Decoding accelerometry for classification and prediction of critically ill patients with severe brain injury

Shubhayu Bhattacharyay[1,2,3,4,*], John Rattray[5], Matthew Wang[3], Peter Dziedzic[1,6], Eusebia Calvillo[7], Han B. Kim[1,7], Eshan Joshi[3], Pawel Kudela[6,8], Ralph Etienne-Cummings[5], Robert D. Stevens[1,6,7,8]

[1]Laboratory of Computational Intensive Care Medicine, Johns Hopkins University, Baltimore, MD, USA.
[2]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK.
[3]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.
[4]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA.
[5]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA.
[6] Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University, Baltimore, MD, USA
[7]Department of Neurosurgery, Johns Hopkins University, Baltimore, MD, USA.
[8]Department of Neurology, Johns Hopkins University, Baltimore, MD, USA.

*Corresponding author: sb2406@cam.ac.uk

ABSTRACT:

The goal of this research is to explore quantitative motor features in critically ill patients with severe brain injury (SBI). We hypothesized that computational decoding of these features would yield important information on underlying neurological states and clinical outcomes. Using wearable microsensors placed on all extremities, we recorded 1,701 hours of continuous, high-frequency accelerometry data from a prospective cohort of patients (*n* = 69) admitted to the ICU with SBI. Models were trained using time-, frequency-, and wavelet-domain motion features and levels of responsiveness and outcome as labels. The two primary tasks were detection of levels of responsiveness, assessed by motor sub-score of the Glasgow Coma Scale (GCSm), and prediction of functional outcome at hospital discharge, measured with the Glasgow Outcome Scale–Extended (GOSE). Detection models achieved significant (AUC: 0.70 [95% CI: 0.53–0.85]) and consistent (observation windows: 12 min – 9 hours) discrimination of SBI patients capable of purposeful movement (GCSm > 4). Prediction models accurately discriminated SBI patients of upper moderate disability or better (GOSE > 5) with 2–6 hours of observation (AUC: 0.82 [95% CI: 0.75–0.90]). Results suggest that computational analysis of time series motor activity in patients with SBI yields clinically important insights on underlying neurologic states and short-term clinical outcomes.

INTRODUCTION

Despite advances in intensive care, the global burden of severe brain injury (SBI), in terms of mortality, long-term disability, and economic costs, is the highest among all major injuries[1]. Existing approaches to predict SBI outcomes, such as recovery of consciousness and functional independence, are imprecise for individual patients[2] and can raise ethical concerns due to the potential for withdrawal of life sustaining-therapies[3]. At the same time, recent developments in artificial intelligence and big data processing represent an opportunity to optimize SBI patient monitoring with high-resolution, longitudinal waveform data and to improve the precision of SBI prognoses with flexible modeling strategies[4]. Hence, a key focus in the care of SBI is the discovery and validation of quantitative monitoring modalities that improve upon the accuracy and reliability of clinical characterization and the reliability of predicted outcomes[5].

For acute neurological disorders, the assessment of motor function provides an important clinical window into neural systems associated with sensorimotor processing, emotion, coordination, planning, and learning[6-8]. Neurological damage and intensive care unit (ICU) practices (e.g., sedation, bedrest) are associated with a dramatic reduction in normal physical activity[9], resulting in systemic pro-inflammatory signaling[10] and an elevated risk of venous thromboembolism, infection, skin and soft tissue damage, delirium, and loss of muscle mass and strength[11-14]. A corollary is that structured programs designed to increase physical activity for SBI patients in the ICU can significantly reduce neurological complications and may lead to improved functional recovery[15]. However, it is uncertain whether the incorporation of continuous motion sensing in the ICU could yield clinically significant gains for SBI monitoring and prognosis.

Wearable accelerometers provide an objective and continuous assessment of motor activity over extended periods of time[16]. In contrast to most other motion sensing modalities[17], integration of accelerometers in the ICU is feasible. Advances in microelectromechanical systems (MEMS) technology have made it possible to construct inexpensive, minimally obtrusive wearable accelerometers that can be optimized for the clinical space[18]. Accelerometers respond to changes in movement frequency and intensity, measure tilt from the gravitational axis, and produce little variation or drift over time[18-22]. The use of accelerometers to monitor gross physical activity in the ICU has already been tested with varying degrees of success[23]. Herein, we aim to more specifically determine whether a relationship exists between motion features derived from triaxial accelerometry time-series and neurological motor states and functional outcomes of SBI patients.

In this pilot study of the Neurological Injury Motion Sensing (NIMS) project, we explore the impact and limitations of high-resolution accelerometry in patients with SBI admitted to the ICU. We developed a matrix of wearable accelerometers to quantitatively capture motor activity from the extremities of SBI patients. Applying techniques from time-series analysis, dimensionality reduction, and logistic regression, we extract interpretable time-, frequency-, and wavelet-domain motion features and assess their performance in motor function detection and short- and long-

term functional outcome prediction models. We then assess relative significance of the extracted features to determine how specific accelerometry profiles relate to clinically evaluated motor function and global outcomes. Finally, we demonstrate how accelerometry-based model outputs can potentially be used to monitor neurological transitions in clinical practice.

RESULTS

Study population characteristics

Of the 72 total SBI patients recruited in the ICU, 3 participants were excluded from the study due to withdrawn consent (n = 2) or corruption of accelerometry data during upload (n = 1), resulting in a study population of n = 69. Five patients were lost from one-year follow-up due to unsuccessful contact, and thus the study population at 12 months post hospital discharge was n = 64. Detailed characteristics of the study population are summarized in **Table 1**.

From each of the study participants, we collected triaxial accelerometry data (sampled at 10 Hz) from a wearable matrix of 6 sensors, placed on each elbow, wrist, and ankle and an additional sensor placed on the bed for external movement correction (**Fig. 1a**). The median recording duration per patient was 24.09 hours (IQR: 22.81–25.11 hours), and accelerometry data was recorded fairly uniformly across the stages of ICU stay in terms of proportion completed (**S. Fig. 1d**). In total, 1,701 hours of multisegmental accelerometry data were recorded.

During their stay in the ICU (median: 19 days, IQR: 11–29 days), study participants were evaluated with the Glasgow Coma Scale (GCS)[24, 25] a median 9.25 times per day (IQR: 7.17–11.50 times per day). In total, we extracted scores from 14,240 GCS evaluations, 13,190 of which (92.63%) took place in the ICU and 653 of which (4.59%) coincided with accelerometry capture times. The trajectory of the motor component scores of the GCS (GCSm), along with corresponding times of accelerometry capture, of each patient included in our analysis is provided in **S. Fig. 2**.

Motor function detection performance

In this work, we use clinically evaluated GCSm scores extracted from electronic health records (EHR) as the primary markers of functional motor states. The scores of the 6-point GCSm are defined by best motor responses to physical stimuli and are outlined in **Table 1**.

We trained and evaluated threshold-level GCSm detection models from automated accelerometry-based motion features extracted from 19 varying observation windows directly preceding the GCSm evaluations (**Fig. 1b**). The count distributions of GCSm scores available for each observation window are listed in **S. Table 1**.

The receiver operating characteristic (ROC) curves of the optimally discriminating models at each GCSm threshold, along with their mean areas under the curves (AUC) and optimal observation windows, are shown in **Fig. 2a**. Based on the 95% confidence intervals of mean AUC, significant discrimination (AUC > 0.5, α = 0.05) was achieved by the extracted features at every threshold of GCSm except for GCSm > 2. However,

only GCSm > 4 detection models achieve significant discrimination from shorter observation window durations (≤ 30 minutes); GCSm > 4 detection models achieve significant discrimination consistently with an observation window of 12 minutes or greater (**Fig. 2b**). The mean AUCs, along with 95% confidence intervals, at each threshold of GCSm is provided for all 19 tested observation windows in **S. Table 2**. As GCSm > 1, GCSm > 3, and GCSm > 5 detection models achieve significant discrimination at less than or equal to 3 different observation windows, only GCSm > 4 detection models achieve significant discrimination at a broad range of observation windows (12 min – 9 hours). Binary classification performance metrics of optimally discriminating motor function detection models are provided in **Table 2**. At none of the GCSm thresholds do the models achieve significantly greater accuracy than the proportion of the most represented class based on 95% confidence intervals. Only the GCSm > 4 detection model achieves a higher mean accuracy (0.71) and a significantly greater F_1 score (0.78 [95% CI: 0.67–0.87]) than its proportion of the most represented (in this case, positive) class (0.66). Only GCSm > 4 and GCSm > 5 detection models achieved both a mean sensitivity and mean specificity over 0.5, but not significantly.

Functional outcome at hospital discharge prediction performance

We used clinically evaluated Glasgow Outcome Scale – Extended (GOSE) scores as the primary markers of functional outcomes, both at hospital discharge and at 12 months post discharge. The scores of the 8-point GOSE are outlined in **Table 1**.

We trained and evaluated threshold-level GOSE at hospital discharge prediction models from automated accelerometry-based motion features extracted from the same 19 varying observation windows directly preceding GCSm evaluations (**Fig. 1b**). The median lead window duration (i.e., time between end of observation window and hospital discharge) was 20 days (IQR: 10–33 days). The count distributions of GOSE scores, at discharge, available for each observation window are listed in **S. Table 3**. Given the low proportion of patients (1.45%) with good recovery (GOSE > 6) at hospital discharge, we limited our threshold-level analysis to GOSE > 1, GOSE > 2, GOSE > 3, GOSE > 4, and GOSE > 5.

The receiver operating characteristic (ROC) curves of the optimally discriminating models at each GOSE threshold, along with their mean areas under the curves (AUC) and optimal observation windows, are shown in **Fig. 3a**. Based on the 95% confidence intervals of mean AUC, significant discrimination (AUC > 0.5, $\alpha = 0.05$) was achieved by the extracted features only at GOSE > 5. GOSE > 5 prediction models achieve significant discrimination at observation windows of two hours or greater, with a peak mean AUC of 0.82 (95% CI: 0.75–0.90) at an observation window duration of 6 hours (**Fig. 3b**). The mean AUCs, along with 95% confidence intervals, at each tested threshold of GOSE is provided for all 19 tested observation windows in **S. Table 4**.

Binary classification performance metrics of optimally discriminating functional outcome prediction models are provided in **Table 2**. At none of the GOSE thresholds do the models achieve a significantly greater F_1 score than the proportion of the positive class or a greater mean accuracy than the proportion of the most represented class. Despite its strong discrimination performance, the GOSE > 5 prediction model achieves near-zero precision and sensitivity. While we observe an ideal precision recall curve for this model (**S. Fig. 3a**), the mean average precision is only 0.08 (95%

CI: 0.02–0.18). This indicates, that while prediction probabilities for true positive cases are, on average, greater than prediction probabilities for true negative cases, they seldom cross the 0.5 threshold for proper classification (**S. Fig. 3b**).

Functional outcome at 12 months post discharge prediction performance

We trained and evaluated threshold-level GOSE at 12 (±1) months post hospital discharge prediction models from automated accelerometry-based motion features extracted from the same 19 varying observation windows directly preceding GCSm evaluations (**Fig. 1b**). The count distributions of GOSE scores, at 12 months, available for each observation window are listed in **S. Table 5**. The receiver operating characteristic (ROC) curves of the optimally discriminating models at each GOSE threshold, along with their mean areas under the curves (AUC) and optimal observation windows, are shown in **S. Fig. 4a**. Based on the 95% confidence intervals of mean AUC, significant discrimination (AUC > 0.5, α = 0.05) was not achieved by the extracted features at any of the GOSE thresholds. Mean AUC is largely independent of observation window duration at each of the thresholds (**S. Fig. 4b**). The mean AUCs, along with 95% confidence intervals, at each threshold of GOSE is provided for all 19 tested observation windows in **S. Table 6**. Binary classification performance metrics of optimally discriminating functional outcome prediction, at 12 months post discharge, models are provided in **Table 2**.

