

Original Paper

Automatic Grading of Stroke Symptoms for Rapid Assessment Using Optimized Machine Learning and 4-Limb Kinematics: Clinical Validation Study

Eunjeong Park¹, PhD; Kijeong Lee², MD; Taehwa Han³, PhD; Hyo Suk Nam⁴, MD, PhD

¹Cerebro-Cardiovascular Disease Research Center, Yonsei University College of Medicine, Seoul, Republic of Korea

²Department of Radiology, Yonsei University College of Medicine, Seoul, Republic of Korea

³Health-IT Center, Yonsei University College of Medicine, Seoul, Republic of Korea

⁴Department of Neurology, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:

Hyo Suk Nam, MD, PhD

Department of Neurology

Yonsei University College of Medicine

50 Yonsei-ro, Seodaemun-gu

Seoul, 03722

Republic of Korea

Phone: 82 222280245

Email: hsnam@yuhs.ac

Abstract

Background: Subtle abnormal motor signs are indications of serious neurological diseases. Although neurological deficits require fast initiation of treatment in a restricted time, it is difficult for nonspecialists to detect and objectively assess the symptoms. In the clinical environment, diagnoses and decisions are based on clinical grading methods, including the National Institutes of Health Stroke Scale (NIHSS) score or the Medical Research Council (MRC) score, which have been used to measure motor weakness. Objective grading in various environments is necessitated for consistent agreement among patients, caregivers, paramedics, and medical staff to facilitate rapid diagnoses and dispatches to appropriate medical centers.

Objective: In this study, we aimed to develop an autonomous grading system for stroke patients. We investigated the feasibility of our new system to assess motor weakness and grade NIHSS and MRC scores of 4 limbs, similar to the clinical examinations performed by medical staff.

Methods: We implemented an automatic grading system composed of a measuring unit with wearable sensors and a grading unit with optimized machine learning. Inertial sensors were attached to measure subtle weaknesses caused by paralysis of upper and lower limbs. We collected 60 instances of data with kinematic features of motor disorders from neurological examination and demographic information of stroke patients with NIHSS 0 or 1 and MRC 7, 8, or 9 grades in a stroke unit. Training data with 240 instances were generated using a synthetic minority oversampling technique to complement the imbalanced number of data between classes and low number of training data. We trained 2 representative machine learning algorithms, an ensemble and a support vector machine (SVM), to implement auto-NIHSS and auto-MRC grading. The optimized algorithms performed a 5-fold cross-validation and were searched by Bayes optimization in 30 trials. The trained model was tested with the 60 original hold-out instances for performance evaluation in accuracy, sensitivity, specificity, and area under the receiver operating characteristics curve (AUC).

Results: The proposed system can grade NIHSS scores with an accuracy of 83.3% and an AUC of 0.912 using an optimized ensemble algorithm, and it can grade with an accuracy of 80.0% and an AUC of 0.860 using an optimized SVM algorithm. The auto-MRC grading achieved an accuracy of 76.7% and a mean AUC of 0.870 in SVM classification and an accuracy of 78.3% and a mean AUC of 0.877 in ensemble classification.

Conclusions: The automatic grading system quantifies proximal weakness in real time and assesses symptoms through automatic grading. The pilot outcomes demonstrated the feasibility of remote monitoring of motor weakness caused by stroke. The system can facilitate consistent grading with instant assessment and expedite dispatches to appropriate hospitals and treatment initiation by sharing auto-MRC and auto-NIHSS scores between prehospital and hospital responses as an objective observation.

KEYWORDS

machine learning; artificial intelligence; sensors; kinematics; stroke; telemedicine

Introduction

Motor weakness is a typical manifestation in various neurological disorders, including stroke, spinal cord injury, and traumatic brain injury. In addition, it is a major obstacle to functional recovery after the treatment of those diseases. As an example of motor weakness, unintentional drift is an indication of arm weakness; this is mainly caused by subtle damages in the motor pathway from the brain to the spinal cord [1]. If the supinator muscles in the upper limb are weaker than the pronator muscles in the presence of upper motor neuron lesion, the arm drifts downward and the palm turns toward the floor. The pathological response is for one of the arms to drift (up, down, or out). Therefore, motor weakness is a major sign in the FAST (face drooping, arm weakness, speech slurring, and time to call) protocol for stroke patients [2].

Rapid detection of such motor weakness is critical because acute treatments, including thrombolysis or thrombectomy, are performed in a constrained time window. More importantly, diagnosis can be established through bedside examination by specialists because it is a qualitative measurement. If the symptom occurs outside a hospital, a substantial time delay can cause poor outcomes for acute stroke patients [3-5]. In addition, the objective and accurate neurological assessments are not possible by mere visual examination because the examiner cannot easily trace the movement using the conventional neurological examination when there are subtle weaknesses. Therefore, systems need to automatically detect motor deficits using sensor data in real time.

However, operating such systems in a real environment requires a significant effort in integrating new systems into an emergency protocol. This is because interruptions caused by the attachment of sensors on patients' bodies and the initiation of the recording process can affect the streamlined structure of emergency protocols. However, evaluation methods are still required to identify stroke patients, as they can be instantly used in the communication among patients or caregivers, emergency call centers, and hospitals. In addition to a sensor-based measurement tool that was demonstrated useful in detecting subtle motor weakness in our previous study [6], the grading of stroke severity can be informed remotely and used in the emergency medical service (EMS) and hospital system.

In the field and in clinical environments, various grading methods exist for identifying ischemic stroke patients with motor weakness [7-10]. The National Institutes of Health Stroke Scale (NIHSS) score [11,12] and Medical Research Council (MRC) score [13,14] have been used as typical assessment indicators for stroke in the clinical environment. The rapid arterial occlusion evaluation scale, the Cincinnati stroke triage assessment tool, and the prehospital acute stroke severity scale are grading methods in the field environment. In this study, we implemented auto-NIHSS and auto-MRC systems to grade the NIHSS and modified MRC scores to assess patients in the clinical environment. We used subdivided MRC scores (10-grade MRC) instead of a 6-grade MRC to define subtle differences, as shown in Table 1.

Table 1. NIHSS and MRC grades for muscle power assessment.

Scale and grade	Description
NIHSS^a	
0	No drift; limb holds 90° (or 45°) angle for full 10 seconds
1	Drift; limb holds 90° (or 45°) angle, but drifts down before full 10 seconds; does not hit bed or other support
2	Some effort against gravity; limb cannot reach or maintain (if cued) 90° (or 45°) angle; drifts down to bed, but has some effort against gravity
3	No effort against gravity; limb falls
4	No movement
MRC^b	
0 (0)	No movement
1 (1)	A flicker of movement is observed or felt in the muscle
2 (1+)	Muscle moves the joint when gravity is eliminated
3 (2)	Muscle moves the joint against gravity, but not through full mechanical range of motion
4 (2+)	Muscle cannot hold the joint against resistance, but moves the joint fully against gravity
5 (3)	Muscle moves the joint fully against gravity and is capable of transient resistance, but collapses abruptly
6 (3+)	Same as grade 4 (on 6-point scale) but muscle holds the joint only against minimal resistance
7 (4)	Muscle holds the joint against a combination of gravity and moderate resistance
8 (4+)	Same as grade 4 (on 6-point scale) but muscle holds the joint against moderate to maximal resistance
9 (5)	Normal strength

^aNIHSS: National Institutes of Health Stroke Scale.

^bMRC: Medical Research Council.

