scientific reports

OPEN

Semi-automated tracking of pain in critical care patients using artificial intelligence: a retrospective observational study

Naoya Kobayashi[1✉], Takuya Shiga[1], Saori Ikumi[1], Kazuki Watanabe[2], Hitoshi Murakami[2] & Masanori Yamauchi[1]

Monitoring the pain intensity in critically ill patients is crucial because intense pain can cause adverse events, including poor survival rates; however, continuous pain evaluation is difficult. Vital signs have traditionally been considered ineffective in pain assessment; nevertheless, the use of machine learning may automate pain assessment using vital signs. This retrospective observational study was performed at a university hospital in Sendai, Japan. Objective pain assessments were performed in eligible patients using the Critical-Care Pain Observation Tool (CPOT). Three machine-learning methods—random forest (RF), support vector machine (SVM), and logistic regression (LR)—were employed to predict pain using parameters, such as vital signs, age group, and sedation levels. Prediction accuracy was calculated as the harmonic mean of sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC). Furthermore, 117,190 CPOT assessments were performed in 11,507 eligible patients (median age: 65 years; 58.0% males). We found that pain prediction was possible with all three machine-learning methods. RF demonstrated the highest AUROC for the test data (RF: 0.853, SVM: 0.823, and LR: 0.787). With this method, pain can be objectively, continuously, and semi-automatically evaluated in critically ill patients.

Patients in intensive care units (ICUs) experience a high incidence of pain, which requires detailed assessments[1–3]. It has been reported that 33% of intubated patients experience pain at rest[4], while an additional 56% experience pain during the procedure[5]. The primary source of pain in surgical patients is the surgical site, whereas that in medical patients includes the back and extremities; notably, the incidence was reported to be no different between surgical and medical patients[1]. Stress caused by pain generally has a detrimental effect on ICU patients. Increased catecholamine levels cause arteriole vasoconstriction and decrease tissue oxygen tension, thus, resulting in tissue perfusion failure[6]. Other reactions caused by pain include increased catabolism, such as lipolysis to provide protein substrates, and muscle loss[7]. Increased catabolism associated with tissue hypoxia delays wound healing and increases the risk of infections. Pain suppresses the activity of natural killer cells[8,9] and lowers the number of cytotoxic T-cells and phagocytic neutrophils[10]. Furthermore, acute pain is a major risk factor for subsequent chronic neuropathic pain.

In the ICU, objective methods of pain assessment, such as the Critical-Care Pain Observation Tool (CPOT)[11] and Behavioral Pain Scale[12] are recommended when patients cannot communicate due to sedatives, intubation, or tracheostomy[13,14]. Regular use of pain assessment scales has been reported to improve the clinical outcomes in ICU patients[15–17]; however, these methods are cumbersome and do not facilitate continuous monitoring of pain. Therefore, a method for continuous and easy assessment of patients that can enable healthcare professionals to perform interventions sooner is highly necessary.

Vital signs are considered indicators of pain; however, changes in vital signs have been reported to be nonreliable indicators of the extent of pain[9,12,18–20]. Notably, the studies that concluded the latter employed time-series data of vital signs at specific points rather than using continuous data because of computational complexity. Such data may not fully reflect the information available from vital signs because they cannot account for continuous changes in the pain.

[1]Department of Anesthesiology and Perioperative Medicine, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan. [2]Hitachi Solutions East Japan, Ltd., Sendai, Miyagi, Japan. ✉email: dynamis_air@yahoo.co.jp

In contrast, artificial intelligence (AI)-based predictive tools are increasingly resulting in automation of diagnostics through comprehensive monitoring of patient health, which is traditionally done by physicians and nurses, based on real-time and continuous recordings, processing, monitoring, and intelligent diagnosis[21]. Notably, AI-based tools have the potential to provide continuous and automated pain assessments that can eliminate the effects of uncertainty caused by discrete decision-making, such as that based on pain assessment scales[22–25]. However, only a few studies have actively employed continuous data, such as vital signs, in pain analysis. Adjei et al. predicted pain by quantifying pain sensitivity by analyzing the electrocardiogram (ECG) waveforms of 17 patients who underwent varicose vein surgery; however, the small number of patients and the use of questionnaires in pain assessment made it difficult to apply this method in unconscious patients[26]. There have also been attempts to detect pain through machine-learning analysis of electroencephalogram (EEG) and images[27–30]; however, they are difficult to interpret and cannot be used for continuous assessment. To the best of our knowledge, no studies have performed objective, continuous, and semi-automatic evaluation of pain in ICU patients. Therefore, this study aimed to investigate if AI analysis of vital signs could be used to automate the assessment of pain.

