Delirium Detection Using Wearable Sensors and Machine Learning in Patients with Intracerebral Hemorrhage

Abdullah Ahmed[1*], Augusto Garcia-Agundez, PhD[1,2*], Ivana Petrovic, PhD[1], Fatemeh Radaei, MS[1], James Fife BS[1], John Zhou[1], Hunter Karas[1], Scott Moody, BS[3], Jonathan Drake, MD[3], Richard N. Jones, ScD[4], Carsten Eickhoff, PhD[1†], Michael E. Reznik, MD[3†]

Author affiliations:
[1]Brown Center for Biomedical Informatics, Brown University, Providence, RI, United States
[2]IMDEA Networks Institute, Madrid, Spain
[3]Department of Neurology, Brown University, Providence, RI, United States
[4]Department of Psychiatry, Brown University, Providence, RI, United States
[*]Equally collaborating authors
[†]Equally collaborating senior authors

Corresponding authors:
Michael Reznik
Division of Neurocritical Care, Rhode Island Hospital
593 Eddy Street, APC 712
Providence, RI 02903
Email: Michael_Reznik@brown.edu

Carsten Eickhoff
233 Richmond St, Room 209
Providence, RI, 02912
Email: Carsten@brown.edu

Keywords: Delirium, Neurocritical Care, Stroke, Intracerebral Hemorrhage, Actigraphy, Machine Learning, Wearable Electronic Devices

Abstract word count: 300
Manuscript word count: 2877
Figures: 3
Tables: 2

Author Contributions:

- Conceptualization and methodology: MER, CE, RNJ
- Funding acquisition: MER, CE
- Investigation and analysis of data: AA, AGA, IP, FR, JF, JZ, HK, SM, JD, CE, MER
- Supervision: MER, CE
- Drafting and/or revising significant portions of the manuscript: AA, AGA, CE, MER

ABSTRACT

Objective. Delirium is associated with worse outcomes in patients with stroke and neurocritical illness, but delirium detection in these patients can be challenging with existing screening tools. To address this gap, we aimed to develop and evaluate machine learning models that detect episodes of post-stroke delirium based on data from wearable activity monitors in conjunction with stroke-related clinical features.

Design. Prospective observational cohort study.

Setting. Neurocritical Care and Stroke Units at an academic medical center.

Patients. We recruited 39 patients with moderate-to-severe acute intracerebral hemorrhage (ICH) and hemiparesis over a 1-year period (mean [SD] age 71.3 [12.20], 54% male, median [IQR] initial NIHSS 14.5 [6], median [IQR] ICH score 2 [1]).

Measurements & Main Results. Each patient received daily assessments for delirium by an attending neurologist, while activity data were recorded throughout each patient's hospitalization using wrist-worn actigraph devices (on both paretic and non-paretic arms). We compared the predictive accuracy of Random Forest, SVM and XGBoost machine learning methods in classifying daily delirium status using clinical information alone and combined with actigraph data. Among our study cohort, 85% of patients (n=33) had at least one delirium episode, while 71% of monitoring days (n=209) were rated as days with delirium. Clinical information alone had a low accuracy in detecting delirium on a day-to-day basis (Accuracy mean [SD] 62% [18%], F1 score mean [SD] 50% [17%]). Prediction performance improved significantly ($p<0.001$) with the addition of actigraph data (Accuracy mean [SD] 74% [10%], F1 score 65% [10%]). Among actigraphy features, night-time actigraph data were especially relevant for classification accuracy.

Conclusions. We found that actigraphy in conjunction with machine learning models improves clinical detection of delirium in patients with stroke, thus paving the way to make actigraph-assisted predictions clinically actionable.

INTRODUCTION

Delirium occurs frequently in critically ill patients and has consistently been associated with higher mortality and worse overall outcomes. [1] However, the diagnosis and detection of delirium remains challenging in many patient populations, as existing screening tools are unreliable in patients with stroke, [2] neurocritical illness, [3] and dementia. [4] Further, because the diagnosis and detection of delirium relies on bedside testing, identifying at-risk patients across all hospital settings is highly labor intensive. Automated methods of delirium screening would therefore fill a critical need in the care of patients whose delirium may otherwise go undetected due to superimposed neurologic deficits, and potentially for critically ill patients as a whole.

We have previously described fluctuations of consciousness as a potential behavioral biomarker that corresponds to delirium in patients with stroke while also potentially identifying cases of delirium that went undetected by conventional screening tools. [5] Additionally, we have found that these fluctuations may correspond to long-term outcomes in a subset of patients with hemorrhagic stroke. [6] However, determining levels of consciousness still depends on frequent bedside assessments and may therefore be prohibitive. Given that motor activity is often heavily factored into clinical measurements of consciousness, [7] and that psychomotor changes are a hallmark of delirium, [8] continuous measurements of motor activity may represent a promising behavioral biomarker to aid in delirium detection.