Methods

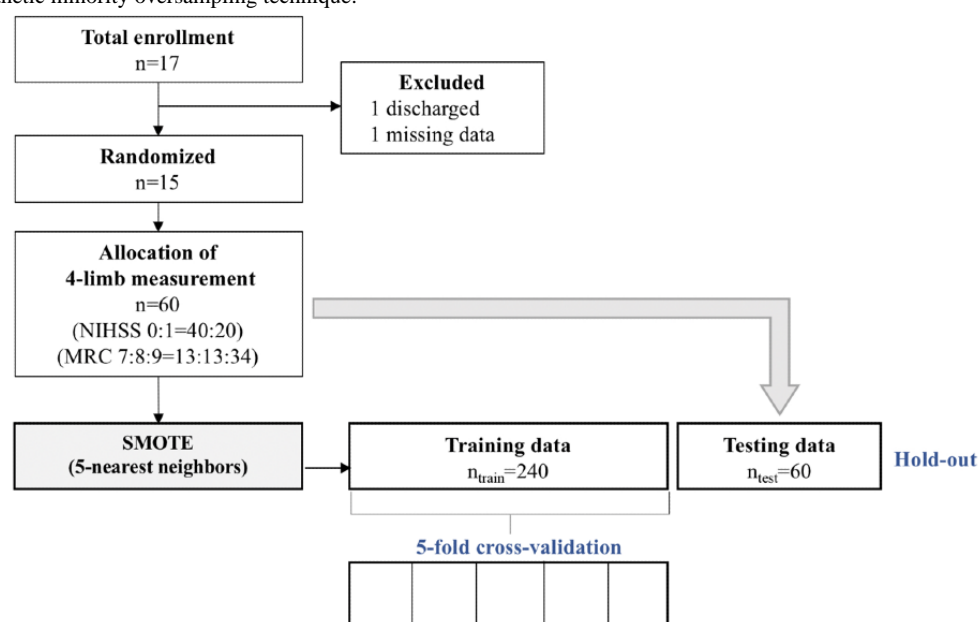
Participants and Data

A total of 17 participants were recruited; 15 participants (10 male and 5 female participants) were finally enrolled and completed 4-limb drift test trials. To estimate the scores of patients with severity, we performed the assessment shortly after admission to a stroke unit. The ages of the participants ranged from 44 to 92 years, with a mean of 68.6 years (SD 16.11). Exclusion criteria were patients (1) who had a substantial

weakness that prevented arm or leg raising against gravity, (2) who were not able to sit and who had bilateral arm weakness or preexisting chronic arm weakness, and (3) who had aphasia, neglect, peripheral neuropathy, myopathy, or joint deformity. This study was approved by the Severance Hospital Institutional Review Board, and informed consent was obtained from all participants.

Figure 1 shows patient enrollment and data preparation for auto-NIHSS and auto-MRC grading. Description of data composition for training, validation and testing is detailed in the section on system design.

Figure 1. Patient enrollment and data set for automatic grading system. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale; SMOTE: synthetic minority oversampling technique.



System Design

The entire process of the system is shown in Figure 2. The system is composed of 2 parts, the measurement and the grading units. The measurement unit sets up sensors and Bluetooth connection with the primary information of patients.

We measured the upper left and upper right limb movements using sensors on both wrists of patients, who were asked to stretch and hold their arms for 20 seconds, as shown in Figure 3. For the lower left and lower right limb drift tests, patients were asked to lift and stretch their left or right leg for 20 seconds.

Figure 2. Automatic grading process. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale.

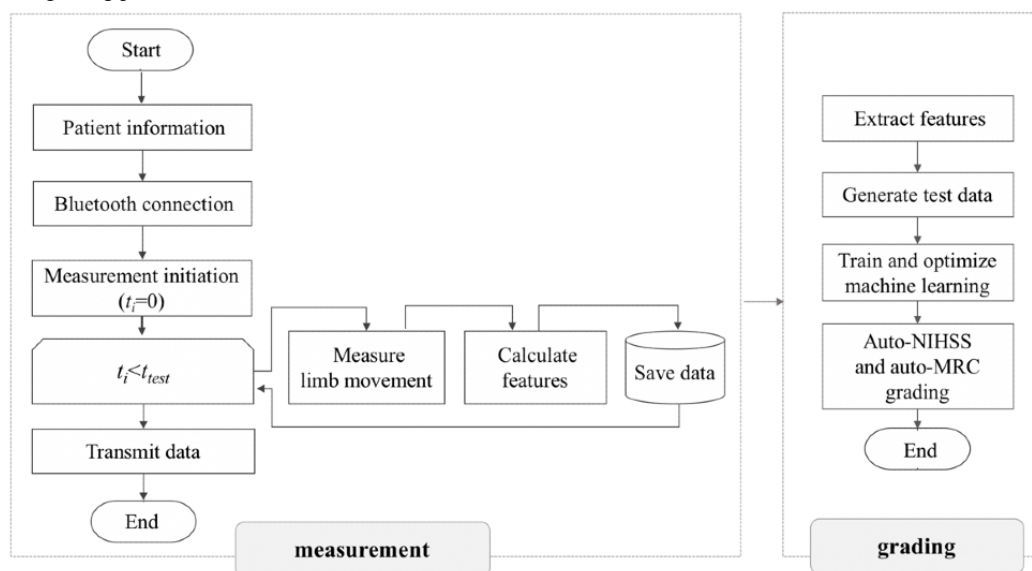


Figure 3. Schematic of upper and lower limb sensors and corresponding segment axes.



The pseudo-code of the measurement unit is shown in [Multimedia Appendix 1](#). For each time frame i , the rotational transformation from the limb into the reference frame xyz is denoted as R_b^i . The corresponding rotation matrices R for each angle are defined using the A_b^i of the accelerometer signals for the i th frame. Subsequently, the degree of drift, θ_{drift} , is calculated and used in key features of machine learning classification.

After collecting the series of 4-limb movements during the test time, the grading unit analyzes the kinematic features. Subsequently, the machine learning algorithm is trained to estimate the NIHSS and MRC scores of each limb. Algorithm 2 (in [Multimedia Appendix 2](#)) shows the process of feature extraction, data generation, and model training for the optimized classification of auto-NIHSS and auto-MRC.

In the feature extraction process, features as predictors of limb paralysis were extracted using a series of measured data. In this study, the duration of the drift test (t_{test}) was set to 20 seconds; however, analysis started 10 seconds after the examination started (t_{start}) to exclude the initial dip. The average, maximum, and oscillation of drift caused by paralysis for each limb and demographic features were fed to train the machine learning algorithms.

In the data generation process, we adopted the synthetic minority oversampling technique (SMOTE) [15], leveraging the K-nearest neighbor (K-NN), to solve the imbalanced problem that is typical in machine learning studies in medicine [16-18]. The SMOTE with K-NN generated n_g samples for each grade.

Therefore, $n_g c$ records were used to construct a grading model with c classes. In this study, n_g was set to 120 for auto-NIHSS ($c=2$) and 80 for auto-MRC ($c=3$) to compose the training data with 240 (t_{train}) instances. Apart from the training data, the original data set with 60 records remained for the test data, as shown in [Figure 1](#).

In the training process, 5-fold cross-validation was applied to reduce overfitting and generalize the model [19]. In the optimization process, the fitted support vector machine (SVM), as well as ensemble models among various SVM kernels and boosting algorithms with tuned hyperparameters, were searched via Bayes optimization in 30 trials for each model [20]. The grading models were implemented and evaluated in MATLAB R2020a (MathWorks Inc) [21].

Results

Sensor Data Characteristics

The system measured the drift of 4 limbs and extracted the kinematic features, as shown in [Multimedia Appendix 3](#). The characteristics of the patients and test data are summarized in [Table 2](#). The grade distribution of clinical scores was not regularized between limbs, as shown in [Figure 4](#). For example, the upper left MRC group had 10 patients graded as MRC 9, 2 patients graded as MRC 8, and 3 patients graded as MRC 7. Among 13 MRC 8 instances, 7 were evaluated as NIHSS 1, whereas 6 were evaluated as NIHSS 0. We constructed auto-MRC, which discriminated instances of grades with a data ratio of 13:13:34, whereas auto-NIHSS performed binary classification with a data ratio of 40:20.

Figure 4. Grade distribution of NIHSS and MRC. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale.