Calibration of motor function detection and functional outcome prediction

The probability calibration curves and associated prediction distributions of the optimally discriminating models at each threshold for GCSm detection and GOSE (at hospital discharge) prediction are provided in **Fig. 4a** and **Fig. 4b** respectively. We observe that the GCSm > 4 detection model achieves the best graphical model calibration of all those tested (E_{max} = 0.30 [95% CI: 0.08–0.64]). However, when considering the prevalence of predicted probabilities in calibration assessment with the integrated calibration index (ICI)[26], we observe that the GOSE > 5 prediction model has the most ideal calibration (ICI = 0.01 [95% CI: 0.00–0.02]). The discrepancy between the weighted and graphical calibration of GOSE > 5 indicates a strong class imbalance, suggesting that more positive cases are necessary to train and recalibrate this model for proper classification. Probability calibration metrics of all optimally discriminating models are provided in **Table 3**.

Extracted feature and sensor placement analysis

At the end of our accelerometry processing pipeline (**Fig. 1**), we extracted eight unique feature types (**Table 4**) from each of the six accelerometers placed around SBI patient joints. For each of these 48 feature-sensor combinations, we calculate a relative significance score equivalent to the mean absolute value of the learned coefficients of supervised dimensionality reduction (i.e., the relative importance in explaining the variance in the dataset stratified by the endpoint) weighted by the absolute value of learned logistic regression coefficients (see **Methods**).

We consider the optimally discriminating configurations of the two most promising model types as representatives for motor function detection and functional outcome prediction respectively: (a) GCSm > 4 with a 6-hour observation window and (b) GOSE

(at hospital discharge) > 5 with a 6-hour observation window. The feature significance scores of these two model types are visualized as heatmaps in **Fig. 5a** and **Fig. 5b** respectively.

For both motor function detection and functional outcome prediction, there is more variation in significance scores across feature types than across sensor placements. For motor function detection, the proportion of dynamic activity (PDA) in the observation window, the frequency-domain entropy (FDE), and the median frequency (MFR) are the three most significant feature types, descending in that order. For functional outcome prediction, the descending order of the three most significant feature types is FDE, MFR, and PDA. PDA is a crude measurement of overall physical activity[27], while FDE enables differentiation between activity profiles which have simple acceleration patterns and those with more complex patterns[16]. From the pair of high-pass-filtered medians (HLF (h)) and low-pass-filtered medians (HLF (l)), HLF (h) has a significantly greater mean significance score than HLF (l) for every sensor placement in both model endpoints based on 95% confidence intervals. This, along with the relative significance of MFR, suggests that finer movements, captured in higher frequencies of accelerometry, can be more clinically significant in discriminating functional motor states and global outcomes from SBI. Moreover, the consistently strong significance of PDA, FDE, and MFR suggests that features of both the time domain (PDA) and the frequency domain (FDE, MFR) in combination may be useful for clinical assessments of functional neurological states.

In detecting motor function, the right wrist (RW) sensor was the most significant placement across the five most significant feature types. The trajectories of mean motion feature values in the six hours preceding GCSm evaluations (**S. Fig. 5**) visually demonstrate that features extracted from the wrist-placed sensors better discriminate cases of GCSm 5 and 6 from the rest of the GCSm scores. This follows clinical observations of a greater frequency of conscious movement in hands and wrists of bedridden SBI patients during ICU stay. Moreover, abnormal profiles of flexion and extension, associated with SBI, are most often observed in the wrists, and thus, the wrist-placed sensors may be more sensitive to abnormal patterns of movement, corresponding to lower levels of consciousness, than the elbow- or ankle-placed sensors.

In functional outcome prediction, we observe the greatest significance scores ascribed to wrist-placed sensors (RW and LW) in the most significant frequency-domain features (FDE and MFR), but the ankle- (RA and LA) and elbow-placed sensors have the greatest significance scores in the most significant time-domain features (PDA, SMA, and HLF (h)). Wrist movements are finer than elbow and ankle movements and may be best distinguished in the frequency-domain in relation to global outcomes.

The correlation of each of the extracted motion features across the six sensor placements is visualized in **S. Fig. 6**, and violin plots of the distributions of motion features, stratified by GCSm, are presented in **S. Fig. 7**.

Retrospective case study analysis of motor function detection in practice

Six patients in our study experienced a transition between GCSm > 4 and GCSm ≤ 4 within the GCS observations coinciding with 6-hour observation windows of

accelerometry recording. For each of these six patients, we trained GCSm > 4 detection models on the remaining patient set with a shorter (27-minute) and a longer (6-hour) observation window. We return predictions with these models on the six case study patients every ten minutes to retrospectively examine the trajectories of probabilities against the recorded times of neurological transition (**Fig. 6**).

In case no. 2, we observed that both model types detect an upward transition in GCSm more than three hours before it was reported in the EHR. Likewise, the 27-min observation window model detected a downward transition in GCSm about an hour before the upcoming evaluation in case no. 4 and about two hours before in case no. 3. In cases no. 3, 4, and 6, we observed that the 6-hour observation window detects the appropriate transition in GCSm, but with a delay of 3–6 hours. In cases no. 1 and 5, in which we observe a shift and resettlement of GCSm within a 3–5-hour span, the 6-hour model fails to detect the transition while the 27-min model uncertainly oscillates above and below the midline. In general, the shorter observation window model was more dynamic and detected GCSm transitions quicker than the longer observation window model. However, persistent transitions, such as the one observed in case no. 6, were detected with more stability and reliability by the longer observation window model.

DISCUSSION

Key findings

We introduce an accelerometry-based based system in critically ill SBI patients that quantitatively captures multisegmental motor patterns correlating with clinical scores of motor responsiveness and functional outcome. The results reveal a significant (AUC = 0.70 [95% CI: 0.53–0.85]), consistent (observation windows: 12 min – 9 hours) association between extracted motion features and the discrimination of SBI patients capable of purposeful movement (GCSm > 4) and those who are not (GCSm ≤ 4) (**Fig. 2a**). A significant discrimination of purposeful movement was achieved with only 12 minutes of accelerometry recording (**Fig. 2b**), and reliable calibration (**Fig. 4a**) and informative classification (**Table 2**) for GCSm > 4 detection suggest that iterations of this system could be clinically useful in automating motor function monitoring. In case studies (**Fig. 6**), we demonstrate that accelerometry-based systems may detect transitions in motor function up to five hours before a clinical evaluation.

The utility of accelerometry-based features for functional outcome prognosis remains ambiguous. While we found no signal between motion features and long-term (12 months post discharge) outcomes (**S. Fig. 4**), the models accurately predicted functional status at hospital discharge (AUC = 0.82 [95% CI: 0.75–0.90]) at a cutoff of GOSE > 5 vs GOSE ≤ 5 for favorable vs unfavorable outcome (**Fig. 3a**). Patients with a GOSE of >5 have upper moderate disability or good recovery and are generally able to resume work or previous activities. However, given the small number of SBI patients with GOSE > 5 at hospital discharge, further validation is necessary to determine the reliability of this result. Conflicting results between the precision recall curve and average precision (**S. Fig. 3a**) and between different calibration metrics (**Table 3**) underline the class imbalance problem of GOSE > 5 in our dataset; at the same time, we find the consistent discrimination (**Fig. 3b**) and difference in outcome distribution (**S. Fig. 3b**) as promising markers for further exploration.

Finally, our analysis of feature significance (**Fig. 5**) reveals that both time-domain and frequency-domain features are important for motor function detection and functional outcome prediction. While sensors placed on the wrist achieved the greatest significance scores overall, particularly for features in the frequency-domain, multisegmental motion capture was validated by comparable significance scores of elbow- and ankle-placed sensors across the feature set.

Relationship with previous studies and future implications

Results presented here represent, to our knowledge, the first approach to relate motion sensor data to neurological states in SBI patients admitted to the ICU. Healthy activity classification with accelerometry-based features has become widespread, especially with advancements in MEMS technology, machine learning, and data sharing[16, 28-30]. However, applications to intrahospital care, and in particular intensive care, have been limited[31, 32] and have largely taken only simple, threshold-based feature approaches to grossly evaluate motor activity (e.g., actigraphy)[33-36]. Reported success in these studies has been variable, but none of them have combined the high-resolution time-domain, feature-domain, and wavelet-domain analysis found in more recent healthy activity classification studies. The focus of our approach, on the relationship between motor profiles of SBI patients over extended periods of time and clinically relevant neurological states, is novel. Yet, it builds upon the developments in time-series analysis, dimensionality reduction, and supervised machine learning from activity classification projects as well as the hypotheses of the clinical validity and utility of accelerometry from applied, medical projects.

A continuous high-frequency motion capture system in the intensive care setting produces a high-dimensional dataset that is also valuable for data-driven research projects. Profiles of motor activity in SBI are poorly understood and decoding specific features of motion in the time-, frequency-, and wavelet-domains can open a window on internal neurological states. Accelerometry-based features may elucidate fundamental mechanisms underlying the strong association between physical activity and clinical outcomes, and we aim to collect more data in the NIMS project to enable the research and development of motion as a quantitative marker of functional recovery for SBI.

More generally, the critical care setting is a fertile ground for the development of advanced computational methods and applications of artificial intelligence for monitoring and decision support[37]. Patients are typically interfacing with physiological monitoring systems that generate a large volume of data whose complexity may overwhelm human interpretation alone but may be ideal for the training of analytical systems[38]. Since critical care specialists typically must make time-sensitive decisions for multiple patients at the same time[39], we expect that a near-real-time computational framework assessing motion features alongside other time-series data continuously could provide valuable decision support. We expect that ongoing and subsequent iterations of this work will enable integration of computational physical activity features into the framework of monitoring and prognostication in the critical care setting.

Study limitations

We recognize several limitations in this work that need to be addressed. Our statistical analyses and retrospective validation of GCSm detection and GOSE prediction were performed on a limited sample size (n = 69 patients) from a single institution and intensive care facility. Further validation will require repeated trials on larger patient populations across multiple centers. There are also improvements to be done to the sensor itself. The planar dimensions of our currently used accelerometer (42 mm × 32 mm) can be reduced further to increase the resolution of localized motion capture. Furthermore, since accelerometry measurements depend on the orientation of the accelerometer with respect to the vertical (gravitational axis), additional modalities of motor output (i.e., gyroscopy and electromyography) could be integrated into the sensor system to inform computational models on the precise arrangement and neural activation of body segments. This would allow us to derive more physiologically relevant features that correspond to validated models of nervous system injury or disease[6]. We also recommend the development of sensors with higher sampling frequencies (≥40 Hz) to capture extremely fine or fast movements of digits or lower extremities. Additionally, GCSm itself has been criticized for lack of standardization among practitioners[40, 41]. GCSm scores for this work were extracted automatically from EHR and were measured from multiple practitioners across the Johns Hopkins Hospital Neurosciences Critical Care Unit (NCCU) staff. Moving forward, we aim to supplement clinical validation of the motion features with multifactorial associations with other consciousness, functional, cognitive, psycho-behavioral, symptomatic, and social outcome scales of SBI patients[42].

METHODS

Study population and experimental protocol

This work was conducted with approval from the Johns Hopkins Medicine Institutional Review Board (IRB00135674) and written informed consent from patients or surrogates. We prospectively enrolled 72 patients admitted to the NCCU who met the following criteria: age ≥ 18 years, SBI defined as an acute brain injury or illness resulting in impaired consciousness, absence of injuries or lesions involving the extremities, and not expected to die or have withdrawal of life-sustaining therapies in the 24 hours following enrolment.

Patients were evaluated daily while in the NCCU, at hospital discharge, and at 12 months post discharge by research team members. All GCS evaluations during each patient's hospital stay were automatically extracted from the institutional EHR system (Epic Systems, Madison, WI, USA). GOSE scores at hospital discharge were obtained by EHR review of discharge reports for patients who survived during hospital stay (n = 53). Patients were contacted by telephone 12 months (±1 month) after hospital discharge, and GOSE scores were obtained using a validated questionnaire[43] (n = 27); in cases where patients could not be contacted, data was extracted from EHR reports (n = 8). Additionally, we identified participants who died between discharge and 12 months post-discharge from national obituary records (n = 12). Thus, we arrived at a 12-month post-discharge sample size of n = 64.

From the first 3 patients, we collected 10 hours of continuous triaxial accelerometry data, and for the remainder of the patients, we augmented our recording duration to between 24 and 48 hours of accelerometry data.

