Journal of the American Heart Association

ORIGINAL RESEARCH

Derivation and Validation of an Algorithm to Detect Stroke Using Arm Accelerometry Data

Steven R. Messé, MD; Scott E. Kasner, MD, MSCE; Brett L. Cucchiara, MD; Michael L. McGarvey, MD; Stephanie Cummings, BS; Michael A. Acker, MD; Nimesh Desai, MD; Pavan Atluri, MD; Grace J. Wang, MD; Benjamin M. Jackson, MD; James Weimer, MS, PhD

BACKGROUND: Early diagnosis is essential for effective stroke therapy. Strokes in hospitalized patients are associated with worse outcomes compared with strokes in the community. We derived and validated an algorithm to identify strokes by monitoring upper limb movements in hospitalized patients.

METHODS AND RESULTS: A prospective case–control study in hospitalized patients evaluated bilateral arm accelerometry from patients with acute stroke with lateralized weakness and controls without stroke. We derived a stroke classifier algorithm from 123 controls and 77 acute stroke cases and then validated the performance in a separate cohort of 167 controls and 33 acute strokes, measuring false alarm rates in nonstroke controls and time to detection in stroke cases. Faster detection time was associated with more false alarms. With a median false alarm rate among nonstroke controls of 3.6 (interquartile range [IQR], 2.1–5.0) alarms per patient per day, the median time to detection was 15.0 (IQR, 8.0–73.5) minutes. A median false alarm rate of 1.1 (IQR. 0–2.2) per patient per day was associated with a median time to stroke detection of 29.0 (IQR, 11.0–58.0) minutes. There were no differences in algorithm performance for subgroups dichotomized by age, sex, race, handedness, nondominant hemisphere involvement, intensive care unit versus ward, or daytime versus nighttime.

CONCLUSIONS: Arm movement data can be used to detect asymmetry indicative of stroke in hospitalized patients with a low false alarm rate. Additional studies are needed to demonstrate clinical usefulness.

Key Words: automation ■ delayed diagnosis ■ in-hospital stroke ■ stroke detection

Proven stroke treatments, including intravenous thrombolysis and mechanical thrombectomy, are highly time dependent. Eligibility for intervention and the probability of good outcome if treated decline continuously as time from onset of symptoms increases.[1–3] Thus, rapid detection of the onset of stroke symptoms is of paramount importance.[4–6]

Of the 800 000 strokes that occur annually in the United States, 5% to 17% develop in patients who are already hospitalized, the majority in patients who recently underwent an intervention or procedure.[7–9] Compared with strokes that occur in the community, in-hospital stroke is associated with delayed detection and assessment, fewer interventions, and worse outcomes.[7–11] Thus, these complications lead to markedly increased cost, length of stay, morbidity, mortality, and medicolegal liability for hospitals and caregivers.[7,8,10–14]

Upper extremity weakness is one of the most common findings in acute stroke.[15] As a result, asymmetric arm strength is used in all screening tools for stroke.[16,17] In addition, neglect is a frequent stroke symptom that also leads to a tendency to move the arm less on the affected side.[18,19] We hypothesized that continuous monitoring for asymmetric arm movement would be

Correspondence to: Steven R. Messé, MD, Hospital of The University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19014.
Email: steven.messe@pennmedicine.upenn.edu

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.028819

For Sources of Funding and Disclosures, see page 9.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Stroke in hospitalized patients often has delayed detection and is associated with poor outcome and high cost.
- Asymmetric arm weakness is one of the most common symptoms of stroke.
- We performed a prospective case control study of bilateral arm accelerometry monitoring in hospitalized patients with and without stroke to derive and validate an algorithm to detect asymmetric weakness.

What Are the Clinical Implications?

- The algorithm provides a median time to detection of stroke of <30 minutes, while maintaining a median false alarm rate of about 1 per patient per day.
- This algorithm may allow for continuous monitoring of patients to detect stroke with lateralized weakness.
- If automated monitoring can detect stroke faster than usual care, this could lead to more and earlier stroke interventions and improved outcomes.

Nonstandard Abbreviations and Acronyms

NIHSS	National Institutes of Health Stroke Scale

a sensitive and practical approach to identify stroke onset. An automated system to identify stroke could facilitate more and earlier acute stroke treatments and improve outcomes. The overarching goal of this project was to develop and validate an alerting algorithm incorporating features from upper extremity accelerometry data to rapidly identify stroke in hospitalized patients.

METHODS

We performed a 2-part prospective case–control study of upper extremity movements of patients admitted to the Hospital of the University of Pennsylvania to first derive and then validate a stroke detection algorithm. The derivation and validation studies were performed by the same research team and approved by the institutional review board at the Hospital of the University of Pennsylvania. All patients or their legally authorized representatives provided informed consent before enrollment. Data may be made available upon reasonable request to the authors.

Subjects

All subjects were recruited from the inpatient setting at the Hospital of the University of Pennsylvania. Because of the relative rarity and frequently uncertain timing of stroke, it is challenging to accumulate a large volume of accelerometry data from individual patients before and after they have a stroke. Thus, a case–control study design was used to derive and then validate the algorithm, which allows for an estimation of the time to detection of asymmetric limb movement in patients with stroke and false alarm rates in patients without stroke. Controls were neurologically normal, with no asymmetric arm weakness, no history of stroke, and no above the wrist amputation (to facilitate wearing the accelerometers). For the derivation cohort, controls included patients with transient ischemic attack without acute infarct on magnetic resonance imaging, patients undergoing workup of transient spells of uncertain cause with normal magnetic resonance imaging, and patients who recently underwent cardiothoracic surgery or vascular surgery without evidence of neurologic complications. For the validation cohort, only patients who underwent recent cardiothoracic or vascular surgery and had no evidence of stroke were included. Cases for both the derivation and validation cohort consisted of patients admitted with acute ischemic or hemorrhagic stroke with a National Institutes of Health Stroke Scale (NIHSS) assessment with at least 1 point for upper extremity weakness on item 5a or 5b of the NIHSS, weaker on the side affected by the stroke. Before the initiation of monitoring, subjects underwent a neurologic evaluation including the NIHSS and a strength assessment, using the Medical Research Council scale to rate the deltoid, biceps, triceps, wrist extension, wrist flexion, intrinsic finger, hip flexor, quadriceps, hamstrings, ankle extension, and ankle flexion ranging from 0 (no movement) to 5 (full strength) on each side.