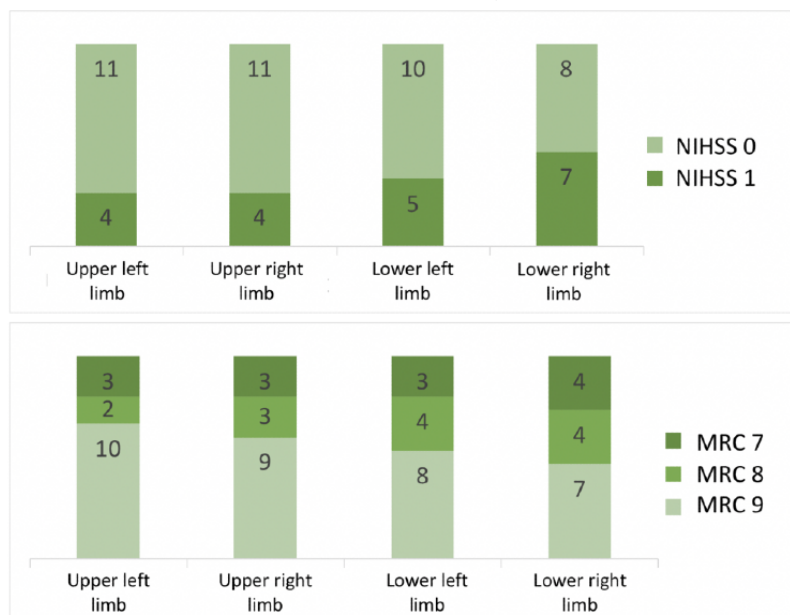


Table 2. Summary of patients and test data.

Diagnosis	Measurement												NIHSS ^a grade (MRC ^b grade)			
	ULL ^c			URL ^d			LLL ^e			LRL ^f			ULL	URL	LLL	LRL
	Mean	Max ^g	Osc ^h	Mean	Max	Osc	Mean	Max	Osc	Mean	Max	Osc				
Lt ⁱ internal capsule infarction	0.82	2.7	14.4	-3	-1.9	15.3	-1.19	2.1	25.1	11.81	17.4	30.8	0 (9)	0 (8)	0 (9)	1 (8)
Lt MCA ^j infarction	-9.33	-12	15.9	-6.47	-9.2	11.7	18.7	10.7	43.5	30.26	27.7	13.6	0 (8)	1 (7)	0 (8)	1 (7)
Lt MCA infarction	0.86	0	7	4.06	1.4	13.8	2.96	0	24.9	9.91	-1.6	61.5	0 (9)	0 (9)	0 (9)	1 (7)
Lt MCA infarction	3.16	4.2	14.5	2.3	3.2	19.1	0.26	1.6	12.9	4.26	8.4	21.7	0 (9)	0 (9)	0 (9)	0 (8)
Lt MCA infarction	1.92	3.6	14.6	3.14	4.2	14.5	1.84	5.7	39.5	0.75	2.9	19.6	0 (9)	0 (9)	0 (9)	0 (9)
Lt pontine infarction	-0.67	0.6	19.5	-1.37	1.3	12.9	-11.93	-10.3	16.7	-4.93	-2.1	17.3	1 (7)	0 (8)	1 (7)	1 (8)
Lt thalamic infarction	2.05	3.5	22.8	8.91	11.4	12.5	4.77	8.8	31.3	1.98	6.8	37.7	0 (9)	1 (7)	0 (9)	1 (8)
Pontine ICH ^k	-1.57	1.5	39.1	0.81	2	18.5	-3	1.2	40	3.18	5.3	16.5	1 (7)	0 (9)	1 (8)	0 (9)
Rt ^l MCA infarction	-9.96	-7.5	17.9	-1.93	-0.6	19	-2.71	0.4	18.5	-1.99	-0.3	17.2	1 (7)	0 (9)	1 (7)	0 (9)
Lt internal capsule infarction	-6	-7.9	14	-0.8	-2	11.6	1.8	0.8	18.6	11	6.5	38.5	0 (9)	0 (9)	0 (8)	0 (9)
Myelitis (no weakness)	1.3	2.9	18.6	-0.56	0.1	11.7	-1.23	1.2	24.1	-1.14	0.7	24	0 (9)	0 (9)	0 (9)	0 (9)
Rt MCA infarction	-4.97	-6.4	19.2	0.7	0	13.1	13.9	7	49.3	6.31	2.3	34.3	0 (9)	0 (9)	1 (7)	0 (9)
Myasthenia gravis	-0.64	1.3	19.2	1.1	2.7	14.4	-1.97	0	18.5	-0.64	2.7	22.6	0 (9)	0 (9)	0 (9)	0 (9)
Lt pontine infarction	15.5	5.4	41.1	23.5	12	54	6.3	2.2	26.1	5.3	0.6	46.2	0 (9)	1 (7)	0 (9)	1 (7)
Pontine hemorrhage	-0.83	1.1	19	-2.72	1.3	26.6	1.69	3.3	13.6	-7.52	-0.8	54.5	1 (8)	1 (8)	1 (8)	1 (7)

^aNIHSS: National Institutes of Health Stroke Scale.^bMRC: Medical Research Council.^cULL: upper left limb.^dURL: upper right limb.^eLLL: lower left limb.^fLRL: lower right limb.^gMax: maximum.^hOsc: oscillation.ⁱLt: left.^jMCA: middle cerebral artery.^kICH: intracerebral hemorrhage.^lRt: right.

Evaluation Outcomes

We evaluated the performance of the system in terms of the accuracy, sensitivity, specificity, precision, F1 score, and area under the receiver operating characteristics curve (AUC) with a confusion matrix.

The statistical plots in Figure 5 show the patterns of the average, maximum, and oscillation of the 4-limb features of each NIHSS grade. Auto-NIHSS discriminated those features, as shown in the confusion matrices in Figure 6. The result shows that the proposed autonomous grading achieved an accuracy of at least 80% and that the overall accuracy was 81.7%, as shown in the

summary of performance in Table 3. The AUC of auto-NIHSS reached 0.912, as depicted in the receiver operating characteristics curves in Figure 6. The sensitivity of the NIHSS grading reached 0.825 with the SVM and 0.875 with the ensemble. The specificity was 0.750 for both models.

Auto-MRC discriminates instances into 3 MRC grades, and the statistical plots of movement features are depicted in Figure 7. The mean AUC was 0.870 for the SVM and 0.877 for the ensemble, as shown in Figure 8. Table 4 shows the summarized performance of auto-MRC; the average accuracy, sensitivity, and specificity for the MRC grading were 0.775, 0.717, and 0.876, respectively.

Figure 5. Statistical plots of 4-limb features of NIHSS grades. NIHSS: National Institutes of Health Stroke Scale.

