University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Datta, S;Karmakar, CK;Rao, AS;Yan, B;Palaniswami, M

Title:
Automated Scoring of Hemiparesis in Acute Stroke From Measures of Upper Limb Co-Ordination Using Wearable Accelerometry.

Date:
2020-04

Citation:
Datta, S., Karmakar, C. K., Rao, A. S., Yan, B. & Palaniswami, M. (2020). Automated Scoring of Hemiparesis in Acute Stroke From Measures of Upper Limb Co-Ordination Using Wearable Accelerometry.. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28 (4), pp.805-816. https://doi.org/10.1109/TNSRE.2020.2972285.

Persistent Link:
https://hdl.handle.net/11343/302066

Automated Scoring of Hemiparesis in Acute Stroke from Measures of Upper Limb Co-ordination using Wearable Accelerometry

Shreyasi Datta, *Student Member, IEEE,* Chandan K. Karmakar, *Member, IEEE,*
Aravinda S. Rao, *Member, IEEE,* Bernard Yan and Marimuthu Palaniswami, *Fellow, IEEE*

***Abstract*—Stroke survivors usually experience paralysis in one half of the body, *i.e.*, hemiparesis, and the upper limbs are severely affected. Continuous monitoring of hemiparesis progression hours after the stroke attack involves manual observation of upper limb movements by medical experts in the hospital. Hence it is resource and time intensive, in addition to being prone to human errors and inter-rater variability. Wearable devices have found significance in automated continuous monitoring of neurological disorders like stroke. In this paper, we use accelerometer signals acquired using wrist-worn devices to analyze upper limb movements and identify hemiparesis in acute stroke patients, while they perform a set of proposed spontaneous and instructed movements. We propose novel measures of time (and frequency) domain coherence between accelerometer data from two arms at different lags (and frequency bands). These measures correlate well with the clinical gold standard of measurement of hemiparetic severity in stroke, the National Institutes of Health Stroke Scale (NIHSS). The study, undertaken on 32 acute stroke patients with varying levels of hemiparesis and 15 healthy controls, validates the use of short length (<10 minutes) accelerometry data to identify hemiparesis through leave-one-subject-out cross-validation based hierarchical discriminant analysis. The results indicate that the proposed approach can distinguish between controls, *moderate* and *severe* hemiparesis with an average accuracy of 91%.**

***Index Terms*—Accelerometry, acute stroke, assistive technology, biomedical devices, hemiparesis scoring, pattern recognition, upper-limb weakness, wearable sensors.**

I. Introduction

STROKE affects 15 million people worldwide every year, leading to death, disability and massive medical expenditures in patient monitoring[1]. Stroke often leads to paralysis in one half of the body *i.e.*, *hemiparesis* [1], which severely affects the upper limbs by limiting movements and co-ordination [2]–[4]. During the *acute* phase of stroke (within the first few days), reperfusion therapy with mechanical thrombectomy or intravenous thrombolysis is administered to resolve clots in the blocked artery [5]. Patients with and without early improvements with such therapy require different forms of treatment and rehabilitation later [6]. Hence, patients are monitored at regular intervals to assess the efficacy of the therapy by monitoring the severity of hemiparesis (improvement or deterioration).

The clinical gold standard for such assessment is the National Institutes of Health Stroke Scale (NIHSS) [7], that is assigned by studying the strength of a patient's arms held against gravity for a predefined amount of time [2]. This process needs to be carried out repeatedly at regular intervals, calling for 24x7 supervision by trained medical personnel. Being dependent on the availability of medical expertise and manual intervention, this is labour-intensive and can be potentially expensive, especially in public hospitals where the doctor-to-patient ratio is skewed [8], [9]. It is also prone to human errors and inter-rater variability [10]. Additionally, being single-point measurements, if there is a delay in the intervention, time-critical identification of deterioration can be missed [3]. Timely detection of motor deterioration is not only essential for monitoring the efficacy of reperfusion therapy, but also necessary to prevent recurrent strokes and post-stroke seizures [3], [11].

Therefore, an automated system for scoring hemiparesis, that is independent of continuous specialized medical attention, has tremendous research importance in acute stroke. Wearable devices with miniaturized sensors, that can unobtrusively and longitudinally measure and track information, can be beneficial towards the development of such a system [12]. Accelerometers are popular wearable motion sensors that measure the applied acceleration along a sensitive axis [13]. They are highly reliable for the measurement of frequency and intensity of human movements with little variability over time. Patterns of upper limb movements in neurological disorders like stroke and epilepsy can be analyzed using accelerometer data acquired from wearable devices [6], [14], [15]. Previous studies have shown that accelerometry can be used to quantify the amount of upper limb use in post-stroke hemiparetic patients using *activity count* [16], which is lower in hemiparetic subjects compared to the healthy counterparts. In addition, the *activity count* is found to correlate well with different clinical measures [16].

Various time and frequency domain features from accelerometer data, along with correlation among the sensor axes have been previously used to predict clinical scores in stroke patients [17], [18]. However, most of such literature focus only on activity monitoring in the rehabilitative phase (after the patient is discharged from the hospital), rather than assessing the severity of hemiparesis in the acute phase. Additionally, several sensors are used, which is not possible in view of the comfort of acute stroke patients. In our earlier work [6], wrist-worn accelerometer data is used to propose novel indices

S. Datta, A. S. Rao and M. Palaniswami are with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC - 3010, Australia, e-mails: shreyasid@student.unimelb.edu.au, aravinda.rao@unimelb.edu.au, palani@unimelb.edu.au

C. K. Karmakar is with the School of Information Technology, Deakin University, Burwood, Melbourne, VIC 3125, Australia, email:karmakar@deakin.edu.au

B. Yan is with Department of Neurology, Melbourne Brain Centre (MBC), Royal Melbourne Hospital, Parkville, VIC - 3052, Australia, e-mail: bernard.yan@mh.org.au

[1]http://www.world-stroke.org/

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

for predicting the severity of hemiparesis, with reference to NIHSS in acute stroke. The metrics comprise *Norm* (average arm activity computed from resultant acceleration), *SMA* (normalized sum of Signal Magnitude Area of three axes) and *Energy* (normalized sum of squared FFT magnitudes over each axis) based measures. Kumar *et al*. [14] introduced a metric based on pairwise angular cross-correlation between accelerometer axes from the two arms in a stroke patient as a measure of hemiparesis. This is motivated by the fact that the stroke affected arm exhibits limited rotatory movements compared to the normal arm. However, both these methods are implemented on a limited patient population and require long recordings ($>$1 hour) and expert annotations of NIHSS for a patient at multiple instants of time from the onset of stroke up to 24 hours after the attack, for validation of the methods. Additionally, in [6], no generalized learning model is developed to predict the severity of hemiparesis for unknown data. Therefore, there is need for formulating a set of upper limb movements for shorter periods of time, developing novel measures of co-ordination between the two arms from them and automatically grading hemiparetic severity in acute stroke from a learnt model. Although accelerometry based analysis have previously shown differences between hemiparetic subjects in acute phase and controls, it has been difficult to differentiate between different levels of hemiparesis [6].

In this paper, we use wrist-worn accelerometers to record upper limb movements in acute stroke patients for automated scoring of the severity of hemiparesis from short-length data. The clinical motivation behind this study is a wearable based automated system that would instruct acute stroke patients to perform short-duration movements (*e.g.*, by audio cues) and record these along with spontaneous movements. The objective behind including instructed movements along with spontaneous movements is that, during periods of rest (*e.g.* sleep) arms will show limited (or no) spontaneous movements even in non-stroke cases; but, that is not indicative of the presence of hemiparesis. Furthermore, acute stroke patients are often heavily sedated by medications, and unless instructed, may not perform any activity at all. Also, the non-dominant arm will always show lesser spontaneous movements. Thus a combination of spontaneous and instructed movements, analyzed at regular intervals can reliably predict the improvement or deterioration in patient's condition and relay the information to a remote doctor. Hence, the system can prove to be beneficial for intermittent analysis of hemiparetic severity avoiding repeated manual intervention by a doctor at the patient's bedside.