Monitoring

The subjects had wrist straps incorporating accelerometers placed on both arms and were asked to keep them on for as long as the battery would last, which varied based on the device used in the 2 cohorts. For the algorithm derivation cohort, we used a commercially available battery-powered Bluetooth-enabled accelerometer/gyroscope, the Wit Motion (Shenzhen City, China) BWT901CL Bluetooth output 9-axis accelerometer gyroscope, synced with an Android tablet to stream the data to a cloud-based server (Heroku, San Francisco, CA). These accelerometry devices had an expected battery life of 2 to 3 hours. To capture more data and allow for comparisons of performance between daytime and nighttime, we required a longer lasting accelerometry device. Thus, for the validation cohort, we used the commercially available Samsung Galaxy Watch Active to collect accelerometry data. An app

collected accelerometry data (Raproto, Philadelphia, PA), which was transmitted via WiFi using a data transfer protocol called message queueing telemetry transport quality of service 1, which ensures that every data point is received and then stored on a cloud-based platform (Thingsboard, New York, NY).[20] The expected battery life of this device was 18 to 24 hours. For both phases of the study, patients and clinical staff were told that the straps could be removed at any time if they were uncomfortable, interfered with clinical treatment, or for any other reason they chose. To ensure conditions were representative of real-world practice, no instructions to limit therapy or passive range motion of the affected limb were given while the patient was being monitored. The neurologic assessments were repeated after monitoring was complete to confirm that there were no changes in neurologic status.

Algorithm Derivation

Although patients with stroke will frequently have weakness on one side leading to reduced movement, a neurologically normal person without weakness will, at times, also demonstrate asymmetric arm movements. The goal of this work, then, is to create an algorithm that can discriminate normal movement from pathological movement patterns and will alarm when there is asymmetric arm movement indicative of acute stroke while minimizing alarms in patients who are neurologically intact. We derived the algorithm using a parameter-invariant method designed to maximize diagnostic performance and generalizability.[21–25] This approach has been previously used to develop multiple medical classifier algorithms requiring high sensitivity and specificity along with stable performance across patients without outliers.[22] The parameter-invariant method uses a statistical first-principle approach to derive algorithms that are invariant to patient-specific parameters (eg, being left or right handed, awake/asleep, restrained/free to move) as well as system anomalies common in accelerometer-based systems (eg, accelerometer bias/drift or device orientation). As a result, the algorithm achieves stable performance across the population without requiring individual tuning.

The algorithm derivation methodology is available in Data S1. Briefly, using the derivation cohort accelerometry data, we identified features invariant to patient-specific parameters and then trained a structured classification tree, combining the features to maximize stability and accuracy for detection of asymmetric movement patterns seen in patients with stroke. Using multiple concurrent threshold tests of varying durations is a common technique in detection theory to balance the trade-off between accuracy and time to detection.[26] Threshold tests, with shorter monitoring durations, provide faster time to detection, whereas longer monitoring durations have increased accuracy. The algorithm simultaneously uses multiple windows of increasing duration of preceding data (when available) and alarms if any window detects the possible presence of a stroke.

In a real-world deployment, an alarm that leads to identification of a stroke triggers a clinical intervention that would include removing the device. Thus, if movement data continue to accrue after an alarm, the algorithm assumes that the prior alarm was a false positive, and no further alarms are generated for 1 hour to allow the monitoring windows to accumulate new data. Every subsequent alarm within 4 hours of the previous alarm extends the alarm pause by an additional hour, up to a maximum of 4 hours. If there is no generated alarm within 8 hours, the alarm pause duration is reset to 1 hour. We note that the proposed strategy results in a maximum false alarm rate of 8 alarms in the first 24 hours, followed by 6 alarms per day from then on. An open source implementation of the described algorithm is available for academic and noncommercial use (https://jamesweimer.net/StrokeDetectAI/).

Validation

The final candidate algorithm was validated using an independent and blinded test data set that was collected separately from the data set used for algorithm derivation using a different, longer-lasting accelerometer as noted above.[27] For this preplanned analysis, the algorithm evaluated individual patient data and was executed every 15 minutes. The performance of the population was then calculated using medians and interquartile ranges of the average performance of the individual subjects. As a result of using message queueing telemetry transport quality of service 1 to transfer data to the cloud, the accelerometry data set had no missing data. However, it remains possible that data will be missing, at least temporarily, in a real-world implementation of a clinical stroke monitor, in which case the algorithm is designed to handle data in the following manner: Data are timestamped by the accelerometry devices and, when evaluated by the algorithm, missing data are treated as missing and not imputed. Examples of how missing data may occur in a real-world implementation include a weak or erratic WiFi network, which would lead to temporary delays in communication of data, although it would eventually be collected and analyzed when the WiFi signal allows. Alternatively, if a wrist strap device runs out of power, stroke detection will not be possible during that time because accelerometry data cannot be collected.

Statistical Analysis

For control subjects without stroke, we evaluated the algorithm performance in terms of false alarms per

patient per day, defined as the number of alarms divided by the monitoring time in days. We report the median false alarms per day by taking the median of the false alarms per day over all control subjects. For each case subject with stroke, we evaluated the algorithm performance in terms of detection rate as time from initiation of monitoring increased. Start times of monitoring were in 15-minute increments throughout the entire duration of monitoring for each patient. Because data for subjects who transition from neurologically intact to having a stroke during monitoring were not available in our study, we used a conservative evaluation for detection rate versus time to detection commonly used in the quickest detection literature.[28] As the time from initiation of monitoring increases, the aggregate test includes only windows of shorter duration, and the detection rate is calculated based on the percentage of aggregate tests that identified stroke.

As noted above, a false alarm will lead to a transient pause in alarm generation. To account for how this feature impacts the time to detection in stroke cases, we calculated the duration that the alarm was paused per day based on the algorithm performance in the control subjects. The median and interquartile range (IQR) of the delay because of pauses was then added to time to detection for the stroke cases. For example, if the median false alarm rate in controls was 1 per day, the alarm would be paused for 1 hour out of 24 hours. Assuming that a stroke can occur at any time during the 24-hour period, there will be 23 hours with no additional delay and 1 hour when the alarm is paused (with a median delay of 30 minutes); therefore, (23/24)×0 minutes+(1/24)*30 minutes=1.25 minutes additional expected delay per day.[27] We also evaluated the correlation between time to detection and false alarm rates across a range of operating points for the algorithm, using a Spearman ρ test. Finally, we evaluated whether patient-specific factors would lead to variations in performance of the algorithm by comparing the median time to detection and false alarm rates by age, sex, race, handedness, nondominant hemisphere involvement, hospital location (intensive care unit versus ward), and whether monitoring occurred during nighttime or daytime using Wilcoxon rank sum testing.

Sample Size

The enrollment of 200 subjects in the derivation cohort was based on prior work on parameter invariant algorithms, which suggested we would need at least 500 hours of bilateral arm movement data to derive the algorithm. The validation cohort sample size was not based on a sample size calculation and instead was determined to be the same number of patients as the derivation cohort but using accelerometers that lasted ≈8 times longer, increasing precision of the estimates of performance.