Results

In 11,527 eligible patients, CPOT assessment was performed 117,190 times and 3,925 patients (3.3%) were determined to experience pain with a score of ≥ 3. Although the level of sedation (Richmond Agitation-Sedation Scale; RASS) was assessed almost simultaneously, 42.8% of the patients were determined to be calm (0 points), 8.3% aggressive (≥ 1 point), and 48.9% sedated (≤ -1 point). Furthermore, 21.7% of the ratings indicated delirium (see Supplementary Table S1 online). As mentioned in the Methods section, we used three machine-learning methods (RF, SVM, and LR) to predict the presence of pain at the time of CPOT evaluation using parameters, such as vital signs, age group, and sedation level. The hyperparameter that derived the highest accuracy in each model was selected via a grid search (SVM: C = 10, gamma = 10, kernel = rbf. RF: max_depth = 14, estimates = 17, criterion = gini, LR: C = 1.0, tol = 0.0001). The results demonstrated that RF provided the highest area under the receiver operating characteristic curve (AUROC) (AUROC for test data: RF: 0.853; SVM: 0.823; LR: 0.787; Fig. 1). A review of the feature importance of the RF model showed that RASS exhibits a higher contribution rate as compared to the other predictors (Supplementary Table S3). Table 1 lists the results obtained by varying the sensitivity and specificity while using the RF model, which demonstrates the highest prediction accuracy. These results are based on the oversampling performed in the proposed method. By contrast, the results obtained using raw data without oversampling are presented in Fig. S2 and Table S4 (Supplementary Information).

Discussion

In this study, we examined if machine learning can be used for continuous and automatic pain assessment and to obtain results similar to those of traditional CPOT assessments performed by nurses using only vital signs and other data. We found that the pain experienced by ICU patients can be assessed objectively, continuously, and semi-automatically using machine learning with an accuracy of up to AUROC 0.853. In our experiments, RF demonstrated the best prediction performance. Our study is novel because we demonstrated the usefulness of vital signs that are measured continuously and constantly in the ICU and were previously considered ineffective in the evaluation of pain.

Severely ill adult patients experience moderate-to-severe pain at rest and during procedures of standard care[1,31]; intense pain can have adverse effects, such as immunosuppression and cardiac and respiratory failure[32]. Currently, an objective index, such as CPOT, is used when an ICU patient cannot express pain; however, continuous evaluation is difficult because of the patient's condition. Particularly, during the night or when the staff is busy, pain evaluation and care tend to be delayed. A consistent approach is important for pain assessment and management[33], especially regarding the use of opioids, the main analgesics in ICUs, which require careful dose adjustments to balance the benefits and potential risks[34–38]. To achieve this goal, continuous evaluation of pain is crucial; however, pain evaluation using the conventional methods only corresponds to a specific time point. The results of this study suggest that RASS, an index of sedation, is the only parameter among all the variables that need to be evaluated each time and that pain evaluation can be semi-automated using vital signs.

Recent research using machine learning and AI has corroborated the usefulness of analyzing data that are recorded continuously over time, such as vital signs, to gain better insights. In line with the present study, a previous machine-learning model analyzed the electronic medical record texts of 2695 patients with breast cancer undergoing chemotherapy and automatically extracted pain and other information with AUROC of 0.82[39]. Furthermore, a median accuracy of 70% was achieved when a deep-learning model—based on a combination of convolutional and recurrent neural networks—that automatically tracks both levels of consciousness and delirium using frontal EEG signals was employed, which allowed the latter model to predict the differences in RASS levels to within 1 point[40]. In a previous study, we continuously analyzed blood pressure values in patients with septic shock and found that both the extent of decrease in blood pressure and the duration of blood pressure affect the prognosis[41]. The use of text and EEGs in medical records may improve the accuracy of pain prediction; however, they cannot always be measured.

