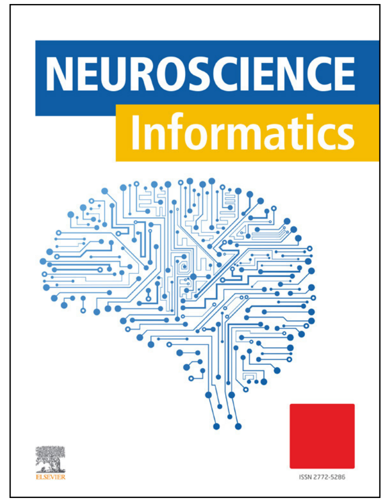
Journal Pre-proof



Detecting the Depth of Sedation in the Intensive Care Unit using a 2-channel Electroencephalogram: An analysis with 2 machine learning models

Esteban A. Alarcón-Braga, Samuel Gruffaz, Cécile Delagarde, Axel Roques, Jean-Clément Riff et al.

PII: S2772-5286(25)00053-6
DOI: https://doi.org/10.1016/j.neuri.2025.100238
Reference: NEURI 100238

To appear in: *Neuroscience Informatics*

Received date: 8 October 2025
Accepted date: 23 October 2025

Highlights

- The 2-channel EEG has a moderate power to discern between sedation states.
- The individual feature with the best performance was Beta Power–EEG2.
- A multimodal framework with variables not only derived from the EEG is needed.

Detecting the Depth of Sedation in the Intensive Care Unit using a 2-channel Electroencephalogram: An analysis with 2 machine learning models

ABSTRACT

Existing methods to detect depth of sedation do not fully adjust to the characteristics of the ICU population. The aim of this study is to evaluate the performance of a two-channel EEG in predicting the depth of sedation in ICU patients. The electroencephalographic signal of 21 patients admitted to the ICU were analyzed, and EEG features were calculated. These served as inputs in 2 machine learning models: Random Forest Classifier (RFC) and Support Vector Machine (SVM). The depth of sedation was assessed using the Richmond Agitation-Sedation Scale (RASS). Patients with RASS scores of -4/-5 were classified as “Deeply Sedated”, otherwise they were classified as “Not Deeply Sedated”. In the general models, all EEG features were used, after which sequential feature selection was conducted to improve performance and reduce the number of variables (reduced models). The general models showed a moderate ability to discriminate between sedation categories (RFC: average F1-score =0.60, SVM: average F1-score =0.59). This ability was improved in the reduced models (RFC: average F1-score =0.65, SVM: average F1-score =0.72). It was observed that decreasing the number of features in the reduced SVM model from 6 to 3 features could achieve a similar performance while simplifying the model (SVM: average F1-score =0.72). An exploratory analysis showed that the individual feature with the best performance was Beta Power–EEG2. Overall, the 2-channel EEG has a moderate power to discriminate between different states of sedation and may not be useful in this purpose if used as a single predictor.

Keywords: Machine learning, Electroencephalogram, Depth of sedation, Intensive Care Unit, Support Vector Machine, Random Forest Classifier.

INTRODUCTION

The administration of analgesics and sedatives is a necessary action in the Intensive Care Unit (ICU), required for conducting both diagnostic and therapeutic procedures, as well as for managing pain and anxiety. Ensuring the correct dosage of sedative drugs is important to achieve the desired depth of sedation (DoS) and avoid oversedation, as this has been associated with serious adverse events, such as: longer length of stay, increased duration of mechanical ventilation, and an overall increase in healthcare costs [1–4].

Currently, there is no gold standard for measuring the DoS, and the most used method worldwide are behavioral scales, such as the Richmond Agitation-Sedation Scale (RASS) [5]. However, these methods are time consuming for the ICU staff and can lack sensitivity to quickly detect transient or slight changes in the DoS. Some of the most used technologies include the following: Bispectral Monitor, Narcotrend Monitor, PSA 4000 Monitor, among others [6]. These monitors, through an algorithm, end up calculating indices that supposedly reflect the DoS.

One of the most commonly used index for assessing the DoS is the bispectral index (BIS). This index was originally developed to assess depth of anesthesia (DoA) in patients under general anesthesia (GA), and therefore it suffers from certain limitations when used in the ICU. For example, in the ICU, patients are not often paralyzed, so contamination with artifacts caused by muscle movements can be problematic for the BIS [7]. Furthermore, the BIS has proven to be slow in detecting changes in the state of consciousness, as well as presenting variations in its values depending on the anesthetic/sedative agent administered to the patient [8,9]. In addition to the BIS, there are also other indices on the market such as the Narcotrend Index (NI) or the Patient State Index (PSI), which also lack extensive validation in the ICU. These three indices are calculated using proprietary algorithms, so physicians do not know how they are calculated, and therefore it is impossible to make modifications or improvements to adjust these indices to the characteristics of the ICU population [10–12].

The use of electroencephalographic (EEG) features for predicting the DoA in the operating room has been studied on multiple occasions in the scientific literature. In an article published by Gu et al. [13], the predictive performance of 4 EEG features (permutation entropy, 95% spectral edge frequency, SynchFastSlow, and Beta Ratio) was analyzed using an artificial neural network. Their findings revealed an 84.4% accuracy in identifying patients under GA and a 14% accuracy in detecting those under "deep anesthesia"[13]. Ortolani et al. [14] conducted a similar study in 2002, where they evaluated the predictive capabilities of 13 EEG features. Here the model showed an accuracy of 95%, and it bore a strong correlation with the BIS (Pearson's correlation coefficient = 0.94) [14]. Apart from this, it has been observed previously by our team [15], that the DoA can be accurately assessed using a 1-channel EEG in the surgical context (Channel F8, AUC: 0.92 ±0.04) [15].

The direct use of EEG features to predict the DoS in the ICU, has not been widely studied. Among the limited available literature, we can find a study conducted by Nagaraj et al. [16] where the method of atomic decomposition is used to derive multiple features of the frontal EEG and thus attempt to predict whether the patient is awake or sedated. The best performance in this experiment resulted in a mean AUC of 0.91. However, this technique suffers from limitations. For instance, the authors did not use RASS scores of -2 or -3 in the assessment of the models, which diminishes its clinical utility. Additionally, they used multiple EEG channels, which complicates its use in the daily practice of the ICU [17].