Towards that direction, in this paper, we investigate a set of upper limb movements and corresponding accelerometry features that can discriminate between different levels of hemiparetic severity. We hypothesize that such movements will be almost the same in the two arms of healthy subjects. But, with increasing hemiparesis, the differences in such movements between the affected and unaffected arm will increase. We propose to quantify these differences using accelerometry features based on relative motion and co-ordination (or the lack of it) between two arms. Finally, we demonstrate that our methods can accurately identify hemiparetic severity in a subject-independent manner from unknown data using pattern recognition algorithms. The main contributions are:

- Novel experimental protocol comprising spontaneous and instructed movements to capture upper limb weakness using *only* wrist-worn accelerometry
- Novel measures of left and right arm co-ordination using time and frequency domain cross-correlation on short length ($<$ 10 minutes) accelerometry data
- Automated classification of wrist-worn accelerometry data into different levels of hemiparesis with reference to the clinical gold standard in a hierarchical fashion

II. Data Acquisition

A. The National Institutes of Health Stroke Scale (NIHSS)

NIHSS is a 42-point scale used as the clinical gold standard for measuring stroke severity through various parameters [7]. NIHSS for quantifying motor weakness in acute stroke involves qualitative monitoring of upper limb movements [2], [6]. It is a 5-point scale that encodes the strength of an upper limb against gravity, with 0 indicating normal strength and 4 indicating no movements or a completely dead arm, as detailed in Table I.

TABLE I: NIHSS for upper limb weakness assessment

Score	Description
0	Subject can lift up their arm, hold it steady against gravity for 10 s, without any drift
1	Subject can lift up their arm, hold it against gravity for 10 s, but there are drifts from the horizontal
2	Subject can lift up their arm, but cannot hold it against gravity for 10 s, arm drops down
3	Subject cannot lift their arms up, but can shrug their shoulders or twist their wrists
4	Subject cannot move any part of the arm in any way, completely dead arm

B. Data Acquisition Protocol

The exercises to be done for NIHSS administration require observing the duration and drift from the horizontal while arms are held up against gravity (Table I). It consists of different phases including elevation of the arm, holding it horizontally for some time and putting the arm down to resting position. Accelerometers capture changes in acceleration related to any movement, and hence a wrist-worn device can easily capture the drifts related to NIHSS exercise [6]. However, the amplitude of this signal depends on the duration of each phase, which can be measured only from the time-stamps of onset and termination of each phase. Unless accurate time-stamps of each phase are available, which requires precise manual marking during data collection, such data from a wrist-worn device cannot be used for developing a model for automated hemiparesis scoring. This manual marking is nearly impossible due to short duration of individual phases such as elevation of arms or putting them down. Therefore, in this paper, we attempt to correlate spontaneous movements and some well-defined repetitive instructed movements with the strength against gravity as indicated by NIHSS score. We hypothesize that with increasing levels of hemiparesis, these movements will dwindle in the affected arm, thereby

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

increasing the difference (and lack of co-ordination) in the movements of the two arms.

All upper limb movements in this study were selected based upon advice from an expert neurologist. The movements were selected such that they are not difficult to be performed by acute stroke patients. An experimental protocol is designed to capture such movements (illustrated in Fig. 1) as follows:

- NIHSS Scoring: Subject is asked to lift up and hold arms horizontally for 10 s during which scoring for each arm is done using the protocol described in Table I. This score is used as the ground truth for the data acquired immediately following this.
- Spontaneous movements or SM (3 mins): Subject is encouraged to move their upper limbs as much as possible in the vertical and horizontal directions and perform simple tasks (*e.g.*, lift a cup, move a pen and use a phone) using both hands.
- Instructed movements or IM (10 s x 2 arms x 5 tasks per arm): Subject is asked to perform five instructed tasks: finger tapping, hand closing/opening, wrist torsion, elbow flexion/extension and finger swiping. A hard surface is provided for tapping and swiping. Each movement lasts for around 10 s and is performed sequentially by the left and right hands.

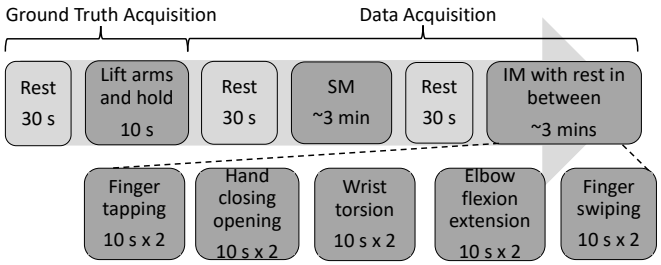


Fig. 1: Flowchart showing data acquisition protocol

The objective behind this study is to investigate the suitability of replacing manual (subjective) intervention in NIHSS administration with a standardized (objective) and automated assessment of hemiparesis in acute stroke using a minimal number of wearable sensors, ensuring patient comfort. The motor component (0-4 scales) comprises only a part of NIHSS administration. Doctors are more interested in getting a broad idea of improvement or deterioration of hemiparetic severity (from mild-to-moderate to severe hemiparesis) indicated by several NIHSS factors [19], [20]. In addition, the clinical granularity is highly subjective and suffers from inter-rater variability [10]. Moreover, wrist-worn accelerometers are limited in picking up minor differences in arm motion related to all classes of motor NIHSS [6]. Additional sensors (and at different locations) for longer duration may be able to pick up more information [21]. But this is not applicable for acute stroke patients in a hospital setting. Furthermore, from Table I, it can be observed that, subjects in NIHSS 3 and 4 cannot lift the arm and those in NIHSS 1 and 2 can lift the arm but for different duration. Based on all these factors, in this preliminary study, we have defined three fundamental levels of hemiparetic severity, namely, *severe*, *moderate* and *control*. Here, NIHSS 3 and 4 are grouped together to indicate severe hemiparesis, NIHSS 1 and 2 are grouped together to indicate moderate hemiparesis and score 0 indicates the control group or healthy subjects.

C. Dataset Description

Data was collected from the stroke units at the Royal Melbourne Hospital, Australia and Sree Chitra Tirunal Institute for Medical Sciences and Technology, India. The data acquisition protocol was approved by the respective institute's human research ethics committee (RMH HREC 2016.146 and SCT/IEC/1081/October-2017). The screening criteria for patient recruitment in this study was the presence of an upper limb weakness due to an ischemic stroke within the past 5 days of data acquisition. The diagnosis of stroke was verified clinically by an expert neurologist through neuroimaging [22]. Informed consent from the patient was ensured prior to data acquisition. During the process of recruitment, 5 patients refused to consent and 3 expressed unwillingness to continue after beginning. Overall, data was acquired from 32 patients (34% female, 67$\pm$13 years old) who had a stroke in the past 5$\pm$4 days resulting in a weak arm with moderate (NIHSS 1 or 2) or severe (NIHSS 3 or 4) hemiparesis. Above 80% of the recruited patients were characterized by Total/Partial Anterior Circulation Infarcts (TACI/PACI) [23] and a history of hypertension. Additionally, 15 healthy participants (NIHSS 0) without any upper limb weakness were also recruited as controls. Hence, a total of 47 subjects were recruited for the study, as illustrated in Table II. For each level of hemiparesis, the affected arm was assigned a score presented in the column labeled NIHSS, whereas, the other arm was labeled as normal, *i.e.*, NIHSS 0. As observed from Table II, the acquired data is highly unbalanced with NIHSS 2, 3 and 4 individually constituting only a small fraction of the entire dataset. Hence grouping similar data into clinically acceptable classes, as discussed before, could lead to the development of an unbiased, well-fitted learning model for pattern recognition.

Data was collected using two low-cost wearable wrist-worn sensors with triaxial accelerometers from Eoxys[2]. The accelerometers acquired data at a sampling rate of 100 Hz and with a full scale range of $\pm 8g$. The sensors communicate via bluetooth to a smartphone application that records data continuously and sends over to a remote server for storage. The average data length across all subjects for SM and IM were 3.28$\pm$0.72 mins and 3.32$\pm$1.33 mins respectively.

TABLE II: Details of acquired data

Motor (Arms) NIHSS	Characteristics				
	Number of Subjects	Left Affected	Female	Age (years)	Total NIHSS(/42)
0	15	N.A.	9	31.3$\pm$3.1	N.A.
1	16	5	6	67.8$\pm$13.6	10.1$\pm$7
2	5	4	1	73.4$\pm$13.5	5.8$\pm$3.9
3	6	4	2	65.6$\pm$8.6	12.5$\pm$2.4
4	5	3	2	62.6$\pm$16.4	15.6$\pm$3.1

[2]https://www.eoxys.com/

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

III. Model for Hemiparesis Identification

Steps of data analysis for identifying the severity of hemiparesis are illustrated through the flowchart in Fig. 2. In order to identify different levels of hemiparesis, the goal is to find accelerometer features that can quantify the difference in activity levels and co-ordination of movements between the normal and affected arms.