RESULTS

From May 8, 2018 through November 23, 2021, we enrolled 405 patients including 200 in the derivation cohort and 205 in the validation cohort. Accelerometry data were not available for 5 control subjects in the validation cohort because of technical difficulties, and they were excluded from the analysis. The algorithm derivation cohort included 77 patients with acute stroke and lateralizing arm weakness and 123 neurologically intact control subjects. In total, 540 hours of bilateral arm accelerometry data were acquired during this phase, with a mean 2.7 hours per subject. The algorithm validation cohort included 33 patients with acute stroke and 167 controls, totaling 4169 hours of bilateral arm accelerometry data with a mean of 20.8 hours per subject. Table 1 presents the clinical and demographic characteristics of the controls for the derivation and validation cohorts,

Table 1. Clinical and Demographic Characteristics of Neurologically Normal Controls in the Algorithm Derivation and Validation Cohorts

Characteristic	Total, n=290	Derivation cohort, n=123	Validation cohort, n=167	*P* value
Age, y, mean±SD	64±15	62±18	65±12	0.06
Female sex	109 (38%)	55 (45%)	54 (32%)	0.03
Non-White race*	42 (14%)	23 (19%)	19 (11%)	0.08
Left handed	39 (13%)	12 (10%)	27 (16%)	0.13
Admission reason				<0.001
TIA	1 (0.3%)	1 (1%)	0	
Epilepsy monitoring	24 (8%)	24 (19%)	0	
Surgery	265 (91%)	98 (80%)	167 (100%)	
Monitoring duration, min, median (IQR)	972 (174–1340)	171 (135–190)	1320 (1216–1404)	<0.001

Continuous variables are presented as median (IQR) unless otherwise specified. IQR indicates interquartile range; and TIA, transient ischemic attack.
*Black, Native American, Asian, or other.

and Table 2 provides these data for the stroke cases. Among the nonstroke controls, subjects in the validation cohort were less likely to be women and were more likely to have recently had surgery. For stroke cases, the validation cohort was similar to the derivation cohort, with the exception of a greater difference in arm strength between the affected and unaffected side, as measured by the sum of the Medical Research Council upper extremity motor scores, although the differential in the NIHSS upper extremity motor score was similar.

Within the validation cohort, stroke cases had similar age (mean 68 versus 65 years, P=0.25), percentage of women (45% versus 32%, P=0.15), and percentage who were right handed (94% versus 84%, P=0.13) compared with the controls, but patients with stroke were more often non-White, self-described as Black, Native American, Asian, or other (45% versus 11%, P<0.001) and were more often in an intensive care unit or step-down unit (85% versus 34%, P<0.001). Overall, stroke cases in the validation cohort were predominantly ischemic (73%) and moderately severe (NIHSS median 14; IQR, 9–18). For both the algorithm derivation and validation cohorts, the wrist straps were well tolerated. None of the patients in the derivation cohort and 2 patients in the validation cohort removed the devices and prematurely terminated the study (after 1 hour and 22 hours of monitoring, respectively). Nurses reported no issues with the straps interfering with clinical care. There were no changes in patient upper extremity strength or presence of neglect comparing the examinations at baseline and study completion.

Algorithm Performance

Figure 1 displays the median and IQR for the percentage of stroke cases that alarm as monitoring time increases using 2 different alarm thresholds. For each, the percentage of patients with stroke who were correctly identified as having a stroke rose as the duration of monitoring increased. With a median false alarm rate among nonstroke controls of 1.1 alarms per patient per day (IQR, 0–2.2 alarms per patient per day), the median time to alarm in stroke cases was 29 minutes (IQR, 11–58 minutes). At 60 minutes, the algorithm is expected to detect 76% of strokes. Among the nonstroke controls, the median time to first alarm was 12.7 hours. With a median false alarm rate of 3.6 alarms per patient per day (IQR, 2.1–5.0 false alarms), the median time to detection in stroke cases was 15 minutes (IQR, 8–74 minutes). At this setting, the algorithm is expected to detect 91% of strokes at 60 minutes. Figure 2 provides the performance of the algorithm at 5 different operating points, demonstrating that as false alarm rates increase, the times to detection decrease (Figure 2; Spearman ρ, −1.0; P<0.001). Importantly, the algorithm was unaffected by patient-specific factors that could theoretically lead to variable performance. Specifically, using the lower sensitivity threshold, there was no significant difference in false alarm rates (median 1.1 versus 1.0 alarms per day, P=0.22) or time to detection

Table 2. Clinical and Demographic Characteristics of Stroke Cases in the Derivation and Validation Cohorts

Characteristic	Total, n=110	Derivation cohort, n=77	Validation cohort, n=33	*P* value
Age, y, mean±SD	68±16	68±15	68±17	0.95
Female sex	53 (48%)	38 (49%)	15 (45%)	0.71
Non-White race*	51 (46%)	36 (47%)	15 (45%)	0.90
Left handed	11 (10%)	9 (12%)	2 (6%)	0.33
Stroke type				0.10
Intracerebral hemorrhage	20 (18%)	11 (14%)	9 (28%)	
Ischemic stroke	90 (82%)	66 (86%)	24 (73%)	
Nondominant hemispheric stroke	40 (36%)	27 (35%)	13 (39%)	0.67
Total NIHSS score at time of monitoring	13 (8–18)	12 (7–16)	14 (9–18)	0.16
Difference in NIHSS upper extremity motor score between affected and unaffected side	4 (3–4)	4 (2–4)	4 (3–4)	0.22
Difference in sum of upper extremity strength scores between affected and unaffected side**	24 (18–30)	24 (16–30)	30 (22–30)	0.01
Weakness from stroke on left side	66 (60%)	47 (61%)	19 (58%)	0.73
Neglect present	54 (49%)	36 (47%)	18 (55%)	0.45
Monitoring duration, min	190 (168–1102)	178 (150–192)	1299 (1235–1408)	<0.001

Continuous variables are presented as median (interquartile range) unless otherwise specified. NIHSS indicates National Institutes of Health Stroke Scale.
*Black, Native American, Asian, or other.
**Upper extremity muscle groups assessed with the Medical Research Council muscle strength score (ranging from 0–5) included deltoid, biceps, triceps, wrist extension, wrist flexion, and intrinsic finger strength with full strength in all 6 muscles tested scoring 30.