The use of a 1-channel EEG to assess DoS, in addition to being simple and precise, promises to be accessible, inexpensive, and addresses the limitations of the previously mentioned indexes. All of this would allow for its widespread use in the clinical setting. However, to date, there are no studies that have evaluated the use of this tool in the ICU. Therefore, the main objective of this study is to assess the accuracy of EEG in predicting the DoS in patients admitted to the ICU. In this case, we decided to use a 2-channel EEG instead of a 1-channel EEG, with the purpose of enhancing the quality of the signal and acquiring more spatial information.

METHODOLOGY

Study Design

This is a single center, retrospective study conducted in the ICU of the Bégin Military Hospital between October 2021 and April 2024. The study has been approved by Pr. JE Bazin, chair of the ethics committee of the French Society of Anesthesiology (SFAR) under the number IRB 00010254-2016–2018.

Study Population

This study included patients admitted to the ICU aged ≥ 18 and ≤ 80, who required sedation. We included patients from both medical and surgical backgrounds. We excluded patients with the following characteristics: 1) Patients admitted to the ICU due to brain injuries (stroke, traumatic brain injury, etc.) or with a history of these, or 2) patients that are chronic users of medication known to influence the central nervous system.

Sedation Protocol

The desired DoS for each patient was determined by the attending physician and the dose of each sedative/analgesic agent was adjusted accordingly. The sedative/analgesic agents used in the ICU were the following: Propofol (0.5-4 mg/kg/h), dexmedetomidine (0.7-1.4 mcg/kg/h), sufentanil (0.1-0.3 mcg/kg/h), midazolam (0.1 -0.3 mg/kg/h) and ketamine (load: 1-4.5 mg/kg, maintenance: 0.1-0.5 mg/kg/h). The exact combination in which this medication was administered varied according to the necessity of each patient and frequently evolved from one day to another.

Assessment of Consciousness

The DoS is routinely assessed in the unit using the RASS and the Glasgow Coma Scale (GCS). The RASS classifies the state of consciousness in 10 different levels that ranges from +4 to -5, where a score of -5 is synonymous of an unarousable patient and a score of +4 is synonymous of a combative one [18]. The GCS is used to describe the extent of impaired consciousness, with values ranging from 3 to 15; it takes into consideration 3 parameters: eye-opening response, verbal response, and motor response [19]. The consciousness assessments were conducted twice (once before a standardized stimulation and once after the stimulation) every 30 minutes by trained personnel (attending physician or nurse in charge). In this case, the stimulation was done by tapping the patient's forehead five times with the tips of the index and middle fingers. At the same time, the dose of medication reported in the medical files for each patient was extracted, as well as any event that could have altered the patient's level of consciousness (e.g. change in position, oral/tracheal suctioning, etc.).

EEG Acquisition

A 2-channel EEG, placed at position Fp1 and C3 with a common reference, was used to record brain activity. The reference and the mass electrodes were placed at position A2 and FpZ, respectively. The impedance of the electrodes remained below 2 kΩ. The EEG signal was recorded at 100 Hz.

Data Collection

After the end of each recording, the obtained data was saved and subsequently transformed into CSV files for its processing and analysis.

EEG Data Preprocessing

For the signal filtering, we removed the spikes and applied a Butterworth bandpass filter, with cut-off frequencies set at 0.5 – 30 Hz. The decision to retain only frequencies within this range stemmed from 2 factor. First, above 20 Hz, there is little neural activity necessary for sedation monitoring, and second, because certain drugs used during the patient’s sedation, such as ketamine, can induce higher-frequency oscillations [16,20].

EEG Processing

With the aim of differentiating each state (awake/light/moderate sedation vs deep sedation) it was decided to extract the following features from the EEG for each channel (EEG1 for Fp1 and EEG2 for C3): Standard deviation, Root mean square, Skewness, Kurtosis, Crest Factor, absolute mean, mean value of power spectrum delta (delta power), mean value of power spectrum theta (theta power), mean value of power spectrum alpha (alpha power), mean value of power spectrum beta (beta power), sample entropy, spectral entropy, renyi entropy, approximate entropy, and permutation entropy.

Data Analysis

The capacity of the EEG features to predict the DoS was assessed using two different machine learning models. In both cases, the initial task to perform was the binary categorization of the DoS (the dependent variable). With this goal in mind, patients with a spontaneous RASS (RASS before the standardized stimulation) equal to or less than -4 were categorized as "deeply sedated", while patients with a spontaneous RASS greater to or equal to -3 were categorized as "not deeply sedated". Therefore, patients that were awake, lightly sedated, or moderately sedated were categorized in this last class. In this case, the RASS was chosen over the GCS as the preferred scale for the categorization of the dependent variable, since unlike the GCS, the RASS was created specifically to assesses the patient's sedation depth. Thus, the predictions made by our models will be contrasted against the spontaneous RASS, which is our gold standard.

After the categorization of the DoS, a Random Forest Classifier (RFC) and Support Vector Machine (SVM) models were developed. For both models, in case of missing values, observations were withdrawn from the dataset. Since the dataset consists of EEG recordings, each patient has multiple observations, and the duration of each recording varied from patient to patient. Additionally, patients in the ICU often require invasive procedures due to the severity of their pathologies. As a result, they are more frequently required to be placed under deep sedation rather than in light or moderate sedation. This caused an imbalance in the classes of the dataset, which was mitigated when developing the predictive models. With the purpose of balancing the dataset, we decided to use the Synthetic Minority Over-sampling Technique (SMOTE) when training the models, as well as employing a stratified group K-fold cross-validation (K = 5) as the resampling method, in order to achieve a representative distribution of the different classes across each fold. It is important to highlight that the stratification was done on patients.

In the initial models we used the 30 EEG features that were calculated, from this moment forward we will refer to these initial models as the general models. The performance of the models was assessed using both, the average accuracy and the average F1 score across all folds. In addition to this, ROC curves with their respective mean areas under the curve (AUC) were calculated, as well as the average confusion matrix across the different test splits (only performed in the reduced SVM model with 3 variables) and the average F1 scores per class. Hyperparameter tuning was performed for each model with the aim of achieving the best possible performance.