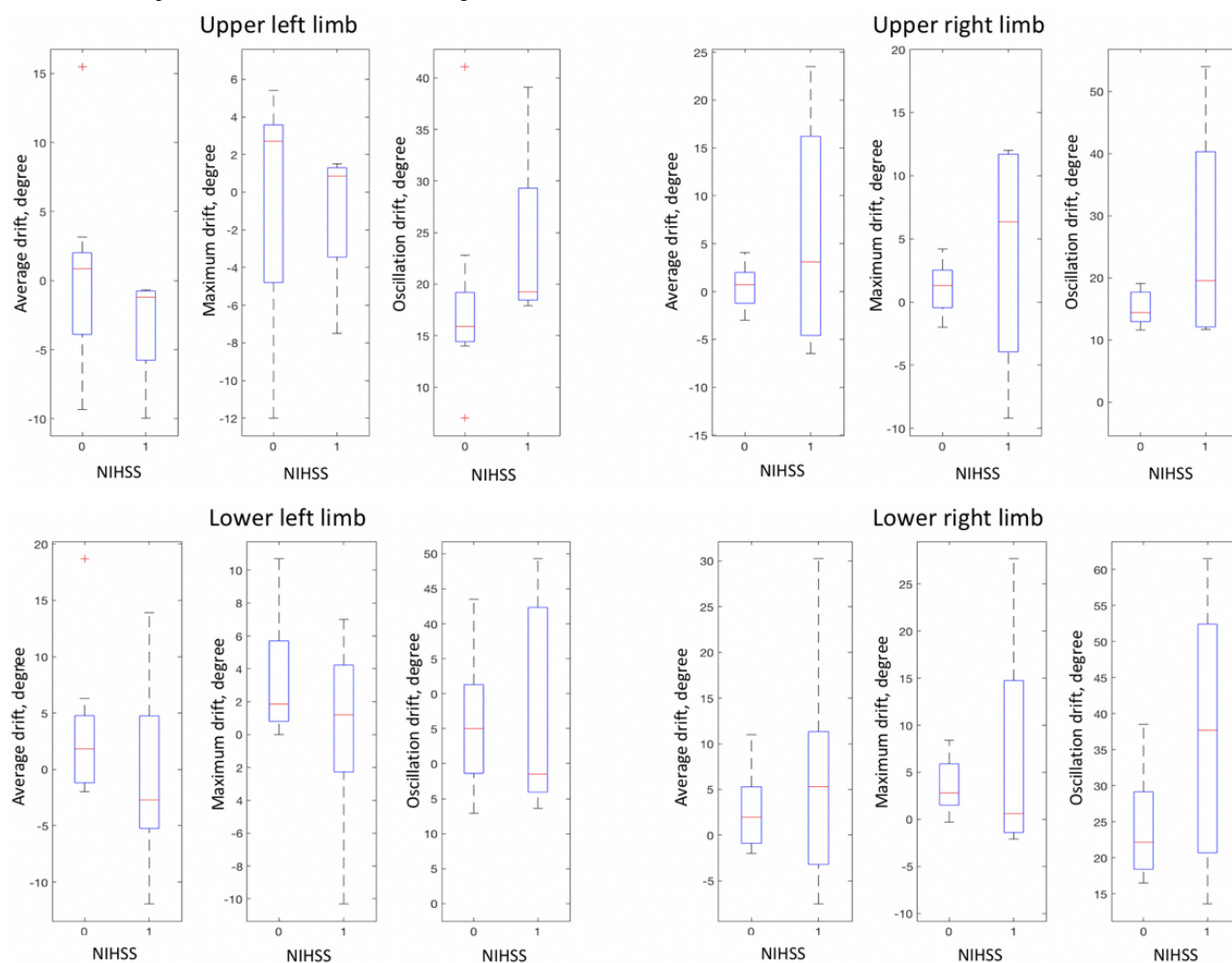


Figure 6. Confusion matrix and receiver operating characteristic of auto-NIHSS grading using (A) support vector machine and (B) ensemble learning. AUC: area under the receiver operating characteristics curve; NIHSS: National Institutes of Health Stroke Scale.

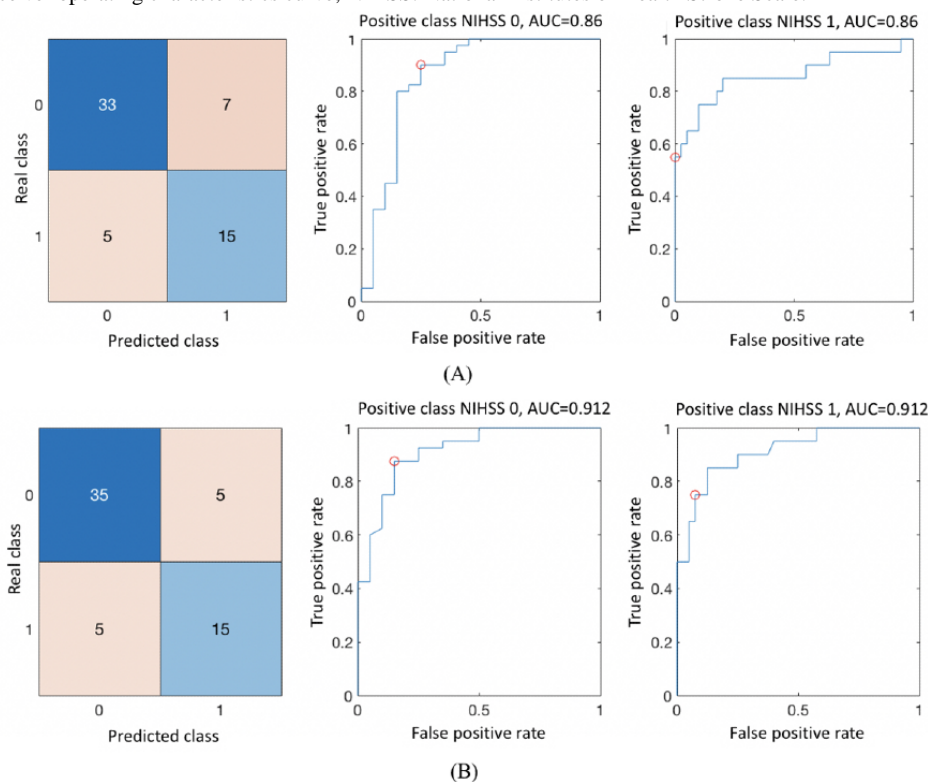


Figure 7. Statistical plots of 4-limb features of MRC grades. MRC: Medical Research Council.

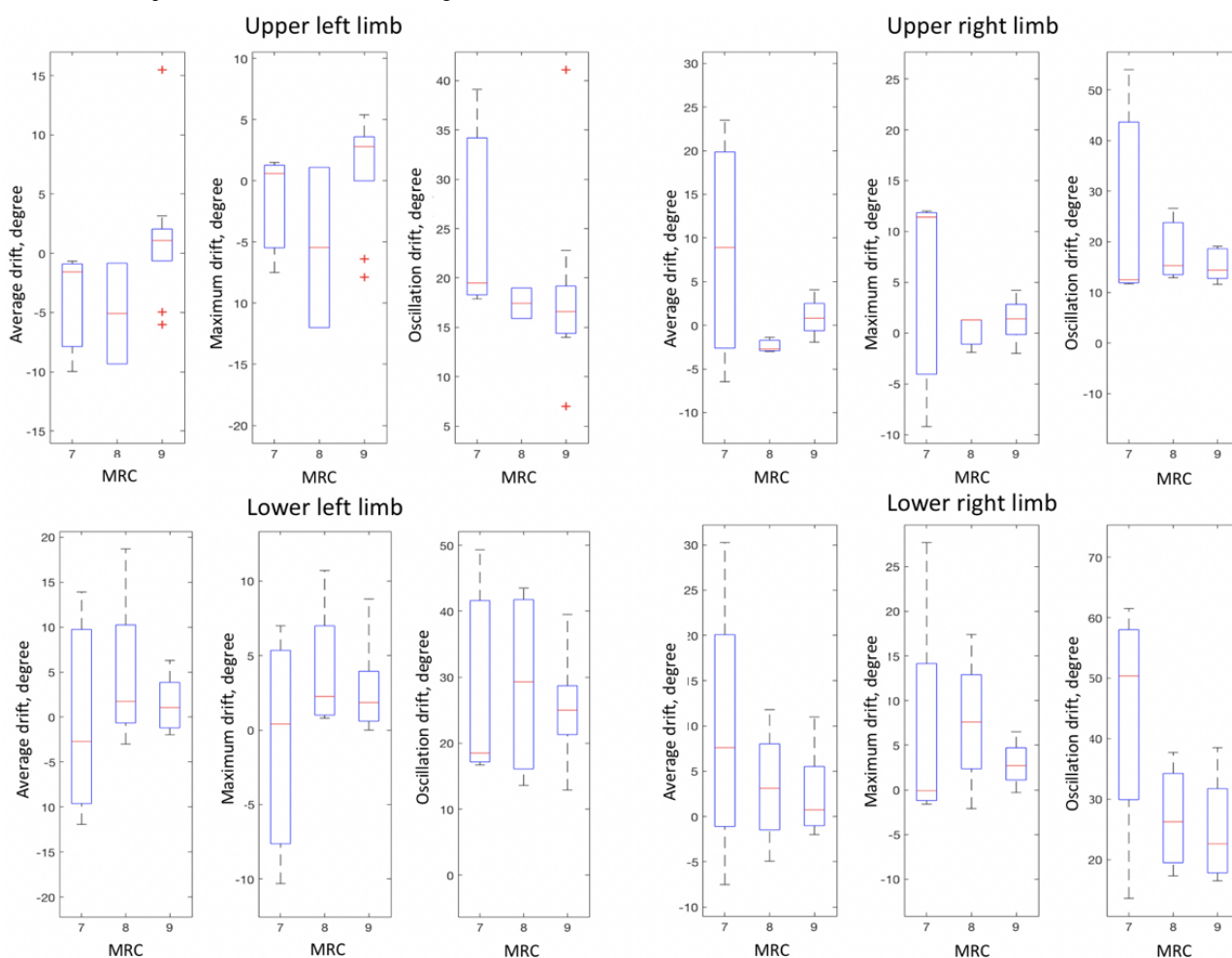


Figure 8. Confusion matrix and receiver operating characteristic of auto-MRC grading using (A) support vector machine and (B) ensemble learning. AUC: area under the receiver operating characteristics curve; MRC: Medical Research Council.