Instrumentation for accelerometry capture

Triaxial sensors (SensorTags CC2650, Texas Instruments, Dallas, TX, USA) were attached with transparent film dressing (Tegaderm Diamond Pattern 1686, 3M, Maplewood, MN, USA) bilaterally near the joints (with common orientation) designated in **Fig. 1a**. An additional sensor was placed vertically on the foot of the patient bed to detect patient-independent bed movements. Sensors were equipped with MEMS, variable capacitance tri-axial accelerometers (MPU-9250 MotionTracking Device, TDK InvenSense, San Jose, CA, USA) with sampling frequency (f_s) set to 10 Hz, the range of measurable amplitude at ±16 g (±157 m/s^2), and sensitivity at ±4,800 least significant bits per g (LSB/g).

The sensors transmitted data via a 2.4-GHz Bluetooth antenna to a portable Linux computer (RPi 3 Model B, Raspberry Pi Foundation, Cambridge, UK) placed in the NCCU room. We would execute a Python script on the computer to collect 3 channels (axes) of accelerometry time series from each of the 7 active accelerometers in parallel. The system would log interruptions on a separate .txt file in the instance of a sensor failure. During each trial, we also recorded a video stream (M1045-LW Network Camera, Axis Communications, Lund, Sweden) of the patient that clearly shows the location of each sensor. In the event of sensor interruptions, irregular movement profiles, or bed-sensor-extracted signal magnitude (SMA) values above 0.135 *g*[27], we would check the footage to identify the source of these results.

Accelerometry processing and motion feature extraction

Each axial component of each sensor was convolved with a 4th-order Butterworth high-pass filter with a critical frequency of f_c = 0.2 Hz (**S. Fig. 8**) to remove the baseline offset of accelerometry readings (**Fig. 1a**) and generally separate the low frequency effect of static orientation from the high frequency effect of active body movement[44].

Filtered time-series were segmented into non-overlapping 5-second windows (~50 data points per window) for motion feature extraction. We selected the motion features listed in **Table 4**, which performed well in physical activity classification tasks[16], to represent three different domains (time frequency, and wavelet). PDA is defined by the proportion of SMA over 0.135 *g* for each sensor in an observation window (**Fig. 1b**). The remaining features are defined by the following formulae for each 5-second window:

$$\mathbf{SMA} = \frac{0.1\ \text{sec}}{2 \times\ 5\ \text{sec}} \sum_{n=1}^{N-1} (|x_{n+1}| + |y_{n+1}| + |z_{n+1}| + |x_n| + |y_n| + |z_n|)$$

$$\mathbf{HLF_h} = \sqrt{(\text{median}\{\boldsymbol{x} * \boldsymbol{b_h}\})^2 + (\text{median}\{\boldsymbol{y} * \boldsymbol{b_h}\})^2 + (\text{median}\{\boldsymbol{z} * \boldsymbol{b_h}\})^2}$$

$$\mathbf{HLF_l} = \sqrt{(\text{median}\{\boldsymbol{x} * \boldsymbol{b_l}\})^2 + (\text{median}\{\boldsymbol{y} * \boldsymbol{b_l}\})^2 + (\text{median}\{\boldsymbol{z} * \boldsymbol{b_l}\})^2}$$

$$\mathbf{MFR} = \sqrt{(\text{median}\{\boldsymbol{X}\})^2 + (\text{median}\{\boldsymbol{Y}\})^2 + (\text{median}\{\boldsymbol{Z}\})^2}$$

$$\mathbf{FDE} = \frac{1}{\log_2 N} \sqrt{\left(\sum_{n=1}^{N} \frac{|X_n|^2}{\sum_{i=1}^{N}|X_i|^2} \log_2\left(\frac{|X_n|^2}{\sum_{i=1}^{N}|X_i|^2}\right)\right)^2 + \left(\sum_{n=1}^{N} \frac{|Y_n|^2}{\sum_{i=1}^{N}|Y_i|^2} \log_2\left(\frac{|Y_n|^2}{\sum_{i=1}^{N}|Y_i|^2}\right)\right)^2 + \left(\sum_{n=1}^{N} \frac{|Z_n|^2}{\sum_{i=1}^{N}|Z_i|^2} \log_2\left(\frac{|Z_n|^2}{\sum_{i=1}^{N}|Z_i|^2}\right)\right)^2}$$

$$\mathbf{BPW} = \frac{1}{3.2\ \text{Hz}}\sqrt{\left(\int_{0.3\ \text{Hz}}^{3.5\ \text{Hz}} |X_f|^2 df\right)^2 + \left(\int_{0.3\ \text{Hz}}^{3.5\ \text{Hz}} |Y_f|^2 df\right)^2 + \left(\int_{0.3\ \text{Hz}}^{3.5\ \text{Hz}} |Z_f|^2 df\right)^2}$$

$$\mathbf{WVL} = \sqrt{\left(\sum_{l=2}^{6} \|\boldsymbol{d}_l^{(x)}\|^2\right)^2 + \left(\sum_{l=2}^{6} \|\boldsymbol{d}_l^{(y)}\|^2\right)^2 + \left(\sum_{l=2}^{6} \|\boldsymbol{d}_l^{(z)}\|^2\right)^2}$$

where:

- x, y, z represent the x-, y- and z- axes vectors, respectively, of the filtered accelerometry time series within the given 5 second window and x_n, y_n, z_n represent the nth elements of these vectors.
- N represents the length of each of the x, y, z vectors.
- $*$ represents the 1-dimensional convolution operator.
- b_h represents a 1-dimensional, 4th-order high-pass Butterworth filter with f_c = 2.5 Hz.
- b_l represents a 1-dimensional, 4th-order low-pass Butterworth filter with f_c = 2.5 Hz.
- X, Y, Z represent the discrete Fourier transforms of the x, y, z vectors respectively where X_n, Y_n, Z_n represent the nth elements of these Fourier transform vectors and X_f, Y_f, Z_f represent the coefficients of the Fourier transforms that correspond to linear frequency *f*.
- $d_l^{(x)}, d_l^{(y)}, d_l^{(z)}$ represent the vector of lth-level detail coefficients of the 5th-order Daubechies wavelet transform of the x, y, z vectors respectively.

Post-capture processing of accelerometry were performed offline using MATLAB (Version 9.8.0, MathWorks, Natick, MA, USA) with the Signal Processing, Wavelet, System Identification, and Symbolic Toolboxes.

Multiple imputation of missing motion features

Due to insufficient battery on the sensors, bedside interventions, interfering equipment, or patient migrations for surgery, imaging, or interunit transfers, a median 1.56% per sensor of each patient's intended recording duration was missing in our dataset. Missing motion features were multiply imputed (m = 9) with a normal (features were normalized with the Box-Cox transform[45]) multivariate time-series algorithm from the 'Amelia II' package (v1.7.6)[46] in R (v4.0.0)[47]. The algorithm exploits both spatial correlation (motion feature correlation across the sensors of the same participant) and temporal correlation (autocorrelation structures within each sensor's time series) to stochastically impute missing time series values in multiple, independently trained runs. We formed subsequent statistical analyses on all 9 imputations to account for variation across imputation.

This model assumes the data is missing at random (MAR) (i.e., the pattern of missingness is independent of unobserved data[48]), which we validated by observing the independence of missingness from sensor placement or time of day (**S. Fig. 9**). A complete characterization of the missing data of each patient can be found in **S. Table 7**.

Correction of gross external movements

At time points where the bed-placed sensor SMA exceeded 0.135 *g* (a proposed threshold between static and dynamic activity[27]) and preceded a spike in extremity feature values (1.33% of the time), the bed sensor values of SMA, HLF, BPW, and WVL were subtracted from the extremity values and the bed sensor values of MFR and FDE were added to the extremity values. If a resulting correction value ended up out of a feasible range of static activity for the feature, we replaced the value with a random value, selected uniformly from the static activity range of that feature (**S. Table 8**).

Repeated *k*-fold cross-validation for unbiased model validation

The study population (n = 69) was partitioned 25 times with repeated *k*-fold cross-validation (5 repeats, 5 folds) into training sets (~80%, $n \approx 55$) and validation sets (~20%, $n \approx 14$) for each of the 19 tested observation windows (**S. Table 1**) for each of the three tested endpoints (**Fig. 2b**). In splits for motor function detection, patients were stratified by median GCSm over their available observations, while in splits for functional outcome detection, patients were stratified by GOSE scores. One of the nine missing value imputations was drawn with replacement for each partition.

Repeated cross-validation partitions were performed with the 'caret' package (v6.0-86)[49] in R.

Motor function detection

We tested 19 unique observation window durations (**S. Table 1**) of accelerometry-derived features directly preceding GCSm evaluations (**Fig. 1b**). At each of these evaluation points, motion features were organized into matrices where each column represents a unique combination of motion feature type (8 total), sensor placement (6 total), and, for non-PDA features, time before the evaluation. Columns were normalized based on distributions of each placement-feature type combination (48 in total) in the training set. Normalized matrices underwent supervised dimensionality reduction with linear optimal low-rank projection (LOL)[50] learned from the training set. Target dimensionality ($d \in [2,20]$) was tested as a model hyperparameter. Low-dimensional vectors of each d then underwent element-wise Yeo-Johnson transforms[51] for scaled normalization (learned from the training set) and were used to train and validate logistic regression ('glm') models with binary endpoints at each GCSm threshold. All of these steps were performed in R.

Functional outcome prediction

The methodology for functional outcome prediction was identical to that of motor function detection except that GOSE thresholds instead of GCSm thresholds were used as endpoints.

Assessment of model performance and calibration on validation sets

Both motor function detection and functional outcome prediction models were trained and validated on each of the 25 repeated cross-validation splits for each of the 19 observation windows for each of the 19 unique target dimensionalities (d) for each of the endpoint thresholds (5 for GCSm, 5 for GOSE at discharge, 7 for GOSE at 12

months). Models returned binary prediction probabilities as well as a classification based on a probability threshold of 0.5 for each validation set observation.

Based on the validation set predictions, we calculated metrics of binary outcome discrimination performance (**S. Table 2, 4,** and **6**), classification performance (**Table 2**), and probability calibration[26, 52] (**Table 3**). We also visualized ROC curves (**Fig. 2a, 3a,** and **S. Fig. 4a**), probability calibration curves (**Fig. 4**), and, in one case, the precision recall curve (**S. Fig. 3a**) of the optimally discriminating (maximal AUC) models to assess discrimination, calibration, and case detection power respectively. We calculated unbiased mean values and 95% confidence intervals for both metrics and curves with bootstrap bias-corrected cross-validation (BBC-CV) with repeats[53] on 1,000 resamples of the patient set across the validation set predictions. In this way, 95% confidence intervals account for the variation across the patient set, across the nine missing value imputations, and across the 25 repeated cross-validation partitions.

Feature significance scores

The coefficients (i.e., loadings) of the trained LOL projection matrix represent the relative importance of each column in explaining the variance in the dataset stratified by the endpoint[50]. Thus, we derived a relative importance score of each sensor-feature type combination for both motor function detection and functional outcome prediction by multiplying the mean absolute value of the loadings per each combination and the absolute value of the trained logistic regression coefficient of the corresponding reduced dimension. This would be performed across all 25 partitions of each combination of observation window, threshold, and endpoint. We then calculated 95% confidence intervals on feature significance scores by bootstrapping 1,000 resamples across the 25 repeated cross-validation folds and nine missing value imputations.

DATA AVAILABILITY

Per our current Johns Hopkins Medicine IRB protocol (IRB00135674), we are not permitted to share the clinical data collected for this study. However, we welcome all forms of collaboration, and urge interested investigators to contact the corresponding author (SB: sb2406@cam.ac.uk) with their institutional affiliation and proposed use of the dataset to submit a new protocol for access. The data may not be used for commercial products or redistributed in any way.

CODE AVAILABILITY

All code used in the data collection and analyses outlined in this manuscript can be found at the following GitHub repository: https://github.com/sbhattacharyay/nims (DOI: 10.5281/zenodo.4765305).