The importance of each feature in the models was assessed using Shapley Additive Explanations (Shapley/Shap Values). One of the most significant advantages of this technique is its applicability across various types of machine learning models, enabling a more consistent comparison between the features of each model. To facilitate this comparison, the absolute normalized mean Shap Values per feature were computed and visualized on a bar graph. In this case, L1-Normalization was used so that the sum of the contributions from all features amounts to a total of 1, allowing the contributions of each feature to the model output to be expressed as percentages.

With the mission of developing the prediction model with the best performance and reducing the number of features, we decided to use the sequential forward feature selection method. The models after the feature selection contained a reduced number of features, thus from this moment forward they will be referred to as the reduced models. The importance of each feature in these reduced models was also assessed using Shapley values. Hyperparameters were also modified to find those that provided the best predictive performance.

Similar to the general model, the average F1 scores and average accuracy were calculated across all folds, as well as the average confusion matrix, mean AUC (with their corresponding ROC curves) and average F1 scores per class.

We decided to conduct an exploratory analysis for each of the features. For this, the individual predictive performance of each feature was evaluated for each model. To do this the average F1 score was used.

RESULTS

General Characteristics

In the database, a total of 21 patients were included, which accounted for 607 observations and an average number of observations per subject of 28.9. Of these 607 observations, 36.7% were classified within the category of "not deeply sedated" and 63.3% within the category of "deeply sedated". The median age of the included patients was 68 years (IQR: 60-71). Of the 21 patients, 13 were male (representing 62% of the total dataset). The primary cause of admission to the ICU, accounting for nearly half of the admissions, were pulmonary pathologies, such as acute respiratory distress syndrome or pneumonia. In second place, we have a tie between patients presenting with shock (hypovolemic and septic) and coma. In third place, we have pathologies of digestive causes such as pancreatitis and gastrointestinal bleeding. The totality of the admission causes to the ICU and details regarding the number of patients related to these are described in Table 1.

Table 1. Patients Characteristics

Age in years (Median [IQR])	68 [60 – 71]
Male (n[%])	13 [62%]
Height in cm (Mean ± SD)	169.9 ± 7.6
Weight in kg (Mean ± SD)	74.0 ± 18.4
BMI (Mean ± SD)	25.7 ± 5.9
Type of ICU admission	
Pneumology (n[%])	10 [47%]
Gastroenterology (n[%])	2 [10%]
Cardiovascular Surgery (n[%])	1 [5%]
Oncological Surgery (n[%])	1 [5%]
Shock (n[%])	3 [14%]
Intoxication (n[%])	1 [5%]
Coma (n[%])	3 [14%]
Medical vs Surgery admission	
Medical (n[%])	19 [90%]
Surgery (n[%])	2 [10%]
Sedatives used	
Propofol (n[%])	12 [57%]
Midazolam (n[%])	9 [42%]
Dexmetomidine (n[%])	3 [14%]
Sufentanil (n[%])	12 [57%]
Ketamine (n[%])	1[5%]

IQR: Interquartile Range, SD: Standard Deviation, CM: Centimeters, KG: Kilograms, BMI: Body Mass Index, ICU: Intensive Care Unit.

In general, 19 of the admissions were due to medical reasons and 2 were due to surgical reasons. The most commonly used sedative/analgesic agents were propofol and sufentanil. It is important to remember that in the ICU, it is sometimes necessary to use more than one sedative agent simultaneously to achieve the desired DoS. This approach allows the medical staff to benefit from the potentiated interactions of each drug and minimize the risk of adverse effects.

General Prediction Models

Regarding the general models (where all features extracted from the EEG were used), the Random Forest Classifier showed an average F1 score of 0.60 and an average accuracy of 0.59. Additionally, this model exhibited a mean AUC of 0.64; this metric along with its respective ROC curve can be seen in Figure 1A. Furthermore, the average F1 score was calculated per class, revealing a score of 0.52 for the "Deeply sedated" class and a score of 0.61 for the "Not deeply sedated" class.

On the other hand, the Support Vector Machine (SVM) presented an average F1 score of 0.59 and an average accuracy of 0.66. In this case, the mean AUC was 0.61; the ROC curve corresponding to this AUC can be visualized in Figure 1B. The average F1 score per class was 0.16 for the "Not deeply sedated" class and 0.77 for the "Deeply sedated" class.

The general SVM model is accurate in predicting "Deeply sedated" patients but has low predictive ability in predicting "Not Deeply sedated" patients. On the other hand, the overall RFC model has a more balanced predictive ability.

A) B)

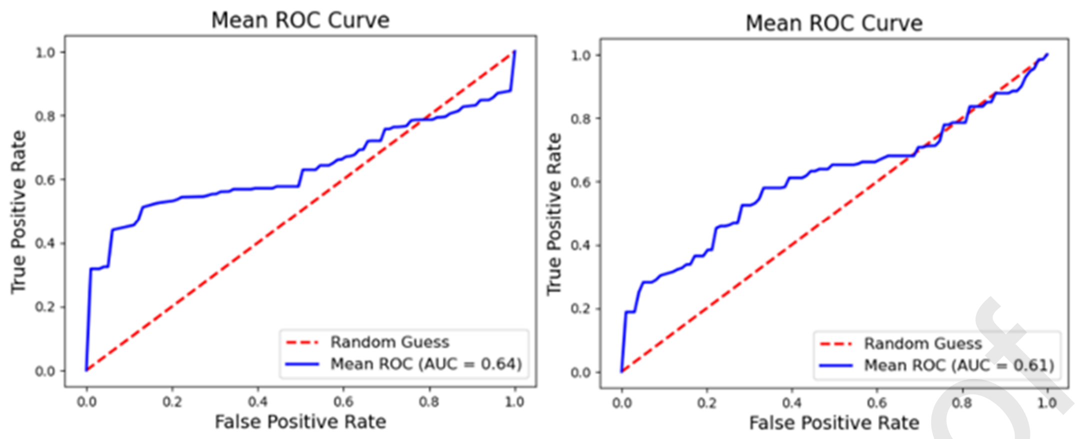
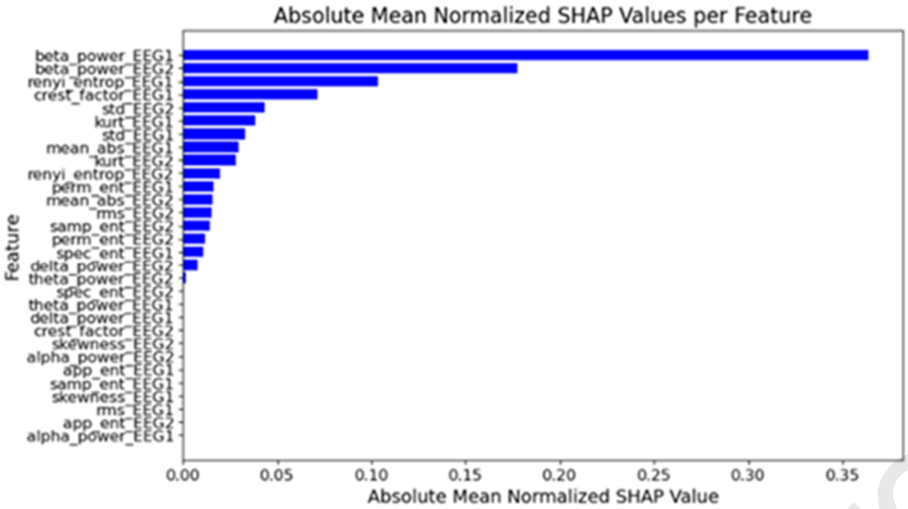