In conclusion, pain can be objectively, continuously, and semi-automatically predicted using machine learning. Of the methods tested, the prediction accuracy was the highest with the RF model. Regarding the limitations of our research and future challenges, it is necessary to improve the prediction accuracy in diverse situations. In this study, we were also able to predict pain with a high degree of accuracy, although not 100%. Clinical limitations in this regard include time differences in pain assessment and recording, and heterogeneity that occurred as a result of the assessments being performed by humans. To address these issues, we plan to develop a bedside device using the prediction algorithm proposed in this study. Subsequently, it is necessary to improve the
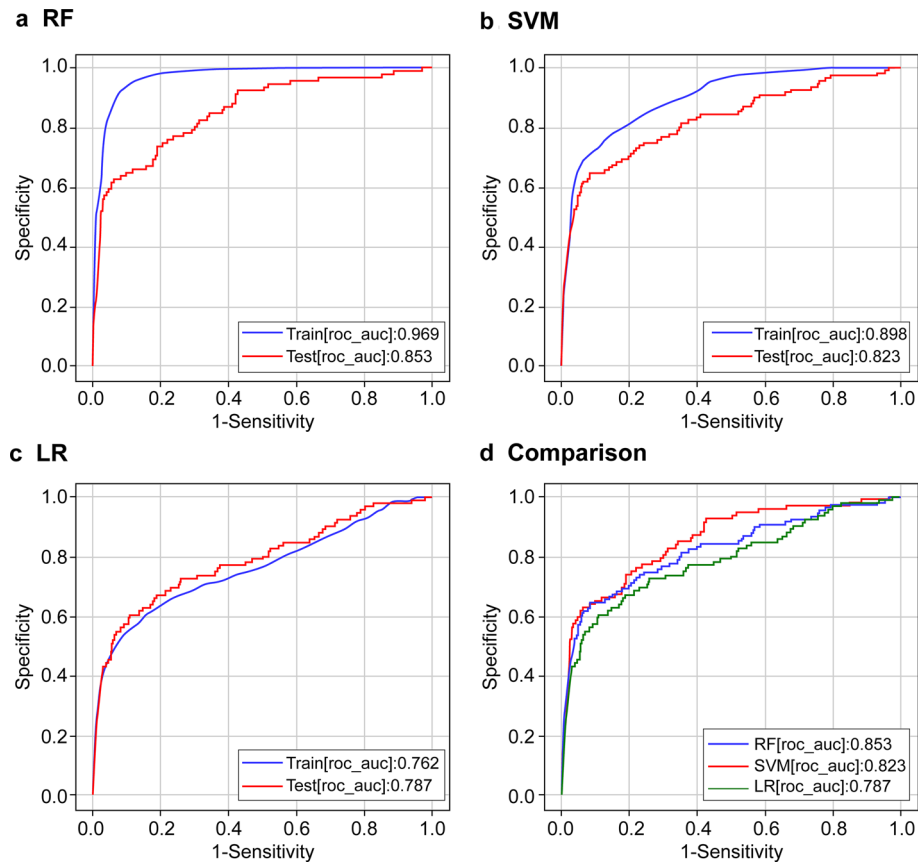
Figure 1. Prediction accuracy. (**a**–**c**) Accuracy of each machine-learning method. The blue line represents the accuracy in the training set, and the red line represents the accuracy in the test set. (**d**) Comparison of the verification data of the three machine-learning methods. The x-axis and y-axis represent the negative sensitivity and specificity in the ROC curve, respectively. The test accuracy depends on the extent to which the machine-learning model can correctly determine whether the CPOT score was < 2 or > 3. Accuracy is represented by AUROC; area of 1 represents the perfect test and that of 0.5 represents an inconclusive test.

Condition	Sensitivity	Specificity	Threshold
Sensitivity at predefined value	0.6	0.942	0.736
	0.7	0.809	0.752
	0.8	0.691	0.742
	0.9	0.576	0.703
Specificity at predefined value	0.868	0.6	0.718
	0.791	0.7	0.742
	0.736	0.8	0.768
	0.637	0.9	0.746
Optimal point	0.736	0.807	0.770

Table 1. Sensitivity, specificity, and threshold values at different operating points for RF-based pain tracking; the AUROC value of the prediction model equaled 0.853.

prediction accuracy in other environments to verify the results. The use of an automated and continuous pain assessment algorithm may allow for continuous pain assessment in all situations. It has the potential to not only immediately help relieve the pain in patients who cannot communicate but also increase their life expectancy. Therefore, more studies are warranted in a wider variety of scenarios for further evidence.

