller hospitals also can make use of this free tool to train a CNN model from scratch. Moreover, once a model is trained, it is not required to retrain the model fully. Instead, fine tuning the fully connected layers will

Table 7
Comparison of previous studies with current proposal.

Literature Study	Method Used	Vital Signs used as input	Result	Time and GPU constraints
[5]	Data Reconstruction: Short Time Fourier Transform Classification Algorithm: Convolutional Neural Network	EEG	93.5 % Accuracy	3.4 h on Nvidia Tesla k40 GPU
[25]	Data Reconstruction: Empirical Mode Decomposition of EEG Classification Algorithm: Artificial Neural Network	EEG, EMG, HR, pulse, SBP, DBP	MAE of 6.54 with 6.69 of standard deviation	10,000 epochs, small learning rate of 0.005. (Computational time consideration due to the epoch number is ignored)
Our Study	Data Reconstruction: None Classification Algorithm: Convolutional Neural Network	ECG and PPG signals	86 % Accuracy	2 h on Google Colab

suffice. This way, the best model can be modified based on new heatmaps generated with less GPU and timing constraints.

In [25], the authors have done a study to predict DoA via Artificial Neural Network using EMD processed EEG signal combined with other mean values of vital signs like EMG, HR, pulse, SBP, DBP, and SQI to evaluate the DoA index, and have concluded that ANN has a lower error than BIS. In order to get the precise model, computational time consideration, due to the large epoch number, was ignored thus increasing the memory and timing constraints. The study conducted in [5] has achieved a 93.5 % accuracy by using CNN with processed EEG signal to predict the DoA with respect to the BIS values. On comparing the results with our study, we see that the accuracy in our case is around 10 % less. This is because basic anesthesia occurs in the brain and spinal cord and so EEG signals are intensively used as a measure to administer anesthesia. However, ECG and PPG signals overcome various shortcomings of the EEG signal, like better feasibility of ECG & PPG than EEG, higher energy levels of ECG signals, better resistance to noise, easy and inexpensive acquisition of the former signals while evaluating the DoA. EEG signals also require more preprocessing like denoising, Empirical Mode Decomposition or short time Fourier Transform, unlike ECG and PPG signals which can be used directly. Moreover, the GPU and time constraints in [5] and [25] are very high and requires expensive hardware. Overall, our study uses inexpensive signals, minimum data reconstruction, minimum memory and timing constrains to achieve a decent accuracy of 86 % to categorize the anesthetic levels to Anesthetic Deep, Anesthetic Light and Anesthetic Ok. Table 7 shows the comparison of our study with the previously done work. Overall, this research shows a good potential in future clinical practices.

With more research into improving the accuracy by using more vital signs like electromyography (EMG), pulse, systolic blood pressure (SBP), diastolic blood pressure (DBP), etc our model can be deployed as a measure of the DoA instead of the currently used BIS Quantro. In the future, the aim is to investigate the degree of analgesia using other measures like Analgesia Nociception Index machine designed by Mdoloris Medical Systems [48] to train this deep learning model via ECG and PPG signals as well. Hence, with ECG and PPG being signals which can detect hypnosis and analgesia, the right amount of anesthesia can be administered to the patient. This can serve as an effective and automated tool in the surgical room. The overfitting can be reduced and hence the accuracy can be further increased by using more patient data, further fine-tuning (Changing learning rates, momentum), train each patient separately, and get the best model to test the other patients. Transfer learning method can be applied for pre-trained models to speed up the training process and hence improve the accuracy. Furthermore, Recurrent Neural Networks or LSTMs can also be deployed instead of CNN to improve accuracy. In all, this study can be improved by training Deep Learning models with a different number of layers and hyperparameters from scratch.

