Identifying Episodes of Hypovigilance in Intensive Care Units Using Routine Physiological Parameters and Artificial Intelligence: a Derivation Study

Raphaëlle Giguère, Victor Niaussat, Monia Noël-Hunter, William Witteman, Tanya S. Paul, Alexandre Marois, Philippe Després, Simon Duchesne, Patrick M. Archambault

Submitted to: JMIR AI
on: May 24, 2024

Table of Contents

Identifying Episodes of Hypovigilance in Intensive Care Units Using Routine Physiological Parameters and Artificial Intelligence: a Derivation Study

Raphaëlle Giguère[1] BSc; Victor Niaussat[2, 3] MSc; Monia Noël-Hunter[4]; William Witteman[4] MISt; Tanya S. Paul[2] MSc; Alexandre Marois[2, 5] PhD; Philippe Després[6] PhD; Simon Duchesne[6, 7*] PhD; Patrick M. Archambault[4, 8, 9, 10*] MD, MSc

[1]Department of Computer Sciences, Faculty of Sciences and Engineering Université Laval Québec CA
[2]Thales Research & Technology Canada (TRT-CA) Québec CA
[3]Mathematics and Computer Science Department École Centrale de Lille Lille FR
[4]Department of Emergency Medicine Centre intégré de santé et de services sociaux de Chaudière-Appalaches Québec CA
[5]École de psychologie Université Laval Québec CA
[6]Département de radiologie et médecine nucléaire, Faculté de médecine Université Laval Québec CA
[7]Centre de recherche de l'Institut universitaire de cardiologie et pneumologie de Québec Québec CA
[8]VITAM - Centre de recherche en santé durable Québec CA
[9]Department of Anesthesiology and Critical Care Medicine Université Laval Québec CA
[10]Department of Family Medicine and Emergency Medicine, Faculty of Medicine Université Laval Québec CA
[*]these authors contributed equally

Corresponding Author:
Patrick M. Archambault MD, MSc
Department of Family Medicine and Emergency Medicine, Faculty of Medicine
Université Laval
Ferdinand Vandry Pavillon, 1050 Av. de la Médecine
Québec
CA

Abstract

Background: Delirium is a prevalent condition in intensive care units (ICUs), often leading to adverse outcomes. Hypoactive delirium is particularly difficult to detect. Despite advancements, timely identification of hypoactive delirium remains challenging due to its dynamic nature, lack of human resources, lack of reliable monitoring tools, and subtle clinical signs that include hypovigilance. Machine learning detection models could support the identification of hypoactive delirium episodes by better detecting episodes of hypovigilance.

Objective: In this study, we aim to develop a machine learning algorithm capable of detecting hypovigilance events using routinely collected physiological data in the ICU.

Methods: This derivation study used prospective observational data collected from eligible ICU patients in Lévis, Québec. We included patients admitted to the ICU between October 2021 and June 2022 who were at least 18 years old and had an anticipated ICU stay of at least 48 hours. ICU nurses identified hypovigilant states every hour using the Richmond Agitation and Sedation Scale (RASS) or the Ramsay Sedation Scale (RSS). Routine vital signs (heart rate, respiratory rate, blood pressure, and oxygen saturation), as well as other physiological and clinical variables (premature ventricular contractions, intubation, use of sedative medication, and fever), were automatically collected using a GE CARESCAPE Gateway or manually collected through chart review. Time series were generated around hypovigilance episodes for analysis. Random Forest, XGBoost, and LightGBM classifiers were then used to detect hypovigilant episodes on the basis of analyzing time series. Hyperparameter optimization was performed using random search in a 10-fold group cross-validation setup. We report the results of this study using the TRIPOD+AI guidelines and potential biases were assessed using PROBAST.

Results: Out of 146 potentially eligible participants, data from 30 patients (mean age: 69 years old; 63% male) were collected for analysis. Of the group, 30% were admitted to the ICU for surgical reasons. Following data preprocessing, the study included 1,493 hypovigilance episodes and 764 non-hypovigilant episodes. Among the three sets of models evaluated, LightGBM

demonstrated the best performance. It achieved an average accuracy of 68% to detect hypovigilant episodes, with a precision of 76%, a recall of 74%, an AUC of 60%, and an F1 score of 69%. Notably, the model was particularly influenced by intubation, respiratory rate, and systolic blood pressure acquired non-invasively.

Conclusions: All of the classifiers showed promising precision and recall, with slightly different yet comparable performance in classifying hypovigilant episodes. Machine learning algorithms designed to detect hypovigilance have the potential to support early detection of hypoactive delirium in ICU patients.

(JMIR Preprints 24/05/2024:60885)
DOI: https://doi.org/10.2196/preprints.60885

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Raphaëlle Giguère[1], Victor Niaussat[2,3], Monia Noël-Hunter[4], William Witteman[4], Tanya S. Paul[3], Alexandre Marois[3,5], Philippe Després[6], Simon Duchesne[*6,7] and Patrick Archambault[**,*4,8,9,10,11]

[1] Department of Computer Sciences, Faculty of Sciences and Engineering, Université Laval, Québec, QC, Canada.

[2] École Centrale de Lille, Mathematics and Computer Science Department, Lille, France.

[3] Thales Research & Technology Canada (TRT-CA), Québec, QC, Canada.

[4] Department of Emergency Medicine, Centre intégré de santé et de services sociaux de Chaudière-Appalaches, Lévis, QC, Canada.

[5] École de psychologie, Université Laval, Québec, QC, Canada.

[6] Département de radiologie et médecine nucléaire, Faculté de médecine, Université Laval, Québec, QC, Canada

[7] Centre de recherche de l'Institut universitaire de cardiologie et pneumologie de Québec, Québec, QC, Canada

[8] Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Québec, QC, Canada.

[9] VITAM - Centre de recherche en santé durable, Québec, QC, Canada.

[10] Department of Anesthesiology and Critical Care Medicine, Université Laval, Québec, QC, Canada.

[11] Centre de recherche intégrée pour un système apprenant en santé et services sociaux, Centre intégré de santé et de services sociaux de Chaudière-Appalaches, Lévis, QC, Canada.

* Both authors contributed equally.
** Correspondance.

Identifying Episodes of Hypovigilance in Intensive Care Units Using Routine Physiological Parameters and Artificial Intelligence: a Derivation Study

Abstract

Background: Delirium is a prevalent condition in intensive care units (ICUs), often leading to adverse outcomes. Hypoactive delirium is particularly difficult to detect. Despite advancements, timely identification of hypoactive delirium remains challenging due to its dynamic nature, lack of human resources, lack of reliable monitoring tools, and subtle clinical signs that include hypovigilance. Machine learning detection models could support the identification of hypoactive delirium episodes by better detecting episodes of hypovigilance.

Objective: In this study, we aim to develop a machine learning algorithm capable of detecting hypovigilance events using routinely collected physiological data in the ICU.

Methods: This derivation study used prospective observational data collected from eligible ICU patients in Lévis, Québec. We included patients admitted to the ICU between October 2021 and June 2022 who were at least 18 years old and had an anticipated ICU stay of at least 48 hours. ICU nurses identified hypovigilant states every hour using the Richmond Agitation and Sedation Scale (RASS) or the Ramsay Sedation Scale (RSS). Routine vital signs (heart rate, respiratory rate, blood pressure, and oxygen saturation), as well as other physiological and clinical variables (premature ventricular contractions, intubation, use of sedative medication, and fever), were automatically collected using a GE CARESCAPE Gateway or manually collected through chart review. Time series were generated around hypovigilance episodes for analysis. Random Forest, XGBoost, and LightGBM classifiers were then used to detect hypovigilant episodes on the basis of analyzing time series. Hyperparameter optimization was performed using random search in a 10-fold group cross-validation setup. We report the results of this study using the TRIPOD+AI guidelines and potential biases were assessed using PROBAST.

Results: Out of 146 potentially eligible participants, data from 30 patients (mean age: 69 years old; 63% male) were collected for analysis. Of the group, 30% were admitted to the ICU for surgical reasons. Following data preprocessing, the study included 1,493 hypovigilance episodes and 764 non-hypovigilant episodes. Among the three sets of models evaluated, LightGBM demonstrated the best performance. It achieved an average accuracy of 68% to detect hypovigilant episodes, with a precision of 76%, a recall of 74%, an AUC of 60%, and an F1 score of 69%. Notably, the model was particularly influenced by intubation, respiratory rate, and systolic blood pressure acquired non-invasively.

Conclusions: All of the classifiers showed promising precision and recall, with slightly different yet comparable performance in classifying hypovigilant episodes. Machine learning algorithms designed to detect hypovigilance have the potential to support early detection of hypoactive delirium in ICU patients.

Keywords: Vigilance; Hypovigilance; Hypoactive Delirium; Intensive Care Unit (ICU); Machine Learning, Artificial Intelligence; Detection model; Physiological Parameters; Automated Monitoring

Introduction

Delirium is defined by the American Psychiatric Association as a transient disturbance of attention and awareness, manifested as a reduced ability to control, focus, maintain, and transfer attention and as a weakened orientation to the environment [1]. A missed or delayed diagnosis of delirium is associated with adverse outcomes, particularly in intensive care units (ICUs) where it can lead to prolonged hospital stays, increased mortality rates, slower rates of overall recovery, and cognitive impairment [2–4].

Lipowski first introduced the concept of delirium subtypes, proposing the terms "hyperactive" and "hypoactive" to characterize patients' motor behavior, distinguishing restlessness and aggression from low vigilance and apathy [5]. Hypoactive delirium, widely recognized as the most common form of delirium, remains a clinically significant but often unrecognized condition, largely due to the challenges of diagnosing this specific variant of the syndrome [6]. Research suggests that hypoactive patients may have a higher risk of mortality compared to other delirium subtypes [7]. Yet despite its clinical relevance, hypoactive delirium is often undetected in routine clinical practice [6,8]. Although improvements in screening and therapy have occurred, the identification of hypoactive delirium still poses a serious challenge, given that the onset of episodes remains difficult to determine, and that its status changes subtly over time [9]. Further, this detection is labor intensive, requiring frequent reevaluation and clinical interpretation using bedside instruments and questionnaires [10]. A systematic review of ICU delirium prediction models [2] found that while many models performed well, they only predicted the condition using transversal data from a single point in time, failing to take into account the dynamic nature of delirium. Additionnally, when patients speak a different language than their healthcare providers, the use of these instruments and questionnaires becomes even more challenging [11].

The main symptom of hypoactive delirium is decreased vigilance, also known as hypovigilance [5]. Vigilance is the ability to remain aware of relevant and unpredictable changes in an individual's surrounding environment, regardless of whether such changes actually occur [12]. This capability comprises two dimensions. The first is related to the level of alertness required to maintain vigilance, and the second arises from the fact that vigilance can vary over time, i.e., the level is known to fluctuate (decrease or increase) over time [12].

Dynamic changes in a patient’s vigilance level can potentially be detected using the collection and analysis of continuous psychophysiological signals. This method involves measuring the physiological activity of a person’s central and autonomic nervous systems to estimate their vigilance level. This approach is based on the hypothesis that the locus coeruleus-norepinephrine system plays a significant role in attention-related activities. This system has been associated with vigilance, attention, arousal, and the sleep-wake cycle [13–16]. This activity is also associated with an increase in autonomic sympathetic nervous system activity and a decrease in parasympathetic nervous system activity [17–19], which is due to the multiple efferent projections of the system in the brain. In short, multiple psychophysiological markers of the hypovigilant state can be collected via proxy measures of the central nervous system and of the peripheral nervous system. The autonomic nervous system modulates heart rate (HR), blood pressure, digestion, respiration, pupillary reactivity, and regulates the functions of internal organs [20]. Heart rate variability (HRV) is considered a convenient tool for monitoring the autonomic nervous system [21,22], but is not routinely collected in all ICUs.