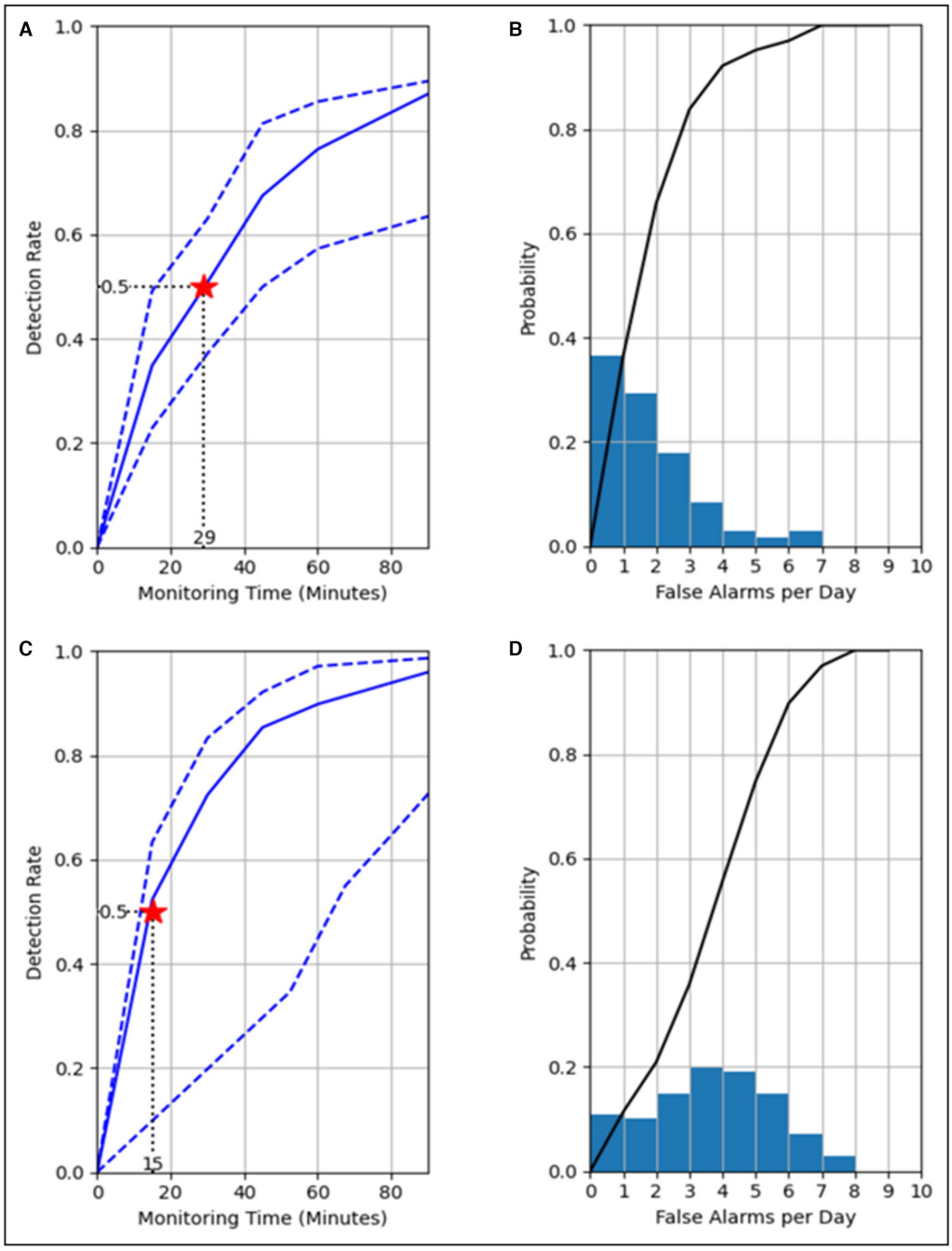

Figure 1. Stroke detection rate over time and false alarm rates per day.
A, Median (solid line) and interquartile range (dashed lines) of the percentage of patients with stroke alarming as the duration of monitoring increases. B, The distribution of false alarms per patient per day in non-stroke controls. The black line represents the cumulative percentage of patients. The time to detection is faster with a lower alarm threshold as shown in (C) with the median (solid line) and interquartile range (dashed lines) providing the percentage of patients with stroke alarming as the duration of monitoring increases. The lower alarm threshold demonstrates more false alarms in (D) with the black line displaying the cumulative percentage of patients.

(median 29 versus 28 minutes, *P*=0.83) comparing right-handed versus left-handed patients. There were similarly no differences in time to detection (31 versus 27 minutes, *P*=0.83) or false alarm rate (median 1.2 versus 1.1 alarms per day, *P*=0.08) comparing patients below or above the median age (68 years), in time to detection (25 versus 31 minutes, *P*=0.51) or false alarm rate (1.1 versus 1.1 alarms per day, *P*=0.91) comparing men to women, in time to detection (29 versus 29 minutes, *P*-1.0) or false alarm rate (1.1 versus 1.1 alarms per

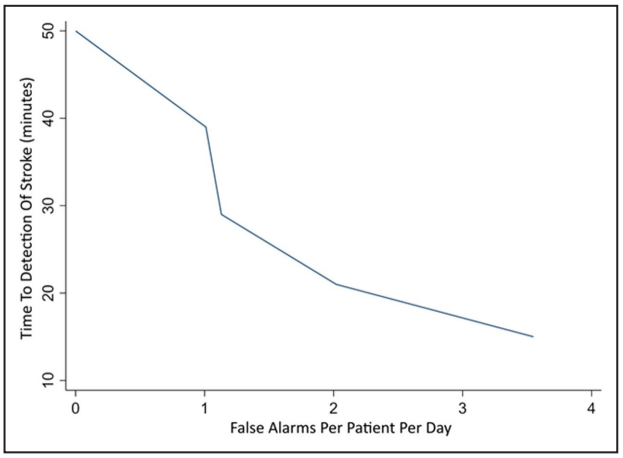

Figure 2. The correlation between false alarm rates in nonstroke controls and speed of detection of stroke in cases across 5 different operating points.

day, *P*=0.72) comparing non-White to White subjects, in time to detection (32 versus 17 minutes, *P*=0.51) or false alarm rate (1.1 versus 1.1 alarms per day, *P*=0.75) comparing patients in the intensive care unit to those on a ward, in false alarm rate (1.1 versus 1.2 alarms per day, *P*=0.73) comparing control patients below the median number of days since their procedure (2 days) compared with those above the median, and in time to detection (median 30 versus 29 minutes, *P*=0.83) if the stroke involved the dominant or nondominant hemisphere. Most importantly, there were no differences in time to detection (median 28 versus 28 minutes, *P*=0.83) or false alarms detected (median 0 versus 0 alarms, *P*=0.80) comparing daytime versus nighttime. These results were similar when evaluated using the threshold with increased sensitivity (data not shown).