	Support vector machine	Random forest	Logistic regression
Explanation	A method of constructing a two-class pattern classifier using a linear input element. This is one of the supervised-learning-based pattern recognition models	A method of constructing numerous decision trees and carrying out a majority vote. It requires a large amount of data but can make highly accurate predictions	Calculates the class membership probability for two categories by fitting the log odds and explanatory variables to a model
Speed	**Low** Since the number of calculations is equal to the square of the number of data, this method can be computationally intensive, and the computational speed can be significantly low	**Moderate** Depending on the number of data and the number of dimensions chosen to build a single tree, it is often faster than SVM when the data size is large	**High** It utilizes stochastic gradient descent, and once the gradient is obtained, it can quickly find the solution and can be easily applied to large datasets
Accuracy	**High for a simplified problem** Classifies the data by subtracting a separation hyper-plane, which is suitable for two classifications; however, for multiple classifications, it creates a separation hyper-plane, which renders it difficult to execute and less accurate	**High when the volume of data is large** When solving simple problems with a small number of data, the accuracy is low owing to unavailability of data required for the large number of calculations. In contrast, when the data size is large, the accuracy is high, and overlearning is unlikely to occur	**Low** Performs calculations by applying a sigmoidal function to the output of a multiple regression. If the target variable tends to be in a single direction, the probability of correct classification is high. However, if there is a skewed class bias in the feature space, classification may be difficult to perform
Calibration	**Difficult** Kernels, regularization penalties, slack variables, etc. need to be adjusted. It is also necessary to regularize and standardize the training data in advance	**Easy** Only the depth and number of decision trees need to be adjusted. Further, regularization and standardization of the training data are not required	**Easy** Only the adjustment of slack variables is necessary. Further, regularization and standardization of the training data are not required

Table 2. Brief schema of the machine learning methods applied in this study. The table presents a comparative representation of the three machine learning methods.

Methods

Dataset. This retrospective observational study was performed in the ICU of Tohoku University Hospital, Sendai, Japan. Ethical approval was obtained from the Ethics Committee of the Tohoku University Graduate School of Medicine (reference number 2018–1–650). The requirement for informed consent was waived due to the retrospective design by the Ethics Committee of Tohoku University Graduate School of Medicine. All methods were performed in accordance with the relevant guidelines and regulations as per the Declaration of Helsinki. The patients' data will be deleted as soon as the current paper is published. The study was registered in a database in advance (ID: R000047019 UMIN000041179, URL: https://www.umin.ac.jp/ctr/index.htm). Patients who met the following inclusion criteria were enrolled at ICU admission between October 2016 and October 2019: (1) age ≥ 20 years; (2) CPOT, RASS, and confusion assessment method for the intensive care unit (CAM-ICU) were evaluated at least five times per patient; and (3) ECG and arterial pressure were monitored for ≥ 30 min before employing CAM-ICU. Data of the following patients whose vital signs differed significantly from those of the general adult population were excluded: (1) pregnant patients; (2) cardiopulmonary bypass recipients; and (3) those with organ transplants (i.e. organ recipients), ventricular assist devices, extracorporeal membrane oxygenation, intra-aortic balloon pumps, or do-not-resuscitate orders. The final dataset included 11,527 patients. The commonest analgesic and sedative were fentanyl and dexmedetomidine. All data were retrieved from the institution's electronic medical record system (PrimeGaia, Nihon Kohden Corporation, Tokyo, Japan). No patient-identifying information was recorded.

CPOT, RASS, and CAM-ICU. To assess the pain levels of all eligible patients, we used CPOT[11] as the target to train the model. CPOT assessment was performed by ICU nurses every 8 h and when obvious pain was observed. CPOT includes nine levels (0–8). RASS[42] was used to assess the sedation level. Delirium was assessed using CAM-ICU[43]. Furthermore, RASS and CAM-ICU assessments were performed concurrently with CPOT assessment. CPOT, RASS, and CAM-ICU evaluations were performed by multiple nurses to ensure agreement between them. If there was a difference in opinion, the intensivist made the final decision and approved the evaluations.