ACKNOWLEDGEMENTS

We graciously acknowledge the patients, families, NCCU nurses, and physicians who participated in and contributed to this study. S.B. would like to thank Kathleen Mitchell-Fox (Univ. of Cambridge) for reviewing and offering comments on the manuscript. We also wish to specifically thank Aditya Joshi (Rowan Univ.), Sanya Yadav (Univ. of Pittsburgh), Tobias Fauser (Univ. of Arizona), Michiru Fredricks (Johns Hopkins

Univ.), Alexander Sigmon (Johns Hopkins Univ.), Shikha Gandhi (Johns Hopkins Univ.), and Joshua Vogelstein (Johns Hopkins Univ.) for their roles in the early development, data curation, and advising of statistical methodologies of the NIMS project.

This work was partially supported by awards from the Johns Hopkins University Office of the Provost and the Hodson Trust, received by S.B. S.B. is currently funded by a Gates Cambridge fellowship.

AUTHOR CONTRIBUTIONS

S.B. aided in the conceptualization of the study, developed the methodology of the experiments, acquired accelerometry data from patients, acquired funding for the project, performed statistical analyses on the data, visualized the results for publication, and wrote the complete manuscript. J.R. and R.E.C. aided in the conceptualization and data collection of this work and revised the manuscript. M.W., H.B.K., and E.J. aided S.B. in the statistical analysis, processing of data, and visualization of results. P.D. extracted neurological assessment scores from electronic health records. E.C. recruited patients for the study, performed clinical surveys, and collected clinical data from patient records. P.K. aided in the conceptualization of the study and the development of the methodology and established the data acquisition infrastructure. R.D.S. served as the principal investigator, conceptualized the study, aided in the development of the methodology, procured IRB approval for data collection from human subjects, aided in data collection, provided access to clinical resources at the Johns Hopkins Hospital, and revised the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no conflicts of interest.

REFERENCES

1. Maas, A. I. R. *et al*. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **16**, 987-1048 (2017).
2. Stevens, R. D. & Sutter, R. Prognosis in Severe Brain Injury. *Crit. Care Med.* **41** (2013).
3. Turgeon, A. F. *et al*. Mortality associated with withdrawal of life-sustaining therapy for patients with severe traumatic brain injury: a Canadian multicentre cohort study. *CMAJ* **183**, 1581-1588 (2011).
4. Alkhachroum, A., Terilli, K., Megjhani, M. & Park, S. Harnessing Big Data in Neurocritical Care in the Era of Precision Medicine. *Current Treatment Options in Neurology* **22**, 15 (2020).
5. Fidali, B. C., Stevens, R. D. & Claassen, J. Novel approaches to prediction in severe brain injury. *Curr. Opin. Neurol.* **33** (2020).
6. Winters, J. M. & Crago, P. E. in *Biomechanics and neural control of posture and movement* 683 (Springer, New York, 2000).
7. Shadmehr, R. & Krakauer, J. W. A computational neuroanatomy for motor control. *Exp. Brain Res.* **185**, 359-381 (2008).
8. Reinkensmeyer, D. J. *et al*. Computational neurorehabilitation: modeling plasticity and learning to predict recovery. *Journal of NeuroEngineering and Rehabilitation* **13**, 42 (2016).
9. Olkowski, B. F. & Shah, S. O. Early Mobilization in the Neuro-ICU: How Far Can We Go? *Neurocritical Care* **27**, 141-150 (2017).
10. Drummond, M. J. *et al*. Short-term bed rest increases TLR4 and IL-6 expression in skeletal muscle of older adults. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **305**, 216 (2013).
11. Parry, S. M. & Puthucheary, Z. A. The impact of extended bed rest on the musculoskeletal system in the critical care environment. *Extrem Physiol. Med.* **4**, 16-7. eCollection 2015 (2015).
12. Topp, R., Ditmyer, M., King, K., Doherty, K. & Hornyak, J. The effect of bed rest and potential of prehabilitation on patients in the intensive care unit. *AACN Clin. Issues* **13**, 263-276 (2002).
13. Bloomfield, S. A. Changes in musculoskeletal structure and function with prolonged bed rest. *Med. Sci. Sports Exerc.* **29**, 197-206 (1997).
14. Fowles, J. R., Sale, D. G. & MacDougall, J. D. Reduced strength after passive stretch of the human plantarflexors. *J. Appl. Physiol. (1985)* **89**, 1179-1188 (2000).
15. Bahouth, M. N. *et al*. Safety and Feasibility of a Neuroscience Critical Care Program to Mobilize Patients With Primary Intracerebral Hemorrhage. *Arch. Phys. Med. Rehabil.* **99**, 1220-1225 (2018).
16. Preece, S. J. *et al*. Activity identification using body-mounted sensors--a review of classification techniques. *Physiol. Meas.* **30**, 1 (2009).

17. Adrian, M. & Cooper, J. M. in *Biomechanics of human movement* 572 (Brown & Benchmark, Madison, Wis., 1995).

18. Mathie, M. J., Coster, A. C., Lovell, N. H. & Celler, B. G. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiol. Meas.* **25**, 1 (2004).

19. Bouten, C. V., Koekkoek, K. T., Verduin, M., Kodde, R. & Janssen, J. D. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans. Biomed. Eng.* **44**, 136-147 (1997).

20. Moe-Nilssen, R. Test-retest reliability of trunk accelerometry during standing and walking. *Arch. Phys. Med. Rehabil.* **79**, 1377-1385 (1998).

21. Hansson, G. A., Asterland, P., Holmer, N. G. & Skerfving, S. Validity and reliability of triaxial accelerometers for inclinometry in posture analysis. *Med. Biol. Eng. Comput.* **39**, 405-413 (2001).

22. Meijer, G. A., Westerterp, K. R., Verhoeven, F. M., Koper, H. B. & ten Hoor, F. Methods to assess physical activity with special reference to motion sensors and accelerometers. *IEEE Trans. Biomed. Eng.* **38**, 221-229 (1991).

23. Verceles, A. C. & Hager, E. R. Use of Accelerometry to Monitor Physical Activity in Critically Ill Subjects: A Systematic Review. *Respir. Care* **60**, 1330-1336 (2015).

24. Teasdale, G. & Jennett, B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* **2**, 81-84 (1974).

25. Teasdale, G. *et al*. The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurol.* **13**, 844-854 (2014).

26. Austin, P. C. & Steyerberg, E. W. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Statistics in Medicine* **38**, 4051-4065 (2019).

27. Lugade, V., Fortune, E., Morrow, M. & Kaufman, K. Validity of using tri-axial accelerometers to measure human movement—Part I: Posture and movement detection. *Medical Engineering & Physics* **36**, 169-176 (2014).

28. Jordao, A., Torres, L. A. B. & Schwartz, W. R. Novel approaches to human activity recognition based on accelerometer data. *Signal, Image and Video Processing* **12**, 1387-1394 (2018).

29. Ignatov, A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing* **62**, 915-922 (2018).

30. Migueles, J. H. *et al*. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Medicine* **47**, 1821-1845 (2017).

31. Verceles, A. C. & Hager, E. R. Use of Accelerometry to Monitor Physical Activity in Critically Ill Subjects: A Systematic Review. *Respiratory Care* **60**, 1330-1336 (2015).

32. Fazio, S. *et al*. Quantifying Mobility in the ICU: Comparison of Electronic Health Record Documentation and Accelerometer-Based Sensors to Clinician-Annotated Video. *Critical Care Explorations* **2** (2020).

33. Montoye, A. H. K., Moore, R. W., Bowles, H. R., Korycinski, R. & Pfeiffer, K. A. Reporting accelerometer methods in physical activity intervention studies: a systematic review and recommendations for authors. *Br. J. Sports Med.* **52**, 1507-1516 (2018).

34. Kanai, M. *et al*. Effect of accelerometer-based feedback on physical activity in hospitalized patients with ischemic stroke: a randomized controlled trial. *Clin. Rehabil.* **32**, 1047-1056 (2018).

35. Grimes, L., Outtrim, J. G., Griffin, S. J. & Ercole, A. Accelerometery as a measure of modifiable physical activity in high-risk elderly preoperative patients: a prospective observational pilot study. *BMJ Open* **9**, e032346 (2019).

36. Stienen, M. N. *et al*. Objective activity tracking in spine surgery: a prospective feasibility study with a low-cost consumer grade wearable accelerometer. *Scientific Reports* **10**, 4939 (2020).

37. Lovejoy, C. A., Buch, V. & Maruthappu, M. Artificial intelligence in the intensive care unit. *Crit. Care* **23**, 7-9 (2019).

38. Gholami, B., Haddad, W. M. & Bailey, J. M. AI in the ICU: In the intensive care unit, artificial intelligence can keep watch. *IEEE Spectrum* **55**, 31-35 (2018).

39. Halpern, N. A., Pastores, S. M., Oropello, J. M. & Kvetan, V. Critical Care Medicine in the United States: Addressing the Intensivist Shortage and Image of the Specialty*. *Crit. Care Med.* **41** (2013).

40. Reith, F. C., Brennan, P. M., Maas, A. I. & Teasdale, G. M. Lack of Standardization in the Use of the Glasgow Coma Scale: Results of International Surveys. *J. Neurotrauma* **33**, 89-94 (2016).

41. Reith, F. C., Van den Brande, R., Synnot, A., Gruen, R. & Maas, A. I. The reliability of the Glasgow Coma Scale: a systematic review. *Intensive Care Med.* **42**, 3-15 (2016).

42. Kean, J. & Malec, J. F. Towards a Better Measure of Brain Injury Outcome: New Measures or a New Metric? *Arch. Phys. Med. Rehabil.* **95**, 1225-1228 (2014).

43. Wilson, J. T., Pettigrew, L. E. & Teasdale, G. M. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. *J. Neurotrauma* **15**, 573-585 (1998).

44. van Hees, V. T. *et al*. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One* **8**, e61691 (2013).

45. Box, G. E. P. & Cox, D. R. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)* **26**, 211-243 (1964).

46. Honaker, J., King, G., Blackwell,M. Amelia II: A Program for Missing Data. *Journal of Statistical Software* **45** (2011).

47. R Core Team. R: A Language and Environment for Statistical Computing. **4.0.0** (2020).

48. Rubin, D. B. Inference and missing data. *Biometrika* **63**, 581-592 (1976).

49. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28** (2008).

50. Vogelstein, J. T. *et al*. Geometric Dimensionality Reduction for Subsequent Classification. *arXiv preprint arXiv:1709.01233* (2017).

51. Yeo, I. & Johnson, R. A. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954-959 (2000).

52. Harrell, F. E., Jr. in *Regression Modeling Strategies* (Springer International Publishing AG, Cham, 2015).

53. Tsamardinos, I., Greasidou, E. & Borboudakis, G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach. Learning* **107**, 1895-1922 (2018).

54. Mathie, M. J., Coster, A. C., Lovell, N. H. & Celler, B. G. Detection of daily physical activities using a triaxial accelerometer. *Med. Biol. Eng. Comput.* **41**, 296-301 (2003).

55. Fahrenberg, J., Foerster, F., Smeja, M. & Muller, W. Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. *Psychophysiology* **34**, 607-612 (1997).

56. Foerster, F. & Fahrenberg, J. Motion pattern and posture: correctly assessed by calibrated accelerometers. *Behav. Res. Methods Instrum. Comput.* **32**, 450-457 (2000).

57. Bao, L. &Intille, S. S. *Activity recognition from user-annotated acceleration data* (International conference on pervasive computing, Springer-Verlag, Berlin, Germany, 2004).

58. Sugimoto, A., Hara, Y., Findley, T. W. & Yoncmoto, K. A useful method for measuring daily physical activity by a three-direction monitor. *Scand. J. Rehabil. Med.* **29**, 37-42 (1997).

59. Wang, N., Ambikairajah, E., Lovell, N. H. & Celler, B. G. *Accelerometry Based Classification of Walking Patterns Using Time-frequency Analysis* (29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, USA, 2007).

FIGURE LEGENDS

Fig 1. Accelerometry processing and feature extraction pipeline and experimental paradigm. a Wearable sensor placement on severe brain injury patients in the ICU (top left) and accelerometry(units: *g*)-to-feature pipeline. Sensor placement acronyms correspond to the right and left elbows (RE and LE), the right and left wrists (RW and LW), and the right and left ankles (RA and LE). f_s represents the sampling rate of accelerometry in Hz. Feature type acronyms are decoded in **Table 4**. **b** Experimental paradigm to derive model probabilities for motor function detection per the motor component score of the Glasgow Coma Scale (GCSm) and functional outcome predictions per the Glasgow Outcome Scale – Extended (GOSE) from extracted motion features. LOL corresponds to linear low-rank projection (LOL).