Further, international guidelines suggest the use of sedatives and analgesics to ensure comfort and permit life-saving but painful procedures [23]. However, when used to induce coma for mechanical ventilation, sedative and analgesic medications such as benzodiazepines and opioids put patients at high risk for delirium [24,25]. Sedative and analgesic therapy is also related to several important side effects, among which are hypotension, bradycardia and other dysrhythmias, and sepsis [26]. Some of the delirium prediction models included the use of either sedatives, benzodiazepines, and antipsychotic 24 hours before the diagnosis of delirium as predictors [2].

Prediction models using artificial intelligence (AI) have been developed to quantify hypovigilance or related concepts (e.g., drowsiness and fatigue) using psychobehavioural correlates of vigilance in laboratory conditions [27,28], but real-world examples are few. A recent scoping review identified 21 psychophysiological models of hypovigilance detection, in which almost all relied on at least one of the following signals, targeting both central and autonomic nervous systems: electrocardiography, photoplethysmography, electroencephalography, electro-oculography, and eye tracking [28]. While sensitive, these systems are resource-intensive in dynamic environments such as the ICU.

Despite existing research on hypovigilance, and the potential applications and clinical implications of developing models for hypovigilance diagnosis, there is no consensus in the literature regarding the most accurate method. There are also significant barriers to the widespread use of more sophisticated diagnostic modalities such as electroencephalography in clinical settings because of poor signal quality [28], specialized and costly equipment requirements [29], and patient discomfort due to extended wear [30]. These factors hinder the adoption of these sensors in routine healthcare settings.

Finally, other emerging and promising sensors capable of monitoring HRV are not universally incorporated into ICU monitors and are not part of routine data collection.

To emphasize the need to develop an AI system for delirium detection, it is important to recognize that delirium detection currently relies on the use of subjective and infrequently used bedside tools that are not applicable to all patients. Current delirium prediction and detection models rely on static data that fail to capture the rapid-cycle fluctuating nature of vigilance [2]. Moreover, vigilance detection models developed in lab environments lack real-world clinical validation [28]. Clinical real-world settings, such as ICUs, can provide a reliable data collection environment where patients often experience frequent episodes of hypovigilance. Further research is needed to identify effective detection methods for ICU patients, including the use of automated delirium assessment tools that could be reliably used on a large scale by non-experts and that could be used in all patients regardless of the language they speak [31]. This highlights why AI technology could be crucial in addressing challenges related to delirium detection.

The objective of this project was to develop an AI algorithm able to continuously detect recurrent episodes of hypovigilance using routinely collected physiological markers in the ICU.

Methods

Design and Setting

We conducted a derivation study using data collected from a prospective observational cohort study carried out in the ICU at the Hôtel-Dieu de Lévis Hospital (Lévis, QC, Canada). The study was approved by the research ethics committee of the Centre Intégré de Soins et Services Sociaux de Chaudière-Appalaches (#2021-771) but the research protocol was not registered in a clinical trials registry. We report our findings using the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis for machine learning models (TRIPOD+AI) guidelines [32,33] (Appendix A). We also used the PRediction model Of Bias ASsessment Tool (PROBAST) to identify potential biases [32,34] (Appendix B). The code and data sets generated for the purposes of preprocessing and training the models are available for download from our GitHub repository. [35].

Participants

Eligible patients were admitted to the Hôtel-Dieu de Lévis Hospital ICU between October 2021 and June 2022. Inclusion criteria were: (1) age ≥18 years; and (2) an anticipated ICU stay ≥ 48 hours from admission. Exclusion criteria were: (1) ICU admission for ≥5 days; (2) inability to obtain informed consent (from patients themselves or their substitute decision-makers); (3) inability to communicate in English or French; (4) severe neurodegenerative diseases (eg, Alzheimer's, Parkinson's, or Huntington's Disease, amyotrophic lateral sclerosis); and (5) unavailability of the data collection device. A total of 30 participants were enrolled in the study.

Data Collection

Despite the inability to blind bedside nurses to the predicted outcome (hypovigilance), they were unaware of the ongoing project. Additionally, all vital signs used as predictors were automatically collected by the GE CARESCAPE Gateway, eliminating any potential for information bias.

Event identification

Bedside ICU nurses completed hourly assessments of the patient's vigilance using the Richmond Agitation and Sedation Scale (RASS) [36] or the Ramsay Sedation Scale (RSS) [37,38]. RASS is a 10-point scale that assesses sedation and agitation based on specific criteria that evaluate the patient's response after verbal stimulation. The RSS categorizes sedation levels across six stages and is widely used in clinical settings [20]. While the RASS is standard for intubated cases, the RSS is used when participants are not intubated. There is a strong correlation between the two scales and they demonstrate good interrater reliability [39]. We identified hypovigilant episodes when RASS scores were <0, indicating a drowsy to unarousable state, and RSS scores > 2, signifying a drowsy to unarousable condition [23]. These specific criteria served as the basis for labeling patients' vigilance state into episodes of hypovigilance vs. non-hypovigilance.

Figure 1: Labeling process to identify episodes of hypovigilance using the corresponding Ramsay Sedation Scale and Richmond Agitation and Sedation Scale.

	Ramsay	RASS	Vigilance state
Combative	1	+4	Non-hypovigilant
Very agitated		+3	Non-hypovigilant
Agitated		+2	Non-hypovigilant
Restless		+1	Non-hypovigilant
Alert and calm	2	0	Non-hypovigilant
Drowsy	3	-1	Hypovigilance
Light sedation		-2	Hypovigilance
Moderate sedation	4	-3	Hypovigilance
Deep sedation	5	-4	Hypovigilance
Unarousable	6	-5	Hypovigilance

We did not capture the sociodemographic characteristics of the nurses who conducted the RASS and RSS assessments. However, vital signs were automatically captured by the GE CARESCAPE Gateway, ensuring no information bias at that level.

Clinical Information

At enrollment, we collected participants' age, gender, height, and comorbidities. We also collected information on the participant's type of admission, history of depression, need for ventilatory support during the ICU stay, and intubation history. We also documented if any intravenously (IV) administered sedative or analgesic agents (eg, midazolam, propofol, hydromorphone, or fentanyl) were being administered at the time of the nurse's vigilance assessment. Use of IV sedation and analgesia was extracted from nursing notes as a binary variable (presence or absence of one of these medications). Multiple medications could be administered simultaneously.

The research team also collected data about: (1) the Glasgow Coma Scale to measure the patient’s level of consciousness ranging from 3 to 15, with lower scores indicating more severe deficits [40,41]; (2) the Pfeffer Functional Activities Questionnaire (FAQ) to operationalize participants' baseline functional capabilities in their daily lives [42]; and (3) the Canadian Frailty Score (CFS) to evaluate the baseline frailty status of participants with scores ranging from 1 (very fit) to 9 (terminally ill) [43]. These questionnaires are described in Appendix C. These tools helped to characterize the patient population, they were not integrated into the developed AI algorithm.

Physiological Time Series Collection

We employed a GE CARESCAPE Gateway (GE Medical Systems, Milwaukee, WI) to streamline and automate continuous data collection. Gateway data was extracted and securely stored in a comma-separated value format on local servers. Vital signs and physiological markers were continuously monitored and recorded at one-minute intervals, allowing for the exploration of indicators associated with hypovigilant episodes. Bedside vital signs and data automatically recorded via the Gateway included heart rate, respiratory rate, premature ventricular complex count, oxygen saturation, body temperature (when an internal body temperature probe was used), invasive arterial blood pressure, noninvasive blood pressure (Table 1). Intubation was automatically derived by the presence of inspired or expired CO2 (Table 1). Occasionally, more than one timestamp's worth of data was gathered by the gateway. To ensure data consistency, we systematically removed all duplicate lines.

Table 1: Bedside vital signs and data recorded with the GE CARESCAPE Gateway.

Feature	Description	Unit
HR	Heart rate	Beats per minute (bpm)
RR	Respiration rate	Breaths per minute
SPO2-%	O2 saturation	Percentage (%)
SPO2-R	Pulse oximeter pulse rate	Beats per minute

NBP-D	Noninvasive diastolic blood pressure	millimeter of mercury (mmHg)
NBP-M	Noninvasive mean blood pressure	mmHg
NBP-S	Noninvasive systolic blood pressure	mmHg
PVC	Premature ventricular complex count	Events per minute
AR-D	Arterial line diastolic pressure	mmHg
AR-S	Arterial line systolic pressure	mmHg
AR-M	Arterial line mean pressure	mmHg
AR-R	Arterial line pulse rate	Beats per minute
CO2-EX	CO2 expired while intubated	mmHg
CO2-IN	CO2 inspired while intubated	mmHg
Temperature	Rectal temperature	Degrees Celsius (°C)

Time Series Selection

Time series of sequential changes in vigilance states were generated by selecting physiological measures data within a 5-minute window before and after each hypovigilant or non-hypovigilant episode, resulting in an 11-point time series spanning 11 minutes (Figure 2). RASS/RSS assessments were made hourly; assuming that if two measurement points were the same, the condition was considered constant throughout the hour, whereas different assessments indicated variability and did not allow for this assumption. Utilizing an 11-minute window enabled us to capture and analyse a broader period around the critical label moment indicating an episode of hypovigilance. This approach provided the temporal context surrounding the vigilance state, as physiological signals may exhibit significant variability over time. The inclusion of an 11-minute window in each time series aimed to maximize clinical relevance, better characterize state changes, and optimize the classification capacity of our AI models.

Figure 2: Label identification process.