Various devices for measuring motor activity exist, though wearable sensors such as wrist actigraphs are especially appealing due to their ease of use and relatively low cost. Such devices are commonly used in outpatient sleep medicine settings, [9] and have increasingly been utilized in studies attempting to measure physical activity [10] and sleep [11] in the intensive care unit (ICU). However, actigraphy in ICU settings may be contaminated by externally mediated activity arising from clinical care. As a result, although wrist actigraphs have also been considered as a potential means of predicting delirium, existing studies have thus far shown mixed results using conventional statistical methods. [12-14]

In this study, we aimed to mitigate some of this externally mediated artifact by collecting actigraphy data from a cohort of critically ill patients with hemiparesis due to acute intracerebral hemorrhage (ICH).

Owing to marked differences in activity profiles between paretic and non-paretic sides, we aimed to leverage actigraphy data from both wrists as a within-patient control. Further, the heterogeneous nature of individual activity profiles and delirium cases lends itself to advanced analysis using machine learning-based techniques. We therefore designed this study to test the feasibility of this novel approach.

MATERIALS AND METHODS

Study Population

We prospectively screened all patients admitted to Rhode Island Hospital's Neurocritical Care Unit (NCCU) or Stroke Unit (SU) with acute intracerebral hemorrhage (ICH) over a 1-year period from 2018-2019 for potential enrollment. We included patients with moderate-to-severe supratentorial ICH (NIH stroke scale ≥ 5) and hemiparesis, specifically focusing on supratentorial ICH because of the higher likelihood of cognitive symptoms occurring in conjunction with motor symptoms as compared to patients with brainstem or cerebellar hematomas. We excluded patients with previous limb amputation or significant pre-morbid functional disability requiring assistance with their daily activities (as assessed using the modified Rankin Scale), as well as patients with devastating strokes considered to have a high likelihood of mortality. All patients were enrolled within 72 hours of admission.

Delirium Assessments

Daily delirium assessments were performed each afternoon by an attending neurocritical care or behavioral neurologist, with the exception of weekends and holidays. Delirium was diagnosed according to DSM-5 criteria [15]: disturbances in attention and awareness (often accompanied by disturbances in other cognitive domains, such as psychomotor slowing or agitation, disorientation, disorganized thinking, impaired executive function, or perceptual disturbance) that develop over a short period of time and tend to fluctuate, represent a change in function, and are due to an underlying medical condition or toxic/withdrawal syndrome. Assessments were supplemented by interviews and history obtained from patients' nurses and

clinical providers, family members, and the medical chart. Joint adjudication sessions were held between the participating neurologists to obtain consensus on delirium diagnoses for each patient.

Actigraph Data Collection

Wrist actigraphs (Micro Motionlogger, Ambulatory Monitoring, Inc., Ardsley, NY) were placed on both wrists for each patient and left in place for the duration of their stay in the NCCU or SU. Actigraphs were otherwise only purposefully removed in anticipation of magnetic resonance imaging scans as a safety precaution and were then replaced thereafter.

Each actigraph was configured to collect activity data in 1-minute epochs. These data were then aggregated by proprietary algorithms from the Action4 software package (Ambulatory Monitoring, Inc.) into two distinct measurements: Zero Crossing Mode (ZCM), which measures the frequency of movement by counting the number of times per epoch that the signal crosses a threshold set near zero; and Proportional Integration Mode (PIM), which calculates the area under the curve for the acceleration signal during each epoch, and therefore discriminates between different intensities of motion. Data were subsequently downloaded using the Action W-2 software package (Ambulatory Monitoring, Inc.).

Additional Clinical Data

All data related to standard clinical stroke care were prospectively collected in a REDCap database (Vanderbilt University, Nashville, TN). [16 17] These data included patient demographics, comorbidities, admission NIH Stroke Scale (NIHSS) score, neuroimaging, and other diagnostic testing. ICH-related clinical predictors, including hematoma location, size, and ICH score were adjudicated by two attending neurologists with board certification in neurocritical care and/or vascular neurology until consensus was achieved.

Data Analysis

The task of recognizing episodes of delirium was delineated as a supervised machine learning problem: Given a set of features *X* derived from patient data, predict delirium status as the dependent outcome variable *Y*. Because delirium status was measured on a 24-hour basis, patient data were preprocessed such that individual features were aggregated into a day-level unit of analysis.