Figure 1A – Mean ROC Curve for the general RFC model; **Figure 1B** – Mean ROC Curve for the general SVM model

The Shapley values were calculated to obtain information regarding the individual importance of each feature for the model. In the general RFC model, as shown in Figure 2A, the 3 most important features for the model's predictive capacity were: beta power - EEG1, beta power - EEG2, and renyi entropy - EEG1. These three features accounted for 63% of the model's predictive ability. In the general SVM model, as seen in Figure 2B, the 3 most important features were: delta power - EEG2, delta power - EEG1, and theta power - EEG1. In this scenario, these three variables represent 99% of the model's output; in fact, only delta activity (delta power - EEG1 and delta power - EEG2) constitutes 97% of the contributions of all variables, while theta activity (theta power - EEG1 and theta power - EEG2) represents the remaining 3%. However, in the general RFC model, although beta activity (beta power EEG1 and beta power EEG2) presents the highest contribution, around 53% of the model's output, there are 16 more features that constitute the remaining 47% of contributions.

A)



B)

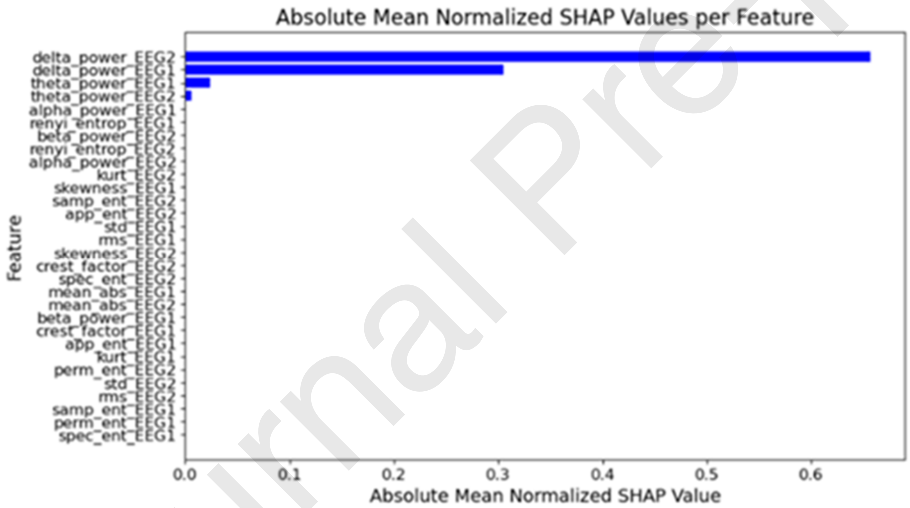


Figure 2A – Absolute mean normalized Shap values per feature for the general RFC model;

Figure 2B – Absolute mean normalized Shap values per feature for the general SVM model

Feature Selection and Reduced Prediction Models

Following the evaluation of the performance of the general models, a sequential feature selection was performed with the aim of finding the model with the best performance and the fewest number of variables. We will refer to these models as the reduced models.

When performing sequential feature selection in the RFC, we found that the features offering the best performance were as follows: beta power - EEG2, standard deviation - EEG1, delta power - EEG1, absolute mean - EEG1, sample entropy - EEG2. The reduced model showed an average accuracy of 0.65, an average F1 score of 0.65, and a mean AUC of 0.75 (Figure 3A). Additionally, the model exhibited an average F1 score per class of 0.57 for "Not deeply sedated" and 0.69 for the "Deeply sedated" class.

In the case of the reduced SVM model, the selected features were as follows: beta power - EEG2, standard deviation - EEG1, absolute mean - EEG1, crest factor - EEG1, skewness - EEG2, and root mean square - EEG1. The reduced model presented an average accuracy of 0.72, an average F1 score of 0.72, and a mean AUC of 0.73 (Figure 3B). When evaluating the performance per class, it was observed that the average F1 score for the "Not deeply sedated" class was 0.60, and for the "Deeply sedated" class, it was 0.76.

The predictive ability of both the general models and the reduced models can be observed in Table 2. This table summarizes the average F1 scores, average accuracy, and mean AUC of each model.

A) B)

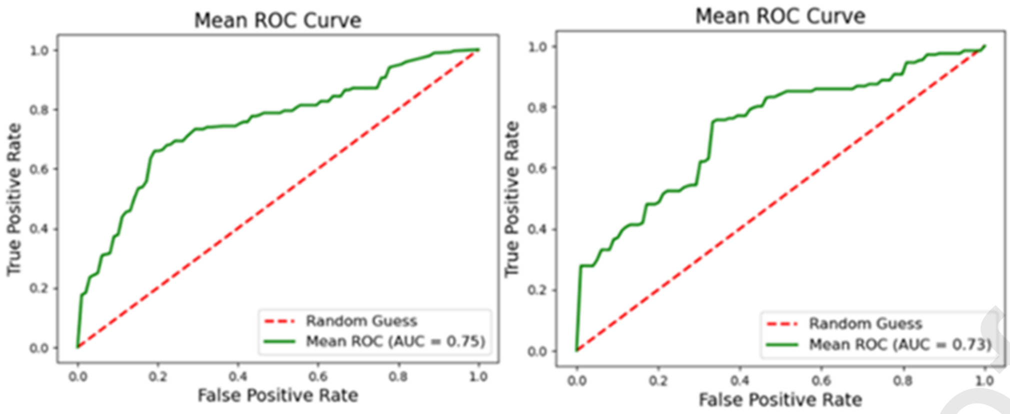


Figure 3A – Mean ROC Curve for the reduced RFC model; **Figure 3B** – Mean ROC Curve for the reduced SVM model