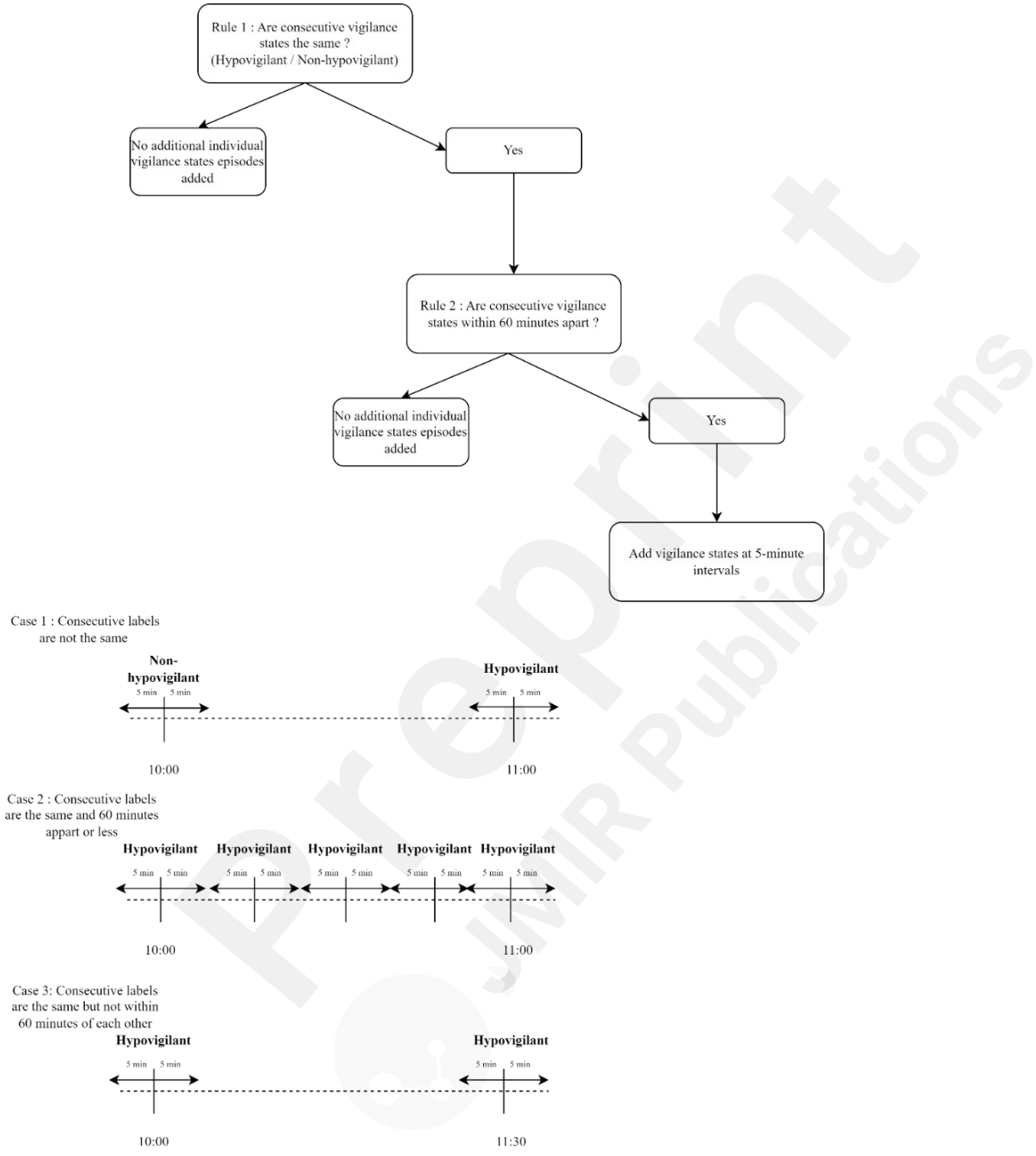


Figure 2 illustrates the two simple rules we followed to label episodes of hypovigilance before and after the hourly vigilance assessments determined by bedside nurse. Labels (hypovigilant vs. non-hypovigilant) were automatically assigned for each separate vigilance state as determined by bedside nurses. Additional labels were assigned to time points before and after each hourly vigilance assessment performed by the nurses based on two simple rules. The first rule determined if two labels were within 60 minutes of each other. If they were, we proceeded to the second rule. If the labels were more than 60 minutes apart, we did not add any new labels for new hypovigilance episodes. The second rule

determined if the consecutive vigilance states (and associated labels) were were identical. If they were identical, we added labels at 5 minute intervals for each episodes of hypovigilance or non-hypovigilance between the original labels based on the value of the vigilance state at both boundaries. If they were not identical, we did not add additional episodes of hypovigilance or non hypovigilance between the consecutive discordant labels because determining the moment when the state changed from hypovigilant to non hypovigilant (or vice-versa) had not been determined by bedside nurses and was impossible to determine retrospectively. For example, in case 1, no additional labels were added because consecutive labels differed. In case 2, labels were added at 5-minute intervals when consecutive labels were identical and less than 60 minutes apart. However, in case 3, no additional labels were added because consecutive states remained the same but were separated by more than 60 minutes. This preserved the temporal structure integrity of our model.

Time series preprocessing

Missing values

We utilized a Backward and Forward Filling strategy to address missing values in the signals' time series [44]. Backward Filling involves filling in missing values in a dataset by using previous data values to complete the gaps. In other words, missing values were filled based on available data preceding them in the time series [44]. In some cases, backward filing was impossible due to the lack of available previous data. In such cases, Forward Filling was performed. Forward filling involves using future values to fill in missing data [44].

Features Extraction

As features for our model, we chose to extract the first, second, and third derivatives for each participant's temporal data stream, to observe global variations or trends across all patient observations. The first derivative represents the rate of change over time, eg, it allows for the identification of rapidly increasing or decreasing blood pressure. The second derivative refers to the acceleration of the rate of change, eg, how the slope of the blood pressure curve evolves over time [45]. For example, a positive second derivative might suggest an acceleration in blood pressure increase, while a negative second derivative could indicate an acceleration in blood pressure decrease. The third derivative captures the variation in the acceleration, meaning the rate at which the acceleration changes [46]. This is particularly useful for detecting unusual changes, eg, blood pressure fluctuations or respiration rate.

We observed that some features (AR-D, AR-M, AR-R, AR-S, Temp, CO2-EX, and CO2-IN) had excessive missing values across participants (more than 40%). As is, they would have induced biases in the classifier. However, since they all related to important clinical information commonly present in more critically ill patient states (eg, presence of an arterial line for a patient in shock, intubation for a patient in respiratory failure, and internal body temperature probe for an intubated hypothermic patient), we elected to replace them with Boolean (Presence/Absence) variables to indicate whether an arterial line was installed, when a patient was intubated, or whether an internal body temperature probe was installed.

Features selection

Our objective was to identify significant differences between the two vigilance states (hypovigilant vs. non-hypovigilant) with a non-normal distribution of the data. We elected to reduce the feature space by only selecting features that were significantly different between both states, on average. To this end, we performed Mann-Whitney U [47] tests between the distribution of features. The data set was divided into training and test sets (see Figure 3), and the Mann-Whitney U test was performed separately on each set. Only the variables that were statistically significant within a particular set were included in the respective model trained on that set. As a result, multiple models were generated, each using a subset of the variables found to be significant in their respective set. This approach ensured that the models were tuned to capture the most relevant features for predicting vigilance states, taking into account the variability observed across sets in the cross-validation process. To provide statistics on the selected features, we counted the number of times a variable was found to be significant across sets.

Machine Learning Models

We employed three distinct AI models for detecting hypovigilance events: Random Forest (RF) from the Python Scikit-learn library[48]: eXtreme Gradient Boosting (XGBoost) from the XGBoost library [49], and the Light Gradient-Boosting Machine classifier (LGBM) from the LightGBM library [50]. We chose these three particular classifiers because RF and XGBoost were used in prior studies to identify delirium and hypovigilance [28]. LightGBM was also used because of its previous application in other ICU databases such as the Medical Information Mart for Intensive Care III database [51].

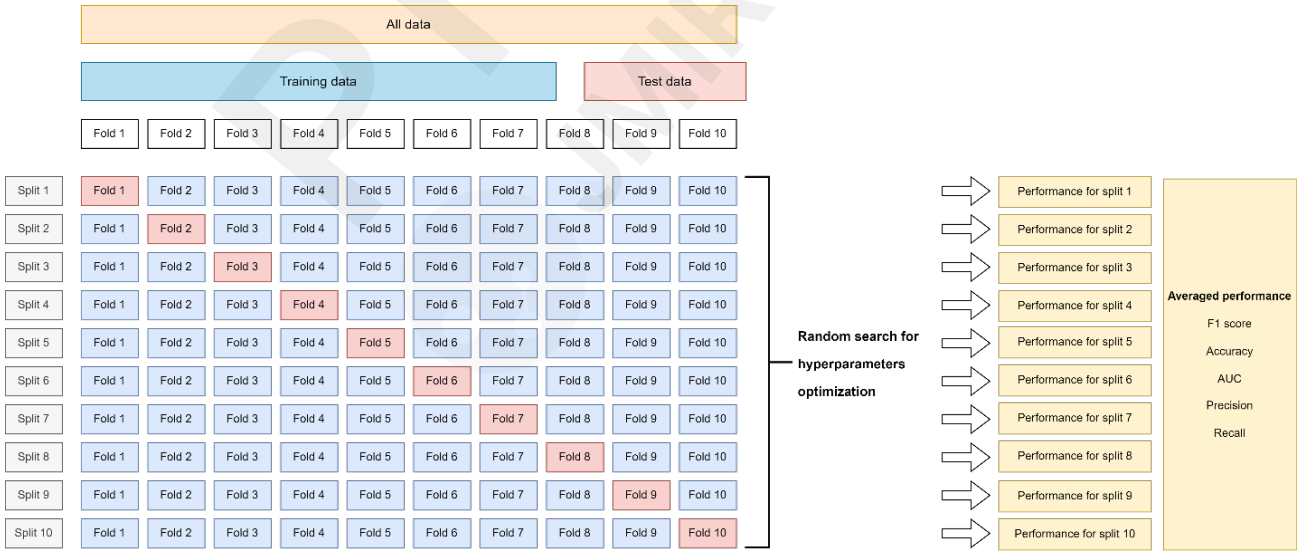
RF is a real-time classification algorithm that excels in capturing non-linear relationships, making it applicable to diverse domains such as genetic effect estimation and clinical outcome prediction [52]. It is composed of a set of data structures characterized by decisions (branches), called trees, with each tree depending on random variables. It creates a forest from a group of decision trees trained by the bagging method. The key notion behind the bagging method is the combination of multiple learning models to improve overall sensitivity [52]. Designed for speed and performance, XGBoost uses gradient-based decision trees. Tailored for classification and regression predictive modeling, preferred for tabular datasets with dominant features [49]. The LightGBM classifier uses a classifier for iterative training to obtain optimal identification models. It employs a gradient-boosting framework with a tree-based learning algorithm, which strategically reduces the computational burden of structure score calculation. The algorithm also uses a histogram-based approach for split point selection and adopts a leaf-wise strategy to streamline computation and improve accuracy [51].

Data Splitting and Hyperparameters Search

To preserve patient data and account for the limited number of patients in our dataset, we employed a 10-fold group cross-validation strategy. The data was partitioned into groups of random size at the patient level. This approach ensured that all within-patient-related information was retained during model evaluation, making models more robust to new participants.

To enhance the performance of our three AI classifiers, we utilized the Random Search technique for hyperparameter tuning for each split. This technique is widely recognized for its computational efficiency compared to traditional grid search methods, as it requires less computational time [53].

Figure 3: Data splitting and hyperparameters optimization to evaluate the models' performances.



Legend for Figure 3: The model was trained and evaluated using group cross-validation. The data was split at the patient level into training and testing subsets for each fold of the cross-validation, with varying group sizes ranging from one to several patients. Across all folds, the Mann-Whitney U test was performed to select only the significant features in each model. Performance metrics (accuracy, precision, and recall) were computed on the testing subset during each fold.

Performance Evaluation

The performance metrics included average accuracy, which measured the proportion of correct predictions made by the model across all iterations of the cross-validation process; precision, which assessed the proportion of true positive predictions among all positive predictions, providing insight into the model's ability to make precise classifications; and average recall (sensitivity), which evaluated the model's capability to correctly identify positive instances from the entire dataset. Additionally, we computed the average Area Under the Curve (AUC), which serves as a measure of our model's ability to distinguish between positive and negative classes, and the average F1 score, which provides a balanced assessment by considering both precision and recall. These metrics collectively indicated the overall performance of our model in classification tasks.

To enhance our understanding of our best models, we utilized SHapley Additive exPlanations (SHAP) values [54]. To calculate SHAP values, we selected one of the top-performing models from the 10-fold group cross-validation and retrained it on the entire dataset of 30 participants. The resulting SHAP values provided insights into the weight of each feature. This analysis offers a representative overview, although the exact feature importance may vary between models.

Sensitivity analyses

We also aimed to evaluate the inclusion of the medication variable (use of sedatives/analgesics) on the performance of our models. We therefore performed two main sets of sensitivity analyses, one without and the second with the medication variable as a variable of interest. We did not consider the specific treatments administered for the management of delirium, we did evaluate the impact of the use but sedative and analgesic medications, which can significantly influence hypovigilance.