Fig 2. Discrimination performance of motor function detection models on validation sets. a Receiver operating characteristic (ROC) curves of models pertaining to the observation windows with the highest achieved area under the ROC curve (AUC) per each detection threshold of the motor component score of the Glasgow Coma Scale (GCSm). Shaded areas represent 95% confidence intervals derived using bias-corrected bootstrapping (1,000 resamples) to represent the variation across repeated cross-validation folds (5 repeats of 5 folds) and nine missing value imputations. The values in each box represent the observation window achieving the highest AUC as well as the corresponding mean AUC (with 95% confidence interval in parentheses). The diagonal dashed line represents the line of no discrimination (AUC = 0.5). **b** AUC vs. observation windows up to 30 minutes per each detection threshold of the motor component score of the Glasgow Coma Scale (GCSm). Points represent observation windows tested and error bars (with the associated shaded region) represent the 95% confidence interval. The horizontal dashed line corresponds to no discrimination (AUC = 0.5).

Fig 3. Discrimination performance of functional outcome at hospital discharge prediction models on validation sets. a Receiver operating characteristic (ROC) curves of models pertaining to the observation windows with the highest achieved area under the ROC curve (AUC) per each tested prediction threshold of the Glasgow Outcome Scale – Extended (GOSE). Shaded areas represent 95% confidence intervals derived using bias-corrected bootstrapping (1,000 resamples) to represent the variation across repeated cross-validation folds (5 repeats of 5 folds) and nine missing value imputations. The values in each box represent the observation window achieving the highest AUC as well as the corresponding mean AUC (with 95% confidence interval in parentheses). The diagonal dashed line represents the line of no discrimination (AUC = 0.5). **b** AUC vs. observation windows up to 6 hours per each tested prediction threshold of the Glasgow Outcome Scale – Extended (GOSE). Points represent observation windows tested and error bars (with the associated shaded region) represent the 95% confidence interval. The horizontal dashed line corresponds to no discrimination (AUC = 0.5).

Fig 4. Probability calibration of optimally discriminating motor function detection and functional outcome prediction models on validation sets. a Validation set probability calibration curves of models pertaining to the observation windows with the highest achieved AUC (**a**) per each detection threshold of the motor component score of the Glasgow Coma Scale (GCSm) as shown in **Fig. 2a** and (**b**)

per each tested prediction threshold of the Glasgow Outcome Scale – Extended (GOSE) as shown in **Fig. 3a**. Shaded areas represent 95% confidence intervals derived using bias-corrected bootstrapping (1,000 resamples) to represent the variation across repeated cross-validation folds (5 repeats of 5 folds) and nine missing value imputations. The distribution of predicted probabilities is shown at the bottom of the graphs, stratified by threshold-level endpoints. The values in each box represent the observation window achieving the highest AUC as well as the corresponding mean maximal absolute difference between observed and predicted probabilities of the endpoint (E_{max}) with 95% confidence interval in parentheses. The diagonal dashed line represents the line of perfect calibration.

Fig 5. Feature significance matrices of optimally discriminating motor function detection and functional outcome prediction models. Significance scores are calculated by weighting linear optimal low-rank projection (LOL) coefficients of sensor-feature type combinations by the logistic regression coefficients of the corresponding LOL component. The feature significance matrix in (**a**) corresponds to the optimally discriminating model configuration (6-hour observation window) for detection of GCSm > 4 (**Fig. 2a**) while the matrix in (**b**) corresponds to the optimally discriminating model configuration (6-hour observation window) for prediction of GOSE > 5 at hospital discharge (**Fig. 3a**). Mean significance scores are listed as well as 95% confidence intervals bootstrapped from 1,000 resamples to represent the variation across repeated cross-validation folds (5 repeats of 5 folds) and nine missing value imputations. Sensor placement acronyms correspond to joints shown in **Fig. 1** and feature type acronyms are decoded in **Table 4**.

Fig 6. Retrospective case study analysis of accelerometry-based detection of motor function in six patients who experienced relevant transition. The red and blue lines correspond to the predicted probabilities returned my models trained on all other patients on short (27 minutes) and long (6 hours) observation windows respectively. Shaded areas represent 95% confidence intervals derived using bootstrapping (10,000 resamples) to represent the variation across nine missing value imputations. Upward triangle markers designate GCSm > 4 while downward triangle markers designate GCSm ≤ 4.

TABLES

Table 1. Study population characteristics.

Characteristic		Severe brain injury patients (*n* = 69)
Age (y)		59 (48–70)
M/F (*n*)		33/36
Types of severe brain injury (*n*)	Intracranial hemorrhage (ICH)	29 (42.03%)
	Subdural or epidural hematoma (SDH/EDH)	18 (26.09%)
	Subarachnoid hemorrhage (SAH)	17 (24.64%)
	Cerebrovascular accident (CVA)	14 (20.29%)
	Brain or spinal tumor (BT/ST)	12 (17.39%)
	Traumatic brain injury (TBI)	8 (11.59%)
Motor component score of the Glasgow Coma Scale (GCSm) at ICU admission	(1) No response	4 (5.80%)
	(2) Abnormal extension	3 (4.35%)
	(3) Abnormal flexion	6 (8.70%)
	(4) Withdrawal from stimulus	4 (5.80%)
	(5) Movement localized to stimulus	17 (24.64%)
	(6) Obeys commands	35 (50.72%)
Motor component score of the Glasgow Coma Scale (GCSm) at ICU discharge	(1) No response	7 (10.14%)
	(2) Abnormal extension	5 (7.25%)
	(3) Abnormal flexion	4 (5.80%)
	(4) Withdrawal from stimulus	4 (5.80%)
	(5) Movement localized to stimulus	11 (15.94%)
	(6) Obeys commands	38 (55.07%)
Net change in GCSm during ICU stay		0 (0–$^{+}$1)
Glasgow Outcome Scale - Extended (GOSE) at hospital discharge	(1) Dead	16 (23.19%)
	(2) Vegetative state	4 (5.80%)
	(3) Lower severe disability	30 (43.48%)
	(4) Upper severe disability	11 (15.94%)
	(5) Lower moderate disability	6 (8.70%)
	(6) Upper moderate disability	1 (1.45%)
	(7) Lower good recovery	1 (1.45%)
	(8) Upper good recovery	0 (0%)
Glasgow Outcome Scale - Extended (GOSE) at 12 months post discharge*	(1) Dead	28 (43.75%)
	(2) Vegetative state	2 (3.12%)
	(3) Lower severe disability	14 (21.88%)
	(4) Upper severe disability	12 (18.75%)
	(5) Lower moderate disability	3 (4.69%)
	(6) Upper moderate disability	0 (0%)
	(7) Lower good recovery	3 (4.69%)
	(8) Upper good recovery	2 (3.12%)
Net change in GOSE in 12 months post discharge*		0 ($^{-}$3.425–$^{+}$3)
Length of stay in ICU (d)		19 (11–29)
Delay between ICU discharge and recording start (d)		7 (2–15)
Recording duration (h)		24.09 (22.81–25.11)
Percentage of ICU stay recorded (%)		5.54 (3.13–8.51)
Percentage of ICU stay elapsed before recording start (%)		43.53 (24.98–62.88)
Recording duration (h)		24.09 (22.81–25.11)

Values represent medians with interquartile ranges (Q1–Q3) in parentheses or counts with percentages (%) in parentheses.

*Total sample size at 12 months post discharge is *n* = 64.

Table 2. Classification performance metrics of optimally discriminating models.

Task	Threshold	*n**	Accuracy	Precision	Recall (Sensitivity)	Specificity	F_1 score
Detection of GCSm	GCSm > 1	15/244 (0.94)	0.91 (0.84–0.97)	0.93 (0.86–0.99)	0.97 (0.95–0.99)	0.03 (0.00–0.10)	0.95 (0.91–0.98)
	GCSm > 2	84/480 (0.85)	0.83 (0.70–0.94)	0.86 (0.73–0.97)	0.96 (0.94–0.99)	0.15 (0.01–0.35)	0.90 (0.83–0.96)
	GCSm > 3	175/424 (0.71)	0.69 (0.57–0.80)	0.75 (0.59–0.90)	0.83 (0.74–0.90)	0.44 (0.28–0.64)	0.79 (0.68–0.88)
	GCSm > 4	166/322 (0.66)	0.71 (0.59–0.83)	0.76 (0.59–0.90)	0.79 (0.66–0.91)	0.54 (0.36–0.73)	0.78 (0.67–0.87)
	GCSm > 5	344/255 (0.57)	0.63 (0.53–0.72)	0.59 (0.38–0.81)	0.58 (0.45–0.69)	0.75 (0.60–0.89)	0.54 (0.40–0.66)
Prediction of GOSE at hospital discharge	GOSE > 1	120/368 (0.75)	0.69 (0.57–0.81)	0.77 (0.61–0.92)	0.87 (0.79–0.94)	0.30 (0.13–0.50)	0.81 (0.70–0.90)
	GOSE > 2	89/245 (0.73)	0.66 (0.52–0.79)	0.74 (0.56–0.89)	0.82 (0.71–0.92)	0.23 (0.03–0.44)	0.78 (0.66–0.87)
	GOSE > 3	451/198 (0.69)	0.64 (0.53–0.75)	0.35 (0.13–0.55)	0.24 (0.13–0.35)	0.84 (0.78–0.90)	0.28 (0.15–0.39)
	GOSE > 4	567/77 (0.88)	0.84 (0.73–0.93)	0.12 (0.00–0.35)	0.11 (0.01–0.24)	0.96 (0.92–0.98)	0.11 (0.01–0.23)
	GOSE > 5	479/9 (0.98)	0.97 (0.95–0.99)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.99 (0.98–1.00)	0.00 (0.00–0.00)
Prediction of GOSE at 12 months post discharge	GOSE > 1	270/339 (0.56)	0.54 (0.45–0.64)	0.58 (0.38–0.78)	0.62 (0.48–0.73)	0.45 (0.33–0.57)	0.60 (0.46–0.72)
	GOSE > 2	288/329 (0.53)	0.51 (0.43–0.59)	0.54 (0.35–0.72)	0.59 (0.50–0.69)	0.47 (0.35–0.59)	0.54 (0.39–0.66)
	GOSE > 3	390/203 (0.66)	0.57 (0.47–0.66)	0.34 (0.14–0.55)	0.33 (0.21–0.44)	0.80 (0.72–0.87)	0.32 (0.18–0.46)
	GOSE > 4	488/105 (0.82)	0.76 (0.63–0.87)	0.12 (0.01–0.29)	0.11 (0.03–0.24)	0.92 (0.87–0.96)	0.08 (0.01–0.18)
	GOSE > 5	538/70 (0.88)	0.85 (0.74–0.93)	0.15 (0.00–0.42)	0.10 (0.02–0.29)	0.96 (0.93–0.98)	0.12 (0.02–0.28)
	GOSE > 6	538/70 (0.88)	0.84 (0.72–0.93)	0.15 (0.00–0.44)	0.07 (0.01–0.17)	0.96 (0.93–0.99)	0.09 (0.00–0.19)
	GOSE > 7	591/24 (0.96)	0.94 (0.92–0.96)	0.04 (0.00–0.12)	0.03 (0.02–0.04)	0.99 (0.96–0.99)	0.02 (0.00–0.06)

Classification metrics [mean (95% confidence interval)] corresponding to models trained on observation windows that maximize the area under the receiver operating characteristic curve (AUC) for each threshold (**Fig. 2a, 3a, and S. Fig. 4a**). Confidence intervals were derived using bias-corrected bootstrapping (1,000 resamples) and represent the variation across repeated cross-validation folds (5 repeats of 5 folds) and nine missing value imputations. Acronyms: motor component score of the Glasgow Coma Scale (GCSm) and Glasgow Outcome Scale – Extended (GOSE).
*Count distribution of negative vs. positive cases with the proportion of the most represented case, equivalent to the no information rate, in parentheses.