Actigraph Data Pre-processing

Raw data for each patient consisted of two parts: raw actigraph measurements, which had minute-to-minute variability, and clinical data, which remained relatively static over the course of each patient's hospitalization (in the case of demographic and stroke-specific data) or over the course of a day (in the case of mechanical ventilation status). Actigraph data for a given 24-hour block were partitioned into four groups: full-day (1PM to 1PM), morning (6 AM to 1 PM), afternoon/evening (1 PM to 10 PM), and night-time (10 PM to 6 AM) epochs. Pre-processing also included normalization of actigraph data prior to classification.

Actigraph Feature Extraction

We then aimed to extract clinically meaningful information from the actigraph data to include as features in our delirium prediction models. **Figure 1** summarizes the feature extraction process, which culminated in our calculation of two key features from the actigraph data: minutes at rest and within-patient dynamic time warping (DTW).

We defined minutes at rest as the daily proportion of PIM measurements equal to zero in both paretic and non-paretic arms. This feature was chosen due to the importance of psychomotor slowing and inactivity in the diagnosis of delirium. [18] It may also provide a helpful estimate of sleep-wake disturbance, a common symptom of delirium, [18] though actigraphy likely overestimates actual sleep time in hospital and ICU settings. [10]

We also implemented a within-patient control feature calculated as the minimal Euclidean distance of actigraph data with DTW. [19] DTW offers a measure of similarity between two temporal sequences that

vary in speed. We hypothesized that warping the actigraph signal would facilitate direct comparison of movements caused by routines occurring at regular intervals (e.g., nursing assessments and nursing care), but not necessarily at identical intervals each time. A higher Euclidean distance suggests a larger difference between the signal in question and the reference. **Figure 2** depicts a DTW example.

We considered two separate within-patient control references: using the non-paretic arm as reference, and using the paretic arm in the first day labeled as non-delirious as reference. For consistently delirious patients without non-delirious days, we used the first day of data instead. Our rationale for using within-patient control references is that a patient would serve as their own best baseline to measure changes in movement, and that assessing movement in both arms could function as a surrogate for whole body movement. For instance, increased movement in both paretic and non-paretic arms would suggest externally mediated whole-body movement (e.g., for nursing care), while increased movement in only the non-paretic arm would be more suggestive of patient-initiated limb movement.

Clinical Features

Finally, we included relevant demographics and stroke-specific clinical features in our delirium prediction models. These variables included age, sex, NIHSS scores, and ICH features including hematoma volume, location, and presence of intraventricular hemorrhage, many of which have been described as risk factors for post-stroke delirium in prior studies. [20 21] In addition, we included day-to-day mechanical ventilation status as a dynamic variable in our models, as this was presumed to affect the amount of patient movement (e.g., from sedation).

Train/Test Split

To account for limited data set size and high variance among enrolled patients' measurements, we used 500 bootstrapping iterations. During each iteration, data were split into non-overlapping training (80% of patients) and test sets (20% of patients), ensuring that no individual patient's observations were included in both the training and test set. Model training and evaluation were separately performed for each of these

500 random train/test splits. Means and standard deviations across the bootstrapping iterations were reported for metrics including accuracy, balanced accuracy, Receive Operating Characteristic Area Under the Curve (ROC-AUC), and F1-score.

Machine Learning Models

A number of alternative machine learning models were trained based on the previously described feature vectors. These included Random Forests, Support Vector Machines (SVM), and XGBoost. We provide the best results obtained by these algorithms, achieved using XGBoost with the following hyperparameters, all of which were tuned via cross-validation on the training set: learning rate 0.03, maximum tree depth 5, minimum child weight 1, subsample fraction 0.8, and column fraction 0.8. Models were trained to detect whether a patient had been delirious at any time during a 24-hour period using different subsets of data sources: 1) clinical data only, 2) clinical and actigraph data using the non-paretic arm as reference, and 3) clinical and actigraph data using the first non-delirious day of the paretic arm as reference.

RESULTS

Patient Characteristics

A total of 40 patients who met eligibility criteria were recruited for this study. To ensure model consistency, we discarded partial actigraph data from the day of enrollment, which also removed the substantial artifact associated with actigraph setup. This led to the exclusion of one patient who was enrolled and discharged prior to recording a full day of actigraph data. As a result, our final cohort comprised 39 patients with a total of 296 days of actigraph monitoring (see **Table 1** for baseline characteristics). Among this cohort, 85% of patients (n=33) had delirium at some point during their hospitalization, including 15 patients who had delirium for the entire duration of monitoring, while 15% of patients (n=6) never had delirium; 71% (n=209) of all monitoring days represented days with delirium.