Patient and Public Involvement

Patients and the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Results

Patients characteristics

Among 146 patients considered for inclusion in our cohort, 30 were eligible (Figure 4). These 30 patients experienced a total of 1,493 hypovigilant episodes and 764 non-hypovigilant episodes. Two participants did not have any hypovigilant episodes. Participants in our cohort were: aged 69 years (mean), male (63%), admitted to the ICU for surgical (30%) or medical (70%) reasons, and mostly intubated receiving IV sedation-analgesia medication (70%).

Table 2. Demographic and clinical characteristics of participants (N=30).

Demographic characteristics	
Age (years), mean (SD); min-max	68.9 (11.0); 35-86
Height (cm), mean (SD); min-max	168.0 (8.9); 152-183
Sex, n (%)	
Women	11 (36.7%)
Men	19 (63.3%)
Length of ICU stay (days), mean (SD); min-max	8.60 (5.29); 1.33-22.24
Comorbidities, n (%)	
Cardiovascular diseases	22 (73.3%)
Respiratory disease	14 (46.7%)
Renal Disease	8 (26.7%)
Diabetes	8 (26.7%)
History of stroke	3 (10.0%)
No comorbidities	1 (3.3%)
Other	21 (70.0%)
Depression in the past, (n, %)	

Yes	1 (3.3%)
No	29 (96.7%)
Type of admission, (n, %)	
Medical	21 (70.0%)
Surgical	9 (30.0%)
Respiratory assistance, (n, %)	
Yes	25 (83.3%)
No	5 (16.7%)
Intubated, n (%)	
Yes	21 (70.0%)
No	9 (30.0%)
Admission assessment	
Glasgow Coma Scale, mean (SD); min-max	14 (1); 8-15
Functional Activity Questionnaire (FAQ), mean (SD); range	3 (5); 0-19
Clinical Frailty Scale, median ; IQR	2 ; 5

Figure 4: Flowchart of the data collection process.

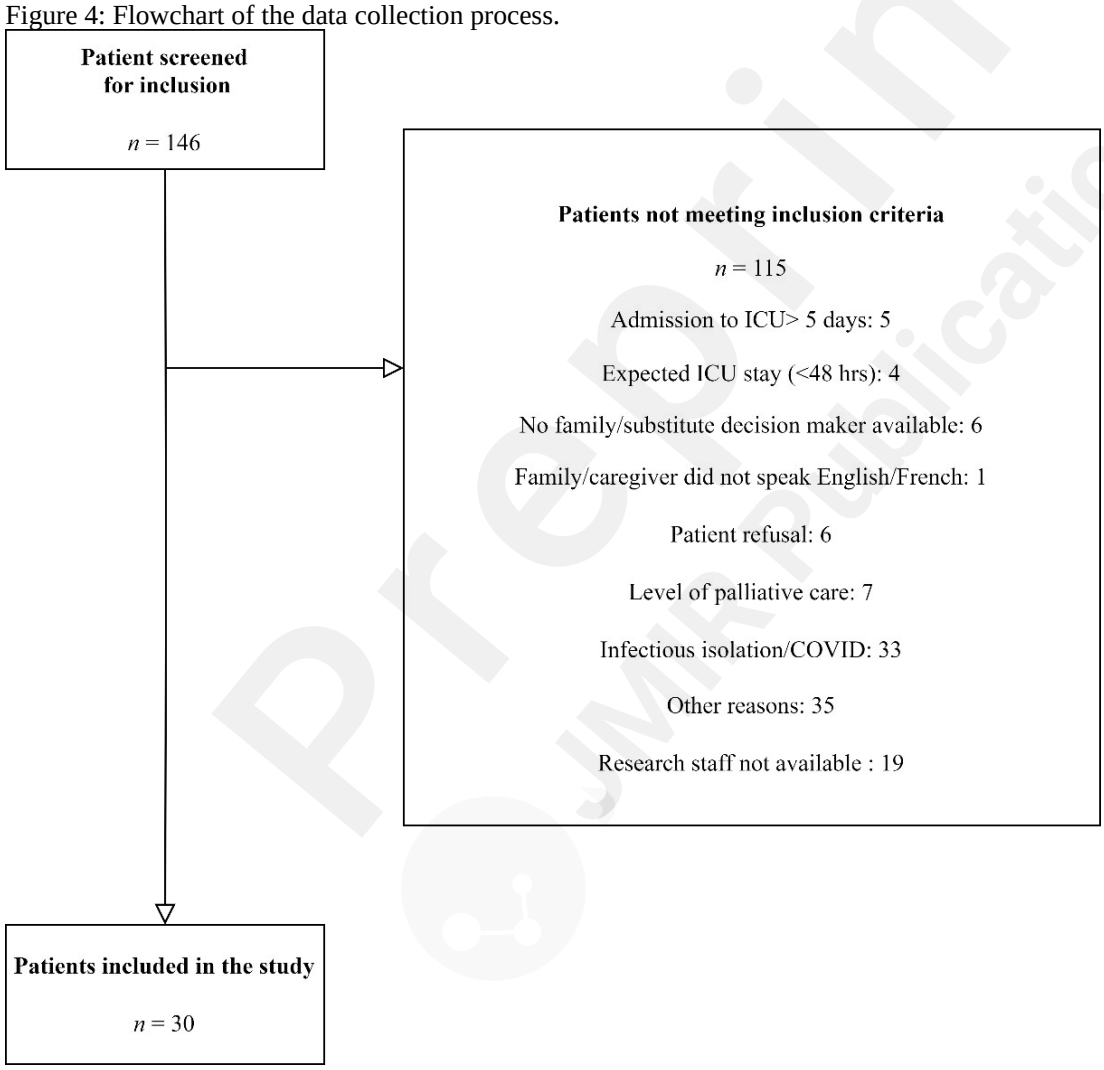


Figure 5: Significant features in the models.

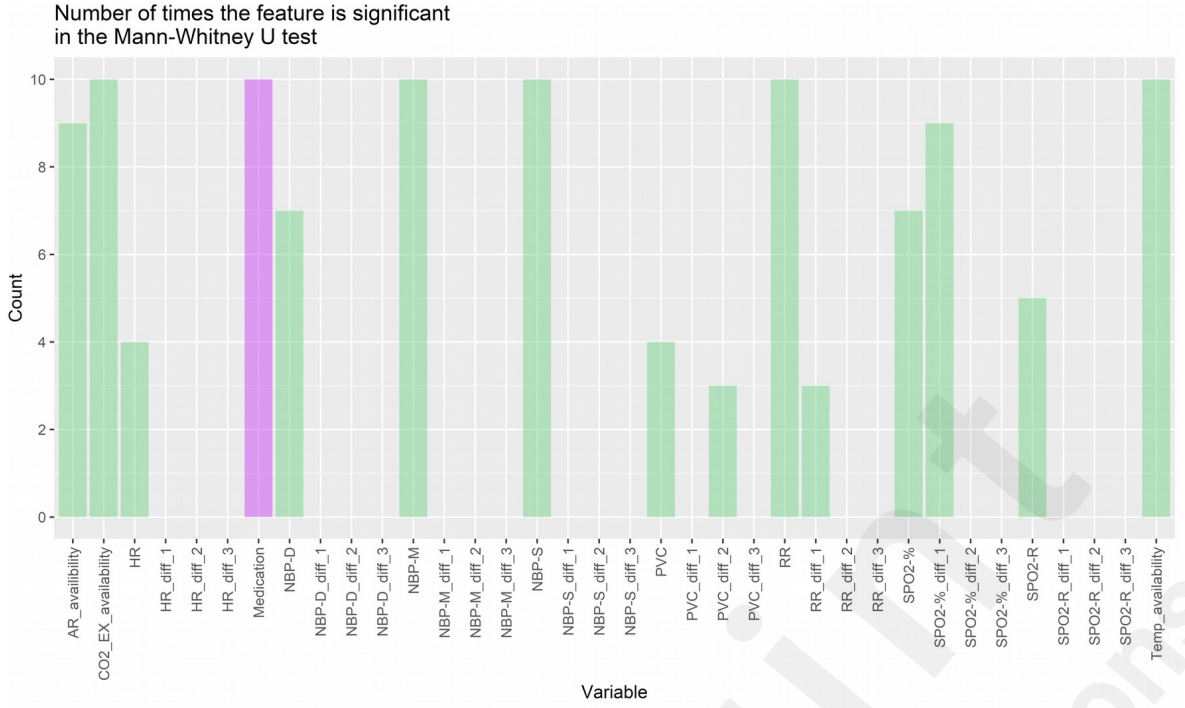


Figure 5 illustrates the frequency of significant features identified through the Mann-Whitney U test conducted during model cross-validation. This is consistent across all of the classifiers, as they all had the same groups. The green bars represent 35 features that showed significance, while the pink bars indicate significance when the medication variable was included in sensitivity analyses. Variables that were consistently significant across multiple folds included CO2-EX_availability (indicating intubation status), NBP-M, NBP-S, RR, and Temp_availability (indicating internal body temperature probe presence). Following these key features, others frequently determined as significant were AR_availability (arterial line presence), NBP-D, PVC, PVC_diff_2, RR_diff_1, SPO2-%, and SPO2-R. These features exhibited varying degrees of importance across cross-validation folds, suggesting their potential relevance in detecting hypovigilance episodes. Notably, the use of IV sedation and analgesia medication variable was significant in 10 instances, highlighting its importance in our models.

Classification Results

The classification results of our three AI classifiers, XGBoost, Random Forest (RF), and LightGBM, along with an additional set of the three models incorporating the sedative/analgesic medication variable, are presented in Table 4.

Table 3: Classification performance metrics for our three AI models.

Model	Average accuracy	Average precision	Average recall (Sensitivity)	Average AUC	Average F1
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)

Models without incorporating the sedative/analgesic medication variable					
XGBoost	0.66 (0.11)	0.75(0.10)	0.73 (0.18)	0.58 (0.09)	0.68 (0.10)
Random Forest	0.67 (0.15)	0.76 (0.11)	0.68 (0.22)	0.60 (0.12)	0.69 (0.12)
LightGBM	0.68 (0.12)	0.76 (0.11)	0.74 (0.18)	0.60(0.12)	0.69 (0.11)
Models with the incorporation of the sedative/analgesic medication variable					
XGBoost	0.70 (0.15)	0.76 (0.13)	0.70 (0.19)	0.62 (0.14)	0.72(0.14)
Random Forest	0.71 (0.15)	0.77 (0.13)	0.64 (0.25)	0.63 (0.13)	0.72 (0.15)
LightGBM	0.70 (0.15)	0.76 (0.13)	0.71 (0.21)	0.62 (0.12)	0.72 (0.14)

For the models without the inclusion of the sedative/analgesic medication variable, the LightGBM model demonstrated the highest average accuracy, average precision, average recall, average AUC, and average F1 score. Furthermore, it exhibited an average sensitivity of 74% with a standard deviation of 18% and a average precision of 76% with a standard deviation of 11%. XGBoost followed as the second-best classifier with an average recall of 73% (SD 18%) and an average sensitivity of 75% (SD 10%). When the sedative/analgesic medication variable was incorporated, LightGBM remained the top performing classifier, closely followed by XGBoost and RF. Their performances are relatively similar except in terms of average recall, where both XGBoost and LightGBM have an average recall of 70% and 71% respectively and RF achieved 64%.

Feature importance

Figure 6: SHapley Additive exPlanations (SHAP) values of the LightGBM model without incorporating the sedative/analgesic medication variable model.