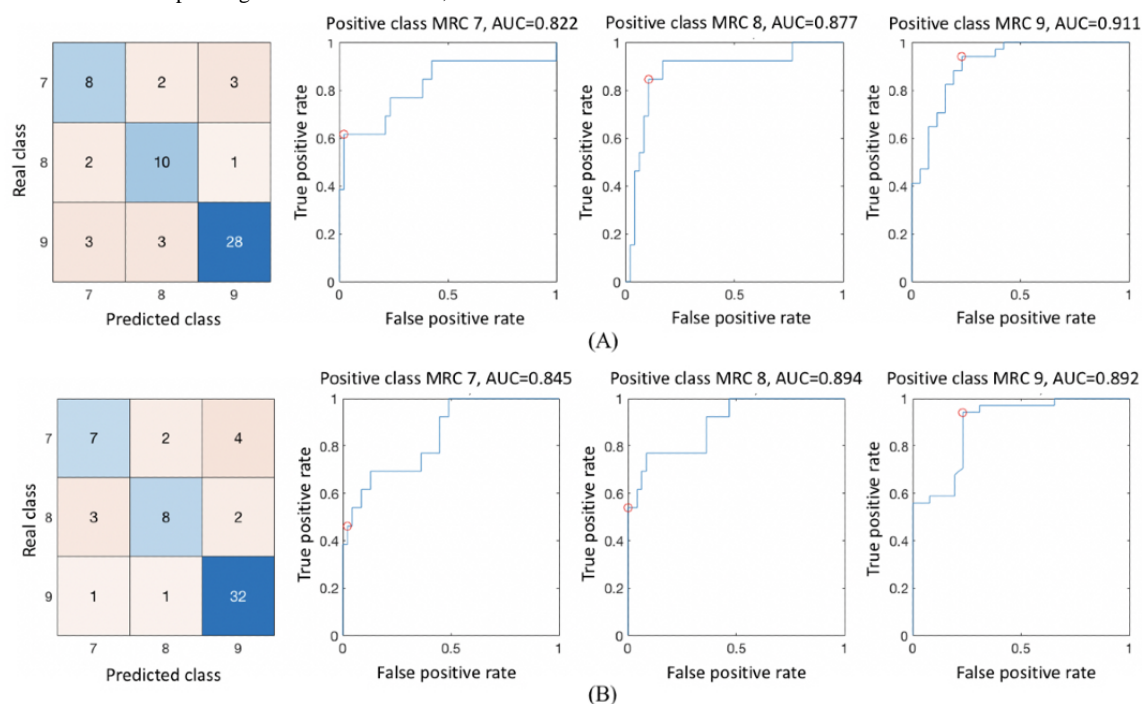


Table 3. Performance of auto-NIHSS grading.

Auto-NIHSS ^a grading	Accuracy	Sensitivity	Specificity	Precision	F1 score
SVM ^b	0.800	0.825	0.750	0.868	0.846
Ensemble	0.833	0.875	0.750	0.875	0.875

^aNIHSS: National Institutes of Health Stroke Scale.

^bSVM: support vector machine.

Table 4. Performance of auto-MRC grading.

Auto-MRC ^a grading	Accuracy	Sensitivity	Specificity	Precision	F1 score
SVM ^b	0.767	0.736	0.878	0.719	0.726
Ensemble	0.783	0.698	0.873	0.735	0.713

^aMRC: Medical Research Council.

^bSVM: support vector machine.

Discussion

Importance of Objective and Fast Assessment of Stroke Severity

The notion “time is brain” is valid in treating stroke patients. Intravenous tissue plasminogen activator (IV tPA) within 4.5 hours of stroke onset is the only therapy for acute ischemic stroke [22]. Subsequently, endovascular thrombectomy (EVT) has been a standard of care for patients with acute ischemic stroke caused by large artery occlusion within 6 to 24 hours of onset, based on successful large randomized clinical trials [23]. Reperfusion therapy, including IV tPA and EVT, for acute ischemic stroke is time sensitive (ie, an earlier treatment yields a better outcome). As the onset-to-intervention time is composed

of prehospital and in-hospital phases, patients who arrive early have more chances of appropriate treatment [24–27]. Delays in hospital admission and the preparation before treatment affect the prognosis of patients [28]. In Goyal et al [24], the authors reported that the most significant issue was getting the correct patient to the correct hospital quickly. In Sukumaran et al [27], strategies for stroke patient workflow optimization were suggested by analyzing and solving prehospital and preprocedural bottlenecks. The interhospital transfer is directly associated with delays in onset to reperfusion time, which results in the poor outcome of stroke patients; therefore, the timely triage of patients is a significant bottleneck [27].

The importance of accurate and objective assessments of stroke severity in telemedicine and telestroke strategies has been

discussed in numerous studies [29]. In particular, the timing constraint in performing reperfusion therapy, which has been shown to significantly reduce mortality, invokes the development of efficient systems and protocols in prehospital care or emergency medical systems. Researchers have addressed the fact that the rapid and accurate evaluation of stroke severity can aid in identifying patients for treatments and accelerate an urgent streamlined process. In the study by Andsberg et al [30], a prehospital ambulance stroke test was performed to score the severity of stroke through commands, answers, and observations. The remote assessment of stroke using smartphones was proposed and compared with bedside examination in calculating the NIHSS score [31]. However, most assessments in those systems used conservative observation or campaigns that were subjective and unreliable between testers. Modern communication, sensor technology, and machine learning can solve this problem through accurate measurements and the fast determination of assessment in a prehospital or remote environment [29,32,33]. A previous study evaluated arm function in activities using kinematic exposure variation analysis and inertial sensors [34]. A mobile-based walk test was developed to report patients' walking ability [35], and upper limb impairments in stroke patients were measured using inertial sensors in the home environment [33]. Such sensor-based testing enables objective evaluation regardless of the testers or place.

Utility of Consistent Grading Method as an Agreement Between Prehospital and Hospital Environment

The necessity of a controlled test is revealed in the results of previous studies for monitoring daily living. Motor recovery was monitored using accelerometers, and the NIHSS motor index was estimated in the study by Gubbi et al [36]. However, the movement in daily living limited the accuracy of estimation to 56% for the low index. Activity monitoring in most sensor-based studies involved trials that were not approved by clinical protocols. Those systems limited extensibility as a standard of remote monitoring systems, although they were efficient in tracking the progress or the treatment outcome.