Analgesic management protocol. All patients were managed using a pain management protocol of the hospital (see Supplementary Fig. S1 online).

Vital signs. Heart rate, pulse oximetry, arterial oxygen saturation, and arterial blood pressure were recorded every minute. Arterial pressure was continuously monitored in either the left or right radial artery. Respiratory rate was measured using electrocardiographic impedance. The noise corresponding to each variable was removed according to the method described in Supplementary Table S2.

Machine learning. We used three machine-learning methods with different characteristics (Table 2) according to the report by Saberioon et al.[44].

Prediction model. Patient pain at the time of CPOT evaluation was predicted using machine learning and vital signs up to that time. The predictors were systolic and diastolic arterial blood pressure, pulse rate, and respiratory rate (one record/min) as vital signs as well as the patient's sex, age group (20–44, 45–64, and 65 years), and RASS score in the last 3 h. Figure 2 illustrates the procedure used to create the training data. CAM-ICU was used for the exploratory analysis but not for the construction of the final model.
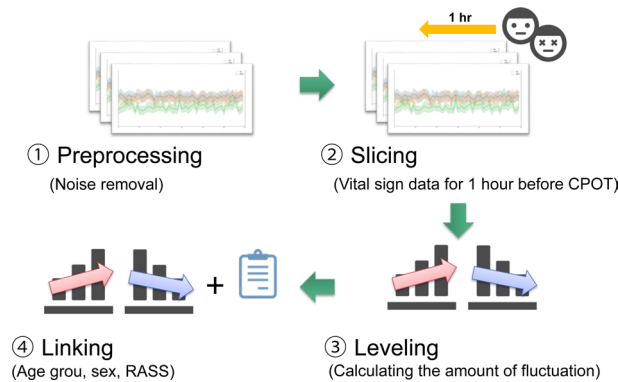
4

Figure 2. Procedure for creating training data.

First, the noise was removed from the data of the vital signs according to the noise requirements in Supplementary Table S2 (preprocessing). Second, the vital signs from one hour before the CPOT evaluation were extracted (slicing). Third, the rising and falling fluctuations were calculated and integrated into the section in which 90% or more of the vital signs were present. To mitigate the influence of individual differences between patients, the integrated value of the amount of fluctuation from 1–2 h after entering the ICU was calculated every minute (60 records) and the fluctuations corresponding to each patient were calculated. The records were normalized based on the maximum and minimum values of the integrated value of the quantity (leveling), which provided a numerical value that indicated the magnitude of the variations in each patient. Finally, the age group, sex, and sedation score were assigned to the training data.

Model evaluation. For predicting pain, CPOT values from 0 to 2 were considered negative and scores of ≥ 3 were considered positive. Since these data were unbalanced, oversampling was performed on the positive group. In terms of the oversampling method, Synthetic Minority Over-sampling Technique (SMOTE), Borderline SMOTE, and Adaptive Synthetic Sampling (ADASYN) were verified, and ADASYN, which showed the highest accuracy, was adopted. In machine learning, precision was compared between random forest (RF), support vector machine (SVM), and logistic regression (LR) models. To evaluate the generalization performance, the data were randomly classified between the training and validation data in a ratio of 9:1 and 10 cross-validations were performed. A grid search was performed on the training data to select the hyperparameter that derived the highest accuracy in each of the models. Furthermore, the harmonic means of the sensitivity, specificity, and AUROC were calculated.

Statistical analysis. Data analysis was performed using JMP v15 (SAS Institute Inc., Cary, NC, USA). Normally distributed data were reported as mean ± standard deviation and non-normally distributed data as median and interquartile range. AUROC was used to compare the accuracy, which was evaluated as low (0.5–0.7), moderate (0.7–0.8), and high (≥ 0.8).

Data availability
The datasets generated during and/or analyzed during the current study are available in the UMIN-ICDR repository at http://www.umin.ac.jp/icdr/index–j.html with permission from the authors.