5. Conclusions

After performing various deep learning models, we can conclude that our study has given us significant results. A maximum accuracy of 86 % is achieved when a 10 layered model is used. It was observed that ECG readings are more important than PPG to predict DoA as the individual accuracy was higher in ECG than PPG. However, a combination of ECG and PPG as input proves to be more efficient than a single signal. Two-channel CNN is an efficient model, however, with more fine-tuning and more patient data we can achieve greater accuracy. Pretrained Imagenet weights do not work well with our image data and it is better to train models from scratch. Being a temporal dataset, better accuracy can be achieved with a larger sample size. By modifying the training step, batch size, optimizer, activation function, kernel size, learning rates, and number of epochs we can have better accuracy. In all, through this research we have found a cost-effective and automated way to predict the depth of anesthesia. Moreover, the vital signs ECG and PPG are not only used to measure the hypnosis aspect of anesthesia, but also the analgesic component in the near future. This method will help anesthesiologists in future clinical practices.

Author contributions

Conceptualization, J.-S.S. and M.F.A.; data curation, R.M. and S.-Z.F.; methodology, M.R.C., R.M, M.F.A, S.-Z.F. and J.-S.S; software, M.R.C. and R.M.; validation, S.-Z.F., J.-S.S. and M.F.A.; formal analysis, J-S.S. and M.F.A.; investigation, M.R.C., R.M. and J.-S.S.; writing—original draft preparation, M.R.C; writing—review and editing, J.-S.S., M.F.A.; supervision, S.-Z.F., J.-S.S. and M.F.A. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgement

This work was financially supported by the Ministry of Science and Technology, Taiwan (Grant number: MOST 107-2221-E-155-009-MY2).

Declaration of Competing Interest

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.

References

[1] C.M. Alwardt, D. Redford, D.F. Larson, General anesthesia in cardiac surgery: a review of drugs and practices, J. Extra-Corp. Technol. 37 (2005) 227–235.
[2] J.Y. Lan, M.F. Abbod, R.G. Yeh, S.Z. Fan, J.S. Shieh, Intelligent modeling and control in anesthesia, J. Med. Biol. Eng. 32 (2012) 293–308.
[3] M. Gruenewald, C. Ilies, J. Herz, T. Schoenherr, A. Fudickar, J. Höcker, B. Bein, Influence of nociceptive stimulation on analgesia nociception index (ANI) during propofol–remifentanil anaesthesia, Br. J. Anaesth. 110 (6) (2013) 1024–1030.