DISCUSSION

This study demonstrates that arm accelerometry data can be used to discriminate patients with weakness caused by acute stroke from neurologically intact hospitalized patients. The algorithm's diagnostic performance achieves a high sensitivity and specificity, such that it could provide a clinically useful monitor to rapidly detect the onset of stroke while maintaining a low false alarm rate. The alarm threshold is modifiable, and a lower threshold demonstrated greater sensitivity and faster time to detection, with a concomitant higher false alarm rate. Importantly, the estimate of the time to detection is conservative and may be faster in clinical use. In the analyses of time to detection, we only included data from patients with stroke. In practice, patients will convert from nonstroke to stroke in the midst of an evaluation window, which may still trigger an alarm, yielding faster times from onset to detection than we report. In addition, stroke cases were cared for in real-world routine clinical practice while they were being monitored, and there were times when the care team or family members would move the patients' weak arms. These time periods were not censored for the validation analysis. In clinical use, we expect that the algorithm will detect over half of strokes within 30 minutes of onset, while maintaining <2 false alarms per day for the vast majority of patients. Of greatest importance, we saw no significant variability in algorithm performance based on handedness, nondominant hemispheric involvement, or whether we were monitoring during daytime versus nighttime. This latter finding suggests that we can detect stroke equally during sleep or wakefulness, which is a critical feature of a useful stroke monitor.

In-hospital stroke is a major public health issue that accounts for a meaningful portion of all strokes and is associated with delayed assessment and treatment, poor outcome, and dramatically increased cost and length of stay.[7–12] Importantly, periprocedural stroke accounts for the majority of cases in most series, and stroke rates for common procedures, such as aortic valve surgery, are much higher than commonly reported when prospective assessments are performed.[7–9,29–31] Given that the algorithm detects asymmetry and is not based on change in movement patterns from a baseline period, it is particularly well suited to detect stroke in the perioperative setting, where patients may awaken from anesthesia with weakness. Prior studies of in-hospital stroke have reported times from last known normal to symptom detection ranging from ≈2 to 10 hours.[29–31] Although proven stroke treatments may have robust benefit, the likelihood of being able to receive these treatments and the response to treatment steadily decline over time.[1–3] Thus, rapid detection of the onset of stroke remains critically important. A device incorporating this algorithm to continuously monitor for stroke onset may be able to reduce the time to assessment, leading to more and faster interventions and better outcomes for patients.

There are several studies that have attempted to identify clinical predictors of in-hospital stroke.[11] However, the algorithm was intentionally designed to not require any patient-specific clinical information to detect asymmetric movement seen in acute stroke. This is analogous to cardiac telemetry monitors that have built-in automated detection of life-threatening arrhythmias but do not rely on clinical parameters such as history of coronary artery disease or congestive heart failure. Upper limb weakness is one of the most common symptoms of acute stroke, seen in ≈75% of patients.[15] For this reason, prehospital stroke screening tools and scales that aim to identify patients with the greatest likelihood of having a large vessel occlusion have all included arm strength.[16,17,32] In addition, attentional neglect is present in 20% to 70% of strokes, and studies of patients with stroke using wrist-worn

accelerometers have demonstrated that neglect is associated with asymmetric movement.[18,19] Importantly, weakness and neglect are both strongly associated with long-term disability from stroke.[33,34] Thus, although upper limb accelerometry monitoring will not capture every stroke, it will identify the vast majority of strokes, including those most likely to result in disability and be most amenable to thrombectomy, which is proven to dramatically improve outcomes.

Patient physiologic monitors are ubiquitous in hospitals in general and in intensive care units in particular, where multimodal monitoring is the standard of care. Unfortunately, these pervasive monitors may result in alarm fatigue, leading to delayed or absent responses.[35] Fatigue is more likely when nonactionable alarms are much more prevalent than actionable alarms that require both clinical awareness and intervention. Stroke is a critical patient event that is both actionable and time sensitive.[6] A study of 461 adults treated in intensive care units annotated a total of 381 560 unique audible alarms over a 31-day study period.[36] Accelerated ventricular arrhythmia alarms, a potentially critical patient abnormality, occurred at an average of 4.5 alarms per patient per day of monitoring, of which only 12 (0.3%) were clinically relevant actionable events. This stroke detection algorithm provides a far lower false alarm rate, while greatly reducing time from symptom onset to stroke detection compared with current clinical practice.

To our knowledge, this study is the first to demonstrate that arm accelerometry data can rapidly identify patients with weakness caused by acute stroke while maintaining clinically acceptable false alarms in neurologically normal hospitalized patients at risk of stroke. Notably, we performed the validation analysis on a separate prospectively acquired cohort of patients, using different accelerometry devices than were used to collect data to derive the algorithm. The performance of the algorithm under these conditions reflects its robustness and generalizability. The validation cohort included control patients who underwent cardiothoracic or vascular surgical procedures, reflecting a population that is high risk for stroke and would benefit from continuous stroke monitoring. As a result of these data, the company using this algorithm to develop a commercially available stroke monitor device was granted Breakthrough Technology status by the US Food and Drug Administration.

Our study did have several limitations. The case–control design leads to risk of spectrum bias. Most stroke cases enrolled were moderate or severe, and it is possible that the algorithm is less sensitive for milder strokes. Importantly, the algorithm appears to perform well in patients with significant weakness, which are the strokes most likely to have a large vessel occlusion amenable to intervention, and most likely to lead to disability or death if untreated. The fact that stroke cases were not specifically postoperative may also lead to an inaccurate assessment of time to stroke detection, although they were all bed bound and many were in an intensive care unit setting. Furthermore, the false alarm rate in the control arm should provide an accurate estimate of what we would expect to see in a clinical trial, because it is the same population of high-risk patients who will be enrolled. The overall size of the validation study may be underpowered to detect clinically important differences in performance for specific subgroups. In particular, only 2 left-handed patients with stroke were enrolled, and it is possible that performance may be different in these patients. However, there is no biological expectation of differential performance, and the results overall suggest there are not likely to be large differences. In addition, subjects were enrolled at a single tertiary referral center, which may limit generalizability, although patient arm movements are not likely to vary meaningfully at different hospitals. The algorithm requires 15 minutes of data initially and uses up to 90 minutes of preceding movement data when assessing for stroke. Given that we test every 15 minutes, there is an unavoidable correlation between successive tests. Importantly, our reported confidence intervals reflect the variance in the population, not the individual, and such an evaluation scenario is consistent with how the algorithm would ultimately be used in practice. Finally, our results do not directly assess how well the algorithm performs when an individual patient converts from being neurologically intact to having a stroke. However, as noted, the algorithm does not rely on a comparison from baseline movements, and this may lead to even faster times to detection that we report here.

CONCLUSIONS

In-hospital stroke is a major public health issue, and a monitor that can rapidly detect the onset of stroke and facilitate expedited assessment and treatment could lead to improved outcomes for patients. We derived a stroke detection algorithm using upper extremity accelerometry data from hospitalized patients that demonstrates promising diagnostic performance in a prospective validation case–control study. The algorithm's performance on speed of detection when a patient develops a stroke and false alarm rates in a real-world clinical setting is not known, and a trial to prospectively monitor patients at risk of stroke is required to demonstrate clinical use and tolerability.

ARTICLE INFORMATION

Received November 15, 2022; accepted December 6, 2022.