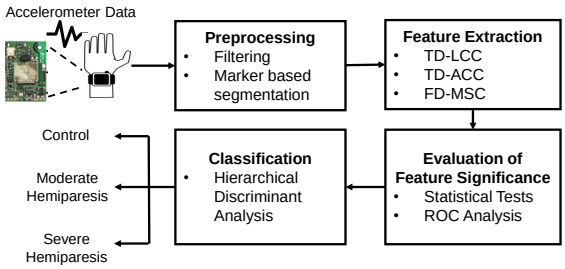


Fig. 2: Block diagram illustrating the proposed steps for identifying hemiparesis from wrist-worn accelerometry

A. Preprocessing

Preprocessing involves removal of artifacts, segmentation of the acquired raw signals and computing resultant acceleration. In order to eliminate gravitational components of acceleration as well as high frequency noise while retaining the movement components of interest [13], data is filtered in 0.25-10 Hz using a 6^{th} order Butterworth bandpass filter. Data is filtered in the forward and reverse directions to obtain zero-phase filtering and avoid phase distortion. Further, a median filter with window length of 3 samples is used to smooth the accelerometer data followed by detrending to obtain a zero-mean signal [13]. Each data session is segmented into the spontaneous and instructed movement portions based on markers noted during data acquisition. Suppose, accelerometer data in *X*, *Y* and *Z* axes are represented by $A_i^x(t)$, $A_i^y(t)$ and $A_i^z(t)$ respectively, where $i \in [l, r]$ indicates left and right arms, the resultant acceleration is computed as

$$A_i(t) = \sqrt{(A_i^x(t))^2 + (A_i^y(t))^2 + (A_i^z(t))^2} \qquad (1)$$

B. Feature Extraction

Traditional features for activity recognition from accelerometry generally use statistical properties of the data in the time and frequency domains to construct models for pattern recognition [12], [13]. These features by themselves cannot be used to quantify the relative motion of a body part with respect to a reference. In this study, the focus is to quantify the differences in the movements of the two arms for identifying hemiparesis, *i.e.*, we aim to analyze the deviation of the movements in the hemiparetic arm with respect to the normal arm (as reference) using accelerometry. Studying the differences in movements across two arms leads to a subject-independent relative approach to pattern recognition, compared to analyzing the absolute motion of the weak arms for all subjects. This section details the proposed measures of co-ordination or *coherence* to quantify such differences using accelerometry in time and frequency domains.

1) Time Domain Lagged Cross-Correlation (TD-LCC): The lack of co-ordination (or correlation) among similar activities of two arms over time is a major indicator of hemiparesis. Healthy subjects (NIHSS 0) are expected to have higher correlation, as both arms are equally equipped to perform any task. However, with increasing hemiparesis (NIHSS 1 to NIHSS 4), correlation is expected to decrease, due to greater disparity in the activity levels of the two arms, contributed by lesser or no movement and more drifts and sways in the affected arm. Because of lack of patient compliance in following instructions, their language impairments and drowsiness due to drugs in the hospital setting, the activities in two arms may not be synchronized at any particular time lag. Hence we study the cross-correlation [24] of accelerometer data from two arms at all possible lags to identify patterns for different levels of hemiparesis. For each segment (SM and IM), cross-correlation R_{lr} at lag m for a length n data is measured by

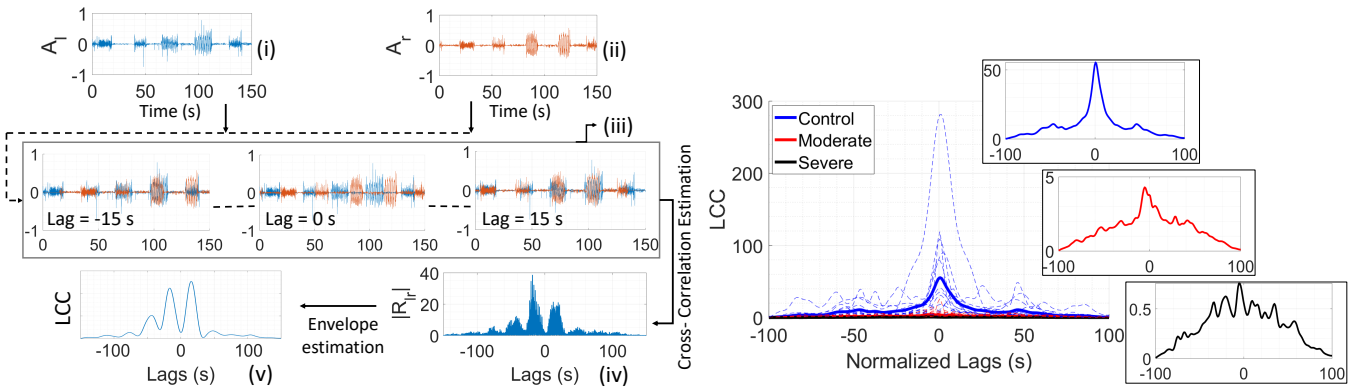
$$R_{lr}(m) = \begin{cases} \sum_{t=1}^{n-m} A_l(t+m)A_r(t), & m \geq 0 \\ R_{rl}(-m), & m < 0 \end{cases} \qquad (2)$$

The absolute value of R_{lr} across all lags is then filtered below 0.1 Hz to produce a low pass envelope, which is termed as the Lagged Cross-Correlation (LCC) signal. The process is illustrated in Fig. 3(a). This envelope exhibits interesting characteristics for each class of data as shown in Fig. 3(b). It is observed that LCC changes from a peaked signal with a higher value of maximum cross-correlation and some prominent peaks to a more spread out signal with lower values of cross-correlation throughout, for increasing levels of hemiparesis, as expected according our hypothesis. Based on these observations, the following characteristics of LCC that indicate measures of cross-correlation and peakedness are used as features:

- Statistical properties: Root Mean Squared (RMS) value (LCC_{rms}), maximum (LCC_{max}) and kurtosis (LCC_{kurt})
- Total area under the LCC curve (LCC_{area})
- *Spread* measured as the normalized length of the signal beyond mean+sd (standard deviation) (LCC_{spread})

All features are computed for the resultant acceleration separately for SM and IM leading to 5 features for each. LCC_{rms} encodes the overall mean squared correlation between the two arms across all times lags, which is expected to decrease with increasing levels of hemiparetic severity. We can easily see from Fig. 3(b) that the maximum value of the correlation is higher in the control subjects and gradually decreases for moderate and severe stroke, thereby, making LCC_{max} an important feature. It is also observed that with increasing stroke severity, the peakedness of LCC curves decreases, that is the curves become flatter throughout, with lesser correlation. As kurtosis is a standard measure of peakedness, we use LCC_{kurt} to quantify is observation. Additionally, we propose to use LCC_{spread} which also measures peakedness in terms of how many data points exceed the mean+sd for that curve, *i.e.*, a less peaked curve will have more number of points exceeding this threshold indicating a more spread-out curve across all lags. Additionally, the area under the LCC curve

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.



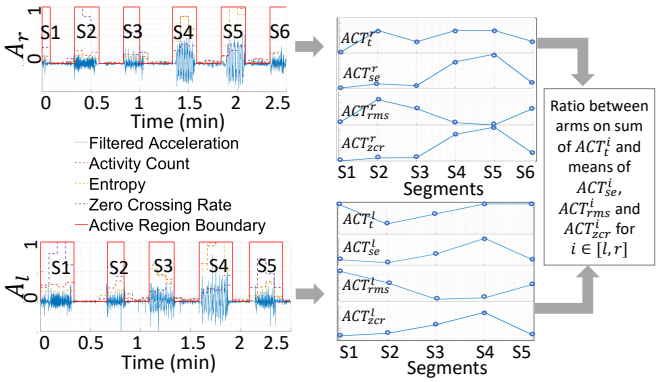
(a) (b)

Fig. 3: Estimating TD-LCC features: (a) Creating LCC signals for (i) Left (A_l) and (ii) Right (A_r) resultant acceleration, (iii) acceleration signals at different lags (only three lags: -15s, 0s and 15s shown), (iv) absolute cross-correlation at all lags, (v) estimated low pass envelope and (b) LCC signals for instructed movements with average of each class in bold: a peaked signal with higher correlation in controls and spread out signals with overall low correlation for increasing hemiparesis

is found to decrease with hemiparesis because of overall low correlations, therefore, we also use LCC_{area} as a feature.