References
1. Chanques, G. *et al.* A prospective study of pain at rest: incidence and characteristics of an unrecognized symptom in surgical and trauma versus medical intensive care unit patients. *Anesthesiology* **107**, 858–860 (2007).
2. Gélinas, C. Management of pain in cardiac surgery ICU patients: have we improved over time?. *Intensive Crit. Care Nurs.* **23**, 298–303 (2007).
3. Stanik-Hutt, J. A., Soeken, K. L., Belcher, A. E., Fontaine, D. K. & Gift, A. G. Pain experiences of traumatically injured patients in a critical care setting. *Am. J. Crit. Care* **10**, 252–259 (2001).
4. Novaes, M. A. *et al.* Stressors in ICU: perception of the patient, relatives and health care team. *Intensive Care Med.* **25**, 1421–1426 (1999).
5. Payen, J. F. *et al.* Current practices in sedation and analgesia for mechanically ventilated critically ill patients: a prospective multicenter patient-based study. *Anesthesiology* **106**, 687–695 (2007) (**quiz 891**).
6. Akça, O. *et al.* Postoperative pain and subcutaneous oxygen tension. *Lancet* **354**, 41–42 (1999).
7. Hedderich, R. & Ness, T. J. Analgesia for trauma and burns. *Crit. Care Clin.* **15**, 167–184 (1999).
8. Beilin, B. *et al.* Effects of anesthesia based on large versus small doses of fentanyl on natural killer cell cytotoxicity in the perioperative period. *Anesth. Analg.* **82**, 492–497 (1996).
9. Pollock, R. E., Lotzová, E. & Stanford, S. D. Mechanism of surgical stress impairment of human perioperative natural killer cell cytotoxicity. *Arch. Surg.* **126**, 338–342 (1991).
10. Peterson, P. K. *et al.* Stress and pathogenesis of infectious disease. *Rev. Infect. Dis.* **13**, 710–720 (1991).