In addition to rapid and accurate measurements, we aimed to increase the utility of the assessment system in the prehospital and hospital environments. At every phase of the prehospital process, consistent methods to conduct assessments can reduce errors and delays in communication among the participants of a community's emergency group. Therefore, automatic scoring can facilitate agreement in assessments among patients, caregivers, paramedics, and medical staff. With regard to bottleneck analysis in acute stroke treatment, the rapid identification of neurological deficits and assessment of motor grading will aid EMS personnel in transporting patients to a comprehensive stroke center because hospitals may be limited in terms of stroke unit availability and resources. In Berglund et al [26], the importance of stroke identification without meeting the patient or without neurological examination was asserted; the time to treatment can be decreased with the high-priority dispatch of ambulances through early identification of stroke from emergency calls. In the hyper acute stroke alarm study [25], researchers observed that higher prehospital priority levels of stroke improved thrombolysis frequency and time to stroke unit. The stroke identification by EMS dispatchers during

emergency calls varied between 31% and 57%, as identifying stroke can be a challenge without examination [26].

Therefore, we developed an automatic grading system, leveraging multiclassification of machine learning using typically performed tests and grading in clinics. Our proposed solution uses controlled observations of drift tests in clinics and can estimate the assessment by neurologists. Therefore, the scores by the automatic grading system can be instantly used for communication in an objective manner.

Data and Techniques for Clinical Scoring by Machine Learning

A considerable number of studies have used artificial intelligence, including machine learning, to estimate clinical scores and assess patients or provide warnings regarding adverse events [37-40]. In those studies, a series of various techniques were used according to the scale of scores, the capacity of collected data, and the skewness of data. Following the significant development of enhanced algorithms, data with significant meaning have gained importance. However, as addressed in Li et al [41], real-world data have a long-tail pattern with a significant imbalance in quality and quantity. Many algorithms have used public big data to develop new algorithms and build models; however, real-world applications have completely different data quality and quantity and cannot directly apply those models. This situation is particularly severe in medicine, as discussed in Hulsén et al [42]. The availability of qualified data differs by disease, severity of disease in patients, and difficulty of collection [43]. Big data from electronic medical records that are already facilitated in hospital information systems can be used in comparatively easy tasks for medical artificial intelligence. The recent success of medical artificial intelligence requires significant effort and cost in collecting and labeling data [44,45]. In addition, machine learning for sporadic events in emergencies or patients with rare diseases is affected by data deficiency. This is because interventions for collecting data can affect the prognosis of treatment due to the possible delay in the rapid streamlining of treatment processes. Previous feasibility studies have stated that the difficulty in real-time capturing of acute neurological disorders was the main limitation in the research [33,46].

The learning models with imbalanced data were affected by low precision or recall in the validation and test phases, although they achieved high accuracy for a large number of data in the majority groups [47]. Recently, techniques to solve this data skewness, including data augmentation, transfer learning, and deep imbalanced learning, were emphasized [48-51]. Studies on deep learning that extract filtered features derived from raw data have attempted to solve the problem by knowledge transfer from pretrained models [52,53] or with data augmentation [54,55]. Machine learning with records can cope with the imbalance problem through sampling, cost-sensitive learning, boosting algorithms, and skew-related performance metrics [47,56]. We used the SMOTE to balance between classes in the training phase and applied techniques, including RUSBoost, in optimized ensemble machine learning. To compare different models according to their precision on each class, the F measure is typically used as a performance metric [57]; additionally, we

validated the performance of the proposed solution using the AUC and F1 scores. Consequently, the performances of auto-NIHSS and auto-MRC indicated the acceptable AUC, sensitivity, specificity, and F1 score as real-world applications with data skewness.

Conclusion

Accurate monitoring and grading of motor weakness are critical for the appropriate assessment of stroke severity, particularly

for reliable and consistent evaluations. We developed an automatic grading system to assess proximal motor weakness using the kinematic features of unintended drift of 4 limbs. We trained optimized machine learning models and obtained promising results in scoring NIHSS and MRC. The objective scoring of neurological deficits can be used to identify stroke patients, dispatch patients to the appropriate medical center, and expedite treatment preparation.

Acknowledgments

This research was supported by a grant funded by the Ministry of Science and ICT (NRF-2020R1A2C1013152) and by a grant of the Korea Health Technology Research and Development Project funded by the Ministry of Health and Welfare (HI19C0481, HC19C0028), Republic of Korea.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Algorithm of measurement unit for automatic grading.

[\[PNG File , 699 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Algorithm of the grading unit for extracting features and training machine learning algorithms. MRC: Medical Research Council; NIHSS: National Institutes of Health Stroke Scale; SMOTE: synthetic minority oversampling technique.

[\[PNG File , 1029 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Sample measurement of unintended drift of limbs.

[\[PNG File , 1259 KB-Multimedia Appendix 3\]](#)

References

1. Darcy P, Moughty AM. Pronator Drift. *N Engl J Med* 2013 Oct 17;369(16):e20. [doi: [10.1056/nejmicm1213343](https://doi.org/10.1056/nejmicm1213343)]
2. Nor AM, McAllister C, Louw S, Dyker A, Davis M, Jenkinson D, et al. Agreement Between Ambulance Paramedic- and Physician-Recorded Neurological Signs With Face Arm Speech Test (FAST) in Acute Stroke Patients. *Stroke* 2004 Jun;35(6):1355-1359. [doi: [10.1161/01.str.0000128529.63156.c5](https://doi.org/10.1161/01.str.0000128529.63156.c5)]
3. Kim J, Fonarow GC, Smith EE, Reeves MJ, Navalkale DD, Grotta JC, et al. Treatment With Tissue Plasminogen Activator in the Golden Hour and the Shape of the 4.5-Hour Time-Benefit Curve in the National United States Get With The Guidelines-Stroke Population. *Circulation* 2017 Jan 10;135(2):128-139. [doi: [10.1161/circulationaha.116.023336](https://doi.org/10.1161/circulationaha.116.023336)]
4. Kodankandath TV, Wright P, Power PM, De Geronimo M, Libman RB, Kwiatkowski T, et al. Improving Transfer Times for Acute Ischemic Stroke Patients to a Comprehensive Stroke Center. *J Stroke Cerebrovasc Dis* 2017 Jan;26(1):192-195. [doi: [10.1016/j.jstrokecerebrovasdis.2016.09.008](https://doi.org/10.1016/j.jstrokecerebrovasdis.2016.09.008)] [Medline: [27743926](https://pubmed.ncbi.nlm.nih.gov/27743926/)]
5. Threlkeld ZD, Kozak B, McCoy D, Cole S, Martin C, Singh V. Collaborative Interventions Reduce Time-to-Thrombolysis for Acute Ischemic Stroke in a Public Safety Net Hospital. *J Stroke Cerebrovasc Dis* 2017 Jul;26(7):1500-1505. [doi: [10.1016/j.jstrokecerebrovasdis.2017.03.004](https://doi.org/10.1016/j.jstrokecerebrovasdis.2017.03.004)] [Medline: [28396187](https://pubmed.ncbi.nlm.nih.gov/28396187/)]
6. Park E, Chang H, Nam HS. Use of Machine Learning Classifiers and Sensor Data to Detect Neurological Deficit in Stroke Patients. *J Med Internet Res* 2017 Apr 18;19(4):e120 [FREE Full text] [doi: [10.2196/jmir.7092](https://doi.org/10.2196/jmir.7092)] [Medline: [28420599](https://pubmed.ncbi.nlm.nih.gov/28420599/)]
7. Singer OC, Dvorak F, du Mesnil de Rochemont R, Lanfermann H, Sitzler M, Neumann-Haefelin T. A Simple 3-Item Stroke Scale. *Stroke* 2005 Apr;36(4):773-776. [doi: [10.1161/01.str.0000157591.61322.df](https://doi.org/10.1161/01.str.0000157591.61322.df)]
8. Pérez de la Ossa N, Carrera D, Gorchs M, Querol M, Millán M, Gomis M, et al. Design and Validation of a Prehospital Stroke Scale to Predict Large Arterial Occlusion. *Stroke* 2014 Jan;45(1):87-91. [doi: [10.1161/strokeaha.113.003071](https://doi.org/10.1161/strokeaha.113.003071)]
9. Katz BS, McMullan JT, Sucharew H, Adeoye O, Broderick JP. Design and Validation of a Prehospital Scale to Predict Stroke Severity. *Stroke* 2015 Jun;46(6):1508-1512. [doi: [10.1161/strokeaha.115.008804](https://doi.org/10.1161/strokeaha.115.008804)]
10. Hastrup S, Damgaard D, Johnsen SP, Andersen G. Prehospital Acute Stroke Severity Scale to Predict Large Artery Occlusion. *Stroke* 2016 Jul;47(7):1772-1776. [doi: [10.1161/strokeaha.115.012482](https://doi.org/10.1161/strokeaha.115.012482)]