2) Time Domain Activity Cross-Coherence (TD-ACC): The activity levels between two arms are expected to be different with the affected arm showing lower energy content [13]. Based on this hypothesis, the cross-coherence between different measures of activity among the two arms are used to quantify hemiparesis through TD-ACC features. This includes segmenting data from each arm into *active* regions, computing features that quantify the amount of *activity* in these regions and formulating a metric to correlate *activity* between two arms. Active regions are segmented based on *adaptive* thresholds on 5s non-overlapping windowed values of three parameters, namely, average amplitude (activity count) [16], Shannon Entropy [25] and Zero Crossing Rate (ZCR) [26]. A window of data is determined to be active if either of the parameters for that window exceeds the threshold. The thresholds for each parameter are computed adaptively, from the mean of the respective parameter over the data from two arms for that session. A session showing the method of segmentation along with the computation of TD-ACC features is illustrated in Fig. 4.

The total duration (ACT_t^i), mean Shannon Entropy (ACT_{se}^i), mean ZCR (ACT_{zcr}^i) and mean RMS (ACT_{rms}^i) across all active regions from the resultant acceleration of the two arms ($i \in [l, r]$) are computed as measures of activity for each arm. While entropy and ZCR encode the chaotic nature and randomness of motion, the RMS value represents the average nature of the signal. All of these are expected to be of higher magnitude for normal subjects and show decreasing trend for the increasing hemiparesis in the affected arm. For severe hemiparesis, often the affected arm exhibits almost no active regions, in which case, the activity measures are assigned an arbitrarily low value (10^{-6}). The coherence in the activity levels between the two arms can be computed as the ratio of these measures in the affected and normal arms. However, as the ultimate objective is to build a model to automatically score hemiparesis from unknown data where which arm is affected might not be known, we use the ratio of the lower value (assumed to represent affected arm) to the larger value (assumed to represent normal arm) of the activity measures as the coherence measure, leading to 4 features ACT_j^{lr} for $j \in [t, se, zcr, rms]$, each for SM and IM.



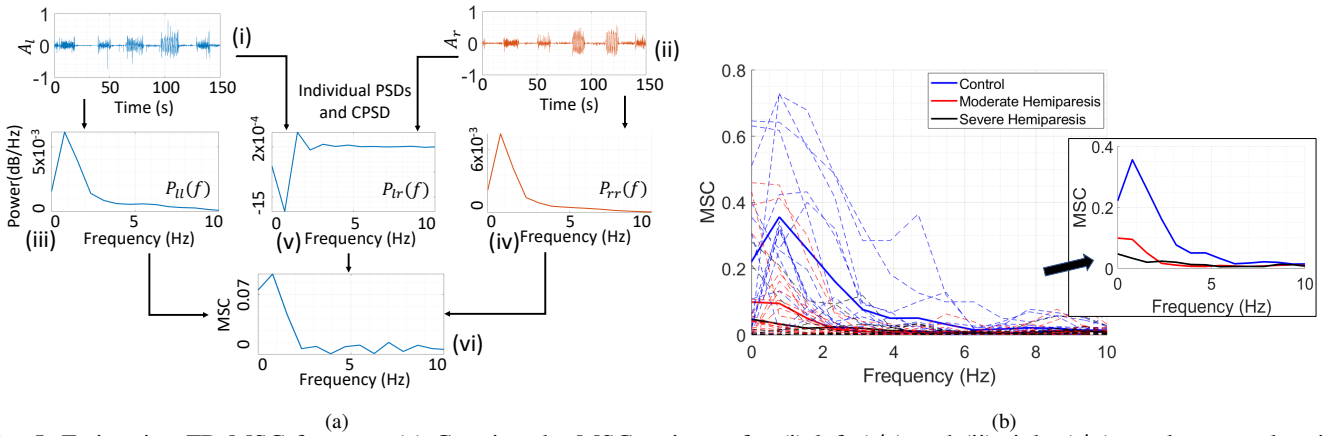
(a) (b)

Fig. 4: Estimating TD-ACC features: (a) Segmentation of active regions followed by (b) Computing activity based features from each segment

3) Frequency Domain Magnitude Squared Coherence (FD-MSC): It is hypothesized that for different levels of hemiparesis, the cross-correlation between two arms will be different at different frequency bands. This is because tasks that can be accomplished easily by subjects of a particular level of hemiparesis might be difficult to be accomplished by subjects at another level of hemiparesis. For frequency domain analysis, Power Spectral Density (PSD) is studied. As the correlation between arms contain information on the degree of hemiparesis, the cross-PSD (CPSD) [24] between the two arms for resultant acceleration is analyzed through Magnitude Squared Coherence (MSC) estimation [27]. Suppose PSD of $A_i(t)$ is given by $P_{ii}(f)$ for $i \in [l, r]$. MSC is a function of individual PSDs $P_{ll}(f)$ and $P_{rr}(f)$ and the CPSD $P_{lr}(f)$ between the left and right arms computed as

$$MSC_{lr}(f) = \frac{|P_{lr}(f)|^2}{P_{ll}(f)P_{rr}(f)} \quad (3)$$

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.



(a) (b)

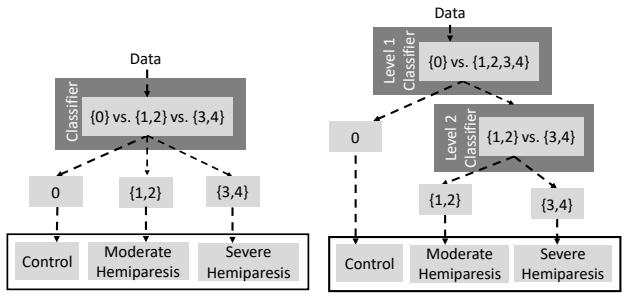
Fig. 5: Estimating FD-MSC features: (a) Creating the MSC estimate for (i) left (A_l) and (ii) right (A_r) resultant acceleration showing individual PSDs (iii) $P_{ll}(f)$ and (iv) $P_{rr}(f)$, (v) CPSD $P_{lr}(f)$ and (vi) computed MSC and (b) MSC estimates for spontaneous movements (mean estimate for each class in bold) showing larger separation at lower frequencies

The MSC estimate is a function of frequency indicating how well the left arm corresponds to the right arm at each frequency, with values from 0 to 1. Because of the time varying nature of accelerometer data, PSD is computed using a non-parametric Welch Method [28]. The signal is split into overlapping segments followed by computation of periodograms of the overlapping segments from their Fourier Transforms, and averaging of the resulting periodograms to produce the PSD estimate. Periodograms are computed using 128 point FFTs with 50% overlap. The process is illustrated in Fig. 5(a) and the MSC computed during spontaneous movements for each class of data from 0 to 10 Hz is presented in Fig. 5(b). It is clearly observed that, at lower frequencies the coherence is higher among the left and the right arms and it decreases at higher frequencies. To quantify this difference, the normalized area under the MSC curve in 3 equally spaced frequency bands 0-3 Hz (MSC_1), 3-6 Hz (MSC_2) and 6-9 Hz (MSC_3) are considered as features and computed separately for SM and IM. The band 9-10 Hz is ignored as all classes are observed to have very low MSC here.

A list of all features that are separately computed for SM and IM is presented in Table III, where $mean(.)$, $min(.)$, $max(.)$, $sd(.)$, $length(.)$ respectively denote the mean, minimum, maximum, standard deviation and length of the input vector, $lags$ denote the vector of all time lags used in computing R_{lr} and $active\_regions$ denote the set of all time instants t identified as active in $A_i(t)$ by the segmentation method.

C. Hierarchical Discriminant Analysis

The set of proposed features can be used to build an automated classification model for detection of hemiparetic severity from accelerometer data. Instead of undertaking only single-level multiclass classification to identify the three classes of hemiparesis, two-level hierarchical classification with binary classifiers based on discriminant analysis at each level has been implemented as illustrated in Fig. 6, where level 1 (L1) and level 2 (L2) respectively classifies controls vs. hemiparetic stroke patients and moderate vs. severe hemiparesis. Compared to a single-level classifier, a multi-layer hierarchical approach to classification offers more flexibility in terms of selecting the most relevant features for each component binary classification module, leading to better performance [29]. In our classification problem, the separation of controls, moderate and severe hemiparesis through this hierarchical architecture is backed by similarity among the certain classes on clinical grounds. As observed from Table I, stroke survivors with NIHSS 3 and 4 both have the minimal amount of arm movements and cannot lift their arms. Though NIHSS 3 exhibits occasional minor twitches in the arm, those are difficult to be identified using wrist-worn accelerometers [6]. On the other hand, both NIHSS 1 and 2 can exhibit stronger movements (and lift their arms) but the degree of co-ordination and activity is lesser than that of controls. Hence we propose to group the data and perform hierarchical classification by selecting a set of most significant features for each level as described in the subsequent subsection. This architecture is also motivated by the skewed class distribution in our dataset (Table II). It is hypothesized that, grouping similar data into clinically acceptable classes can aid in the development of an unbiased, well-fitted pattern recognition model by mitigating problems associated with unbalanced class distributions [30].