5

11. Gélinas, C., Fillion, L., Puntillo, K. A., Viens, C. & Fortier, M. Validation of the critical-care pain observation tool in adult patients. *Am. J. Crit. Care* **15**, 420–427 (2006).
12. Payen, J. F. *et al.* Assessing pain in critically ill sedated patients by using a behavioral pain scale. *Crit. Care Med.* **29**, 2258–2263 (2001).
13. Barr, J. *et al.* Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Crit. Care Med.* **41**, 263–306 (2013).
14. Chanques, G. *et al.* The measurement of pain in intensive care unit: comparison of 5 self-report intensity scales. *Pain* **151**, 711–721 (2010).
15. Chanques, G. *et al.* Impact of systematic evaluation of pain and agitation in an intensive care unit. *Crit. Care Med.* **34**, 1691–1699 (2006).
16. Gélinas, C., Arbour, C., Michaud, C., Vaillant, F. & Desjardins, S. Implementation of the critical-care pain observation tool on pain assessment/management nursing practices in an intensive care unit with nonverbal critically ill adults: a before and after study. *Int. J. Nurs. Stud.* **48**, 1495–1504 (2011).
17. Payen, J. F., Bosson, J. L., Chanques, G., Mantz, J. & Labarere, J. Pain assessment is associated with decreased duration of mechanical ventilation in the intensive care unit: a post Hoc analysis of the DOLOREA study. *Anesthesiology* **111**, 1308–1316 (2009).
18. Aïssaoui, Y., Zeggwagh, A. A., Zekraoui, A., Abidi, K. & Abouqal, R. Validation of a behavioral pain scale in critically ill, sedated, and mechanically ventilated patients. *Anesth. Analg.* **101**, 1470–1476 (2005).
19. Gélinas, C. & Arbour, C. Behavioral and physiologic indicators during a nociceptive procedure in conscious and unconscious mechanically ventilated adults: similar or different?. *J. Crit. Care* **24**(628), e7-628.e17 (2009).
20. Gélinas, C. & Johnston, C. Pain assessment in the critically ill ventilated adult: validation of the critical-care pain observation tool and physiologic indicators. *Clin. J. Pain* **23**, 497–505 (2007).
21. Christodoulou, L. Why AI vital signs processing, monitoring and diagnosing systems are crucial for in hospital ICUs and in home ICUs? Predicting acute deterioration, heart attacks, brain strokes, and the state of the art in AI health technology and beyond! (2017).
22. Houthooft, R. *et al.* Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif. Intell. Med.* **63**, 191–207 (2015).
23. Kuo, R. J., Huang, M. H., Cheng, W. C., Lin, C. C. & Wu, Y. H. Application of a two-stage fuzzy neural network to a prostate cancer prognosis system. *Artif. Intell. Med.* **63**, 119–133 (2015).
24. Liu, N. T., Holcomb, J. B., Wade, C. E., Darrah, M. I. & Salinas, J. Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock* **42**, 108–114 (2014).
25. Mohktar, M. S. *et al.* Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. *Artif. Intell. Med.* **63**, 51–59 (2015).
26. Adjei, T. *et al.* Pain prediction from ECG in vascular surgery. *IEEE J. Transl. Eng. Health Med.* **5**, 2800310 (2017).
27. Haxby, J. V. Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage.* **62**, 852–855 (2012).
28. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* **37**, 435–456 (2014).
29. Lindquist, M. A. *et al.* Group-regularized individual prediction: theory and application to pain. *Neuroimage* **145**, 274–287 (2017).
30. Naselaris, T. & Kay, K. N. Resolving ambiguities of MVPA using explicit models of representation. *Trends Cogn. Sci.* **19**, 551–554 (2015).
31. Puntillo, K. A. *et al.* Determinants of procedural pain intensity in the intensive care unit. The EUROPIAN study. *Am. J. Respir. Crit. Care Med.* **189**, 39–47 (2014).
32. Skrobik, Y. *et al.* Protocolized intensive care unit management of analgesia, sedation, and delirium improves analgesia and sub-syndromal delirium rates. *Anesth. Analg.* **111**, 451–463 (2010).
33. Devlin, J. W. *et al.* Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. *Crit. Care Med.* **46**, e825–e873 (2018).
34. De Jong, A. *et al.* Decreasing severe pain and serious adverse events while moving intensive care unit patients: a prospective interventional study (the NURSE-DO project). *Crit. Care* **17**, R74 (2013).
35. Georgiou, E., Hadjibalassi, M., Lambrinou, E., Andreou, P. & Papathanassoglou, E. D. The impact of pain assessment on critically ill patients' outcomes: a systematic review. *Biomed Res. Int.* **2015**, 503830 (2015).
36. Macintyre, P. E., Huxtable, C. A., Flint, S. L. & Dobbin, M. D. Costs and consequences: a review of discharge opioid prescribing for ongoing management of acute pain. *Anaesth. Intensive Care* **42**, 558–574 (2014).
37. Puntillo, K. A. & Naidu, R. Chronic pain disorders after critical illness and ICU-acquired opioid dependence: two clinical conundra. *Curr. Opin. Crit. Care* **22**, 506–512 (2016).
38. Yi, P. & Pryzbylkowski, P. Opioid induced hyperalgesia. *Pain Med.* **16**(Suppl 1), S32–S36 (2015).
39. Forsyth, A. W. *et al.* Machine learning methods to extract documentation of breast cancer symptoms from electronic health records. *J. Pain Symptom Manag.* **55**, 1492–1499 (2018).
40. Sun, H. *et al.* Automated tracking of level of consciousness and delirium in critical illness using deep learning. *NPJ Digit. Med.* **2**, 89 (2019).
41. Kobayashi, N. *et al.* Arterial blood pressure correlates with 90-day mortality in sepsis patients: a retrospective multicenter derivation and validation study using high-frequency continuous data. *Blood Press. Monit.* **24**, 225–233 (2019).
42. Sessler, C. N. *et al.* The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am. J. Respir. Crit. Care Med.* **166**, 1338–1344 (2002).
43. Ely, E. W. *et al.* Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA* **286**, 2703–2710 (2001).
44. Saberioon, M. *et al.* Comparative performance analysis of support vector machine, random forest, logistic regression and k-nearest neighbours in rainbow trout (*Oncorhynchus mykiss*) classification using image-based features. *Sensors (Basel, Switzerland)* **18**, 1027 (2018).

Acknowledgements

Author contributions

N.K. prepared the main manuscript text, K.W. and H.M. analyzed the data and prepared Figs. 1 and 2. T.S. and S.I. provided advice regarding data analysis, and M.Y. supervised the entire study. All authors have reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.