Table 1. Baseline characteristics and delirium features for patients with intracerebral hemorrhage (ICH) enrolled in this study.

Demographics	
Age, years, mean (SD)	71.4 (12.2)
Male, n (%)	21 (53.9%)
White, n (%)	38 (97.5%)
ICH characteristics	
Admission NIHSS score, median (IQR)	14.5 (6)
ICH score, median (IQR)	2 (1)
ICH volume, cc, mean (SD)	38.4 (24.1)
Intraventricular hemorrhage, n (%)	21 (53.9%)
Location, n (%)[a]	
Lobar	26 (66.7%)
Deep	14 (38.5%)
Mechanically ventilated, n (%)	8 (20.5%)
Ever delirious, n (%)	33 (84.6%)
Always delirious, n (%)	15 (38.5%)
Delirium days, median (IQR)	3 (6)

[a]One patient presented with both lobar and deep ICH

Delirium Detection

Clinical data alone had low accuracy in detecting delirium on a day-to-day basis (**Table 2**). However, the addition of actigraph data yielded a significant improvement in accuracy ($p<0.001$), with the highest recognition performance reaching an accuracy score of 74%. Using the first non-delirious day as DTW reference resulted in better accuracy than using the non-paretic arm as reference.

In a post-hoc analysis, we attempted to further filter activity that was presumed to be externally mediated (e.g., from nursing care) by removing certain actigraph data outliers. We defined these outliers as PIM values from the paretic arm that were 10 or more standard deviations higher than the median paretic arm PIM value. However, this procedure did not improve accuracy (accuracy 0.74, balanced accuracy 0.68, F1 0.65, ROC AUC 0.68).

Table 2. Model performance for same-day delirium detection

Data	Balanced Accuracy	Accuracy	F1	ROC AUC
Clinical Only	0.56 +/- 0.14	0.62 +/- 0.18	0.5 +/- 0.17	0.56 +/- 0.14
Clinical + Minutes at Rest	0.61 +/- 0.1	0.69 +/- 0.1	0.58 +/- 0.11	0.61 +/- 0.1
Clinical + Non-paretic arm DTW	0.57 +/- 0.1	0.63 +/- 0.11	0.53 +/- 0.11	0.57 +/- 0.1
Clinical + Non-paretic arm DTW + Minutes at Rest	0.63 +/- 0.1	0.71 +/- 0.1	0.61 +/- 0.1	0.63 +/- 0.1
Clinical + Reference day DTW	0.66 +/- 0.12	0.68 +/- 0.12	0.61 +/- 0.12	0.66 +/- 0.12
Clinical + Reference day DTW + Minutes at Rest	0.65 +/- 0.11	0.71 +/- 0.11	0.62 +/- 0.11	0.65 +/- 0.11
Clinical + Non-paretic arm DTW+ Reference day DTW + Minutes at Rest	**0.68 +/- 0.11**	**0.74 +/- 0.1**	**0.65 +/- 0.1**	**0.68 +/- 0.11**

We were also interested in determining which individual features most contributed to model accuracy. This was done by analyzing each feature and the associated information gain it produced, averaged across each run of cross validation. This analysis suggests that night-time DTW actigraph data was an especially important feature, with minutes at rest during afternoon/evening and night-time periods also meaningfully contributing to prediction power (**Figure 3**).

DISCUSSION

We found that detecting delirium using actigraphy and machine learning-based analysis was feasible and provided valuable information that significantly improved upon the accuracy of clinical data alone. Given the increasingly recognized impact of delirium on patients with stroke, with consequences ranging from withdrawal of life-sustaining treatment to decreased rates of rehabilitation utilization [22], the early and accurate recognition of delirium in these patients is paramount. Because existing delirium screening tools