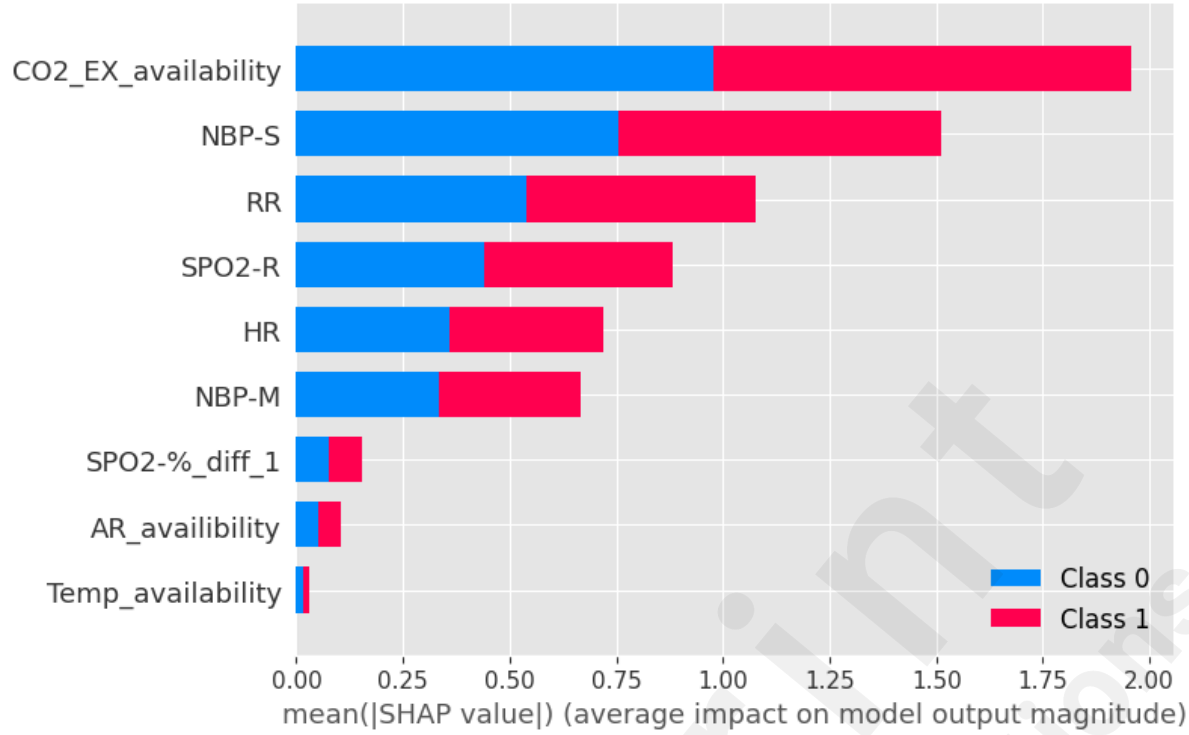


Figure 6 depicts that the CO2-EX availability (intubation variable), NBP-S and RR were particularly influential features in the model. SPO2-R, HR, and the NBP-M were also influential. The *y*-axis displays the feature names in order of importance from greatest to least importance. The *x*-axis represents the average SHAP value, which indicates the impact of each feature on the model. The model used to generate this figure is a LightGBM generated in the cross-validation that used only these five features as inputs because the Mann-Whitney was significant only for these features.

Discussion

This study aimed to develop an AI model to identify hypovigilance episodes in patients using data from a single ICU. Our results demonstrate that the differentiation of episodes of hypovigilance from non-hypovigilance episodes is possible with three different classifiers using routinely acquired clinical ICU data.

Principal Results

Our LightGBM classifier showed slightly better results than XGBoost and RF across multiple evaluation criteria. However, all the classifiers show significant variability in correctly identifying positive instances (true positives) across different folds of the cross-validation process. The standard deviation of our results may be due to differences in participant characteristics or class imbalances during training. The precision of 76% with a standard deviation of 11% indicates the ability to correctly identify positive instances among all instances classified as positive across different folds. LightGBM classifiers generally outperformed both XGBoost and Random Forest in terms of average accuracy, precision, recall, AUC, and F1 score. These results demonstrate the potential for future prospective external validation with larger data sets. An average recall rate of 74% indicates effective detection of hypovigilant episodes. The XGBoost algorithm achieved a recall rate of 73% and an average precision rate of 75%. A screening tool needs to be sensitive. This is best addressed by a model with a high recall [55]. Although these values may not be exceptionally high individually, collectively they demonstrate promise for potential future refinement and validation for clinical use.

Our analysis also revealed a clear and anticipated improvement in model performance with the inclusion of IV sedative and analgesic medications, correlating with the occurrence of hypovigilant episodes. This suggests the importance of incorporating data about IV sedative and analgesic medications into a predictive model for monitoring and predicting hypovigilant episodes, and eventually a continuous delirium prediction model.

Comparison With Prior Work

The majority of research on hypovigilance has previously been conducted in laboratory settings, which offer a highly regulated setting but might not accurately reflect real-world ICU situations [28]. The field of vigilance research has been

hindered by inconsistent and poorly defined terminology, making it hard to compare our results to all of the existing literature on the subject [27]. The existing research is also mostly focused on driving and flight simulations, during which operators do experience hypovigilance, but it may not present the same as ICU patients [28]. Our results from the ICU setting are interesting, as bedside nurses monitor patients' vigilance hourly, and medical devices automatically capture physiological markers used to monitor patients. These markers can be used to monitor vigilance automatically.

Comparable delirium prediction models, which also use non-invasive features show similar or slightly better performances. For example, the model developed by Oh et al. (2018) [55] achieved a balanced accuracy of 70%, with a maximum accuracy of 71.5%, only using automatically collected variables, notably HRV. These results are in line with other delirium studies using EEG and ECG [57]. Despite our lower performance compared to Oh et al. (2018), our model still shows promise considering that we derived our algorithm on a smaller dataset without HRV data.

Current delirium prediction models rely on static data collected by clinicians, such as age, mechanical ventilation, and sedative administration. Most models use a single snapshot in time, typically within 24 hours of admission, and do not account for fluctuations in a patient’s condition during their ICU stay [2]. This approach is inconsistent with critical illness and delirium pathophysiology [2]. To improve the accuracy of delirium diagnostic and prediction models, we considered the dynamic nature of a patient's condition and incorporated real-time data into our models [2]. A future enhanced and more accurate automated model could potentially offer real-time patient monitoring throughout their ICU stay. Such a model could utilize data that is generally accessible across all ICUs. Further, the variability of the baseline "gold standard" in hypovigilance detection or prediction models is a significant challenge. Different studies use diverse measures, some lacking prior validation [28]. To address this, our study used two validated sedation scales (RSS and RASS) in a clinical setting, incorporating a validated gold standard to enhance the labeling process of our hypovigilant state.

Strengths

Our study has some strengths. First, by conducting our research in a real-world ICU setting and using routinely collected vital signs and physiological markers in ICUs around the world, we ensured that our findings are relevant and transferable to similar healthcare settings. In addition, we employed rigorous data collection methods using an automated vital sign data collection system. This ensures the consistency and accuracy of our data set, minimizing the risk of classification bias. Our data collection and classification methods are entirely non-invasive and exclude procedures such as blood tests or EEGs, thereby emphasizing the integration potential into machine learning-based autonomous systems. Given that no baseline sociodemographic variables such as age, gender, past medical history, or other clinical variables were included in our model, our model without the IV sedative/analgesic medication variable could stand alone without any human-collected data. Additionally, our AI-derived detection model based on automatically collected vital signs is agnostic to language which would make our model more equitable for patients who speak languages different from their healthcare providers.

The study analyzed numerous episodes of hypovigilance and non-hypovigilance in ICU patients. These episodes often lasted for extended periods, making them suitable for detailed analysis. The experienced nurses' regular assessments of vigilance using the RASS or RSS provided a rich dataset for training machine learning models. The expertise and familiarity of the bedside nurses with these standard assessment tools contributed to the reliability and credibility of our outcome measures.

Our study also took advantage of the routine use of IV sedatives and analgesia medications in the ICU, such as propofol, hydromorphone, fentanyl, and midazolam that induce prolonged states of deep sedation. This provided valuable opportunities to detect episodes of hypovigilance, allowing us to refine our model and improve its effectiveness in identifying clinically relevant conditions. Identification of hypovigilance has important implications for healthcare settings for screening persons at risk of delirium. Delirium screening is a time-consuming task that requires completing multicomponent screening tools such as the Confusion Assessment Method for the Intensive Care Unit. Motivated by the growing sophistication of AI models in the medical domain, our project investigated the possibility of using machine learning to enhance the screening capacity in the context of nursing shortages. This potentially represents a viable path towards improving patient outcomes and decreasing the workload of healthcare professionals [58]. More work is still needed to develop a practical and reliable tool for hypovigilance assessment in the context of detecting and predicting delirium episodes. The creation of AI algorithms capable of timely detecting the onset of hypovigilant episodes may allow clinicians to apply treatments or mitigation measures, thereby enhancing patient outcomes in the ICU setting.

Limitations

We utilized PROBAST to assess the risk of bias in our model [34]. The PROBAST checklist is provided in Appendix B. Based on this assessment, we identified several limitations to our study. First, our study only recruited a small cohort consisting of only 30 participants. To mitigate this limitation, we employed a 10-fold group cross-validation approach. Future studies that include a larger sample in other ICUs will help identify new patterns in physiological marker fluctuations that help identify hypovigilant states. The cross-validation method used to evaluate the model's performance across multiple patient groups helped make use of all our small dataset. The cross-validation strategy also helped identify stable and reliable model performance metrics, minimizing overfitting risk and providing more accurate estimates of the model's true performance on the whole patient dataset. Our model's wide standard deviations are attributed to our small dataset. Nonetheless, our models showed moderate discriminative power, surpassing chance levels which suggests a hopeful path for future refinement and improvement.

A second limitation comes from the fact that our models were built using physiological data captured at low frequency using one-minute intervals. Low-frequency data collection limited our ability to capture the subtle changes in high-frequency variations that could be manifested in the transition from non-hypovigilant to hypovigilant states, such as changes in HRV. HRV would have been a valuable characteristic – as shown in other contexts [28] - but we could not collect it with our current GE CARESCAPE Gateway setup. To address this lack of HRV, we investigated the use of derivative heart rate metrics in our study and the use of other cardiac measures such as PVC.

While other, more sophisticated hypovigilance detection models incorporate electroencephalography (EEG) data [28], continuous EEG data was not available in our study. In fact, few ICUs have access to continuous EEG monitoring for all their patients at the same time, underscoring the importance of developing a model that does not rely on these complicated sensors to ensure its generalizability to all patients. Hence, our approach enables the evaluation of hypovigilance detection in a broader context, where EEG may not be available.

Finally, medication administration is not automatically or digitally captured in our ICU. Charting of medication administration was manually documented by bedside nurses in patients charts. Our study also limited its data collection only to intravenous (IV) sedatives and analgesics administered, ignoring for the moment the potential impact of orally administered sedatives, analgesics and psychoactive medications. Future AI algorithms that use data from digital medication administration charting tools would offer new opportunities to build better AI algorithms to detect and predict hypovigilant states and hypoactive delirium episodes.

Conclusion

We developed an automatic machine learning algorithm to detect hypovigilance in ICU patients using only routine and easily captured physiological parameters. The classifiers presented in this study demonstrated that hypovigilance could be distinguished from non-hypovigilance cases with modest results. Our models exhibited consistency and potential for improvement. Our study underscores the potential of machine learning algorithms in real-world clinical settings and identifies avenues for future research to enhance the detection of hypovigilance and improve patient outcomes.