Table 3. Probability calibration metrics of optimally discriminating models.

Task	Threshold	*n**	E_{max}	E_{50}	E_{90}	ICI
Detection of GCSm	GCSm > 1	15/244 (0.94)	0.92 (0.63–1.00)	0.04 (0.01–0.10)	0.04 (0.01–0.11)	0.07 (0.02–0.12)
	GCSm > 2	84/480 (0.85)	0.91 (0.55–1.09)	0.09 (0.01–0.18)	0.21 (0.09–0.36)	0.11 (0.04–0.19)
	GCSm > 3	175/424 (0.71)	0.55 (0.21–0.86)	0.10 (0.02–0.21)	0.25 (0.08–0.45)	0.13 (0.05–0.22)
	GCSm > 4	166/322 (0.66)	0.30 (0.08–0.64)	0.10 (0.02–0.21)	0.22 (0.06–0.44)	0.11 (0.03–0.22)
	GCSm > 5	344/255 (0.57)	0.43 (0.17–0.78)	0.10 (0.03–0.23)	0.31 (0.11–0.55)	0.14 (0.06–0.24)
Prediction of GOSE at hospital discharge	GOSE > 1	120/368 (0.75)	0.81 (0.46–1.07)	0.17 (0.04–0.32)	0.38 (0.19–0.61)	0.20 (0.11–0.32)
	GOSE > 2	89/245 (0.73)	0.65 (0.27–1.01)	0.18 (0.04–0.34)	0.39 (0.18–0.66)	0.20 (0.10–0.32)
	GOSE > 3	451/198 (0.69)	0.60 (0.24–0.95)	0.14 (0.06–0.23)	0.29 (0.14–0.45)	0.15 (0.08–0.23)
	GOSE > 4	567/77 (0.88)	0.80 (0.38–1.03)	0.07 (0.02–0.17)	0.15 (0.06–0.30)	0.10 (0.05–0.18)
	GOSE > 5	479/9 (0.98)	0.89 (0.00–1.00)	0.00 (0.00–0.01)	0.00 (0.00–0.01)	0.01 (0.00–0.02)
Prediction of GOSE at 12 months post discharge	GOSE > 1	270/339 (0.56)	0.63 (0.39–0.86)	0.21 (0.11–0.32)	0.49 (0.30–0.68)	0.24 (0.15–0.34)
	GOSE > 2	288/329 (0.53)	0.66 (0.43–0.88)	0.19 (0.10–0.30)	0.44 (0.29–0.59)	0.22 (0.14–0.31)
	GOSE > 3	390/203 (0.66)	0.66 (0.44–0.88)	0.24 (0.11–0.38)	0.57 (0.35–0.82)	0.27 (0.19–0.35)
	GOSE > 4	488/105 (0.82)	0.96 (0.76–1.06)	0.14 (0.02–0.28)	0.27 (0.10–0.47)	0.17 (0.08–0.28)
	GOSE > 5	538/70 (0.88)	0.87 (0.56–1.03)	0.09 (0.01–0.20)	0.22 (0.09–0.38)	0.12 (0.04–0.24)
	GOSE > 6	538/70 (0.88)	0.86 (0.53–1.03)	0.09 (0.01–0.21)	0.22 (0.08–0.36)	0.13 (0.04–0.24)
	GOSE > 7	591/24 (0.96)	0.92 (0.82–1.00)	0.04 (0.01–0.07)	0.04 (0.01–0.07)	0.05 (0.02–0.09)

Probability calibration metrics[26, 52] [mean (95% confidence interval)] corresponding to models trained on observation windows that maximize the area under the receiver operating characteristic curve (AUC) for each threshold (**Fig. 2a, 3a, and S. Fig. 4a**). Confidence intervals were derived using bias-corrected bootstrapping (1,000 resamples) and represent the variation across repeated cross-validation folds (5 repeats of 5 folds) and nine missing value imputations. Acronyms: motor component score of the Glasgow Coma Scale (GCSm), Glasgow Outcome Scale – Extended (GOSE), and Integrated Calibration Index (ICI).

*Count distribution of negative vs. positive cases with the proportion of the most represented case, equivalent to the no information rate, in parentheses.

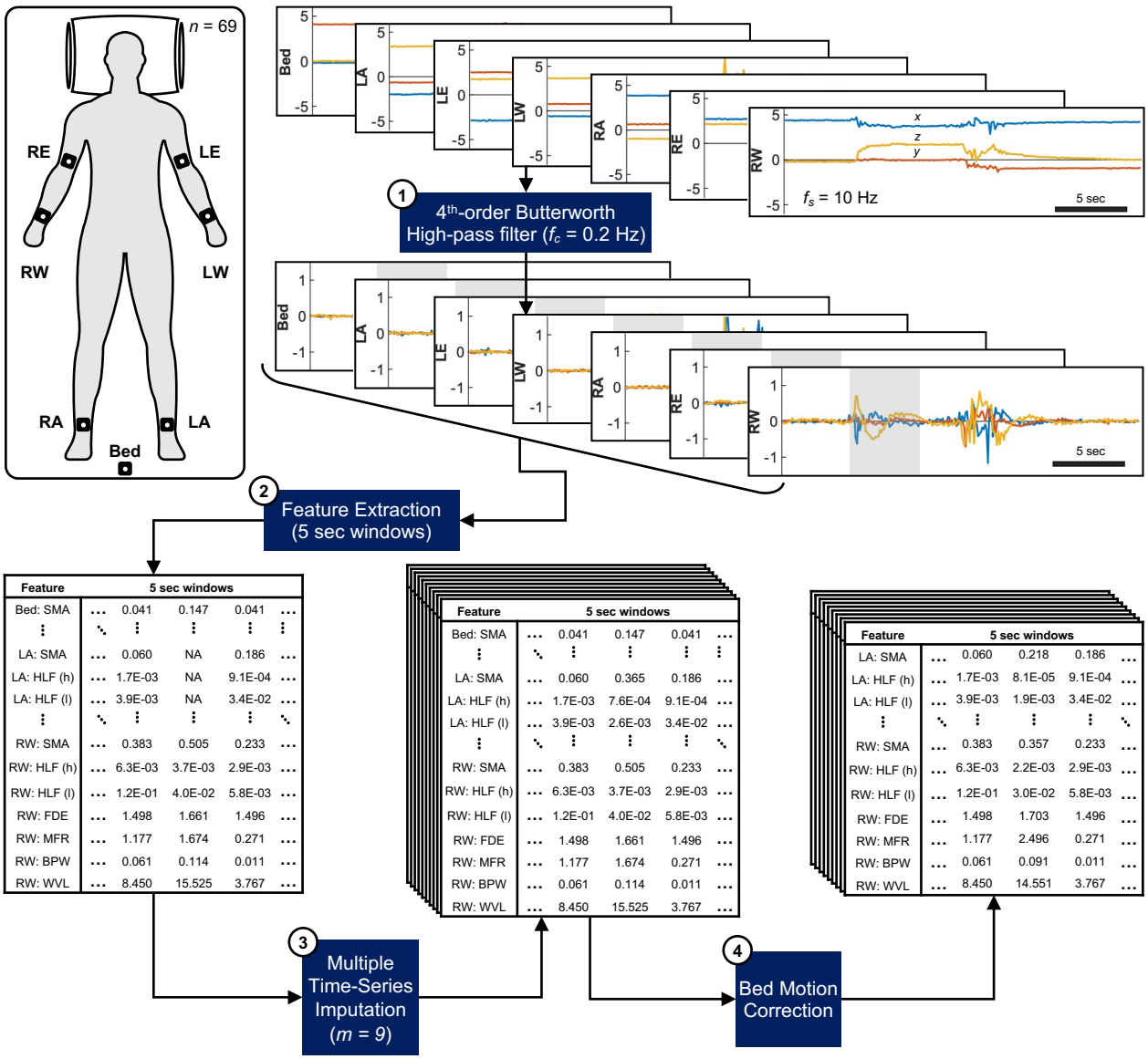
Table 4. Overview of extracted motion feature types.

Acronym	Feature description	Domain	Reference
PDA	Proportion of dynamic activity (SMA ≥ 0.135*g*) in observation window	Time	27
SMA	Signal magnitude area	Time	54
HLF (h)	Median of high-pass-filtered (4th-order Butterworth, f_c = 2.5 Hz) signal	Time	55
HLF (l)	Median of low-pass-filtered (4th-order Butterworth, f_c = 2.5 Hz) signal	Time	
MFR	Median frequency according to Fourier transform coefficients	Frequency	56
FDE	Frequency-domain entropy	Frequency	57
BPW	Band power between 0.3 and 3.5 Hz	Frequency	58
WVL	Level 2 – 6 detail coefficients of the 5th-order Daubechies wavelet transform	Wavelets	59

Each feature, except PDA, was extracted from non-overlapping 5 second windows and root-sum-of-squares leveled across the three axes of accelerometry measurement (x, y, z). A single PDA value was calculated per each sensor for the entire observation window. f_c represents the critical frequency of the Butterworth filter. The references point to original works in which the features (or similar variants) were used in physical activity recognition. Explicit formulae for each feature can be found in **Methods**.

Figure 1

a)



Feature	5 sec windows				
Bed: SMA	…	0.041	0.147	0.041	…
⋮	⋱	⋮	⋮	⋮	⋮
LA: SMA	…	0.060	NA	0.186	…
LA: HLF (h)	…	1.7E-03	NA	9.1E-04	…
LA: HLF (l)	…	3.9E-03	NA	3.4E-02	…
⋮	⋱	⋮	⋮	⋮	⋱
RW: SMA	…	0.383	0.505	0.233	…
RW: HLF (h)	…	6.3E-03	3.7E-03	2.9E-03	…
RW: HLF (l)	…	1.2E-01	4.0E-02	5.8E-03	…
RW: FDE	…	1.498	1.661	1.496	…
RW: MFR	…	1.177	1.674	0.271	…
RW: BPW	…	0.061	0.114	0.011	…
RW: WVL	…	8.450	15.525	3.767	…

Feature	5 sec windows				
Bed: SMA	…	0.041	0.147	0.041	…
⋮	⋱	⋮	⋮	⋮	⋮
LA: SMA	…	0.060	0.365	0.186	…
LA: HLF (h)	…	1.7E-03	7.6E-04	9.1E-04	…
LA: HLF (l)	…	3.9E-03	2.6E-03	3.4E-02	…
⋮	⋱	⋮	⋮	⋮	⋱
RW: SMA	…	0.383	0.505	0.233	…
RW: HLF (h)	…	6.3E-03	3.7E-03	2.9E-03	…
RW: HLF (l)	…	1.2E-01	4.0E-02	5.8E-03	…
RW: FDE	…	1.498	1.661	1.496	…
RW: MFR	…	1.177	1.674	0.271	…
RW: BPW	…	0.061	0.114	0.011	…
RW: WVL	…	8.450	15.525	3.767	…

Feature	5 sec windows				
LA: SMA	…	0.060	0.218	0.186	…
LA: HLF (h)	…	1.7E-03	8.1E-05	9.1E-04	…
LA: HLF (l)	…	3.9E-03	1.9E-03	3.4E-02	…
⋮	⋱	⋮	⋮	⋮	⋱
RW: SMA	…	0.383	0.357	0.233	…
RW: HLF (h)	…	6.3E-03	2.2E-03	2.9E-03	…
RW: HLF (l)	…	1.2E-01	3.0E-02	5.8E-03	…
RW: FDE	…	1.498	1.703	1.496	…
RW: MFR	…	1.177	2.496	0.271	…
RW: BPW	…	0.061	0.091	0.011	…
RW: WVL	…	8.450	14.551	3.767	…

b)