11. Heldner MR, Jung S, Zubler C, Mordasini P, Weck A, Mono M, et al. Outcome of patients with occlusions of the internal carotid artery or the main stem of the middle cerebral artery with NIHSS score of less than 5: comparison between thrombolysed and non-thrombolysed patients. *J Neurol Neurosurg Psychiatry* 2015 Jul;86(7):755-760. [doi: [10.1136/jnnp-2014-308401](https://doi.org/10.1136/jnnp-2014-308401)] [Medline: [25266203](https://pubmed.ncbi.nlm.nih.gov/25266203/)]
12. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. *Stroke* 2000 May;31(4):858-862. [doi: [10.1161/01.str.31.4.858](https://doi.org/10.1161/01.str.31.4.858)] [Medline: [10753988](https://pubmed.ncbi.nlm.nih.gov/10753988/)]
13. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999 Jul;54(7):581-586 [FREE Full text] [doi: [10.1136/thx.54.7.581](https://doi.org/10.1136/thx.54.7.581)] [Medline: [10377201](https://pubmed.ncbi.nlm.nih.gov/10377201/)]
14. Paternostro-Sluga T, Grim-Stieger M, Posch M, Schuhfried O, Vacariu G, Mittermaier C, et al. Reliability and validity of the Medical Research Council (MRC) scale and a modified scale for testing muscle strength in patients with radial palsy. *J Rehabil Med* 2008 Aug;40(8):665-671 [FREE Full text] [doi: [10.2340/16501977-0235](https://doi.org/10.2340/16501977-0235)] [Medline: [19020701](https://pubmed.ncbi.nlm.nih.gov/19020701/)]
15. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *Jair* 2002 Jun 01;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]
16. Verbiest N, Ramentol E, Cornelis C, Herrera F. Preprocessing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection. *Applied Soft Computing* 2014 Sep;22:511-517. [doi: [10.1016/j.asoc.2014.05.023](https://doi.org/10.1016/j.asoc.2014.05.023)]
17. Yijing L, Haixiang G, Xiao L, Yanan L, Jinling L. Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. *Knowledge-Based Systems* 2016 Feb;94:88-104. [doi: [10.1016/j.knsys.2015.11.013](https://doi.org/10.1016/j.knsys.2015.11.013)]
18. Douzas G, Bacao F, Last F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences* 2018 Oct;465:1-20. [doi: [10.1016/j.ins.2018.06.056](https://doi.org/10.1016/j.ins.2018.06.056)]
19. Burman P. A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika* 1989 Sep;76(3):503. [doi: [10.2307/2336116](https://doi.org/10.2307/2336116)]
20. Snoek J, Larochelle H, Adams R. Practical Bayesian optimization of machine learning algorithms. 2012 Presented at: Annual Advances in Neural Information Processing Systems 2012; Dec 3-6, 2012; Lake Tahoe, NV URL: <https://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf>
21. MATLAB R2020a. URL: <https://mathworks.com> [accessed 2020-08-25]
22. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. Thrombolysis with Alteplase 3 to 4.5 Hours after Acute Ischemic Stroke. *N Engl J Med* 2008 Sep 25;359(13):1317-1329. [doi: [10.1056/nejmoa0804656](https://doi.org/10.1056/nejmoa0804656)]
23. Goyal M, Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *The Lancet* 2016 Apr;387(10029):1723-1731. [doi: [10.1016/s0140-6736\(16\)00163-x](https://doi.org/10.1016/s0140-6736(16)00163-x)]
24. Goyal M, Jadhav AP, Wilson AT, Nogueira RG, Menon BK. Shifting bottlenecks in acute stroke treatment. *J Neurointerv Surg* 2016 Dec;8(11):1099-1100. [doi: [10.1136/neurintsurg-2015-012151](https://doi.org/10.1136/neurintsurg-2015-012151)] [Medline: [26674088](https://pubmed.ncbi.nlm.nih.gov/26674088/)]
25. Berglund A, Svensson L, Sjöstrand C, von Arbin M, von Euler M, Wahlgren N, et al. Higher Prehospital Priority Level of Stroke Improves Thrombolysis Frequency and Time to Stroke Unit. *Stroke* 2012 Oct;43(10):2666-2670. [doi: [10.1161/strokeaha.112.652644](https://doi.org/10.1161/strokeaha.112.652644)]
26. Berglund A, von Euler M, Schenck-Gustafsson K, Castrén M, Böhm K. Identification of stroke during the emergency call: a descriptive study of callers' presentation of stroke. *BMJ Open* 2015 May 28;5(4):e007661 [FREE Full text] [doi: [10.1136/bmjopen-2015-007661](https://doi.org/10.1136/bmjopen-2015-007661)] [Medline: [25922106](https://pubmed.ncbi.nlm.nih.gov/25922106/)]
27. Sukumaran M, Cantrell D, Ansari S, Huryley M, Shaibani A, Potts M. Stroke patient workflow optimization. *Endovasc Today* 2019 Feb;18(2):46-50 [FREE Full text]
28. Itrat A, Taqui A, Cerejo R, Briggs F, Cho S, Organek N, Cleveland Pre-Hospital Acute Stroke Treatment Group. Telemedicine in Prehospital Stroke Evaluation and Thrombolysis: Taking Stroke Treatment to the Doorstep. *JAMA Neurol* 2016 Mar;73(2):162-168. [doi: [10.1001/jamaneurol.2015.3849](https://doi.org/10.1001/jamaneurol.2015.3849)] [Medline: [26641366](https://pubmed.ncbi.nlm.nih.gov/26641366/)]
29. Hess DC, Audebert HJ. The history and future of telestroke. *Nat Rev Neurol* 2013 Jul;9(6):340-350. [doi: [10.1038/nrneurol.2013.86](https://doi.org/10.1038/nrneurol.2013.86)] [Medline: [23649102](https://pubmed.ncbi.nlm.nih.gov/23649102/)]
30. Andsberg G, Esbjörnsson M, Olofsson A, Lindgren A, Norrving B, von Euler M. PreHospital Ambulance Stroke Test - pilot study of a novel stroke test. *Scand J Trauma Resusc Emerg Med* 2017 May 11;25(1):37 [FREE Full text] [doi: [10.1186/s13049-017-0377-x](https://doi.org/10.1186/s13049-017-0377-x)] [Medline: [28399897](https://pubmed.ncbi.nlm.nih.gov/28399897/)]
31. Anderson ER, Smith B, Ido M, Frankel M. Remote assessment of stroke using the iPhone 4. *J Stroke Cerebrovasc Dis* 2013 May;22(4):340-344. [doi: [10.1016/j.jstrokecerebrovasdis.2011.09.013](https://doi.org/10.1016/j.jstrokecerebrovasdis.2011.09.013)] [Medline: [22018507](https://pubmed.ncbi.nlm.nih.gov/22018507/)]
32. Alber M, Buganza Tepole A, Cannon WR, De S, Dura-Bernal S, Garikipati K, et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digit Med* 2019;2:115 [FREE Full text] [doi: [10.1038/s41746-019-0193-y](https://doi.org/10.1038/s41746-019-0193-y)] [Medline: [31799423](https://pubmed.ncbi.nlm.nih.gov/31799423/)]
33. Held JPO, Klaassen B, Eenhoorn A, van Beijnum BJF, Buurke JH, Veltink PH, et al. Inertial Sensor Measurements of Upper-Limb Kinematics in Stroke Patients in Clinic and Home Environment. *Front Bioeng Biotechnol* 2018;6:27 [FREE Full text] [doi: [10.3389/fbioe.2018.00027](https://doi.org/10.3389/fbioe.2018.00027)] [Medline: [29707537](https://pubmed.ncbi.nlm.nih.gov/29707537/)]