Acknowledgments

We would like to thank the research team at the CISSS-CA for their help in implementing this project.
We would like to express our gratitude to Pascal Smith, Emilie Côté, Stéphane Turcotte, Jean-François Gagnon, Maxime Huot-Lavoie, Laurie Couture, and Bao Tran Tran for their valuable contributions to this project at the CISSS-CA. Special thanks to Nathalie Germain for her assistance during the writing process.

PA obtained funding to conduct the study. WW, MNH, PA, and SD actively contributed to the research design, coordination, and data acquisition. RG and VN conducted data analysis and coding. All authors played a crucial role in writing the manuscript and all approved the submitted version. Additionally, PA, SD, and PD were RG’s master’s thesis supervisors.

Funding

This study was funded by a MITACS grant attributed to Raphaelle Giguère. Thales Research & Technology provided an unrestricted grant to acquire the GE CARESCAPE Gateway. The Fondation de l’Hôtel-Dieu de Lévis

provided an operating grant. Patrick Archambault held a Fonds de recherche du Québec – Santé Clinical Scholar Award during this project. No funding organization influenced the interpretation or the reporting of the results. Raphaëlle Giguère received funding from the NSERC CREATE Program.

Conflicts of Interest

None declared.

Abbreviations

AUC: Area Under the Curve
AI: artificial intelligence
AR-D: arterial pressure diastolic
AR-M: arterial pressure mean
AR-R: arterial pressure pulse rate
AR-S: arterial pressure systolic
CFS: Canadian Frailty Score
CO2-EX: CO2 expired while intubated
CO2-IN: CO2 inspired while intubated
DSM-5: American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EEG: Electro-encephalography
ECG: Electro-cardiography
FAQ: Pfeffer Functional Activities Questionnaire
GE: General electric
HR: Heart rate
HRV: Heart rate variability
ICU: Intensive care unit
LightGBM: light gradient-boosting machine classifier
NBP-D: Non invasive blood pressure diastolic
NBP-M: Non invasive blood pressure mean
NBP-S: Non invasive blood pressure systolic
PNS: parasympathetic autonomic nervous system
PROBAST: Prediction model Risk Of Bias ASsessment Tool
PVC: Premature ventricular contraction
RASS: Richmond Agitation and Sedation Scale
RF: Random forest
RR: Respiratory rate
RSS: Ramsay sedation scale
SHAP: SHapley Additive exPlanations
SPO2-%: Oxygen saturation
SPO2-R: Pulse oximeter pulse rate
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
XGBoost: eXtreme Gradient Boosting

Multimedia Appendix

TRIPOD + AI (grid)
PROBAST (grid)
Questionnaires used to describe the cohort (docfile)

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. American Psychiatric Association; 2013.

2. Ruppert MM, Lipori J, Patel S, Ingersent E, Cupka J, Ozrazgat-Baslanti T, Loftus T, Rashidi P, Bihorac A. ICU Delirium-Prediction Models: A Systematic Review. Crit Care Explor 2020 Dec 16;2(12):e0296. PMID:33354672

3. Ely EW, Margolin R, Francis J, May L, Truman B, Dittus R, Speroff T, Gautam S, Bernard GR, Inouye SK. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit Care Med 2001 Jul;29(7):1370–1379. PMID:11445689

4. Fiest KM, Soo A, Hee Lee C, Niven DJ, Ely EW, Doig CJ, Stelfox HT. Long-Term Outcomes in ICU Patients with Delirium: A Population-based Cohort Study. Am J Respir Crit Care Med 2021 Aug 15;204(4):412–420. PMID:33823122

5. Lipowski ZJ. Transient cognitive disorders (delirium, acute confusional states) in the elderly. Am J Psychiatry 1983 Nov;140(11):1426–1436. PMID:6624987

6. Pisani MA, Murphy TE, Araujo KLB, Van Ness PH. Duration of ICU Delirium, Severity of the Underlying Disease, and Mortality. American Journal of Respiratory and Critical Care Medicine 2010;181(4):420–421. doi: 10.1164/ajrccm.181.4.420

7. Kiely DK, Jones RN, Bergmann MA, Marcantonio ER. Association between psychomotor activity delirium subtypes and mortality among newly admitted post-acute facility patients. J Gerontol A Biol Sci Med Sci 2007 Feb;62(2):174–179. PMID:17339642

8. Fang C-K, Chen H-W, Liu S-I, Lin C-J, Tsai L-Y, Lai Y-L. Prevalence, Detection and Treatment of Delirium in Terminal Cancer Inpatients: A Prospective Survey. Japanese Journal of Clinical Oncology 2008 Jan 1;38(1):56–63. doi: 10.1093/jjco/hym155

9. Pandharipande PP, Ely EW, Arora RC, Balas MC, Boustani MA, La Calle GH, Cunningham C, Devlin JW, Elefante J, Han JH, MacLullich AM, Maldonado JR, Morandi A, Needham DM, Page VJ, Rose L, Salluh JIF, Sharshar T, Shehabi Y, Skrobik Y, Slooter AJC, Smith HAB. The intensive care delirium research agenda: a multinational, interprofessional perspective. Intensive Care Med 2017 Sep;43(9):1329–1339. PMID:28612089

10. Kotfis K, Marra A, Ely EW. ICU delirium - a diagnostic and therapeutic challenge in the intensive care unit. Anaesthesiol Intensive Ther 2018;50(2):160–167. PMID:29882581

11. Reppas-Rindlisbacher C, Shin S, Purohit U, Verma A, Razak F, Rochon P, Sheehan K, Rawal S. Association between non-English language and use of physical and chemical restraints among medical inpatients with delirium. Journal of the American Geriatrics Society 2022;70(12):3640–3643. doi: 10.1111/jgs.17989

12. van Schie MKM, Lammers GJ, Fronczek R, Middelkoop HAM, van Dijk JG. Vigilance: discussion of related concepts and proposal for a definition. Sleep Med 2021 Jul;83:175–181. PMID:34022494

13. Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 2005;28:403–450. PMID:16022602

14. Bouret S, Sara SJ. Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. Eur J Neurosci 2004 Aug;20(3):791–802. PMID:15255989

15. Nieuwenhuis S, De Geus EJ, Aston-Jones G. The anatomical and functional relationship between the P3 and autonomic components of the orienting response. Psychophysiology 2011 Feb;48(2):162–175. PMID:20557480

16. Rajkowski J, Majczynski H, Clayton E, Aston-Jones G. Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task. J Neurophysiol 2004 Jul;92(1):361–371. PMID:15028743

17. Elam M, Svensson TH, Thorén P. Locus coeruleus neurons and sympathetic nerves: activation by cutaneous sensory afferents. Brain Res 1986 Feb 26;366(1–2):254–261. PMID:3697682

18. Sara SJ, Bouret S. Orienting and reorienting: the locus coeruleus mediates cognition through arousal. Neuron 2012 Oct 4;76(1):130–141. PMID:23040811

19. Wang C-A, Munoz DP. A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. Curr Opin Neurobiol 2015 Aug;33:134–140. PMID:25863645

20. Arslan D, Ünal Çevik I. Interactions between the painful disorders and the autonomic nervous system. Agri 2022 Jul;34(3):155–165. PMID:35792695

21. Cygankiewicz I, Zareba W. Heart rate variability. Handb Clin Neurol 2013;117:379–393. PMID:24095141

22. Xhyheri B, Manfrini O, Mazzolini M, Pizzi C, Bugiardini R. Heart Rate Variability Today. Progress in Cardiovascular Diseases 2012;55(3):321–331. doi: https://doi.org/10.1016/j.pcad.2012.09.001

23. Mistraletti G, Mantovani ES, Cadringher P, Cerri B, Corbella D, Umbrello M, Anania S, Andrighi E, Barello S, Carlo AD, Martinetti F, Formenti P, Spanu P, Iapichino G. Enteral vs. intravenous ICU sedation management: study protocol for a randomized controlled trial. Trials BMC; 2013;14:92. PMID:23551983

24. Pandharipande PP, Pun BT, Herr DL, Maze M, Girard TD, Miller RR, Shintani AK, Thompson JL, Jackson JC, Deppen SA, Stiles RA, Dittus RS, Bernard GR, Ely EW. Effect of sedation with dexmedetomidine vs lorazepam on acute brain dysfunction in mechanically ventilated patients: the MENDS randomized controlled trial. JAMA 2007 Dec 12;298(22):2644–2653. PMID:18073360

25. Ouimet S, Kavanagh BP, Gottfried SB, Skrobik Y. Incidence, risk factors and consequences of ICU delirium. Intensive Care Medicine 2007 Jan 1;33(1):66–73. doi: 10.1007/s00134-006-0399-8

26. Riker RR, Fraser GL. Adverse events associated with sedatives, analgesics, and other drugs that provide patient comfort in the intensive care unit. Pharmacotherapy 2005 May;25(5 Pt 2):8S-18S. PMID:15899744

27. Oken BS, Salinsky MC, Elsas SM. Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin Neurophysiol 2006 Sep;117(9):1885–1901. PMID:16581292

28. Marois A, Kopf M, Fortin M, Huot-Lavoie M, Martel A, Boyd JG, Gagnon J-F, Archambault PM. Psychophysiological models of hypovigilance detection: A scoping review. Psychophysiology 2023;60(11):e14370. doi: 10.1111/psyp.14370

29. Alkhachroum A, Appavu B, Egawa S, Foreman B, Gaspard N, Gilmore EJ, Hirsch LJ, Kurtz P, Lambrecq V, Kromm J, Vespa P, Zafar SF, Rohaut B, Claassen J. Electroencephalogram in the intensive care unit: a focused look at acute brain injury. Intensive Care Med 2022 Oct;48(10):1443–1462. PMID:35997792

30. Zhang J, Li J, Huang Z, Huang D, Yu H, Li Z. Recent Progress in Wearable Brain–Computer Interface (BCI) Devices Based on Electroencephalogram (EEG) for Medical Applications: A Review. Health Data Science 2023;3:0096. doi: 10.34133/hds.0096

31. Pendlebury ST. Delirium screening in older patients. Age and Ageing 2018 Jul;47(5):635–637. doi: 10.1093/ageing/afy103

32. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015 Jan 6;162(1):55–63. PMID:25560714

33. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Calster BV, Ghassemi M, Liu X, Reitsma JB, Smeden M van, Boulesteix A-L, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L, Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ British Medical Journal Publishing Group; 2024 Apr 16;385:e078378. PMID:38626948

34. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med United States; 2019 Jan 1;170(1):51–58. PMID:30596875

35. Giguère R, Niaussat V, Noël-Hunter M, Witteman W, Paul TS, Marois A, Després P, Duchesne S, Archambault P. Identifying Episodes of Hypovigilance in Intensive Care Units Using Routine Physiological Parameters and Artificial Intelligence: a Derivation Study. Open Code and Dataset. Zenodo; 2024. doi: 10.5281/zenodo.11241914

36. Ely EW, Truman B, Shintani A, Thomason JWW, Wheeler AP, Gordon S, Francis J, Speroff T, Gautam S, Margolin R, Sessler CN, Dittus RS, Bernard GR. Monitoring Sedation Status Over Time in ICU PatientsReliability and Validity of the Richmond Agitation-Sedation Scale (RASS). JAMA 2003 Jun 11;289(22):2983–2991. doi: 10.1001/jama.289.22.2983

37. Sessler CN, Grap MJ, Brophy GM. Multidisciplinary management of sedation and analgesia in critical care. Semin Respir Crit Care Med 2001;22(2):211–226. PMID:16088675

38. Ramsay MAE, Savege TM, Simpson BRJ, Goodwin R. Controlled Sedation with Alphaxalone-Alphadolone. Br Med J 1974 Jun 22;2(5920):656–659. PMID:4835444