(a) (b)

Fig. 6: Schemes of classification (a) Single-level multiclass and (b) Two-level hierarchical for NIHSS scores 1 to 5

The classification model for discriminant analysis [31] assumes multivariate normal distribution of the data X for each class y, where the means and covariances of each class can vary. The predicted class $\hat{y}$ among K classes is obtained by

$$\hat{y} = \underset{y=1,\ldots,K}{\operatorname{argmin}} \sum_{k=1}^{K} \hat{P}(k|x)C(y|k) \tag{4}$$

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

TABLE III: List of Features

Set	ID	Feature	Formulation
TD-LCC	1	LCC_{rms}	$\sqrt{mean(LCC^2)}$ for LCC computed $\forall t \in [-max(lags), max(lags)]$
	2	LCC_{max}	$max(LCC)$ for LCC computed $\forall t \in [-max(lags), max(lags)]$
	3	LCC_{spread}	$length(LCC_{t_k})/length(LCC)$, where LCC_{t_k} is computed for $t_k \subset t$ and $LCC(t_k) > mean(LCC) + sd(LCC)$ $\forall t$
	4	LCC_{kurt}	$E[(LCC(t) - mean(LCC))^4]/E[(LCC(t) - mean(LCC)^2]^2$ $\forall t$
	5	LCC_{area}	$\int_{-max(lags)}^{max(lags)} LCC(t)dt$
TD-ACC	6	ACT_t^{lr}=$min(ACT_t^l, ACT_t^r)/max(ACT_t^l, ACT_t^r)$	$ACT_t^i = length(t)$, $t \in active\_regions$, $i \in [l, r]$
	7	ACT_{se}^{lr}=$min(ACT_{se}^l, ACT_{se}^r)/max(ACT_{se}^l, ACT_{se}^r)$	$ACT_{se}^i = mean(A_i^2(t)log(A_i^2(t)))$, $t \in active\_regions$, $i \in [l, r]$
	8	ACT_{zcr}^{lr}=$min(ACT_{zcr}^l, ACT_{zcr}^r)/max(ACT_{zcr}^l, ACT_{zcr}^r)$	$ACT_{zcr}^i = \frac{1}{A_L}\sum_t Z(A_i(t), A_i(t-1))$, $A_L = length(A_i)$ $Z(x, y) = 1$ if $xy < 0$, $i \in [l, r]$
	9	ACT_{rms}^{lr}=$min(ACT_{rms}^l, ACT_{rms}^r)/max(ACT_{rms}^l, ACT_{rms}^r)$	$ACT_{rms}^i = \sqrt{mean(A_i^2(t))}$, $t \in active\_regions$, $i \in [l, r]$
FD-MSC	10	MSC_1	$\int_0^3 MSC_{lr}(f)df$
	11	MSC_2	$\int_3^6 MSC_{lr}(f)df$
	12	MSC_3	$\int_6^9 MSC_{lr}(f)df$

where $\hat{P}(k|x)$ is the posterior probability of class k for observation x and $C(y|k)$ is the cost of classifying an observation as y when its true class is k. Assuming same feature covariance matrices of both classes, this produces a linear decision boundary resulting in Linear Discriminant Analysis (LDA). In contrast, allowing different feature covariance matrices for different classes leads to a quadratic decision boundary in Quadratic Discriminant Analysis (QDA). Combining the hierarchical architecture with such binary classifiers at each level, we describe these approaches as H-LDA and H-QDA.

Classification performance is evaluated by a leave-one-subject-out method of cross-validation where the training model is built using data from $N-1$ subjects for testing on the N^{th} subject. Performance metrics, *i.e.*, Accuracy (Acc), Sensitivity ($Sens$), Specificity ($Spec$) and F-score are computed for each class over all subjects [32]. This method of validation ensures generalization and subject-independence.

D. Analysis of Feature Significance and Feature Selection

The significance of the proposed features is determined by statistical tests and the Area Under the Receiver Operating Characteristics (ROC) curve (AUC) [33]. For determining the statistical significance of each feature, the non-parametric Kruskal-Wallis test [34] is used and features with $p < 0.05$ are determined as statistically significant. However, Kruskal-Wallis test marks those features as significant that can divide the feature set into at least two different groups. In order to measure the capability of each feature in differentiating control from hemiparesis and moderate hemiparesis from severe, a multiple comparison test is used for statistical significance in each pair of classes. Non-parametric tests are used since the number of data samples is relatively small.

To further analyse feature significance and select features for building the hierarchical classification model, AUC is calculated separately from the ROCs for binary classifications at each level of the hierarchy, *i.e.*, control vs. stroke cases in L1 and moderate vs. severe hemiparesis in L2. ROC curves highlight the goodness of each feature at different thresholds, in maximizing the probability of detection and minimizing the possibility of false alarms, thereby assisting in developing an automatic classification model.

In order to build a classification model with the limited sample size in our dataset, we only select the top 5 features for each level to prevent overfitting [32]. These features are selected from the pool of the statistically significant features ($p < 0.05$) for that level, based on their rank determined by the AUC value (high to low) for binary classification at that level. Only those statistically significant features are considered for ranking which have AUC>0.8. All computations are performed on MATLAB R2018a.

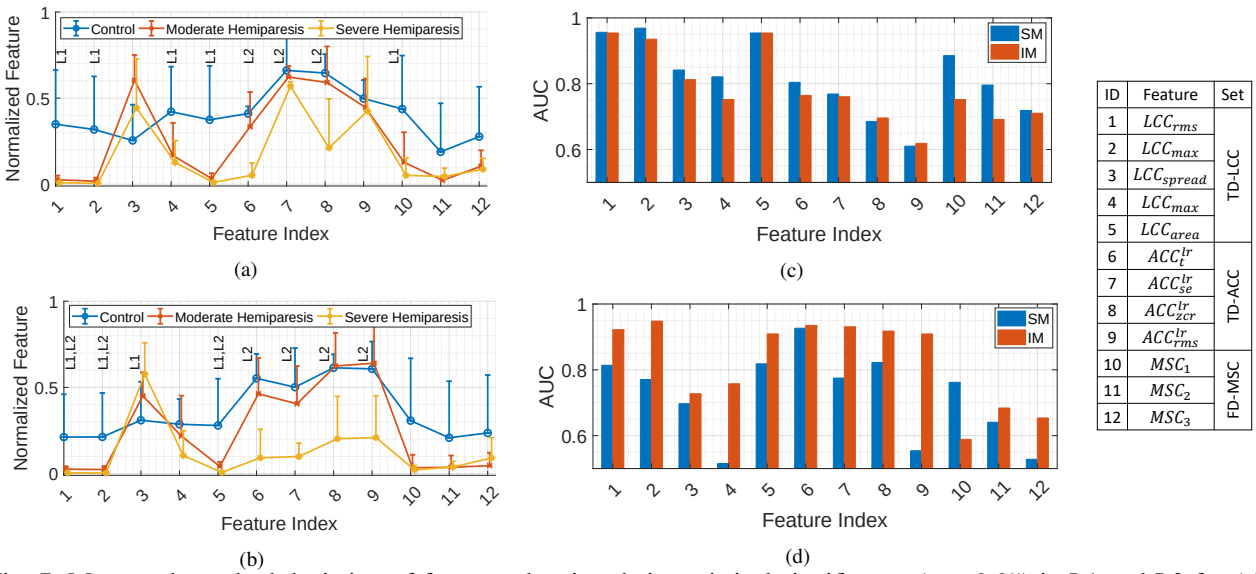
IV. Results

A. Analysis of Feature Significance

The mean and standard deviation of each feature across three classes are presented in Fig. 7(a) and Fig. 7(b), for SM and IM respectively. In these figures, the features are normalized in [0,1] by mapping the minimum and maximum values to 0 and 1 respectively, for the ease of visibility only. The statistical significance of each feature for the two levels of hierarchy, determined by the p-value from the Kruskal Wallis and multiple comparison tests is indicated by marking L1 or L2 against each feature.