34. Ertzgaard P, Öhberg F, Gerdle B, Grip H. A new way of assessing arm function in activity using kinematic Exposure Variation Analysis and portable inertial sensors--A validity study. *Man Ther* 2016 Mar;21:241-249. [doi: [10.1016/j.math.2015.09.004](https://doi.org/10.1016/j.math.2015.09.004)] [Medline: [26456185](https://pubmed.ncbi.nlm.nih.gov/26456185/)]
35. Salvi D, Poffley E, Orchard E, Tarassenko L. The Mobile-Based 6-Minute Walk Test: Usability Study and Algorithm Development and Validation. *JMIR Mhealth Uhealth* 2020 Jan 03;8(1):e13756 [FREE Full text] [doi: [10.2196/13756](https://doi.org/10.2196/13756)] [Medline: [31899457](https://pubmed.ncbi.nlm.nih.gov/31899457/)]
36. Gubbi J, Rao AS, Fang K, Yan B, Palaniswami M. Motor recovery monitoring using acceleration measurements in post acute stroke patients. *BioMed Eng OnLine* 2013;12(1):33. [doi: [10.1186/1475-925x-12-33](https://doi.org/10.1186/1475-925x-12-33)]
37. Lei H, Huang Z, Zhang J, Yang Z, Tan E, Zhou F, et al. Joint detection and clinical score prediction in Parkinson's disease via multi-modal sparse learning. *Expert Systems with Applications* 2017 Sep;80:284-296. [doi: [10.1016/j.eswa.2017.03.038](https://doi.org/10.1016/j.eswa.2017.03.038)]
38. Campanella G, Hanna MG, Geneslaw L, Mirafior A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019 Aug;25(8):1301-1309 [FREE Full text] [doi: [10.1038/s41591-019-0508-1](https://doi.org/10.1038/s41591-019-0508-1)] [Medline: [31308507](https://pubmed.ncbi.nlm.nih.gov/31308507/)]
39. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Critical Care Medicine* 2018;46(4):547-553. [doi: [10.1097/ccm.0000000000002936](https://doi.org/10.1097/ccm.0000000000002936)]
40. Kim J, Chae M, Chang H, Kim Y, Park E. Predicting Cardiac Arrest and Respiratory Failure Using Feasible Artificial Intelligence with Simple Trajectories of Patient Data. *J Clin Med* 2019 Aug 29;8(9):1336 [FREE Full text] [doi: [10.3390/jcm8091336](https://doi.org/10.3390/jcm8091336)] [Medline: [31470543](https://pubmed.ncbi.nlm.nih.gov/31470543/)]
41. Li Q, Li Y, Gao J, Su L, Zhao B, Demirbas M, et al. A confidence-aware approach for truth discovery on long-tail data. *Proc VLDB Endow* 2014 Dec;8(4):425-436. [doi: [10.14778/2735496.2735505](https://doi.org/10.14778/2735496.2735505)]
42. Hulsén T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From Big Data to Precision Medicine. *Front Med (Lausanne)* 2019;6:34 [FREE Full text] [doi: [10.3389/fmed.2019.00034](https://doi.org/10.3389/fmed.2019.00034)] [Medline: [30881956](https://pubmed.ncbi.nlm.nih.gov/30881956/)]
43. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017 Dec;2(4):230-243 [FREE Full text] [doi: [10.1136/svn-2017-000101](https://doi.org/10.1136/svn-2017-000101)] [Medline: [29507784](https://pubmed.ncbi.nlm.nih.gov/29507784/)]
44. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016 Dec 13;316(22):2402-2410. [doi: [10.1001/jama.2016.17216](https://doi.org/10.1001/jama.2016.17216)] [Medline: [27898976](https://pubmed.ncbi.nlm.nih.gov/27898976/)]
45. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 Jan 25;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)]
46. Qassem T. Emerging technologies for dementia patient monitoring. In: Xhafa F, Moore P, Tadros G, editors. *Advanced Technological Solutions for E-Health and Dementia Patient Monitoring*. Dallas, TX: IGI Global; 2015:62-104.
47. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A Hybrid Approach to Alleviating Class Imbalance. *IEEE Trans Syst, Man, Cybern A* 2010 Jan;40(1):185-197. [doi: [10.1109/tsmca.2009.2029559](https://doi.org/10.1109/tsmca.2009.2029559)]
48. Ammar H, Eaton E, Ruvolo P, Taylor M. Unsupervised cross-domain transfer in policy gradient reinforcement learning via manifold alignment. 2015 Presented at: Twenty-Ninth AAAI Conference on Artificial Intelligence; Jan 25-30, 2015; Austin, TX.
49. Pan SJ, Yang Q. A Survey on Transfer Learning. *IEEE Trans Knowl Data Eng* 2010 Oct;22(10):1345-1359. [doi: [10.1109/tkde.2009.191](https://doi.org/10.1109/tkde.2009.191)]
50. Mikolajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. 2018 Presented at: 2018 International interdisciplinary PhD workshop (IIPhDW); May 9-12, 2018; Swinoujskie, Poland. [doi: [10.1109/iiphdw.2018.8388338](https://doi.org/10.1109/iiphdw.2018.8388338)]
51. Huang C, Li Y, Chen CL, Tang X. Deep Imbalanced Learning for Face Recognition and Attribute Prediction. *IEEE Trans Pattern Anal Mach Intell* 2019;1-1. [doi: [10.1109/tpami.2019.2914680](https://doi.org/10.1109/tpami.2019.2914680)]
52. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med* 2019 Jan;81(1):116-128 [FREE Full text] [doi: [10.1002/mrm.27355](https://doi.org/10.1002/mrm.27355)] [Medline: [29774597](https://pubmed.ncbi.nlm.nih.gov/29774597/)]
53. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imag* 2016 Jul 01;3(3):034501. [doi: [10.1117/1.jmi.3.3.034501](https://doi.org/10.1117/1.jmi.3.3.034501)]
54. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. 2017 Presented at: AMIA 2017 Annual Symposium Proceedings; Nov 4-8, 2017; Washington, DC.
55. Shin H, Tenenholtz NA, Rogers JK, Schwarz CG, Senjem ML, Gunter JL, et al. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. 2018 Presented at: International workshop on simulation and synthesis in medical imaging; Sep 6, 2018; Granada, Spain.
56. Jeni L, Cohn J, De LTF. Facing imbalanced data--recommendations for the use of performance metrics. 2013 Presented at: Humaine association conference on affective computing and intelligent interaction; Sep 2-5, 2013; Geneva, Switzerland. [doi: [10.1109/acii.2013.47](https://doi.org/10.1109/acii.2013.47)]
57. López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences* 2013 Nov;250:113-141. [doi: [10.1016/j.ins.2013.07.007](https://doi.org/10.1016/j.ins.2013.07.007)]

Abbreviations

AUC: area under the receiver operating characteristic curve
EMS: emergency medical service
EVT: endovascular thrombectomy
IV tPA: intravenous tissue plasminogen activator
K-NN: K-nearest neighbor
MRC: Medical Research Council
NIHSS: National Institutes of Health Stroke Scale
SMOTE: synthetic minority oversampling technique
SVM: support vector machine

Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by Y Arai, KL Ong; comments to author 18.06.20; revised version received 13.08.20; accepted 13.08.20; published 16.09.20

Please cite as:

Park E, Lee K, Han T, Nam HS

Automatic Grading of Stroke Symptoms for Rapid Assessment Using Optimized Machine Learning and 4-Limb Kinematics: Clinical Validation Study

J Med Internet Res 2020;22(9):e20641

URL: <http://www.jmir.org/2020/9/e20641/>

doi: [10.2196/20641](https://doi.org/10.2196/20641)

PMID:

©Eunjeong Park, Kijeong Lee, Taehwa Han, Hyo Suk Nam. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.09.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.