39. Namigar T, Serap K, Esra AT, Özgül O, Can ÖA, Aysel A, Achmet A. The correlation among the Ramsay sedation scale, Richmond agitation sedation scale and Riker sedation agitation scale during midazolam-remifentanil sedation. Brazilian Journal of Anesthesiology (English Edition) 2017 Jul 1;67(4):347–354. doi: 10.1016/j.bjane.2016.07.002

40. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974 Jul 13;2(7872):81–84. PMID:4136544

41. Jain S, Iverson LM. Glasgow Coma Scale. StatPearls Treasure Island (FL): StatPearls Publishing; 2024. PMID:30020670

42. Pfeffer RI, Kurosaki TT, Harrah CHJ, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol United States; 1982 May;37(3):323–329. PMID:7069156

43. Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, Mitnitski A. A global clinical measure of fitness and frailty in elderly people. CMAJ 2005 Aug 30;173(5):489–495. PMID:16129869

44. Chan B, Sedghi A, Laird P, Maslove D, Mousavi P. Predictive Modeling using Intensive Care Unit Data: Considerations for Data Pre-processing and Analysis. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019. p. 3429–3432. doi: 10.1109/EMBC.2019.8857564

45. Stewart J 1941-. Student solutions manual for Stewart's Calculus : early transcendentals, third edition. Pacific Grove, Calif. : Brooks/Cole Pub., c1995-; Available from: https://search.library.wisc.edu/catalog/999826642402121

46. Gibs P. Third derivative of displacement. 1996. Available from: https://math.ucr.edu/home/baez/physics/General/jerk.html [accessed Mar 27, 2024]

47. McKnight PE, Najab J. Mann-Whitney U Test. The Corsini Encyclopedia of Psychology John Wiley & Sons, Ltd; 2010. p. 1–1. doi: 10.1002/9780470479216.corpsy0524

48. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12(85):2825–2830.

49. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining New York, NY, USA: Association for Computing Machinery; 2016. p. 785–794. doi: 10.1145/2939672.2939785

50. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems Curran Associates, Inc.; 2017. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html [accessed May 9, 2024]

51. Qian Q, Wu J, Wang J, Sun H, Yang L. Prediction Models for AKI in ICU: A Comparative Study. Int J Gen Med 2021;14:623–632. PMID:33664585

52. Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, Han B, Wang H, Duan X, Sun T. A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. Front Public Health 2021 Oct 15;9:754348. PMID:34722452

53. Putatunda S, Rama K. A Comparative Analysis of Hyperopt as Against Other Approaches for Hyper-Parameter Optimization of XGBoost. Proceedings of the 2018 International Conference on Signal Processing and Machine Learning New York, NY, USA: Association for Computing Machinery; 2018. p. 6–10. doi: 10.1145/3297067.3297080

54. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems Curran Associates, Inc.; 2017. Available from: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html [accessed Apr 19, 2024]

55. Bhattacharyya A, Sheikhalishahi S, Torbic H, Yeung W, Wang T, Birst J, Duggal A, Celi LA, Osmani V. Delirium prediction in the ICU: designing a screening tool for preventive interventions. JAMIA Open 2022 Jul 1;5(2):ooac048. doi: 10.1093/jamiaopen/ooac048

56. Oh J, Cho D, Park J, Na SH, Kim J, Heo J, Shin CS, Kim J-J, Park JY, Lee B. Prediction and early detection of delirium in the intensive care unit by using heart rate variability and machine learning. Physiol Meas 2018 Mar 27;39(3):035004. doi: 10.1088/1361-6579/aaab07

57. van den Boogaard M, Pickkers P, Slooter AJC, Kuiper MA, Spronk PE, van der Voort PHJ, van der Hoeven JG, Donders R, van Achterberg T, Schoonhoven L. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. BMJ 2012 Feb 9;344:e420. PMID:22323509

58. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44–56. PMID:30617339

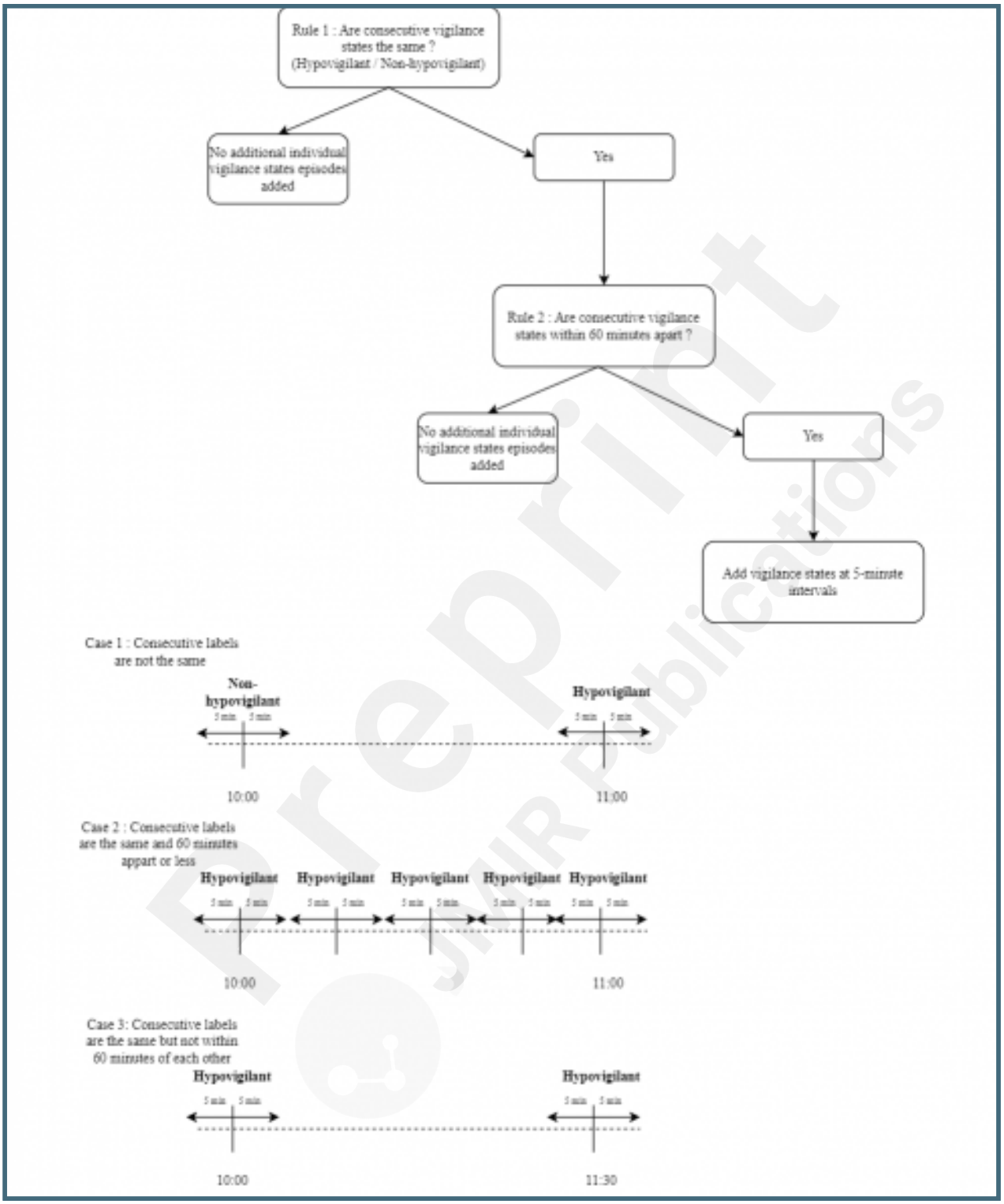
Supplementary Files

Figures

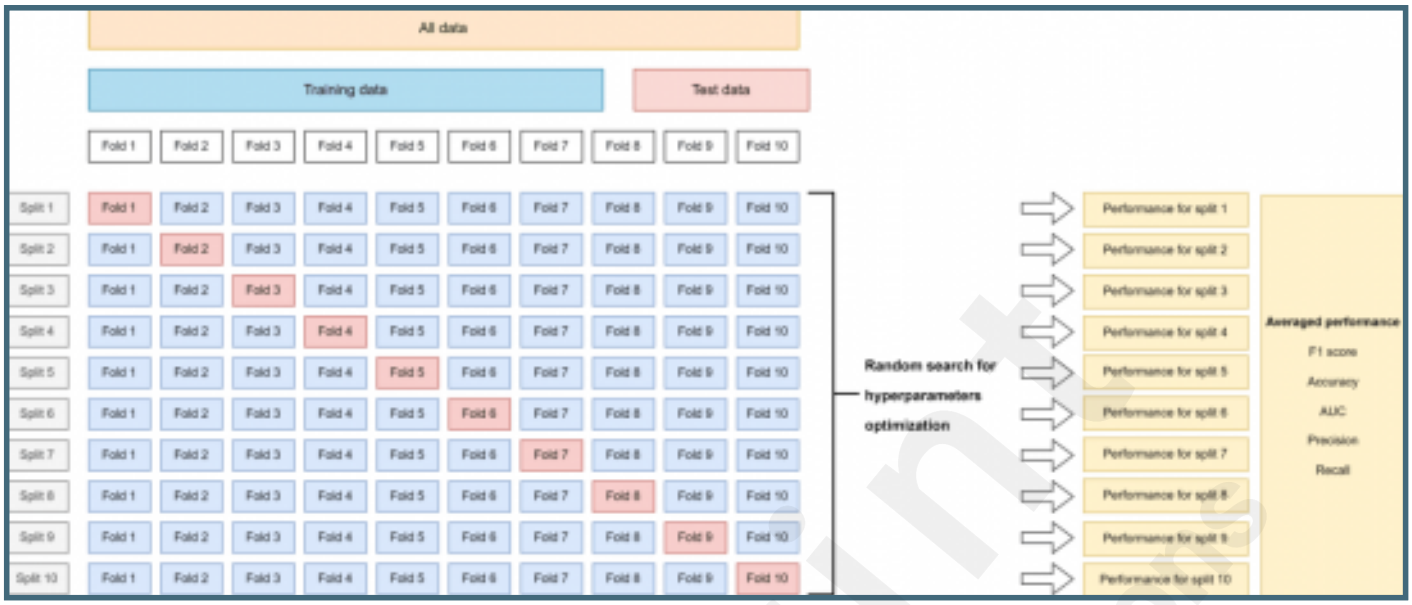
Labeling process to identify episodes of hypovigilance using the corresponding Ramsay Sedation Scale and Richmond Agitation and Sedation Scale.