[4] H.-H. Huang, Y.-H. Lee, H.-L. Chan, Y.-P. Wang, C.-H. Huang, S.-Z. Fan, Using a short-term parameter of heart rate variability to distinguish awake from isoflurane anesthetic states, Med. Biol. Eng. Comput. 46 (10) (2008) 977–984.
[5] Q. Liu, J. Cai, S.-Z. Fan, M.F. Abbod, J.-S. Shieh, Y. Kung, L. Lin, Spectrum analysis of EEG signals using CNN to model patient's consciousness level based on anesthesiologists' experience, IEEE Access 7 (2019) 53731–53742.
[6] C. Prys-Roberts, Anaesthesia: a practical or impractical construct? Br. J. Anaesth. 59 (1987) 1341–l345.
[7] Kissin, Depth of anesthesia and bispectral index monitoring, Anesth. Analg. 90 (5) (2000) 1114–1117.
[8] Medical Advisory Secretariat, Bispectral index monitor: an evidence-based analysis, Ontario Health Technol. Assess. Ser. 2004 4 (9) (2004) 1–70.
[9] Q. Liu, Y.-F. Chen, S.-Z. Fan, M.F. Abbod, J.-S. Shieh, Quasi-periodicities detection using phase-rectified signal averaging in EEG signals as a depth of anesthesia monitor, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (10) (2017) 1773–1784.
[10] Q. Liu, L. Ma, R.C. Chiu, S.Z. Fan, M.F. Abbod, J.S. Shieh, HRV-derived data similarity and distribution index based on ensemble neural network for measuring depth of anaesthesia, PeerJ 2017 (5) (2017) 4067.
[11] I.J. Rampil, A primer for EEG signal processing in anesthesia, Anesthesiology 89 (4) (1998) 980–1002.
[12] R. Shalbaf, H. Behnam, J.W. Sleigh, L.J. Voss, Using the Hilbert–Huang transform to measure the electroencephalographic effect of propofol, Physiol. Meas. 33 (2012) 271–285.
[13] R. Shalbaf, H. Behnam, H. Jelveh Moghadam, Monitoring depth of anesthesia using combination of EEG measure and hemodynamic variables, Cogn. Neurodyn. 9 (1) (2015) 41–51.
[14] L. Dou, H. Gao, L. Lu, W. Chang, Bispectral index in predicting the prognosis of patients with coma in intensive care unit, World J. Emerg. Med. 5 (1) (2014) 53–56.
[15] G. Schneider, K. Wagner, W. Reeker, F. Hänel, C. Werner, E. Kochs, Bispectral index (BIS) may not predict awareness reaction to intubation in surgical patients, J. Neurosurg. Anesthesiol. 14 (1) (2002) 7–11.
[16] S. Kreuer, A. Biedler, R. Larsen, S. Schoth, S. Altmann, W. Wilhelm, The narcotrendTM - a new EEG monitor designed to measure the depth of anaesthesia - a comparison with bispectral index monitoring during propofol-remifentanil-anaesthesia, Anaesthesist 50 (12) (2001) 921–925.
[17] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377.
[18] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (1) (2019) 65–69.
[19] A. Ortiz, J. Munilla, M. Martínez-Ibañez, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, Parkinson's disease detection using isosurfaces-based features and convolutional neural networks, Front. Neuroinform. 13 (2019) 48.
[20] F. Andreotti, O. Carr, M.A.F. Pimentel, A. Mahdi, M. de Vos, in: Comparing Feature Based Classifiers and Convolutional Neural Networks to Detect Arrhythmia from Short Segments of ECG, 2017 Couputing in Cardiology, Rennes, France, 24-27 September, 2017.
[21] S. Parvaneh, J. Rubin, S. Babaeizadeh, M. Xu-Wilson, Cardiac arrhythmia detection using deep learning: a review, J. Electrocardiol. 57S (2019) S70–S74.
[22] G. Ruffini, D. Ibañez, M. Castellano, L. Dubreuil-Vall, A. Soria-Frisch, R. Postuma, J.-F. Gagnon, J. Montplaisir, Deep learning with EEG spectrograms in rapid eye movement behavior disorder, Front. Neurol. 10 (2019) 806.
[23] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (2018) 270–278.
[24] M. Tjepkema-Cloostermans, J. Hofmeijer, B. Ruijter, A. Beishuizen, F. Bosch, M. van Putten, O77 Deep learning for EEG classification for outcome prediction of postanoxic coma, Clin. Neurophysiol. 128 (9) (2017) e203.
[25] M. Sadrawi, S.Z. Fan, M.F. Abbod, K.K. Jen, J.S. Shieh, Computational depth of anesthesia via multiple vital signs based on artificial neural networks, Biomed Res. Int. 2015 (2015), 536863.