Affiliations

Department of Neurology (S.R.M., S.E.K., B.L.C., M.L.M., S.C.) Department of Surgery (M.A.A., N.D., P.A., G.J.W., B.M.J.) and Department of Computer and Information Science (J.W.).

Sources of Funding

This study was funded by a University of Pennsylvania Department of Neurology Translational Center of Excellence Grant. The development of the StrokeDetectAI algorithm was supported in part by that National Science Foundation (NSF-1915398) and the Air Force Research Laboratory and Defense Advanced Research Projects Agency (AFRL and DARPA FA8750-18-C-0090). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Air Force Research Laboratory, the Defense Advanced Research Projects Agency, or the Department of Defense, or the United States Government.

Disclosures

Drs Messé and Weimer are cofounders and have equity in a company named Neuralert Technologies, which has licensed the algorithm described in this study to develop a monitor to rapidly detect stroke in hospitalized patients. The remaining authors have no disclosures to report.

Supplemental Material

Data S1

References[37–40]

REFERENCES

1. Saver JL. Time is brain–quantified. *Stroke.* 2006;37:263–266. doi: 10.1161/01.STR.0000196957.55928.ab
2. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, Davalos A, Majoie CB, van der Lugt A, de Miquel MA, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet.* 2016;387:1723–1731. doi: 10.1016/S0140-6736(16)00163-X
3. Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, Campbell BC, Nogueira RG, Demchuk AM, Tomasello A, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA.* 2016;316:1279–1288. doi: 10.1001/jama.2016.13647
4. Alberts MJ, Perry A, Dawson DV, Bertels C. Effects of public and professional education on reducing the delay in presentation and referral of stroke patients. *Stroke.* 1992;23:352–356. doi: 10.1161/01.STR.23.3.352
5. Lecouturier J, Rodgers H, Murtagh MJ, White M, Ford GA, Thomson RG. Systematic review of mass media interventions designed to improve public recognition of stroke symptoms, emergency response and early treatment. *BMC Public Health.* 2010;10:784. doi: 10.1186/1471-2458-10-784
6. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, Biller J, Brown M, Demaerschalk BM, Hoh B, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2019;50:e344–e418. doi: 10.1161/STR.0000000000000211
7. Dulli D, Samaniego EA. Inpatient and community ischemic strokes in a university hospital. *Neuroepidemiology.* 2007;28:86–92. doi: 10.1159/000098551
8. Park HJ, Cho HJ, Kim YD, Lee DW, Choi HY, Kim SM, Heo JH. Comparison of the characteristics for in-hospital and out-of-hospital ischaemic strokes. *Eur J Neurol.* 2009;16:582–588. doi: 10.1111/j.1468-1331.2009.02538.x
9. Saltman AP, Silver FL, Fang J, Stamplecoski M, Kapral MK. Care and outcomes of patients with in-hospital stroke. *JAMA Neurol.* 2015;72:749–755. doi: 10.1001/jamaneurol.2015.0284
10. Cumbler E, Wald H, Bhatt DL, Cox M, Xian Y, Reeves M, Smith EE, Schwamm L, Fonarow GC. Quality of care and outcomes for in-hospital ischemic stroke: findings from the national get with the guidelines-stroke. *Stroke.* 2014;45:231–238. doi: 10.1161/STROKEAHA.113.003617
11. Nouh A, Amin-Hanjani S, Furie KL, Kernan WN, Olson DM, Testai FD, Alberts MJ, Hussain MA, Cumbler EU; American Heart Association Stroke Council, et al. Identifying best practices to improve evaluation and management of in-hospital stroke: a scientific statement from the American Heart Association. *Stroke.* 2022;53:e165–e175. doi: 10.1161/STR.0000000000000402
12. Puskas JD, Winston AD, Wright CE, Gott JP, Brown WM III, Craver JM, Jones EL, Guyton RA, Weintraub WS. Stroke after coronary artery operation: incidence, correlates, outcome, and cost. *Ann Thorac Surg.* 2000;69:1053–1056. doi: 10.1016/S0003-4975(99)01569-6
13. Bhatt A, Safdar A, Chaudhari D, Clark D, Pollak A, Majid A, Kassab M. Medicolegal considerations with intravenous tissue plasminogen activator in stroke: a systematic review. *Stroke Res Treat.* 2013;2013:562564. doi: 10.1155/2013/562564
14. Haslett JJ, Genadry L, Zhang X, LaBelle LA, Bederson J, Mocco J, Kellner CP. Systematic review of malpractice litigation in the diagnosis and treatment of acute stroke. *Stroke.* 2019;50:2858–2864. doi: 10.1161/STROKEAHA.119.025352
15. Lawrence ES, Coshall C, Dundas R, Stewart J, Rudd AG, Howard R, Wolfe CD. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke.* 2001;32:1279–1284. doi: 10.1161/01.STR.32.6.1279
16. van Gaal SC, Kamal N, Betzner MJ, Vilneff RL, Mann B, Lang ES, Demchuk A, Buck B, Jeerakathil T, Hill MD. Approaches to the field recognition of potential thrombectomy candidates. *Int J Stroke.* 2017;12:698–707. doi: 10.1177/1747493017724585
17. Duvekot MHC, Venema E, Rozeman AD, Moudrous W, Vermeij FH, Biekart M, Lingsma HF, Maasland L, Wijnhoud AD, Mulder L, et al. Comparison of eight prehospital stroke scales to detect intracranial large-vessel occlusion in suspected stroke (presto): a prospective observational study. *Lancet Neurol.* 2021;20:213–221. doi: 10.1016/S1474-4422(20)30439-7
18. Chen P, Hreha K, Fortis P, Goedert KM, Barrett AM. Functional assessment of spatial neglect: a review of the catherine bergego scale and an introduction of the kessler foundation neglect assessment process. *Top Stroke Rehabil.* 2012;19:423–435. doi: 10.1310/tsr1905-423
19. Trejo-Gabriel-Galan JM, Rogel-Melgosa V, Gonzalez S, Sedano J, Villar JR, Arenaza-Basterrechea N. Rehabilitation of hemineglect of the left arm using movement detection bracelets activating a visual and acoustic alarm. *J Neuroeng Rehabil.* 2016;13:79. doi: 10.1186/s12984-016-0191-0
20. Watson A, Choi H, Lee I, Weimer J. Raproto: an open source platform for rapid prototyping of wearable medical devices. Proceedings of the Workshop on Medical Cyber Physical Systems and Internet of Medical Things. 2021;1–6.
21. Ivanov R, Weimer J, Simpao A, Rehman M, Lee I. Early detection of critical pulmonary shunts in infants. Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems. 2015;110–119.
22. Weimer J, Chen S, Peleckis A, Rickels MR, Lee I. Physiology-invariant meal detection for type 1 diabetes. *Diabetes Technol Ther.* 2016;18:616–624. doi: 10.1089/dia.2015.0266
23. Weimer J, Ivanov R, Chen S, Roederer A, Sokolsky O, Lee I. Parameter-invariant monitor design for cyber–physical systems. *Proceedings of the IEEE.* 2017;106:71–92. doi: 10.1109/JPROC.2017.2723847
24. Nguyen H, Jang S, Ivanov R, Bonafide CP, Weimer J, Lee I. Reducing pulse oximetry false alarms without missing life-threatening events. *Smart Health (Amst).* 2018;9-10:287–296. doi: 10.1016/j.smhl.2018.07.002
25. Weimer J, Ivanov R, Roederer A, Chen S, Lee I. Parameter-invariant design of medical alarms. *IEEE Design & Test.* 2015;32:9–16. doi: 10.1109/MDAT.2015.2451083
26. Wilsky A. A survey of design methods for failure detection in dynamic systems. *Automatica.* 1976;12:601–611. doi: 10.1016/0005-1098(76)90041-8
27. Bishop C. *Pattern Recognition and Machine Learning.* Springer; 2006.
28. Poor H, Hadjiliadis O. *Quickest detection.* Cambridge University Press; 2008. doi: 10.1017/CBO9780511754678
29. Kassardjian CD, Willems JD, Skrabka K, Nisenbaum R, Barnaby J, Kostyrko P, Selchen D, Saposnik G. In-patient code stroke: a quality improvement strategy to overcome knowledge-to-action gaps in response time. *Stroke.* 2017;48:2176–2183. doi: 10.1161/STROKEAHA.117.017622
30. Manners J, Khandker N, Barron A, Aziz Y, Desai SM, Morrow B, Delfyett WT, Martin-Gill C, Shutter L, Jovin TG, et al. An interdisciplinary approach to inhospital stroke improves stroke detection and treatment time. *J Neurointerv Surg.* 2019;11:1080–1084. doi: 10.1136/neurintsurg-2019-014890
31. Cummings S, Kasner SE, Mullen M, Olsen A, McGarvey M, Weimer J, Jackson B, Desai N, Acker M, Messe SR. Delays in the identification and assessment of in-hospital stroke patients. *J Stroke Cerebrovasc Dis.* 2022;31:106327. doi: 10.1016/j.jstrokecerebrovasdis.2022.106327
32. Brandler ES, Sharma M, Sinert RH, Levine SR. Prehospital stroke scales in urban environments: a systematic review. *Neurology.* 2014;82:2241–2249. doi: 10.1212/WNL.0000000000000523