	Ramsay	RASS	Vigilance state
Combative	1	+4	Non-hypovigilant
Very agitated		+3	Non-hypovigilant
Agitated		+2	Non-hypovigilant
Restless		+1	Non-hypovigilant
Alert and calm	2	0	Non-hypovigilant
Drowsy	3	-1	Hypovigilance
Light sedation		-2	Hypovigilance
Moderate sedation	4	-3	Hypovigilance
Deep sedation	5	-4	Hypovigilance
Unarousable	6	-5	Hypovigilance

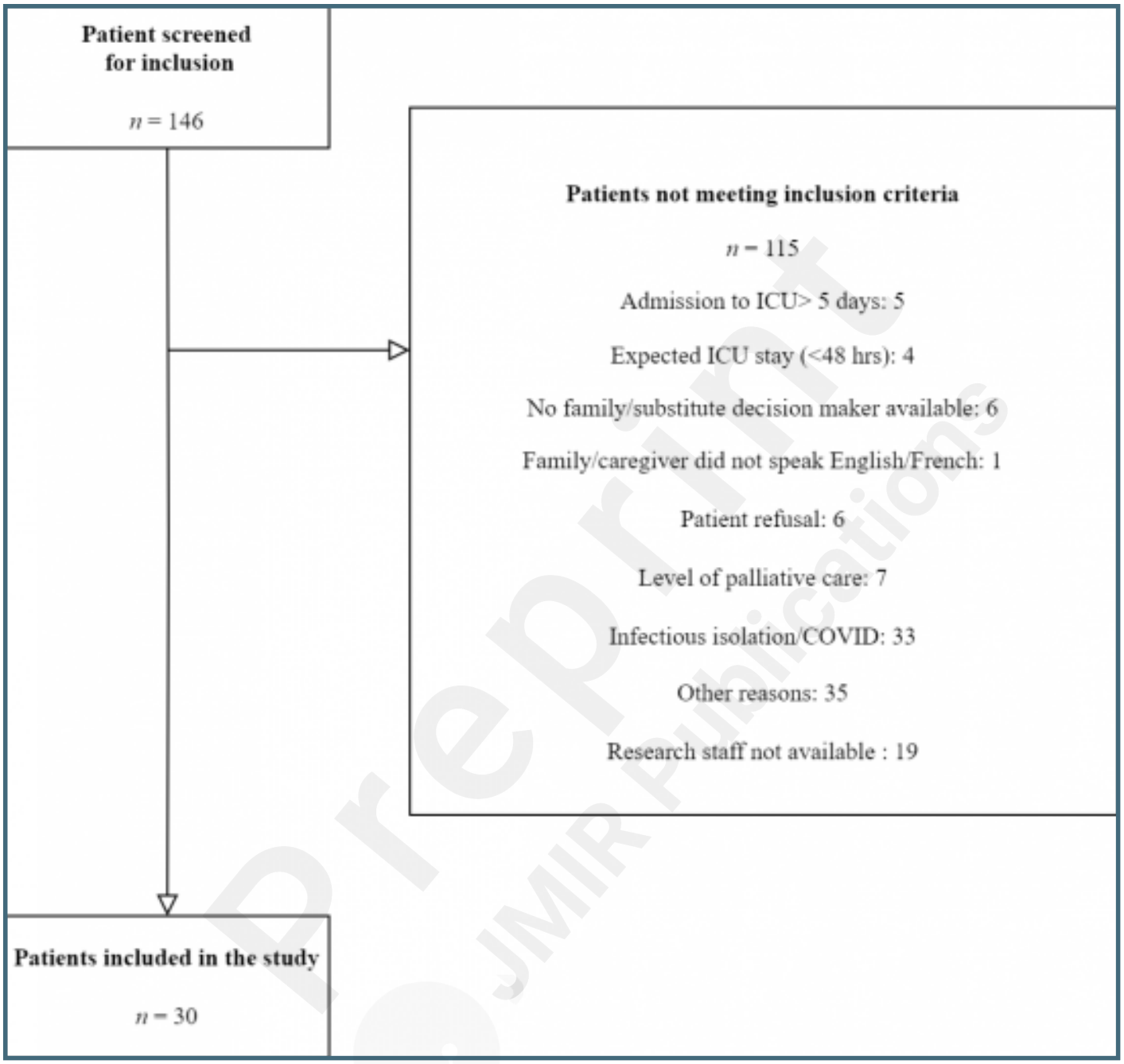
Label identification process.



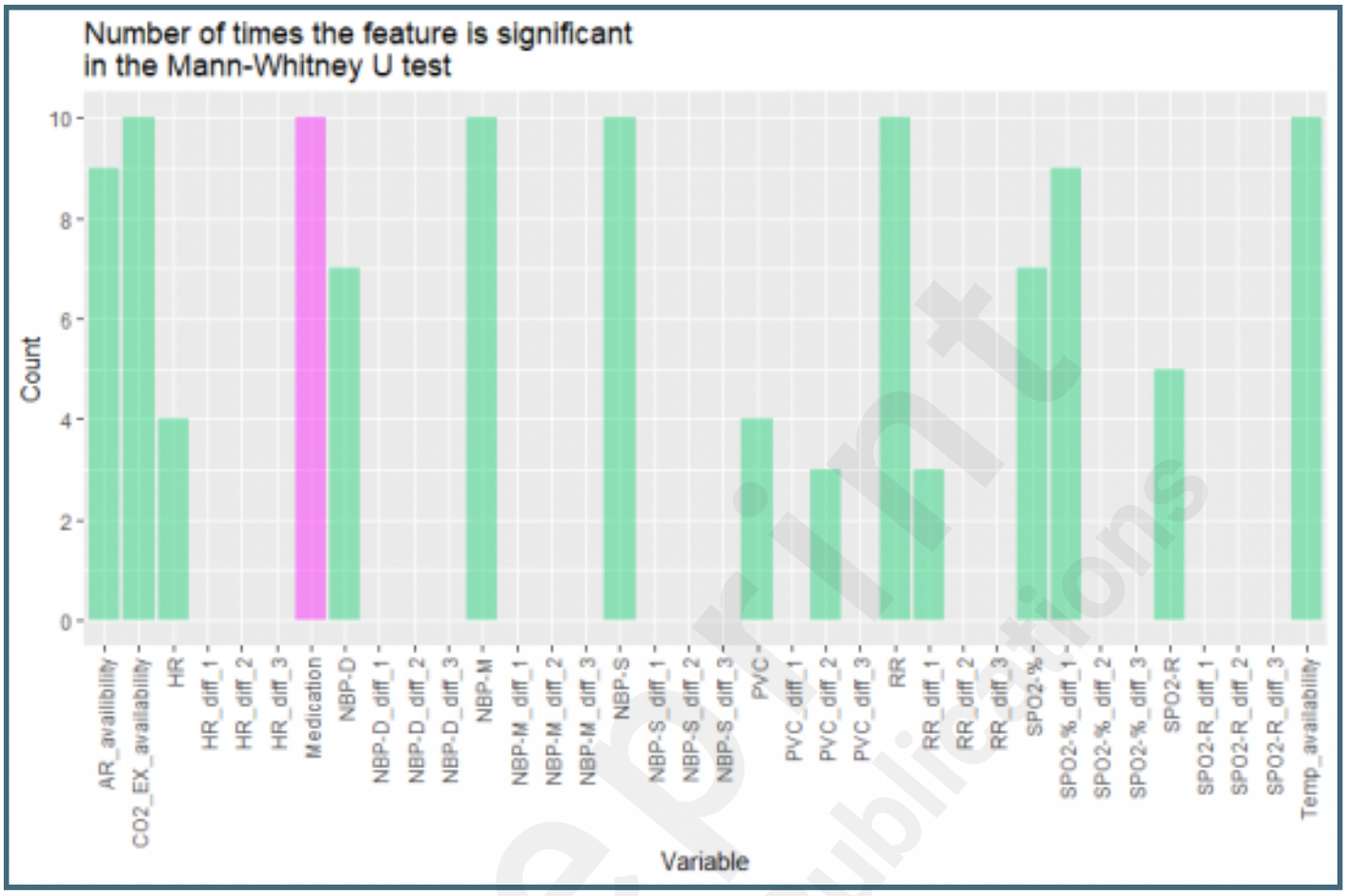
Data splitting and hyperparameters optimization to evaluate the models' performances.



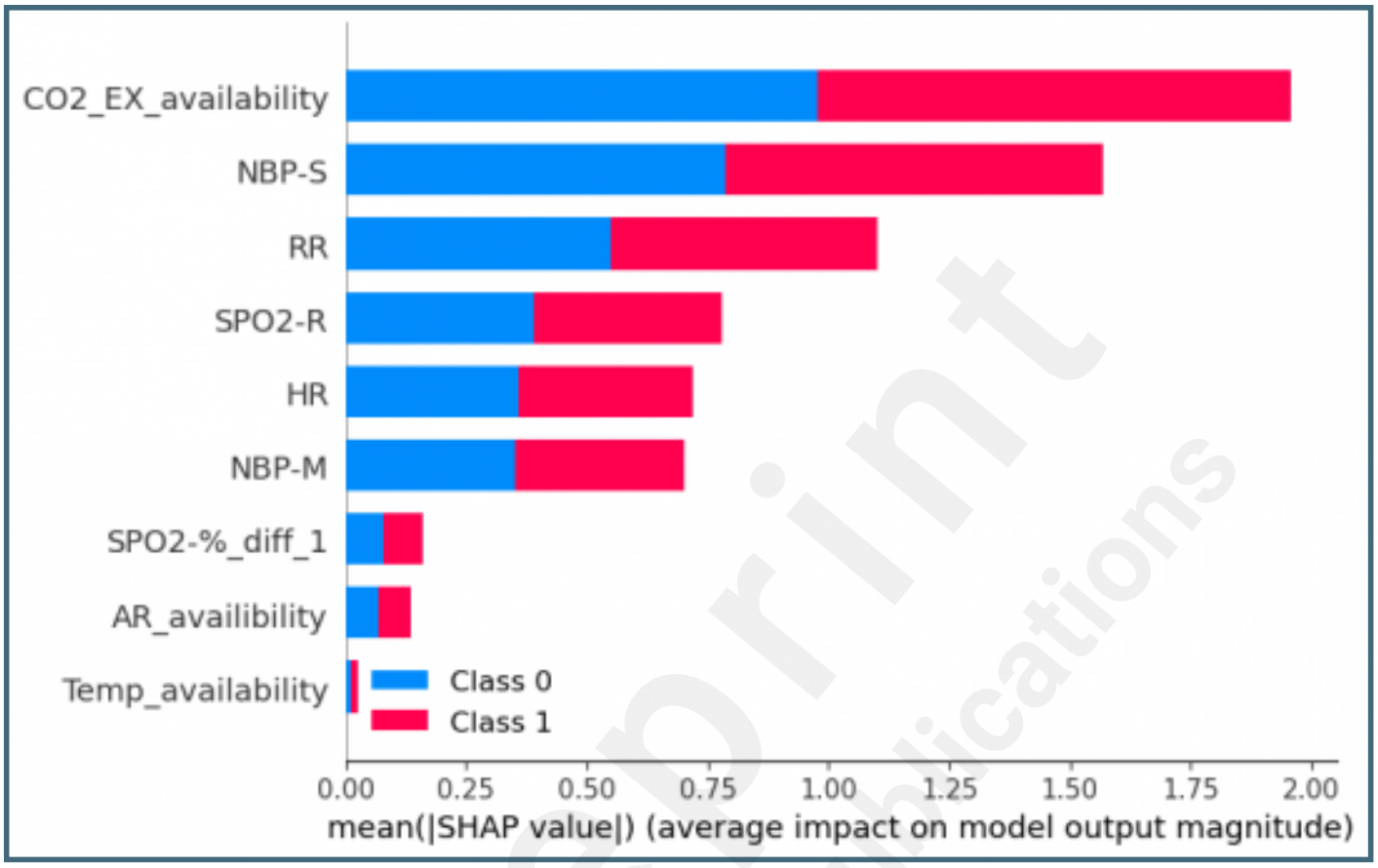
Flowchart of the data collection process.



Significant features in the models.



SHapley Additive exPlanations (SHAP) values of the LightGBM model without incorporating the sedative/analgesic medication variable model.



Multimedia Appendixes

Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis for machine learning models.
URL: http://asset.jmir.pub/assets/5e660326db0cd3d8883d475566a62452.pdf

PRediction model Of Bias ASsessment Tool (PROBAST) to identify potential biases.
URL: http://asset.jmir.pub/assets/35cf033305bbe1737c758a81d7799e48.docx

Questionnaires used to describe the cohort.
URL: http://asset.jmir.pub/assets/e7550d9b07ce6438e41deaffa27d75c2.docx