Fig. 7(a) indicates that, for SM, all TD-LCC features except LCC_{spread} can clearly discriminate between control and hemiparetic patients, thereby being relevant for L1. A similar characteristic is observed for FD-MSC features for SM as well, though only MSC_1 is found statistically significant for L1 in the current dataset. For TD-ACC features the differences are larger between moderate and severe hemiparesis. Hence these features (except ACT_{rms}^{lr}), are found to be significant for L2. For IM, as observed from Fig. 7(b), it is found that TD-LCC and FD-MSC features are again visibly different for controls and hemiparetic subjects. However, out of these, only four TD-LCC features are found to be statistically significant for the same. TD-LCC features LCC_{rms}, LCC_{max} and LCC_{area} for IM are also significant in L2. All the TD-ACC features for IM are particularly significant for separating moderate and severe hemiparesis as observed from Fig. 7(b) as well

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.



ID	Feature	Set
1	LCC_{rms}	TD-LCC
2	LCC_{max}	
3	LCC_{spread}	
4	LCC_{max}	
5	LCC_{area}	
6	ACC_t^{lr}	TD-ACC
7	ACC_{se}^{lr}	
8	ACC_{zcr}^{lr}	
9	ACC_{rms}^{lr}	
10	MSC_1	FD-MSC
11	MSC_2	
12	MSC_3	

Fig. 7: Mean and standard deviation of features showing their statistical significance ($p < 0.05$) in L1 and L2 for (a) SM and (b) IM and variation in AUCs in differentiating (c) controls vs. stroke *i.e.*, L1 and (d) moderate vs. severe hemiparesis *i.e.*, L2

as from the outcome of statistical tests. Additionally, most of the features exhibit monotonic changes in the patterns of their means from control to severe hemiparesis. For example, both Fig. 7(a) and Fig. 7(b) show that TD-LCC features LCC_{rms}, LCC_{max}, LCC_{kurt} and LCC_{area}, TD-ACC features ACT_t^{lr} and ACT_{se}^{lr} and FD-MSC feature MSC_1, decrease in their average values for increasing hemiparesis, whereas, LCC_{spread} increases monotonically for increasing hemiparesis for IM.

The variations in AUC for each feature are illustrated in Fig. 7(c) and Fig. 7(d) respectively for L1 and L2, for both SM and IM. For L1 classifications with spontaneous movements, the maximum AUC (0.97) is observed for LCC_{max}, closely followed by LCC_{rms} (0.96) and LCC_{area} (0.95). For instructed movements, the same features lead to the highest AUCs (0.95 for LCC_{rms} and LCC_{area}). Additionally, both sets of movements show similar trends in differentiating between controls and hemiparesis, with TD-LCC features being the most relevant in terms of AUC values, as clearly observed in Fig. 7(c). However, for each set of features, spontaneous movements fare better for L1 classifications in terms of AUC values. Upon close observation, it can be seen that while almost all TD-LCC features produce high AUCs ($>$ a threshold of 0.8) for both SM and IM, the FD-MSC features are particularly good for spontaneous movements only, with only MSC_1 exceeding the threshold. However, a completely different trend is observed in Fig. 7(d) for the AUCs in L2 classifications. In this case, the maximum AUC (0.93) for spontaneous movements is achieved for ACT_t^{lr}. The same for instructed movements is observed for $LCCmax$ (0.95), closely followed by ACT_t^{lr} (0.94), ACT_{se}^{lr} (0.93), LCC_{rms} (0.92) and ACT_{zcr}^{lr} (0.92). All TD-ACC features lead to AUCs>0.9 for IM. For FD-MSC features, both SM and IM lead to AUCs below 0.7 for L2 thereby indicating that such features are not discriminative enough for L2. Therefore, L2 classifications are characterized by higher AUCs for IM across both TD-LCC and TD-ACC features, and a larger variability and lower performance for both these feature sets for SM. Hence, instructed movements are particularly useful, especially with TD-ACC features for L2 classification, (as also concluded from the feature distributions in Fig. 7(a) and 7(b)). 7 features across TD-LCC and TD-ACC have AUC>0.8 for IM compared to only 4 features for SM.

TABLE IV: Overview of Significant Features (Mean±sd) with AUCs in the Corresponding Levels of Hierarchy

ID		Control	Moderate Hemiparesis	Severe Hemiparesis	AUC
SM	1	20.01±18.00	1.55±1.48	0.57±0.34	0.96 (L1)
	2	74.51±72.02	4.26±4.85	1.45±0.86	0.97 (L1)
	4	7.99±3.90	4.19±2.80	3.6±1.89	0.82 (L1)
	5	14.33±11.94	1.32±1.16	0.47±0.28	0.95 (L1)
	6	1.06±0.10	0.86±0.52	0.13±0.18	0.93 (L2)
	8	1.07±0.18	0.98±0.34	0.35±0.47	0.82 (L2)
	10	0.28±0.19	0.08±0.11	0.03±0.06	0.89 (L1)
IM	1	15.56±18.14	1.96±1.21	0.39±0.37	0.95 (L1), 0.92 (L2)
	2	60.18±71.91	6.74±5.61	1.05±1.03	0.94 (L1), 0.95 (L2)
	3	0.16±0.06	0.19±0.03	0.23±0.04	0.81 (L1)
	5	9.69±9.35	1.48±0.94	0.32±0.30	0.95 (L1), 0.90 (L2)
	6	1.00±0.26	0.84±0.37	0.17±0.29	0.94 (L2)
	7	1.11±0.54	0.88±0.52	0.15±0.18	0.93 (L2)
	8	1.06±0.13	1.07±0.33	0.35±0.42	0.92 (L2)
	9	1.01±0.26	1.07±0.36	0.35±0.41	0.91 (L2)

B. Feature Selection

It is found that for spontaneous (and instructed) movements, the requirements of statistical significance with AUC>0.8 are met by 5 (and 4) features for L1 and 2 (and 7) features for L2. The list of these significant features with the average values across each class and AUCs for the level of significance in the hierarchy is presented in Table IV. It clearly shows that the average feature value has larger difference in controls and any type of hemiparesis when selected for L1 whereas the same is true for moderate and severe hemiparesis when selected for L2. Additionally, all these significant features show monotonic increase/decrease for increasing hemiparesis, thereby exhibiting a pattern for classification of the data. The

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

TABLE V: Classification Performance Metrics with Leave-One-Subject-Out Cross-Validation

Classifier	Hierarchical						Multi-class				
	Classes		Acc	Sens	Spec	F-score	Classes	Acc	Sens	Spec	F-score
LDA	L1	Control	0.89	0.67	1	0.80	Control	0.85	0.60	0.96	0.73
		Hemiparetic	0.89	1	0.67	0.93	Moderate	0.72	0.76	0.69	0.86
	L2	Moderate	0.79	0.86	0.73	0.78					
		Severe	0.89	0.81	0.92	0.78	Severe	0.83	0.73	0.71	0.67
QDA	L1	Control	0.92	0.93	0.91	0.88	Control	0.89	0.80	0.83	0.93
		Hemiparetic	0.92	0.91	0.93	0.94	Moderate	0.83	0.86	0.80	0.82
	L2	Moderate	0.87	0.86	0.88	0.86					
		Severe	0.96	0.82	1	0.9	Severe	0.93	0.82	0.97	0.86
kNN	L1	Control	0.85	0.73	0.91	0.76	Control	0.87	0.80	0.91	0.80
		Hemiparetic	0.85	0.91	0.73	0.89	Moderate	0.70	0.67	0.73	0.67
	L2	Moderate	0.74	0.71	0.77	0.71					
		Severe	0.89	0.82	0.92	0.78	Severe	0.83	0.64	0.89	0.64
Decision Tree	L1	Control	0.89	0.87	0.91	0.81	Control	0.83	0.73	0.87	0.73
		Hemiparetic	0.89	0.91	0.87	0.92	Moderate	0.72	0.67	0.77	0.68
	L2	Moderate	0.72	0.67	0.77	0.68					
		Severe	0.83	0.64	0.89	0.64	Severe	0.81	0.64	0.86	0.61

selected top features in each level ranked according to AUC values, as discussed in Section III-D, comprise:

- L1: TD-LCC 1,2,5 (SM) & TD-LCC 1,5 (IM)
- L2: TD-ACC 6 (SM) & TD-LCC 1,2, TD-ACC 6,7 (IM)

Hence, the final set of features selected for L1 classification comprise solely of TD-LCC features across both SM and IM. For L2, TD-ACC features dominate across SM and IM, however, IM based features are found to be more prevalent. Overall, a set of TD-LCC and TD-ACC features, LCC_{rms}, LCC_{max}, LCC_{area}, ACT_t^{lr} and ACT_{se}^{lr}, are therefore popular in both the levels and across both types of movements.