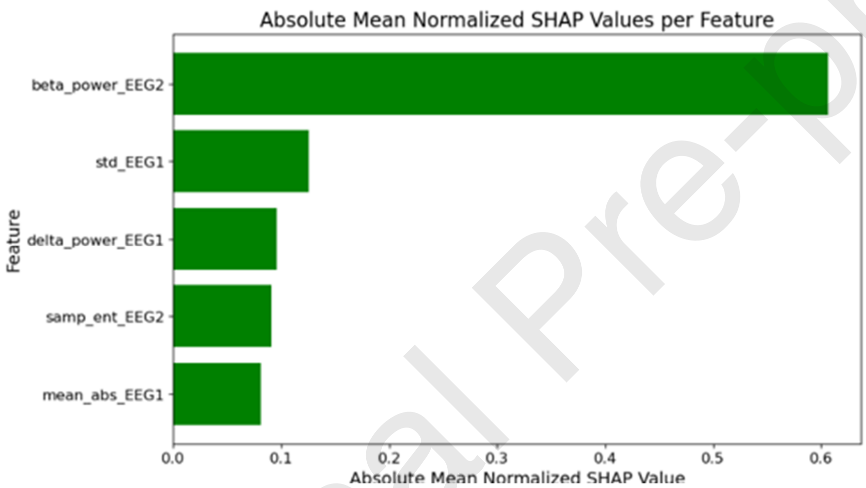
Table 2. Model Performance

Model	Average F1 score	Average Accuracy	Mean AUC
General Model - RFC	0.60	0.59	0.64
General Model - SVM	0.59	0.66	0.61
Reduced Model - RFC	0.65	0.65	0.75
Reduced Model - SVM	0.72	0.72	0.73
Reduced Model – SVM (3 variables)	0.72	0.73	0.74

RFC: Random Forest Classifier, SVM: Support Vector Machine, AUC: Area Under the Curve

Shapley values were also calculated for the reduced models. In both machine learning models, as observed in Figure 4A and Figure 4B, the feature that made the greatest contribution to the predictive capacity was beta power - EEG2. In the case of the reduced RFC model, beta power - EEG2 contributed 61% to the predictive output, and in the reduced SVM model, this same feature contributed 80% to the output. Apart from this, the remaining features of the reduced RFC model accounted from 8% to 13% of the performance. However, in the reduced SVM model, crest factor - EEG1 represented 15%, and skewness - EEG2 accounted for 4%; the remaining 3 variables contributed less than 1% to the performance of the model.

A)



B)

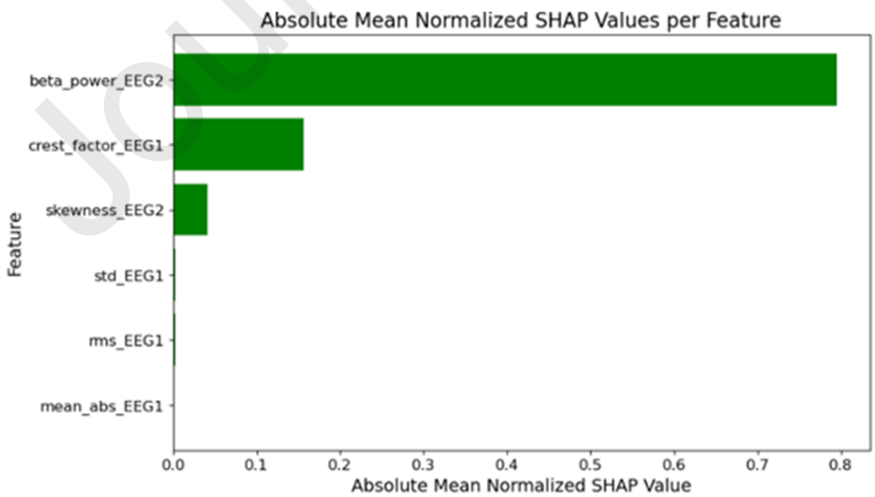
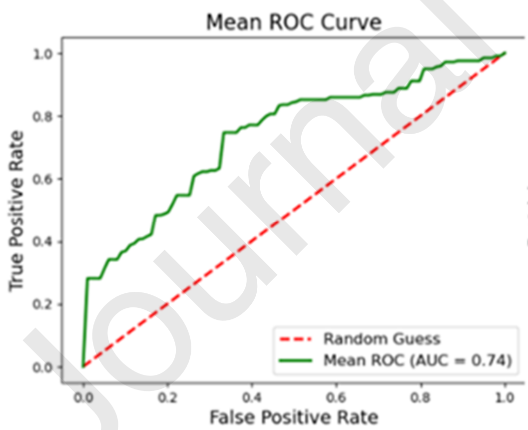


Figure 4A – Absolute mean normalized Shap values per feature for the reduced RFC model;
Figure 4B – Absolute mean normalized Shap values per feature for the reduced SVM model

Given these results, in order to explore further this algorithm, we decided to remove these 3 variables with low contributions (absolute mean - EEG1, root mean square - EEG1, and standard deviation - EEG1) from the reduced SVM model. After this, the model continued to show a performance comparable, and in some metrics slightly better, than the initial reduced model (average F1 score: 0.72, average accuracy: 0.73, mean AUC: 0.74 [Figure 5A]), even in the assessment per class (average F1 score - "Not deeply sedated": 0.61, average F1 score - "Deeply sedated": 0.76 [figure 5B]). Subsequently, an ablation study was conducted, where one of the three variables included in the model was repeatedly removed, one at a time. Upon removing the crest factor – EEG1 variable, the model showed an average F1 score of 0.67; when removing skewness – EEG2, the average F1 score was 0.71; and when removing beta power – EEG2, the average F1 score was 0.61.