[26] T. Horiguchi, T. Nishikawa, Heart rate response to intravenous atropine during propofol anesthesia, Anesth. Analg. 95 (2) (2002) 389–392.
[27] T. Horiguchi, T. Nishikawa, Propofol-nitrous oxide anesthesia enhances the heart rate response to intravenous isoproterenol infusion, Anesth. Analg. 96 (1) (2003) 132–135.
[28] E.N. Brown, P.L. Purdon, C.J. van Dort, General anesthesia and altered states of arousal: a systems neuroscience analysis, Annu. Rev. Neurosci. 34 (2011) 601–628.
[29] S. Ching, E.N. Brown, Modeling the dynamical effects of anesthesia on brain circuits, Curr. Opin. Neurobiol. 25 (2014) 116–122.
[30] E. Niedermeyer, F.L. Da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Lippincott Williams & Wilkins, Baltimore, 2005.
[31] A.-H. Samarkandi, The bispectral index system in pediatrics-is it related to the end-tidal concentration of inhalation anesthetics? Middle East J. Anaesthesiol. 18 (4) (2006) 769–778.
[32] M.S. Avidan, L. Zhang, B.A. Burnside, K.J. Finkel, A.C. Searleman, J.A. Selvidge, L. Saager, M.S. Turner, S. Rao, M. Bottros, C. Hantler, E. Jacobsohn, A.S. Evers, Anesthesia awareness and the bispectral index, N. Engl. J. Med. 358 (11) (2008) 1097–1108.
[33] J.W. Sleigh, J.P.M. Barnard, Entropy is blind to nitrous oxide. Can we see why? Br. J. Anaesth. 92 (2) (2004) 159–161.
[34] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, H. Nazeran, A review on wearable photoplethysmography sensors and their potential future applications in health care, Int. J. Biosens. Bioelectron. 4 (4) (2018) 195–202.
[35] E.K. Naeini, I. Azimi, A.M. Rahmani, P. Liljeberg, N. Dutt, A real-time PPG quality assessment approach for healthcare internet-of-things, Procedia Comput. Sci. 151 (2019) 551–558.
[36] P. Dash, Electrocardiogram monitoring, Indian J. Anaesth. 46 (2002) 251–260.
[37] Y.-H. Byeon, S.-B. Pan, K.-C. Kwak, Intelligent deep models based on scalograms of electrocardiogram signals for biometrics, Sensors 19 (4) (2019) 935.
[38] A.F. Merry, J.B. Cooper, O. Soyannwo, I.H. Wilson, J.H. Eichhorn, International standards for a safe practice of anesthesia 2010, Can. J. Anesthesia 57 (11) (2010) 1027–1034.
[39] M. Oji, Y. Terao, T. Toyoda, T. Kuriyama, K. Miura, M. Fukusaki, K. Sumikawa, Differential effects of propofol and sevoflurane on QT interval during anesthetic induction, J. Clin. Monit. Comput. 27 (3) (2013) 243–248.
[40] H.-M. Wang, S.-C. Huang, Musical rhythms affect heart rate variability: algorithm and models, Adv. Electr. Eng. 2014 (2014), 851796.
[41] M. Elgendi, On the analysis of fingertip photoplethysmogram signals, Curr. Cardiol. Rev. 8 (1) (2012) 14–25.
[42] R. Sahni, Noninvasive monitoring by photoplethysmography, Clin. Perinatol. 39 (3) (2012) 573–583.
[43] L.Ø. Høiseth, I.E. Hoff, Ø. Skare, K.A. Kirkebøen, S.A. Landsverk, Photoplethysmographic and pulse pressure variations during abdominal surgery, Acta Anaesthesiol. Scand. 55 (10) (2011) 1221–1230.
[44] L. Zhang, L. Xu, J. Zhu, Y. Gao, Z. Luo, H. Wang, Z. Zhu, Y. Yu, H. Shi, H. Bao, To clarify features of photoplethysmography in monitoring balanced anesthesia, compared with cerebral state index, Med. Sci. Monit. 20 (2014) 481–486.
[45] M. Huiku, K. Uutela, M. van Gils, I. Korhonen, M. Kymäläinen, P. Meriläinen, M. Paloheimo, M. Rantanen, P. Takala, H. Viertiö-Oja, A. Yli-Hankala, Assessment of surgical stress during general anaesthesia, Br. J. Anaesth. 98 (4) (2007) 447–455.
[46] Q. Liu, L. Ma, S.-Z. Fan, M.F. Abbod, C.-W. Lu, T.-Y. Lin, K.-K. Jen, S.-J. Wu, J.-S. Shieh, Design and evaluation of a real time physiological signals acquisition system implemented in multi-operating rooms for anesthesia, J. Med. Syst. 42 (8) (2018) 148.
[47] R. Ke, W. Li, Z. Cui, Y. Wang, Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact, Transp. Res. Rec. (2020) 1–12.
[48] R. Abdullayev, O. Uludag, B. Celik, Índice de Analgesia/Nocicepção: avaliação da dor aguda pós-operatória, Braz. J. Anesthesiol. 69 (4) (2019) 396–402.