253 are unreliable in the setting of severe neurologic deficits, novel tools are needed in the clinical setting, and
254 unconventional methods such as actigraph monitoring may be a promising way to address this gap [21].
255 Our study may also have relevance for non-neurologic patients. Although validated delirium
256 screening tools exist in general critical care and hospitalized populations, they can be resource intensive
257 and demand that nurses and other providers be appropriately trained in their use. Automated methods of
258 delirium monitoring could therefore help offset the burden of a strained healthcare staff, especially methods
259 that help monitor for fluctuations in activity and arousal that are characteristic of delirium. Additionally,
260 adjunctive methods of delirium monitoring may also improve overall detection rates in other challenging
261 patient populations, such as those with dementia.
262 Monitoring patients at night may be especially important in detecting delirium, a concept that is
263 underscored by the significant contribution of night-time actigraph data to the accuracy of our delirium
264 prediction models. Although the high incidence of nocturnal symptoms and sleep-wake disturbances
265 associated with delirium is well known, these symptoms may often go undetected until they reach a point
266 where they become obvious (i.e., severe agitation). It is possible that nocturnal symptoms could be detected
267 sooner via frequent clinical assessments, which are the basis of neurological monitoring in neurocritically ill
268 patients, but overly frequent neurological exams at night may also contribute to delirium by leading to sleep
269 fragmentation and overstimulation. On the other hand, the relatively unobtrusive nature of wearable sensors
270 may mitigate these concerns, while providing valuable information that could help detect delirium during
271 especially high-risk time periods.
272 Our study is notable for its innovative techniques, including the use of machine-learning to analyze
273 actigraph data and the use of within-patient controls via actigraphs worn on both paretic and non-paretic
274 limbs. However, the study does have several limitations. First, actigraph data are limited by noise and
275 artifacts caused by external movements such as nurse or provider-initiated movements (e.g. during
276 repositioning or clinical examination). Although we excluded artifactual data associated with actigraph
277 initiation from the day of admission and incorporated measurements from both paretic and non-paretic limbs
278 to mitigate potential confounding, we could not definitively filter these externally-mediated movements

further. However, their influence may have been relatively modest, as a post-hoc analysis using an outlier filtering method did not result in a meaningful difference in accuracy. Second, because we assessed delirium status only once per day, we may have missed shorter periods of delirium or non-delirium that would have allowed for closer correlation with actigraph data. Finally, our sample size was relatively small and had a class imbalance in favor of delirium-positive days, as many patients were rated as either always (or almost always) delirious or never delirious. Because deep learning requires that voluminous amounts of data be available for optimal results, future studies are needed to analyze data from larger cohorts of patients and further evaluate machine learning-based methods for detecting and predicting delirium.

CONCLUSIONS

We found that actigraphy in conjunction with machine learning models improves clinical detection of delirium in patients with stroke, thus paving the way to make actigraph-assisted predictions clinically actionable.

DECLARATIONS

Ethical Approval

This study was approved by Lifespan Institutional Review Board 2 (ID #1126240), and standard written informed consent procedures were undertaken with patients and/or designated surrogates who agreed to participate in the study. A STROBE checklist is provided in **Supplementary File 1**.

Availability of Data and Materials

All relevant data are presented within the article and its supporting tables and figures. Additional information can be obtained upon reasonable request to the corresponding author.

Funding

This study was supported by the Rhode Island Foundation, Brown University's Office of the Vice President for Research (OVPR) via a Big Data Collaborative Seed Award, and the Global Individual Fellowship Marie Skłodowska-Curie Action H2020-MSCA-IF-2020 MAESTRO (Grant Number 101027770).

Competing Interests

None declared.

References

1. Salluh JI, Wang H, Schneider EB, et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. BMJ 2015;**350**.
2. Reznik ME, Drake J, Margolis SA, et al. Deconstructing poststroke delirium in a prospective cohort of patients with intracerebral hemorrhage. Crit. Care. Med. 2020;**48**(1):111-18.
3. van Eijk MM, van den Boogaard M, van Marum RJ, et al. Routine use of the confusion assessment method for the intensive care unit: a multicenter study. Am. J. Respir. Crit. Care Med. 2011;**184**(3):340-44.
4. Morandi A, Davis D, Bellelli G, et al. The diagnosis of delirium superimposed on dementia: an emerging challenge. J. Am. Med. Dir. Assoc. 2017;**18**(1):12-18.
5. Reznik ME, Daiello LA, Thompson BB, et al. Fluctuations of consciousness after stroke: associations with the confusion assessment method for the intensive care unit (CAM-ICU) and potential undetected delirium. J. Crit. Care 2020;**56**:58-62.
6. Reznik ME, Mahta A, Schmidt JM, et al. Duration of agitation, fluctuations of consciousness, and associations with outcome in patients with subarachnoid hemorrhage. Neurocrit. Care 2018;**29**(1):33-39.
7. Ely EW, Truman B, Shintani A, et al. Monitoring sedation status over time in ICU patients: reliability and validity of the Richmond Agitation-Sedation Scale (RASS). JAMA 2003;**289**(22):2983-91.
8. Meagher DJ, Moran M, Raju B, et al. Motor symptoms in 100 patients with delirium versus control subjects: comparison of subtyping methods. Psychosomatics 2008;**49**(4):300-08.
9. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med. Rev. 2011;**15**(4):259-67.
10. Schwab KE, To AQ, Chang J, et al. Actigraphy to measure physical activity in the intensive care unit: a systematic review. J. Intensive Care Med. 2020;**35**(11):1323-31.
11. Schwab KE, Ronish B, Needham DM, To AQ, Martin JL, Kamdar BB. Actigraphy to evaluate sleep in the intensive care unit. A systematic review. Ann. Am. Thorac. Soc. 2018;**15**(9):1075-82.
12. Osse RJ, Tulen JH, Hengeveld MW, Bogers AJ. Screening methods for delirium: early diagnosis by means of objective quantification of motor activity patterns using wrist-actigraphy. Interact. Cardiovasc. Thorac. Surg. 2009;**8**(3):344-48.
13. Eeles E, Tahir TA, Johansen A, Bisson JI, Hubbard RE. Comparison of clinical assessment and actigraphy in the characterization of delirium. J. Psychosom. Res. 2009;**67**(1):103-04.
14. Maybrier HR, King CR, Crawford AE, et al. Early postoperative actigraphy poorly predicts hypoactive delirium. J. Clin. Sleep Med. 2019;**15**(1):79-87.