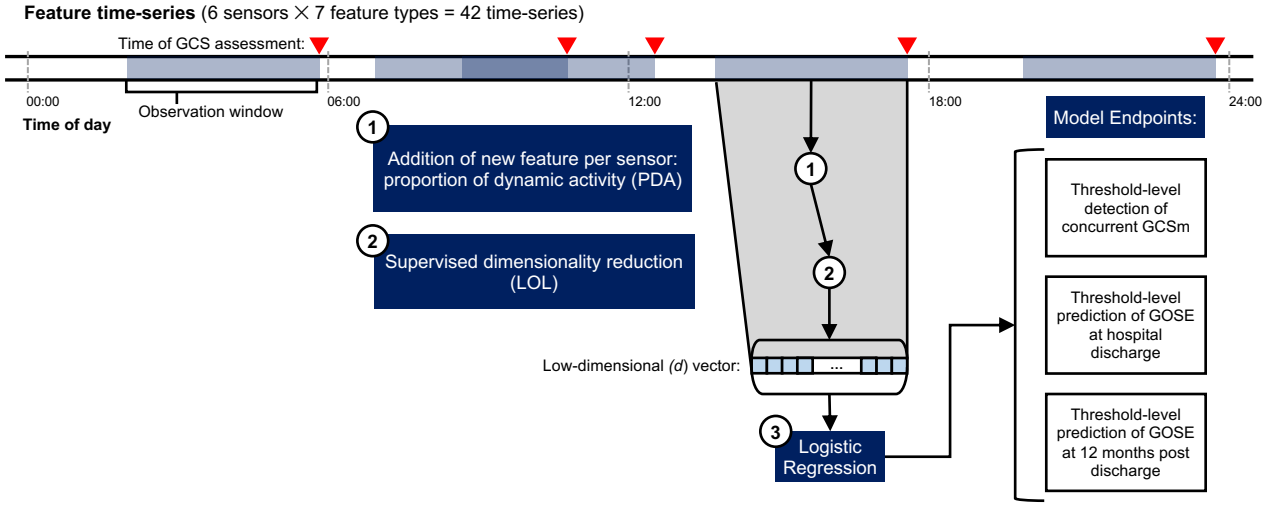


Figure 2

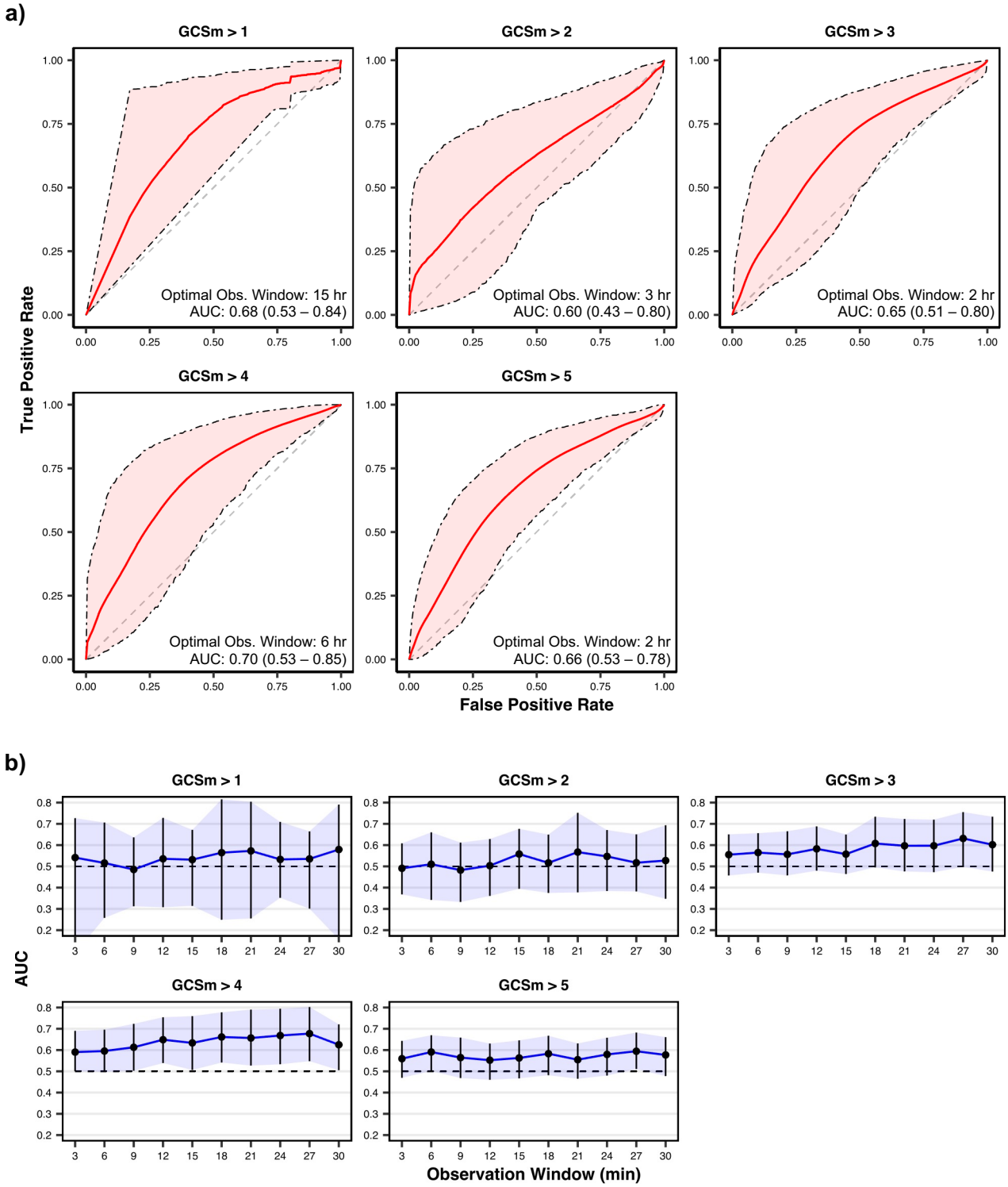

a)
GCSm > 1
GCSm > 2
GCSm > 3
GCSm > 4
GCSm > 5
True Positive Rate
False Positive Rate
Optimal Obs. Window: 15 hr
AUC: 0.68 (0.53 – 0.84)
Optimal Obs. Window: 3 hr
AUC: 0.60 (0.43 – 0.80)
Optimal Obs. Window: 2 hr
AUC: 0.65 (0.51 – 0.80)
Optimal Obs. Window: 6 hr
AUC: 0.70 (0.53 – 0.85)
Optimal Obs. Window: 2 hr
AUC: 0.66 (0.53 – 0.78)
b)
AUC
Observation Window (min)

Figure 3

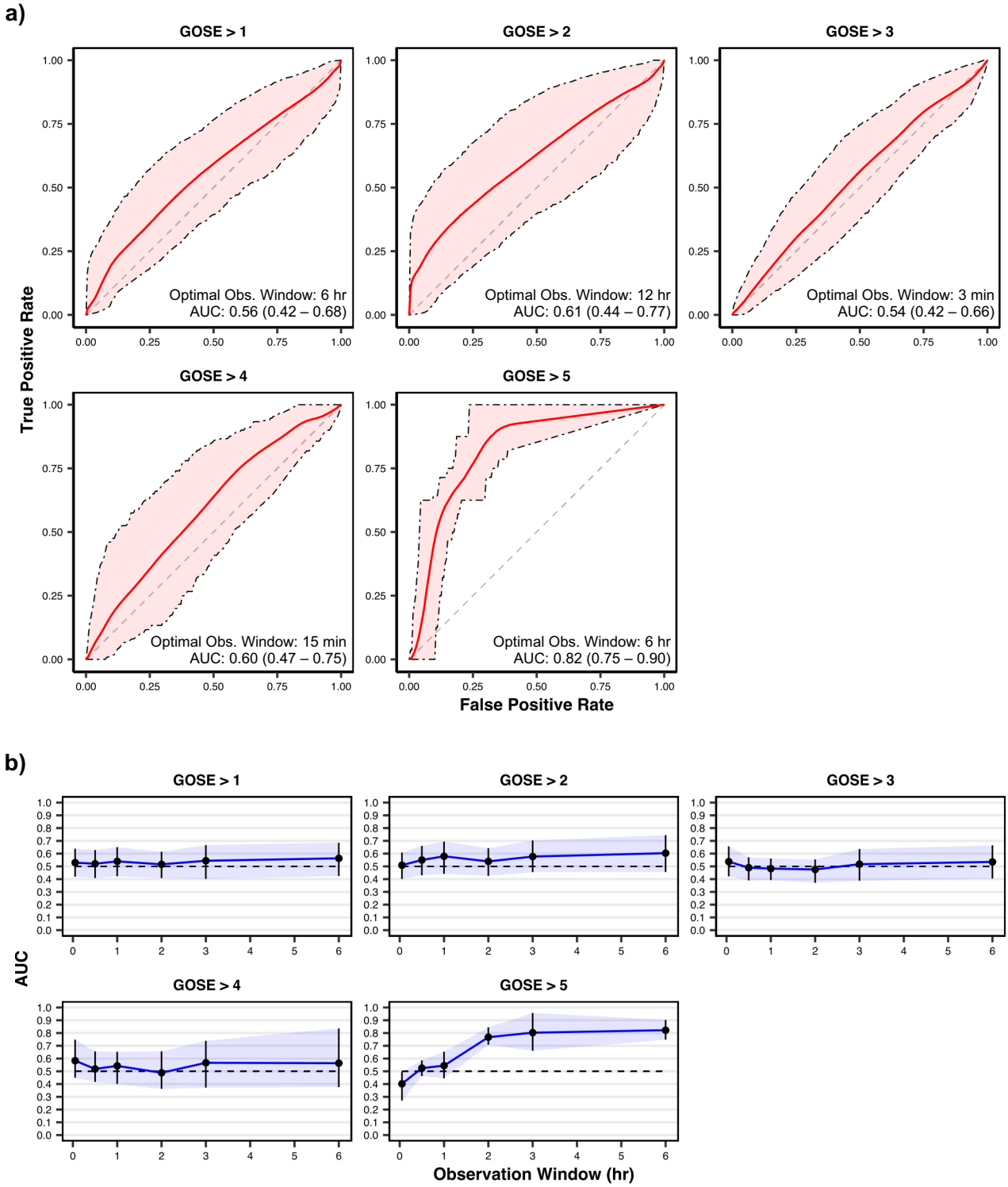

a)
GOSE > 1
GOSE > 2
GOSE > 3
GOSE > 4
GOSE > 5
True Positive Rate
False Positive Rate
Optimal Obs. Window: 6 hr
AUC: 0.56 (0.42 – 0.68)
Optimal Obs. Window: 12 hr
AUC: 0.61 (0.44 – 0.77)
Optimal Obs. Window: 3 min
AUC: 0.54 (0.42 – 0.66)
Optimal Obs. Window: 15 min
AUC: 0.60 (0.47 – 0.75)
Optimal Obs. Window: 6 hr
AUC: 0.82 (0.75 – 0.90)
b)
GOSE > 1
GOSE > 2
GOSE > 3
GOSE > 4
GOSE > 5
AUC
Observation Window (hr)