A)



B)

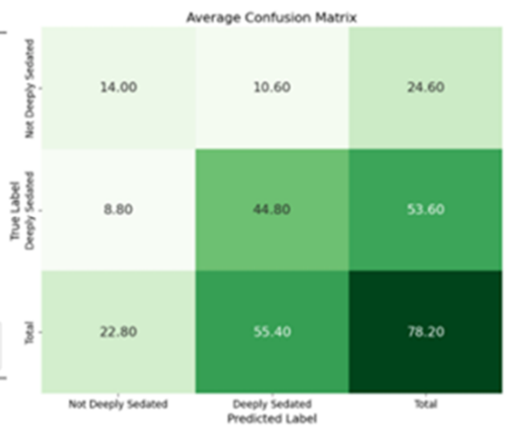


Figure 5A – Mean ROC Curve for the reduced SVM model with 3 variables;
Figure 5B – Average confusion matrix for the reduced SVM model with 3 variables

Exploratory Analysis of Individual Features

An exploratory analysis was conducted where the performance of each model (RFC and SVM) was tested using each of the features individually. In the case of RFC, the top 5 features that performed the best (evaluated with averaged F1 scores) were as follows: beta power - EEG2, sample entropy - EEG2, renyi entropy - EEG1, beta power - EEG1, and kurtosis - EEG1. For SVM, the top 5 features were as follows: beta power -EEG2, sample entropy - EEG2, beta power - EEG1, approximate entropy - EEG2, and alpha power - EEG2. The specific F1 scores for each feature are summarized in Table 3.

The feature with the best performance in both models was the beta power - EEG2. In the case of the RFC, it had an average F1 score of 0.65, while in the SVM, the beta power - EEG2 had an average F1 score of 0.70.

Table 3. Individual Feature Performance per Model – Top 5 highest values

Model	**Feature**	**Individual F1 score**
	Beta Power – EEG2	0.65
	Sample Entropy - EEG2	0.63
RFC	Renyi Entropy – EEG1	0.60
	Beta Power – EEG1	0.58
	Kurtosis – EEG1	0.58
	Beta Power – EEG2	0.70
	Sample Entropy - EEG2	0.68
SVM	Beta Power – EEG1	0.65
	Approximate Entropy – EEG2	0.65
	Alpha Power – EEG2	0.64

RFC: Random Forest Classifier, SVM: Support Vector Machine

DISCUSSION

General Aspects

This study investigated the accuracy of a 2-channel EEG for predicting the DoS in the ICU. Two machine learning models were used for this purpose, the first being an RFC and the second an SVM. Overall, the two general models showed, through the average F1 score (RFC: 0.60, SVM: 0.59), average accuracy (RFC: 0.59, SVM: 0.66), and mean AUC (RFC: 0.64, SVM: 0.61), a moderate ability for predicting the DoS. It is important to highlight that when evaluating the predictive capacity per class, we can see that despite the RFC model showing moderate ability for the prediction of DoS (average F1 score: “Deeply Sedated” = 0.52, “Not deeply sedated” = 0.61), the SVM model exhibits high to moderate performance for the "Deeply Sedated" class, although it is poor for the "Not deeply sedated" class (average F1 score: “Deeply Sedated” = 0.77, “Not deeply sedated” = 0.16).

The reduced RFC model exhibits a superior predictive ability compared to the general model, showing an increase of 0.05 in the average F1 score (reduced RFC: 0.65), 0.06 in the average accuracy (reduced RFC: 0.65), and 0.11 in the mean AUC (reduced RFC: 0.75). However, despite this improvement, the model still has a moderate predictive ability, even when evaluating the average F1 scores per class (average F1 score: “Deeply Sedated” = 0.69, “Not deeply sedated” = 0.57). Similar to the reduced RFC model, the reduced SVM model showed an overall improvement in performance, with an increase of 0.13 in the average F1 score (reduced SVM: 0.72), 0.06 in average accuracy (reduced SVM: 0.72), and 0.12 in the mean AUC (reduced SVM: 0.73). However, viewed in general terms, the model's performance remained moderate. When evaluating performance per class, a significant improvement in the model's ability to classify "Not deeply sedated" patients was observed, with an increase of 0.44, while the ability to classify "Deeply sedated" patients remained almost unchanged (average F1 score: “Deeply Sedated” = 0.76, “Not deeply sedated” = 0.60). The best performance among the models, was that shown by the reduced SVM model with 3 variables (average F1 score: 0.72, average accuracy: 0.73, mean AUC: 0.74).

As seen in the results section, the feature with the best individual performance was the beta power – EEG2 (average F1 score for RFC: 0.65, average F1 score for SVM: 0.70). The reasons why a model with only one feature can outperform one with multiple features, like the general model which has 30 features, are multiple. However, the most likely scenario is that the models with multiple features are suffering from overfitting. In other words, the algorithms are capturing noise instead of true underlying patterns necessary to make accurate predictions on unseen data. This same reason also explains why the reduced models have better predictive ability than the general models and could also explain why the reduced SVM model with 3 variables has a slightly better performance in some metrics than the reduced SVM model with 6 variables.

As previously mentioned, the EEG's capacity for predicting sedation/anesthesia depth has been extensively studied in the context of surgical practice. Examples include studies like Shi et al., where an artificial neural network was used to predict the DoA using electroencephalographic signals from 4 channels, or the study by our team, that aimed to select the channel with the best predictive capacity from a 32-channel EEG [15,21]. However, there is limited literature on this specific topic regarding patients in ICU.

The direct use of EEG features, without the signal passing through proprietary processing algorithms, has been less extensively evaluated in the ICU compared to indices like the BIS. However, within the existing literature, we have the study by Weber et al. [22], where in addition to the NI, parameters directly extracted from the EEG were evaluated, such as the spectral edge frequency (95%), the relative power (in the alpha, beta, theta, and delta band), the median frequency, among others. The EEG parameter that demonstrated the best predictive capacity was the relative power of the beta band, with a prediction probability of 0.80 in the 3-level RASS and 0.75 in the 6-level RASS, followed in second place by the spectral edge frequency [22]. Additionally, we have the study by Nagaraj et al. [16], whose objective was to assess the detection ability of deep sedation by EEG features extracted using the atomic decomposition (AD) method, and its comparison with spectral and entropy features. In this study, the combination of these three sets (AD, spectral, and entropy) achieved better discriminatory capacity (mean AUC: 0.91 (IQR: 0.81-0.98)) than these separately. Finally, there are authors that have attempted to use the raw EEG to assess the DoS, one of these authors is Sun et al. [23], who used raw EEG spectrograms without extracting any features for the identification of deeply sedated vs. non-sedated patients through the use of a recurrent neural network, obtaining an average AUC of 0.8 [16,23].