C. Classification

The results of leave-one-subject out classification are illustrated in Table V. Here, the hierarchical approach is compared with single-level multi-class approach and along with LDA and QDA, some standard classifiers such as k-Nearest Neighbours (kNN, with k=3) and Decision Trees (DT) [32] are also studied. Multi-class classifications are implemented with all statistically significant features ($p < 0.05$). QDA shows the best performance across all the classifier models, and H-QDA with feature selection leads to higher scores for the metrics across all classes. In L1, controls and hemiparetic subjects can be distinguished with a value of more than 0.9 for almost all metrics. The overall average Acc, $Sens$, $Spec$ and F-score across all classes are 91%, 87%, 93% and 88% respectively after final classification in L2. For multi-class QDA, overall average Acc, $Sens$, $Spec$ and F-score across all classes are 88.33%, 82.67%, 90% and 83.67% respectively. In both cases, the metrics are higher in identifying controls and severe hemiparesis than for moderate hemiparesis. The performances are relatively poor for LDA, kNN and DT. However, across all classifiers, the hierarchical models fare better than the single-level multi-class models in terms of average accuracy across all classes.

V. Discussions

Clinical research shows evidence of worsening of motor weakness in upper limbs due to hemiparesis in acute stroke [2], [3]. Though wearable motion sensors have proved to be suitable to capture motor functionalities for identifying hemiparesis [6], [15], there is lack of research on the quantification of the process and the development of an automatic scoring system for the same using short-length data from a minimal number of sensors. The results of this work indicate that a combination of the proposed set of spontaneous and instructed movements quantified by time and frequency domain measures of upper limb co-ordination using wearable accelerometry can be used for this purpose. Further, the proposed measures can be used to score hemiparesis using a subject-independent pattern recognition model with reference to clinical NIHSS scores.

From the statistical analysis of features, it is observed that control subjects can be easily separated from stroke patients with hemiparesis, which is clinically very evident. This is supported by the higher value of AUCs across all features for L1 in Fig. 7(c) compared to L2 in Fig. 7(d). For the separation of moderate and severe hemiparesis, instructed movements and TD-ACC features are found to be more significant. This can be attributed to the fact that spontaneous movements for both moderate and severe hemiparesis are very low, but patients with moderate hemiparesis can make attempts to perform the instructed movements. Hence, the measures of coherence in activity show larger difference for instructed movements in moderate and severe hemiparesis. All features found to be significant (Table IV), decrease in their average values for increasing hemiparesis, except feature ID 3, LCC_{spread} for IM. This is justified by the fact that the features other than LCC_{spread} quantify the level of activity and co-ordination which is expected to be lower in severe hemiparesis. Whereas, LCC_{spread} is a measure of lack of co-ordination throughout the signal and is expected to be higher for severe hemiparesis. The features finally selected for classification, described in Section IV-B also bear clinical significance. It is found that TD-LCC features dominate for both SM and IM in L1, for discriminating control from hemiparesis. This can be explained from the fact that control subjects should have high overall correlation between the two arms as the same activities are attempted to be performed by both hands [4]. However, for hemiparetic patients, there is lack of co-ordination between two arms leading to a very different pattern in the LCC signal. Features selected for L2, for discriminating moderate and severe hemiparesis are dominated by instructed movements.

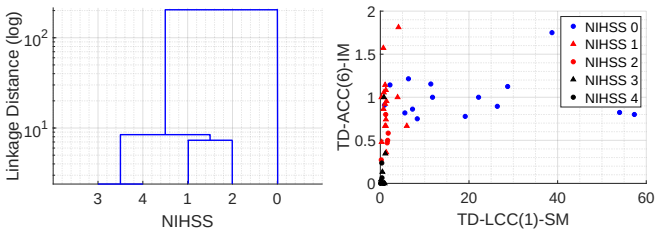
This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

This is because the spontaneous movements for stroke patients (whether moderate or severe hemiparetic) tend to be limited. In this case, other than TD-LCC features, TD-ACC features also prove to be vital, supported by the fact that the degree and duration of activities differentiate hemiparetic severity.

As a comparison of these features with the measures of hemiparesis for acute stroke from accelerometry data proposed in earlier works, we performed statistical analysis on the three activity based metrics proposed in [6] and the three axes correlation based metrics in [14] for both spontaneous and instructed movements acquired with our experimental protocol. It is found that the SMA and energy based metrics proposed in [6] have $p < 0.05$, but they are found to be significant only for classification of controls vs. hemiparesis, and control and moderate hemiparesis vs. severe hemiparesis respectively, thereby being unsuitable for scoring hemiparesis from short length accelerometer data.

Prior to classification, we employ clustering to group the feature space into meaningful *clusters* [35] in order to analyse the suitability of the proposed classification method based on data characteristics. We use simple visual analysis of *hierarchical agglomerative* clustering for recognizing the similarity among classes and assessing the formation of a hierarchical classifier on this dataset [36]. This form of hierarchical clustering starts with each data point as an individual cluster, and these clusters are sequentially combined into bigger clusters. Two clusters closest in the sense of a well defined set distance are joined at each step in this process. We use *single linkage* with Euclidean distance between feature vectors, and the distance between two clusters is defined as the distance between the pair of data points (one in each cluster), that are closest to each other [35]. The results can be visualized using a dendrogram, which shows the sequence of cluster fusion and the set distance at which each fusion took place, thereby illustrating the hierarchical process. Fig. 8(a) illustrates this process using the average of the significant features obtained for each NIHSS level. Logarithm scale is used for the linkage distance only for the ease for visibility. The dendrogram clearly indicates a hierarchical structure in the data with controls completely separated out and NIHSS {1,2} and NIHSS {3,4} grouped together. Further, it can be seen that NIHSS 3 and 4 have lesser distance among them compared to NIHSS 1 and 2. Based on these observations, we also investigate the scatter plots of the significant features across all classes, one of which is presented in Fig. 8(b). The features are also representative of the fact that there is an inherent hierarchy in the dataset, with the control and hemiparetic groups separated by a larger distance than the moderate and severe hemiparetic groups, while there is significant overlap between NIHSS 1 and 2 and also NIHSS 3 and 4.

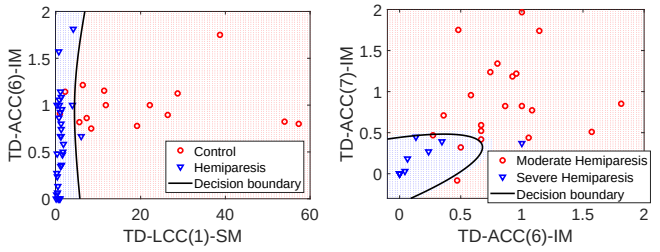
It is observed from Table V, that the hierarchical approach with the most relevant features selected at each level results in better performance metrics for each class for all classification methods. The performance metrics are higher for the first level showing better separability between control and hemiparetic subjects. The relatively lower performance in identifying moderate hemiparesis as observed from the performance metrics can be attributed to the fact that it has overlap in feature spaces



(a) (b)

Fig. 8: Analysis of hierarchical structure in the dataset (a) Dendrogram from Single Linkage clustering on significant features shows similar grouping as hypothesized in Fig. 6(b) and (b) Scatter plots using two of the selected features also highlights similarities among hemiparetic data, with overlap in NIHSS {1,2} and NIHSS {3,4}

partly with controls and partly with severe hemiparesis. The hierarchical architecture captures the wider difference between controls and hemiparetic subjects (Table IV) and propagates a small error from the first level to the next leading to an overall smaller error than the single-level multi-class case. H-QDA leads to best performance and additionally, there is a balance between $Sens$ and $Spec$ resulting in high F-scores and implying an unbiased classifier for each of the classes.