Figure 4

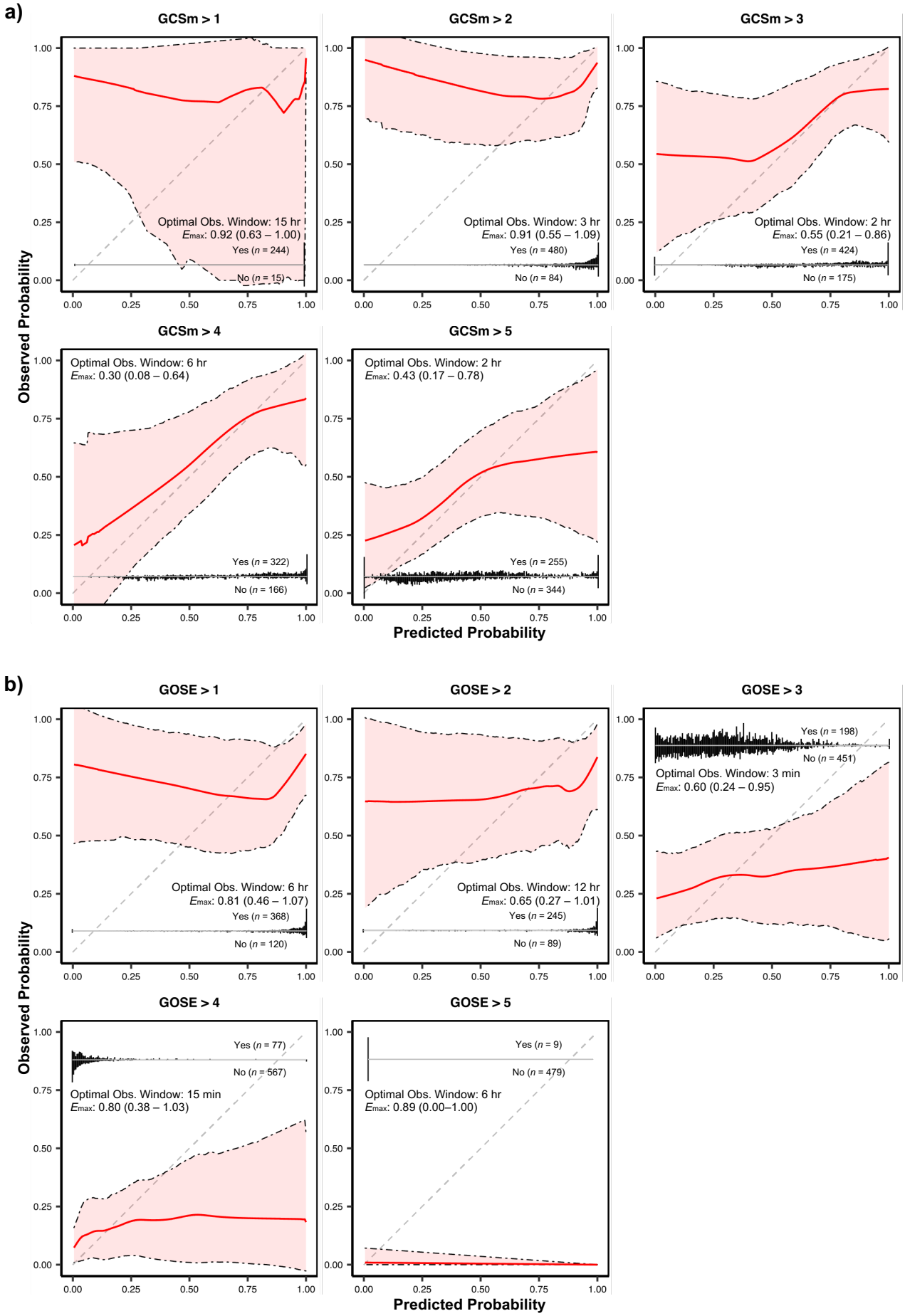

a)
GCSm > 1
Optimal Obs. Window: 15 hr
Emax: 0.92 (0.63 – 1.00)
Yes (n = 244)
No (n = 15)
GCSm > 2
Optimal Obs. Window: 3 hr
Emax: 0.91 (0.55 – 1.09)
Yes (n = 480)
No (n = 84)
GCSm > 3
Optimal Obs. Window: 2 hr
Emax: 0.55 (0.21 – 0.86)
Yes (n = 424)
No (n = 175)
GCSm > 4
Optimal Obs. Window: 6 hr
Emax: 0.30 (0.08 – 0.64)
Yes (n = 322)
No (n = 166)
GCSm > 5
Optimal Obs. Window: 2 hr
Emax: 0.43 (0.17 – 0.78)
Yes (n = 255)
No (n = 344)
Observed Probability
Predicted Probability
b)
GOSE > 1
Optimal Obs. Window: 6 hr
Emax: 0.81 (0.46 – 1.07)
Yes (n = 368)
No (n = 120)
GOSE > 2
Optimal Obs. Window: 12 hr
Emax: 0.65 (0.27 – 1.01)
Yes (n = 245)
No (n = 89)
GOSE > 3
Yes (n = 198)
No (n = 451)
Optimal Obs. Window: 3 min
Emax: 0.60 (0.24 – 0.95)
GOSE > 4
Yes (n = 77)
No (n = 567)
Optimal Obs. Window: 15 min
Emax: 0.80 (0.38 – 1.03)
GOSE > 5
Yes (n = 9)
No (n = 479)
Optimal Obs. Window: 6 hr
Emax: 0.89 (0.00–1.00)
Observed Probability
Predicted Probability

Figure 5

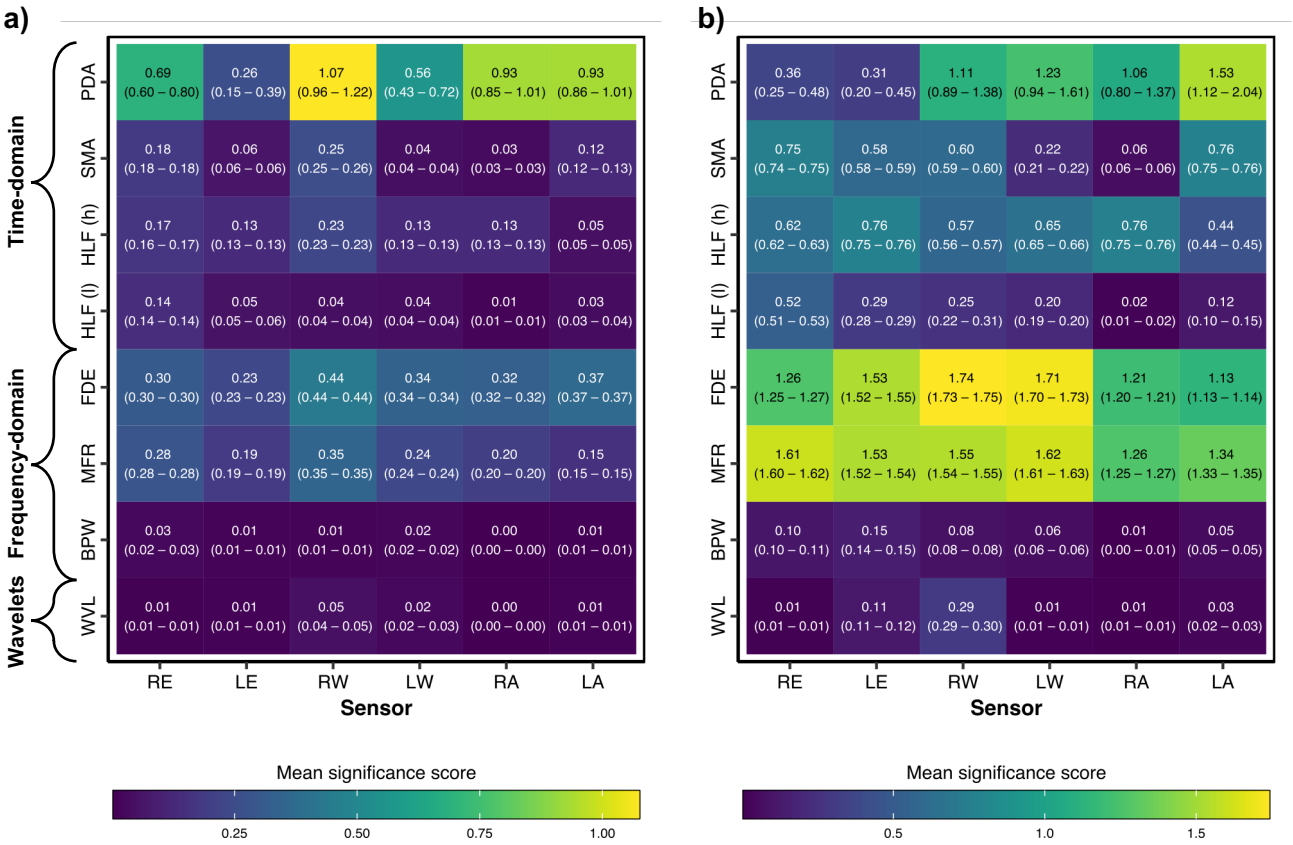
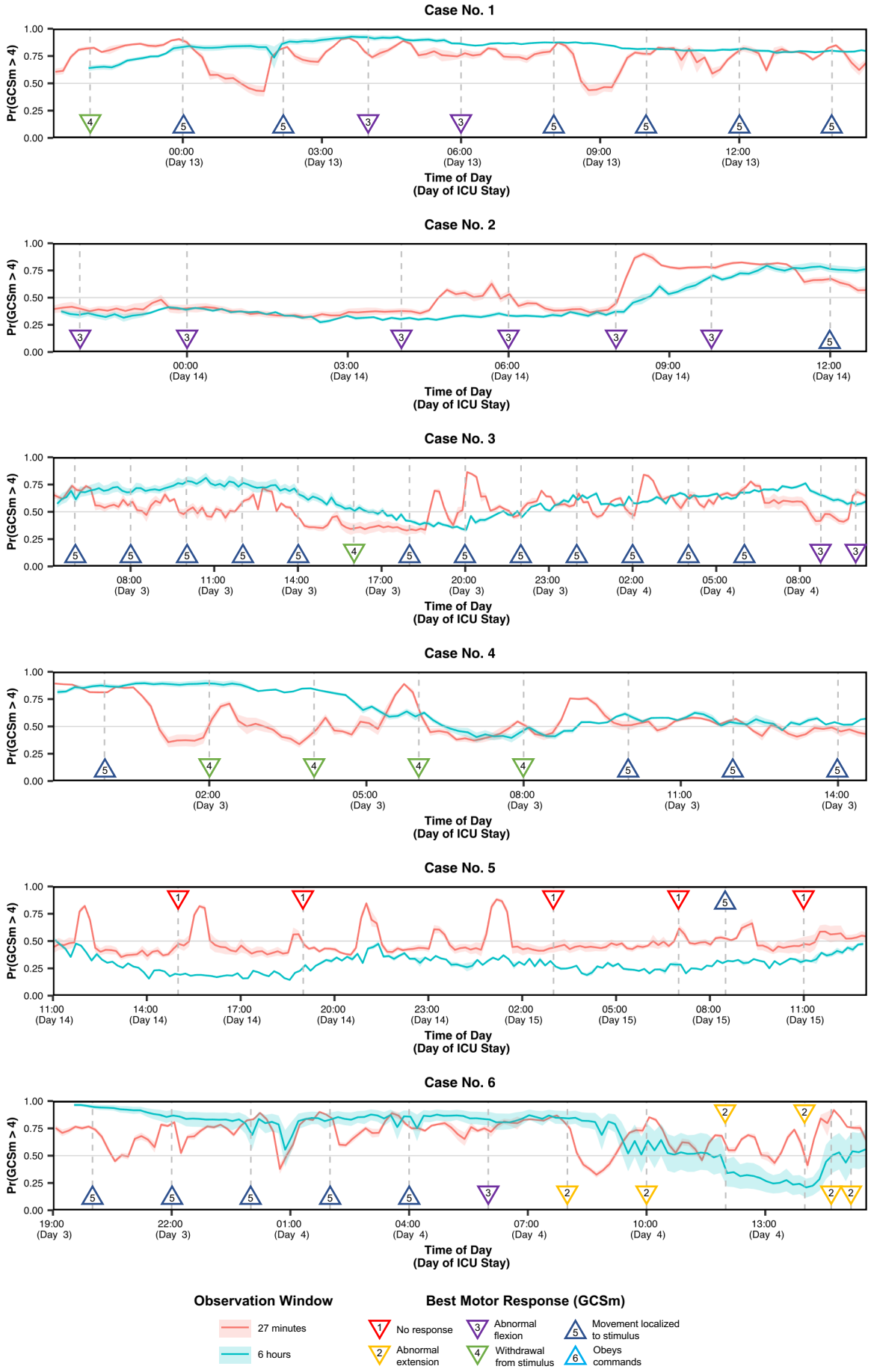

a)
Time-domain
Frequency-domain
Wavelets
PDA
SMA
HLF (h)
HLF (l)
FDE
MFR
BPW
WVL
RE
LE
RW
LW
RA
LA
Sensor
Mean significance score
0.25
0.50
0.75
1.00
b)
Sensor
Mean significance score
0.5
1.0
1.5

Figure 6


Case No. 1
Case No. 2
Case No. 3
Case No. 4
Case No. 5
Case No. 6
Pr(GCSm > 4)
Time of Day
(Day of ICU Stay)
Observation Window
27 minutes
6 hours
Best Motor Response (GCSm)
No response
Abnormal extension
Abnormal flexion
Withdrawal from stimulus
Movement localized to stimulus
Obeys commands