15. *Diagnostic and statistical manual of mental disorders : DSM-5*: Fifth edition. Arlington, VA : American Psychiatric Association, [2013], 2013.
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics 2009;**42**(2):377-81 doi: https://doi.org/10.1016/j.jbi.2008.08.010.
17. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics 2019;**95**:103208 doi: https://doi.org/10.1016/j.jbi.2019.103208.
18. Meagher DJ, Moran M, Raju B, et al. Phenomenology of delirium: assessment of 100 adult cases using standardised measures. Br. J. Psychiatry 2007;**190**(2):135-41.
19. Using dynamic time warping to find patterns in time series. KDD workshop; 1994. Seattle, WA, USA:.
20. Oldenbeuving AW, de Kort PL, van Eck van der Sluijs JF, Kappelle LJ, Roks G. An early prediction of delirium in the acute phase after stroke. Journal of neurology, neurosurgery, and psychiatry 2014;**85**(4):431-4 doi: 10.1136/jnnp-2013-304920 [published Online First: 2013/06/08].
21. Reznik ME, Margolis SA, Mahta A, et al. Impact of Delirium on Outcomes After Intracerebral Hemorrhage. Stroke 2021:Strokeaha120034023 doi: 10.1161/strokeaha.120.034023 [published Online First: 2021/10/06].
22. Reznik ME, Moody S, Murray K, et al. The impact of delirium on withdrawal of life-sustaining treatment after intracerebral hemorrhage. Neurology 2020;**95**(20):e2727-e35 doi: 10.1212/wnl.0000000000010738 [published Online First: 2020/09/12].

FIGURE CAPTIONS

Figure 1: **Feature extraction and classification process**. Elements in parentheses were initially explored but ultimately discarded because they did not contribute to higher prediction accuracy

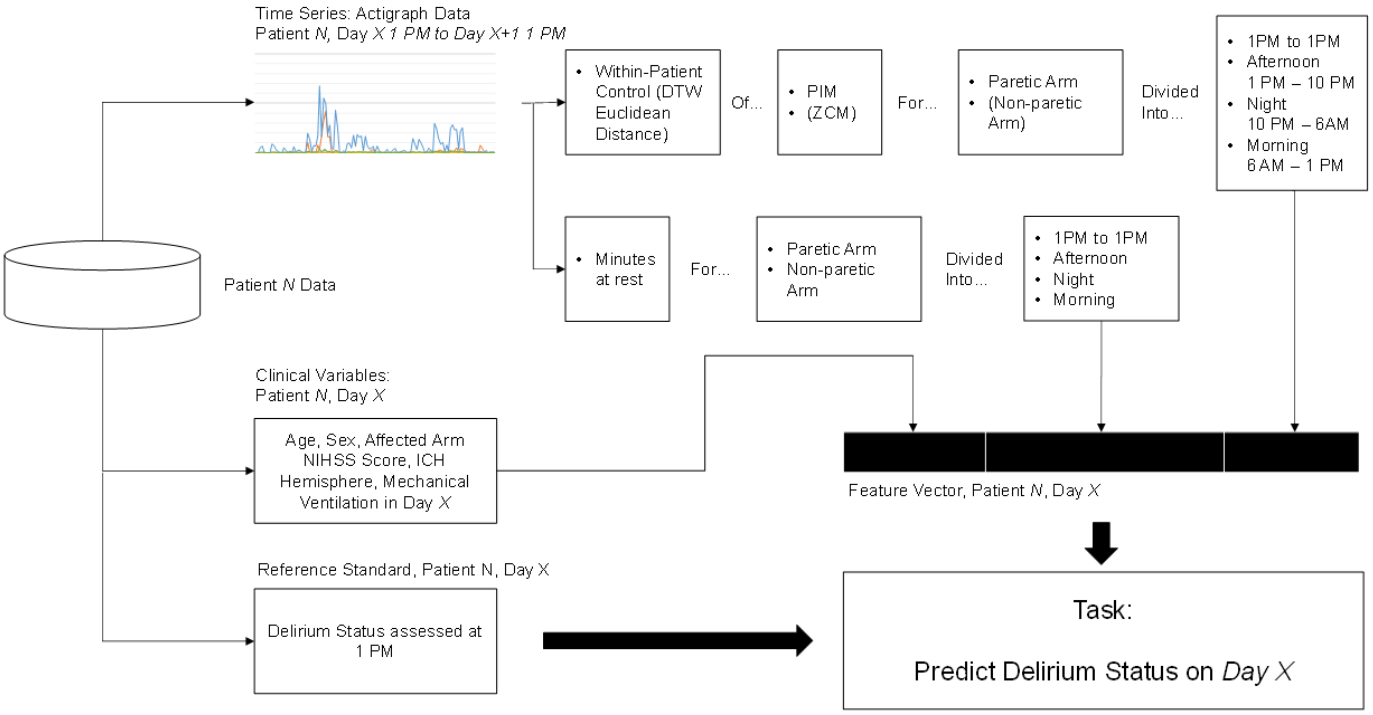
Figure 2: **Dynamic Time Warping (DTW) example**. Original (left) and minimal Euclidean distance warped (right) signals of two hours of actigraph data