When comparing the results of our study with similar ones, such as those by Nagaraj et al. or Sun et al. [16,23], where the RASS was used as the chosen method to evaluate DoS, we can observe that the results reported by these articles indicate a better discriminative ability than that of our models [16,26]. However, upon closer analysis of these articles, we can see that the two groups categorized according to the RASS were "Deeply sedated" (RASS = -4 and -5) vs "Awake" (RASS = 0 and -1). This categorization of the RASS excludes RASS scores of -2 and -3 from the assessment, which facilitates the predictive task of the models, as these are consciousness states found at the extremes of the continuum. This characteristic may explain why the previously mentioned studies show a better discriminative capacity than what we have found.

However, the problem with conducting this type of categorization is that it diminishes the clinical utility of prediction models, as these algorithms were created with the purpose of assisting physicians in distinguishing between states of consciousness that are difficult to discern clinically. Our model takes this clinical need into consideration, and therefore, when categorizing, we include RASS scores of -2 and -3 within the category of "Not deeply sedated," as these are consciousness states that are more difficult to clinically differentiate from deep sedation (RASS scores = -4 and -5).

In our study, the individual feature that performed best in both models (SVM and RFC) was the beta power - EEG2. Beta power also contributed the most to the predictions in the majority of the models. The study by Weber et al. [22], reports that among the studied variables, the relative power of the beta band had the best predictive capacity [22]. This may suggest that beta activity in the EEG plays an important role in detecting sedation depth. The beta band frequency is associated with alertness and active thinking [24]. Thus, it seems that the shift of activity from this frequency to lower frequencies such as theta and delta (associated with states of decreased consciousness such as sleep or sedation) could be providing the greatest predictive capacity to our models. It is this shift that appears to be captured by features calculated from the beta band frequency. The case of the general SVM model is different from the rest of the models, as this model, instead of taking beta activity as the main predictor, gives primary importance to delta activity and theta activity. This may explain why, in terms of predictive capacity per class, the general SVM model is very good at predicting patients who are "deeply sedated" but poor at identifying those categorized as "Not deeply sedated". Since, it seems to focus on the presence of these frequencies, associated with deep sleep and low brain activity, to perform the discriminatory task between classes rather than on the transition of activity between frequencies [24].

Limitations

This investigation encountered certain limitations during its development. The first limitation comes from the size of the database. The dataset consists of information collected from 21 patients, however, similar studies such as those by Nagaraj et al. or Weber et al. included more than double the number of patients [16,22]. This makes the models more prone to overfitting, limiting the generalization of the findings [25]. Additionally, another limitation was the imbalance of our dataset, with more observations classified as "deeply sedated" than "not deeply sedated." This reflects the frequent need for deep sedation in ICU procedures. To address this, we applied an oversampling technique.

CONCLUSION

This study assessed the predictive ability of 2 machine learning models on DoS in patients admitted to the ICU. The algorithm that showed the best performance in differentiating between deeply sedated patients and non-deeply sedated patients was the SVM with 3 variables (beta power - EEG2, crest factor - EEG1, skewness - EEG2). The EEG feature with the best individual predictive capacity was the beta power – EEG2. These models show promising results, as they represent one of the first steps towards the development of a multimodal framework.

CONTRIBUTIONS

Esteban A. Alarcón Braga: Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. **Samuel Gruffaz**: Formal analysis, Investigation, Data Curation, Writing - Review & Editing. **Cécile Delagarde**: Investigation, Data Curation. **Axel Roques**: Investigation, Data Curation. **Jean-Clément Riff**: Investigation, Data Curation. **Laurent Oudre**: Conceptualization, Resources, Writing - Review & Editing, Supervision. **Clément Dubost**: Conceptualization, Resources, Writing - Review & Editing, Supervision.

BIBLIOGRAPHY

1. Hughes CG, McGrane S, Pandharipande PP. Sedation in the intensive care setting. Clin Pharmacol [Internet]. 2012 [cited 2024 May 6];4(1):53–63. Available from: https://pubmed.ncbi.nlm.nih.gov/23204873/

2. Simpson JR, Katz SG, Laan T Vander. Oversedation in Postoperative Patients Requiring Ventilator Support Greater than 48 Hours: A 4-year National Surgical Quality Improvement Program-driven Project. https://doi.org/101177/000313481307901031 [Internet]. 2013 Oct 1 [cited 2024 May 6];79(10):1106–10. Available from: https://journals.sagepub.com/doi/abs/10.1177/000313481307901031?journalCode=asua

3. Katherine R, Fletcher S. Sedation in the intensive care unit. Continuing Education in Anaesthesia Critical Care & Pain [Internet]. 2008 Apr 1 [cited 2024 Feb 13];8(2):50–5. Available from: https://dx.doi.org/10.1093/bjaceaccp/mkn005

4. Ramsay MAE. Intensive care: problems of over- and undersedation. Best Pract Res Clin Anaesthesiol. 2000 Jun 1;14(2):419–32.

5. Bruhn J, Myles PS, Sneyd R, Struys MMRF. Depth of anaesthesia monitoring: what's available, what's validated and what's next? Br J Anaesth. 2006 Jul 1;97(1):85–94.

6. Musizza B, Ribaric S. Monitoring the Depth of Anaesthesia. Sensors (Basel) [Internet]. 2010 Dec [cited 2024 Feb 13];10(12):10896. Available from: /pmc/articles/PMC3231065/

7. Nasraway SAJ, Wu EC, Kelleher RM, Yasuda CM, Donnelly AM. How reliable is the Bispectral Index in critically ill patients? A prospective, comparative, single-blinded observer study. Crit Care Med. 2002 Jul;30(7):1483–7.

8. Alsafy I, Diykh M. Developing a robust model to predict depth of anesthesia from single channel EEG signal. Phys Eng Sci Med [Internet]. 2022 Sep 1 [cited 2024 May 6];45(3):793. Available from: /pmc/articles/PMC9448694/

9. Kim J, Hyub H, Yoon SZ, Choi HJ, Kim KM, Park SH. Analysis of EEG to quantify depth of anesthesia using Hidden Markov Model. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014. 2014 Nov 2;4575–8.