33. Kwakkel G, Wagenaar RC, Kollen BJ, Lankhorst GJ. Predicting disabil-ity in stroke–a critical review of the literature. *Age Ageing.* 1996;25:479– 489. doi: 10.1093/ageing/25.6.479
34. Cherney LR, Halper AS, Kwasnica CM, Harvey RL, Zhang M. Recovery of functional status after right hemisphere stroke: relationship with uni-lateral neglect. *Arch Phys Med Rehabil.* 2001;82:322–328. doi: 10.1053/apmr.2001.21511
35. Hravnak M, Pellathy T, Chen L, Dubrawski A, Wertz A, Clermont G, Pinsky MR. A call to alarms: current state and future directions in the battle against alarm fatigue. *J Electrocardiol.* 2018;51:S44–S48. doi: 10.1016/j.jelectrocard.2018.07.024
36. Drew BJ, Harris P, Zegre-Hemsey JK, Mammone T, Schindler D, Salas-Boni R, Bai Y, Tinoco A, Ding Q, Hu X. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS One.* 2014;9:e110274. doi: 10.1371/journal.pone.0110274
37. McGaughey G, Walters WP, Goldman B. Understanding covariate shift in model performance. *F1000Res.* 2016;5:597.
38. Kolomogorov A. Sulla determinazione empirica di una legge di dis-tribuzione. *G Ist Ital Attuari.* 1933;4:83–91.
39. Smirnov N. Table for estimating the goodness of fit of empiri-cal distributions. *Ann Math Stat.* 1948;19:279–281. doi: 10.1214/aoms/1177730256
40. Jang S, Lee I, Weimer J. Improving classifier confidence using lossy label-invariant transformations. Proceedings of the 24th international confer-ence on artificial intelligence and statistics. *PMLR.* 2021;130:4051–4059.

SUPPLEMENTAL MATERIAL

Data S1.

Supplemental Methods

Algorithm derivation methodology

Data pre-processing

To design a low-cost lightweight comfortable wrist-worn device for stroke detection, we sought to utilize only accelerometry data in our stroke detection analysis. While incorporating additional sensors, such as gyroscopes and magnetometers, would theoretically enable device orientation and arm position estimation, they would also increase device cost, power consumption, battery size, and weight. Consequently, this work aimed to utilize off-the shelf low-power accelerometers to detect stroke and the pre-processing considered herein assumed only accelerometry data were available.

Low-cost low-power accelerometers common in wrist-worn devices produced, at time k, 3-dimensional data, $a_x(k)$, $a_y(k)$, and $a_z(k)$, but were also susceptible to bias and rotation/sliding on the wrist. We denoted the constant bias as c_x, c_y, and c_z, and removed their effect by utilizing the first-derivative of acceleration (known as "jerk") since $J_x(k) = \frac{d}{dk}(a_x(k) + c_x) = \frac{d}{dk}a_x(k)$, and similarly for the y and z dimensions. Once the bias was removed, we removed the effect of rotation/sliding on the wrist by only considering the magnitude of jerk, written for the left-arm motion data as $x_L(k) = \sqrt{J_{x,L}^2(k) + J_{y,L}^2(k) + J_{z,L}^2(k)} \in X$, where X denotes the feature space and a similar equation exists for the right arm, $x_R(k)$. This pre-processing step served to eliminate inherent system biases that are likely to occur during real-world deployments and are consistent

with other data pre-processing techniques for accelerometry data without access to gyroscopes and magnetometers.

Test statistic engineering

To engineer a test statistic for discriminating the between stroke and neurologically intact subjects, we began by writing $D_X = \{f : X \to P\}$, to be a space of probability distributions mapping the feature space to a probability. In an (idealized) controlled evaluation environment, where a subject performs a prescribed sequence of actions/motions, the distribution for the left arm, $f_L \in D_X$, and right arm, $f_R \in D_X$, can discriminate between neurologically intact subjects (*i.e.,* $f_L = f_R$) and stroke subjects (i.e., $f_L \neq f_R$). While this idealized scenario can yield highly sensitive and specific stroke detection, in practice it would be far too invasive -- requiring frequent neurological assessments to timely detect stroke.