Figure 3: **Individual feature importance for the final model according to XGBoost**

APPENDICES

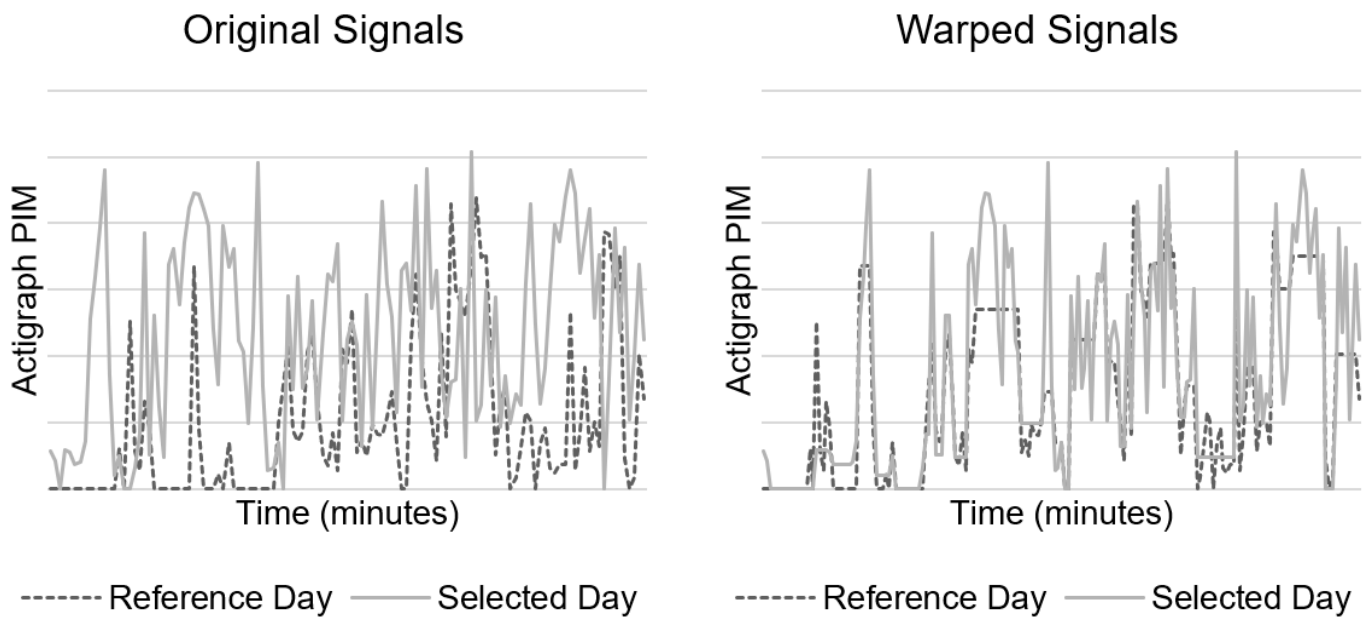
Supplementary File 1: STROBE Statement

377 *Figure 1*



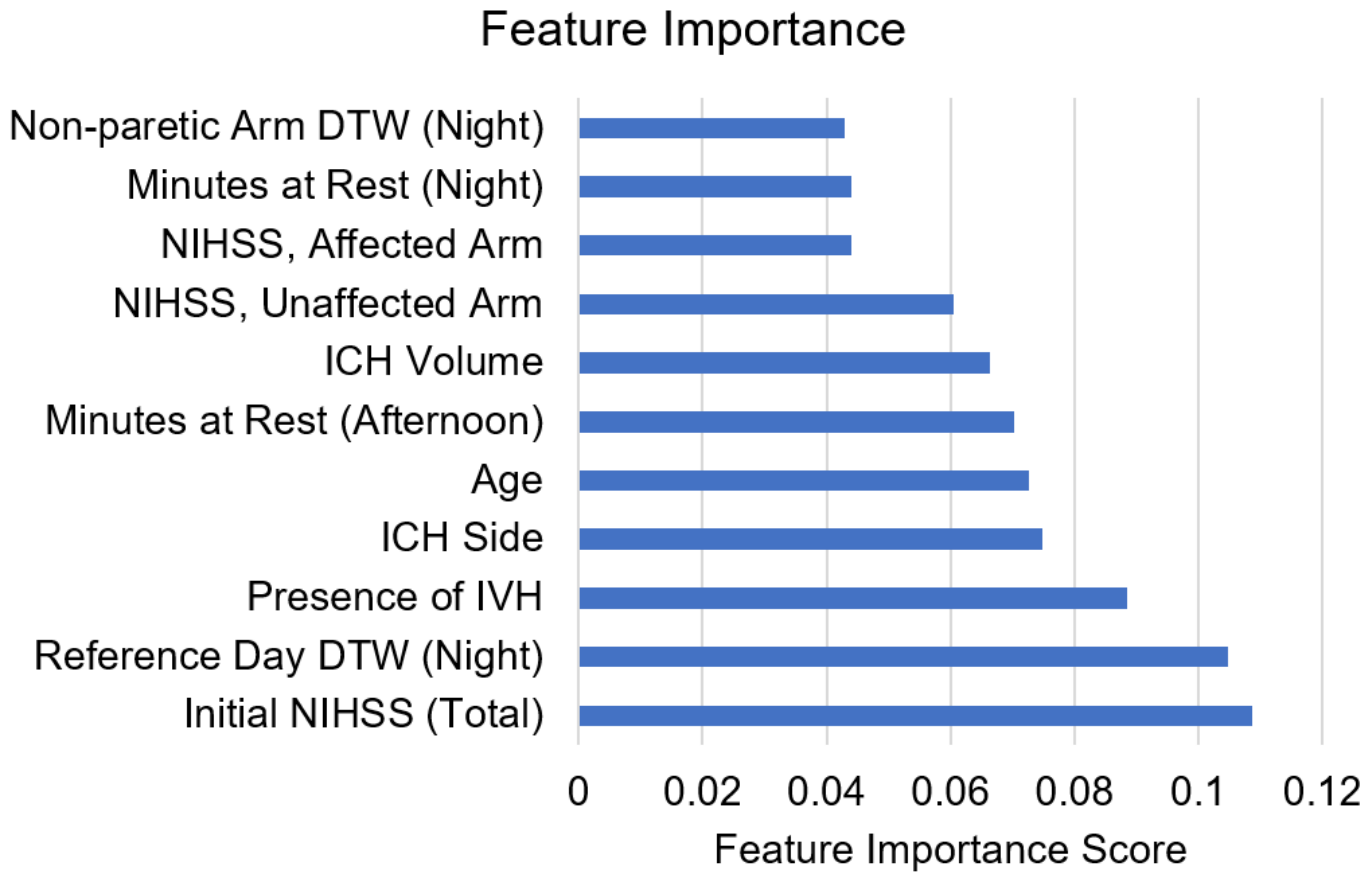
378

379 *Figure 2*



380

Figure 3


Feature Importance
Non-paretic Arm DTW (Night)
Minutes at Rest (Night)
NIHSS, Affected Arm
NIHSS, Unaffected Arm
ICH Volume
Minutes at Rest (Afternoon)
Age
ICH Side
Presence of IVH
Reference Day DTW (Night)
Initial NIHSS (Total)
0
0.02
0.04
0.06
0.08
0.1
0.12
Feature Importance Score

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-8
Bias	9	Describe any efforts to address potential sources of bias	6, 9
Study size	10	Explain how the study size was arrived at	10
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	9
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	10
		(*d*) If applicable, explain how loss to follow-up was addressed	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	10
		(b) Give reasons for non-participation at each stage	10
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	Table 1
		(c) Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	Report numbers of outcome events or summary measures over time	Table 1

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	1

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.