10. Lee HC, Ryu HG, Park Y, Yoon S Bin, Yang SM, Oh HW, et al. Data Driven Investigation of Bispectral Index Algorithm. Scientific Reports 2019 9:1 [Internet]. 2019 Sep 24 [cited 2024 May 6];9(1):1–8. Available from: https://www.nature.com/articles/s41598-019-50391-x

11. Zhao T yuan meng, Chen D, Xu Z xin, Wang H liang, Sun H. Comparison of bispectral index and patient state index as measures of sedation depth during surgeries using remimazolam tosilate. BMC Anesthesiol [Internet]. 2023 Dec 1 [cited 2024 May 6];23(1). Available from: /pmc/articles/PMC10268360/

12. Fehrlin ES, Hight D, Kaiser HA, Luedi MM, Huber M, Zubler F, et al. A Pilot Investigation Evaluating Relative Changes in Fronto-Occipital Alpha and Beta Spectral Power as Measurement of Anesthesia Hypnotic Depth. Anesth Analg [Internet]. 2023 Sep 1 [cited 2024 May 6];137(3):656–64. Available from: https://pubmed.ncbi.nlm.nih.gov/36961823/

13. Gu Y, Liang Z, Hagihira S. Use of Multiple EEG Features and Artificial Neural Network to Monitor the Depth of Anesthesia. Sensors (Basel) [Internet]. 2019 Jun 1 [cited 2024 May 6];19(11). Available from: /pmc/articles/PMC6603666/

14. Ortolani O, Conti A, Di Filippo A, Adembri C, Moraldi E, Evangelisti A, et al. EEG signal processing in anaesthesia. Use of a neural network technique for monitoring depth of anaesthesia. Br J Anaesth. 2002 May 1;88(5):644–8.

15. Dubost C, Humbert P, Benizri A, Tourtier JP, Vayatis N, Vidal PP. Selection of the Best Electroencephalogram Channel to Predict the Depth of Anesthesia. Front Comput Neurosci. 2019 Oct 1;13:445878.

16. Nagaraj SB, McClain LM, Boyle EJ, Zhou DW, Ramaswamy SM, Biswal S, et al. Electroencephalogram based Detection of Deep Sedation in ICU Patients Using Atomic Decomposition. IEEE Trans Biomed Eng [Internet]. 2018 Dec 1 [cited 2024 May 6];65(12):2684. Available from: /pmc/articles/PMC6424570/

17. Mu W, Fang T, Wang P, Wang J, Wang A, Niu L, et al. EEG Channel Selection Methods for Motor Imagery in Brain Computer Interface. International Winter Conference on Brain-Computer Interface, BCI. 2022;2022-February.

18. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal P V., Keane KA, et al. The Richmond Agitation-Sedation Scale: Validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med [Internet]. 2002 Nov 15 [cited 2024 May 6];166(10):1338–44. Available from: www.atsjournals.org

19. Shobhit J, Lindsay MI. Glasgow Coma Scale [Internet]. 2023 [cited 2024 Feb 23]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK513298/

20. Akeju O, Song AH, Hamilos AE, Pavone KJ, Flores FJ, Brown EN, et al. Electroencephalogram signatures of ketamine anesthesia-induced unconsciousness. Clin Neurophysiol [Internet]. 2016 Jun 1 [cited 2024 May 6];127(6):2414–22. Available from: https://pubmed.ncbi.nlm.nih.gov/27178861/

21. Shi M, Huang Z, Xiao G, Xu B, Ren Q, Zhao H. Estimating the Depth of Anesthesia from EEG Signals Based on a Deep Residual Shrinkage Network. Sensors 2023, Vol 23, Page 1008 [Internet]. 2023 Jan 15 [cited 2024 May 6];23(2):1008. Available from: https://www.mdpi.com/1424-8220/23/2/1008/htm

22. Weber F, Steinberger M, Ritzka M, Prasser C, Bein T. Measuring depth of sedation in intensive care patients with the electroencephalographic Narcotrend Index. Eur J Anaesthesiol [Internet]. 2008 Feb [cited 2024 May 6];25(2):123–8. Available from: https://journals.lww.com/ejanaesthesiology/fulltext/2008/02000/measuring_depth_of_sedation_in_intensive_care.6.aspx

23. Sun H, Nagaraj SB, Akeju O, Purdon PL, Westover BM. Brain Monitoring of Sedation in the Intensive Care Unit Using a Recurrent Neural Network. Conf Proc IEEE Eng Med Biol Soc [Internet]. 2018 Oct 26 [cited 2024 May 6];2018:1. Available from: /pmc/articles/PMC6775766/

24. Roohi-Azizi M, Azimi L, Heysieattalab S, Aamidfar M. Changes of the brain's bioelectrical activity in cognition, consciousness, and some mental disorders. Med J Islam Repub Iran [Internet]. 2017 [cited 2024 May 6];31(1):53. Available from: /pmc/articles/PMC5804435/

25. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics [Internet]. 2023 Dec 1 [cited 2024 May 6];24(1):1–17. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-023-05156-9

Declaration of interests

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

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Detecting the Depth of Sedation in the Intensive Care Unit using a 2-channel Electroencephalogram: An analysis with 2 machine learning models

Esteban A. ALARCÓN-BRAGA[a,b] (ORCID: 0000-0003-3125-5949), Samuel GRUFFAZ[a], Cécile DELAGARDE[c], Axel ROQUES[a], Jean-Clément RIFF[c], Laurent OUDRE[a], Clément DUBOST[a, c, d]

Affiliations :

a : Centre Borelli UMR 9010/Université Paris-Saclay, ENS Paris-Saclay, CNRS, SSA, Université de Paris-Cité, Inserm, 91190 Gif-sur-Yvette, France

b : Carrera de Medicina Humana, Facultad de Ciencias de la Salud, Universidad Científica del Sur, Lima, Peru

c : Intensive care unit, Begin military hospital, 94160 SAINT-MANDE, France,

d: Ecole du Val-de-Grâce, Place Alphonse Laveran, 75005 Paris, France

Corresponding Author:

Esteban A. ALARCÓN-BRAGA: aealarco@gmail.com

Postal Address: Escuela de Medicina, Universidad Científica del Sur, Lima, Peru, Carretera Panamericana Sur Km 19, Distrito de Villa El Salvador, 15067