(a) (b)

Fig. 9: Classification decision boundaries with H-QDA using two of the selected top features in (a) L1 and (b) L2

Visual representation of the decision boundaries obtained by the H-QDA model on the entire dataset for two of the selected features in each level of the hierarchy is shown in Fig. 9. The figures indicate that the features are non-linearly separable. This justifies the better performance of QDA instead of a linear classification model (such as LDA), such that the non-linear decision boundaries at each level of the proposed model can successfully differentiate between pairs of classes with minor miss-classifications. Additionally, among popular pattern classification methods, discriminant analysis (LDA or QDA) leads to more stable classifiers for small sample size and overlapping classes, as in this study [37]. The relatively poor performance of kNN and DT can be attributed to the small sample size, as both these classifiers inherently perform better with increasing training data. Also these classifiers are highly affected by the choice of parameters, (*i.e.*, k for kNN, number of leaves, information gain metric etc. for DT), which need tuning after analysis on bigger datasets [32].

VI. Conclusion

In this paper, a novel experimental paradigm is proposed for identifying different levels of hemiparetic severity in acute

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.

stroke using short length wrist-worn accelerometer data comprising upper limb movements made during spontaneous and instructed activities. Time and frequency domain coherence analysis of accelerometer data has been used to estimate novel measures of co-ordination between two arms. While such measures of co-ordination during spontaneous movements can differentiate between controls and hemiparetic subjects, more features based on instructed movements are found significant to identify differences in moderate and severe hemiparesis. Further, a classification model is also developed based on discriminant analysis that can identify controls and the two levels of hemiparesis for automatic scoring in a hierarchical fashion. Future studies in this direction include the use of multiple sensors and different modalities to accurately identify hemiparetic severity with respect to each level in the clinical gold standard. Analyzing continuously recorded spontaneous data to identify improvement or deterioration in the condition of acute stroke patients, without any instruction, is another challenging future implication of this study.

REFERENCES

[1] J. R. Villar, C. Chira, J. Sedano, S. Gonzlez, and J. M. Trejo, "A hybrid intelligent recognition system for the early detection of strokes," *Integrated Computer-Aided Engineering*, vol. 22, no. 3, pp. 215–227, 2015.

[2] C. Weimar, T. Mieck, J. Buchthal, C. E. Ehrenfeld, E. Schmid, and H.-C. Diener, "Neurologic worsening during the acute phase of ischemic stroke," *Archives of neurology*, vol. 62, no. 3, pp. 393–397, 2005.

[3] A. Dávalos, D. Toni, F. Iweins, E. Lesaffre, S. Bastianello, and J. Castillo, "Neurological deterioration in acute ischemic stroke: potential predictors and associated factors in the european cooperative acute stroke study (ecass) I," *Stroke*, vol. 30, no. 12, pp. 2631–2636, 1999.

[4] D. Rose and C. Winstein, "The co-ordination of bimanual rapid aiming movements following stroke," *Clinical rehabilitation*, vol. 19, no. 4, pp. 452–462, 2005.

[5] C. A. Molina and J. L. Saver, "Extending reperfusion therapy for acute ischemic stroke: emerging pharmacological, mechanical, and imaging strategies," *Stroke*, vol. 36, no. 10, pp. 2311–2320, 2005.

[6] J. Gubbi, A. S. Rao, K. Fang, B. Yan, and M. Palaniswami, "Motor recovery monitoring using acceleration measurements in post acute stroke patients," *Biomedical engineering online*, vol. 12, no. 1, p. 33, 2013.

[7] P. Lyden, M. Lu, C. Jackson, J. Marler, R. Kothari, T. Brott, and J. Zivin, "Underlying structure of the national institutes of health stroke scale: results of a factor analysis," *Stroke*, vol. 30, no. 11, pp. 2347–2354, 1999.

[8] A. Dávalos, J. Castillo, and E. Martinez-Vila, "Delay in neurological attention and stroke outcome," *Stroke*, vol. 26, no. 12, pp. 2233–2237, 1995.

[9] K. Nagao, A. Koschel, H. Haines, L. Bolitho, and B. Yan, "Rural victorian telestroke project," *Internal medicine journal*, vol. 42, no. 10, pp. 1088–1095, 2012.

[10] C. D. Bushnell, D. C. Johnston, and L. B. Goldstein, "Retrospective assessment of initial stroke severity: comparison of the nih stroke scale and the canadian neurological scale," *Stroke*, vol. 32, no. 3, pp. 656–660, 2001.

[11] J. Burn, M. Dennis, J. Bamford, P. Sandercock, D. Wade, and C. Warlow, "Long-term risk of recurrent stroke after a first-ever stroke. the oxfordshire community stroke project." *Stroke*, vol. 25, no. 2, pp. 333–337, 1994.

[12] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 386–403, 2017.

[13] M. J. Mathie, A. C. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiological measurement*, vol. 25, no. 2, p. R1, 2004.

[14] D. Kumar, J. Gubbi, B. Yan, and M. Palaniswami, "Motor recovery monitoring in post acute stroke patients using wireless accelerometer and cross-correlation," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 6703–6706.

[15] C. Le Heron, K. Fang, J. Gubbi, L. Churilov, M. Palaniswami, S. Davis, and B. Yan, "Wireless accelerometry is feasible in acute monitoring of upper limb motor recovery after ischemic stroke," *Cerebrovascular Diseases*, vol. 37, no. 5, pp. 336–341, 2014.

[16] C. E. Lang, J. M. Wagner, D. F. Edwards, and A. W. Dromerick, "Upper extremity use in people with hemiparesis in the first few weeks after stroke," *Journal of Neurologic Physical Therapy*, vol. 31, no. 2, pp. 56–63, 2007.

[17] T. Hester, R. Hughes, D. M. Sherrill, B. Knorr, M. Akay, J. Stein, and P. Bonato, "Using wearable sensors to measure motor abilities following stroke," in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN)*, 2006, pp. 4 pp.–8.

[18] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato, "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 450–461, 2010.

[19] F. S. Vahidy, W. J. Hicks, I. Acosta, H. Hallevi, H. Peng, R. Pandurengan, N. R. Gonzales, A. D. Barreto, S. Martin-Schild, T.-C. Wu *et al.*, "Neurofluctuation in patients with subcortical ischemic stroke," *Neurology*, vol. 83, no. 5, pp. 398–405, 2014.

[20] O. Ozdemir, V. Beletsky, R. Chan, and V. Hachinski, "Thrombolysis in patients with marked clinical fluctuations in neurologic status due to cerebral ischemia," *Archives of neurology*, vol. 65, no. 8, pp. 1041–1043, 2008.

[21] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. Dy, and P. Bonato, "Tracking motor recovery in stroke survivors undergoing rehabilitation using wearable technology," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2010, pp. 6858–6861.

[22] J. R. Sims, G. Rordorf, E. E. Smith, W. J. Koroshetz, M. H. Lev, F. Buonanno, and L. H. Schwamm, "Arterial occlusion revealed by ct angiography predicts nih stroke score and acute outcomes after iv tpa treatment," *American Journal of Neuroradiology*, vol. 26, no. 2, pp. 246–251, 2005.

[23] J. Bamford, P. Sandercock, M. Dennis, C. Warlow, and J. Burn, "Classification and natural history of clinically identifiable subtypes of cerebral infarction," *The Lancet*, vol. 337, no. 8756, pp. 1521–1526, 1991.

[24] P. Stoica, R. L. Moses *et al.*, "Spectral analysis of signals," 2005.

[25] J. Gubbi, S. Kusmakar, A. S. Rao, B. Yan, T. OBrien, and M. Palaniswami, "Automatic detection and classification of convulsive psychogenic nonepileptic seizures using a wearable device," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1061–1072, 2016.

[26] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[27] S. Kay, "Modern spectral estimation (englewood cliffs)," 1988.

[28] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[29] A. Mukherjee, A. D. Choudhury, S. Datta, C. Puri, R. Banerjee, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, and S. Khandelwal, "Detection of atrial fibrillation and other abnormal rhythms from ecg using a multi-layer classifier architecture," *Physiological measurement*, vol. 40, no. 5, p. 054006, 2019.

[30] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[31] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: data mining, inference, and prediction, springer series in statistics," 2009.

[32] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[33] T. Fawcett, "Roc graphs: Notes and practical considerations for researchers," *Machine learning*, vol. 31, no. 1, pp. 1–38, 2004.

[34] W. W. Daniel, "Applied nonparametric statistics pws," 1990.

[35] J. C. Bezdek, *A primer on cluster analysis: 4 basic methods that (usually) work*. First Edition Design Publishing, 2017.

[36] S. Datta, J. C. Bezdek, and M. Palaniswami, "Experiments with dissimilarity measures for clustering waveform data from wearable sensors," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019.

[37] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.

This article is accpted for publication. The final published version is available at https://doi.org/10.1109/TNSRE.2020.2972285.