Rather than require patients to perform a set of prescribed tasks at set intervals, we sought to engineer a test statistic that is suitable for passive monitoring scenarios. Such a test statistic must be robust to changes in the underlying patient motion distribution, referred to in the statistical literature as a *covariate shift*.[37] Motion distribution covariate shift is common in passive monitoring scenarios and captures the effect of any patient-specific tendency in the data (e.g., dominant hand, comorbidities, etc.). However, the impact of the motion covariate shift will be limited by the patient's neurological state, which is presumed to be unknown at the time of testing. Consequently, we modeled the family of motion covariate shifts as a group of distribution nuisance transformations applied to the patient's (unknown) neurological state, namely for $f \in \{f_L, f_R\}$,

$$G_f = \{g: \mathrm{D_X} \to \mathrm{D_X} | \forall x \in X,\ f(x) \neq 0 \leftrightarrow g(f(x)) \neq 0\}$$

where $g \in G_f$ denotes a potential motion covariate shift. Consequently, we seek a test statistic that can assess the neurological state robust to motion distribution covariate shift.

A promising approach to realize a robust test statistic utilizes parameter invariant (PAIN) statistics – which have been previously applied in multiple domains.[21-24] Given a group of nuisance transformations, a PAIN statistic, t, seeks to provide invariance to the nuisance transformations (*i.e.,* is invariant: $\forall f \in D_X, \forall g \in \mathrm{G}, t(g(f)) = t(f)$) while only eliminating information affected by the nuisance transformations, (*i.e.,* is maximal $\forall f, f' \in \mathrm{D_X},\ \exists g \in G,\ t(f) = t(f') \to g(f) = f'$). Thus, we considered a candidate PAIN statistic,

$$t: d \in D_X \mapsto d' \in D_X: \exists c, \forall x \in X, c1(d(x) \neq 0)$$

and proved it to be invariant since, $\forall f \in D_X, \forall g \in G$

$$\forall g \in G, t(g(f)) = d' \in D_X: \exists c, \forall x \in X, c1(g(f(x)) \neq 0) = d' \in D_X: \exists c, \forall x \in X, c1(f(x)) \neq 0) = t(f)$$

and maximal since, $\forall f, f' \in D_X, \exists g \in G$,

$$t(f) = t(f') \to d \in D_X: \exists c, \forall x \in X, c1(f(x) \neq 0) = d' \in D_X: \exists c', \forall x \in X, c'1(f'(x) \neq 0) \to g(f) = f'$$

Moreover, we note that $t(f_L)$ and $t(f_R)$ have an attractive property, namely if $f_L = f_R$ (as is the case in neurologically intact subjects in the idealized scenario), then $t(f_L) = t(f_R)$, stated formally as $f_L = f_R \to t(f_L) = t(f_R)$. This means that in the idealized monitoring scenario, if subjects are neurologically intact, then in the passive monitoring scenario they should also appear neurologically intact.

Thus, we aimed to generate a test statistic, that discriminated between neurologically intact subjects (i.e., $t(f_L) = t(f_R)$) and stroke subjects (i.e., $t(f_L) \neq t(f_R)$). In this scenario, we utilized the Kolmogorov-Smirnov (KS) statistic,[38,39] denoted by letting $t_L = t(f_L)$ and $t_R = t(f_R)$, and writing the test statistic

$$s = sup_{z \in X} \left| \int_{-\infty}^{z} t_L(x) - t_R(x)\, dx \right|$$

which, represents a non-parametric statistic of distribution equality that equals the maximum absolute deviation of the cumulative distribution functions corresponding to the probability mass functions $t_L$and t_R. The KS statistic is a widely used test of distribution equality when the underlying test distribution family is unknown or non-parameterized (i.e., non-parametric).

Test generation

We then developed a threshold test for the test statistic, s, derived in the previous section. The test statistic requires the cumulative distribution functions corresponding to the probability mass functions t_L and t_R. Unfortunately, these are not generally known and must be estimated from a recent history (1 hour) of the pre-processed sampled data, $X(k) = \{(x_L(k), x_R(k)), (x_L(k-1), x_R(k-2)), \ldots\}$. Utilizing sampled data estimates in place of the actual distribution presents two potential concerns. First, when there is significant missing data the amount of information contained in the sampled data decreases. Second, anytime the patient has no motion (i.e., laying perfectly still) while the data is not technically missing, it provides no discriminatory information for testing stroke versus neurologically intact. Consequently, we write $s(k)$ to be the test statistic

estimated using $X(k)$, and write $r_1(k) = |X(k)|$ to be the number of data points in $X(k)$ and $r_2(k) = |\{(x_L, x_R)|(x_L, x_R) \in X(k), x_L \neq 0 \vee x_R \neq 0\}|/|X(k)|$ to be the percentage of $X(k)$ with patient movement.

To derive a threshold test we leveraged r_1 and r_2 to adapt a threshold such that the resulting test has a constant false alarm rate, $\alpha \in [0,1]$. To achieve this, we grouped the data using kmeans with k=100 on $[(r_1(1), r_2(1)), (r_1(2), r_2(2)), \dots]$ and generated a corresponding threshold for each group to achieve a constant false alarm rate α. To achieve maximal distributional accuracy when tuning the false alarm rate the threshold test was calibrated prior to threshold selection.[40] At runtime, a new $s(k)$ was generated with corresponding $r_1(k)$ and $r_2(k)$. The decision threshold utilized for testing $s(k)$ corresponds to the group containing $(r_1(k), r_2(k))$. In the following, we refer to the threshold test described above as $\phi \in \{0,1\}$, where ϕ=0 predicts the absence of stroke and ϕ=1 predicts the presence of stroke.

To improve sensitivity to the onset of stroke, we ran multiple threshold tests, $\phi_1, \dots, \phi_L$, simultaneously with different monitoring durations, $d_1, \dots, d_L$, respectively. For example, for each $l \in \{1, \dots, L\}$ at time t, ϕ_1 utilized data in the time range $[t - d_1, t]$. Leveraging the multiple threshold tests, we defined an aggregate threshold test, $\phi = \max\{\phi_1, \dots, \phi_L\}$, that predicts the presence of stroke if and only if one of the L monitoring durations predicts the presence of a stroke. We note that the false alarm rate of the aggregate test is always greater than α. Consequently, we select α in the threshold test design to be small enough such that the aggregate test achieves our desired false alarm rate.

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	
	5b	Describe eligibility criteria for participants.	
	5c	Give details of treatments received, if relevant.	
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	Explain how the study size was arrived at.	
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	
Risk groups	11	Provide details on how risk groups were created, if done.	
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	
Model development	14a	Specify the number of participants and outcome events in each analysis.	
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	
	15b	Explain how to the use the prediction model.	
Model performance	16	Report performance measures (with CIs) for the prediction model.	
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	
Implications	20	Discuss the potential clinical use of the model and implications for future research.	
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	
Funding	22	Give the source of funding and the role of the funders for the present study.	

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.