Title	**Recognizing Upper Limb Movements with Wrist Worn Inertial Sensors using k-means Clustering Classification**

Author names	**Dwaipayan Biswas[1], Andy Cranny[1], Nayaab Gupta[1], Koushik Maharatna[1], Josy Achner[2], Jasmin Klemke[2], Michael Jöbges[2] and Steffen Ortmann[3]**

Affiliations	[1] Faculty of Physical Sciences and Engineering, University of Southampton, Hampshire, UK
	[2] Brandenburg Klinik, Bernau, Berlin, Germany
	[3] IHP, Frankfurt (Oder), Germany

Email	{db9g10, awc, ng2g10, km3}@ecs.soton.ac.uk,
	{josy.achner, ergo_n1, joebges}@brandenburgklinik.de
	ortmann@ihp-microelectronics.com.

Corresponding author	**Dwaipayan Biswas**
	Building 53, room 4033, School of Electronics and Computer Science, University of Southampton, Highfield Campus, SO17 1EP, UK.
	Tel: +447405462045,
	Email: db9g10@ecs.soton.ac.uk

Abstract: In this paper we present a methodology for recognizing three fundamental movements of the human forearm (extension, flexion and rotation) using pattern recognition applied to the data from a single wrist-worn, inertial sensor. We propose that this technique could be used as a clinical tool to assess rehabilitation progress in neurodegenerative pathologies such as stroke or cerebral palsy by tracking the number of times a patient performs specific arm movements (e.g. prescribed exercises) with their paretic arm throughout the day. We demonstrate this with healthy subjects and stroke patients in a simple proof of concept study in which these arm movements are detected during an archetypal activity of daily-living (ADL) - 'making-a-cup-of-tea'. Data is collected from a tri-axial accelerometer and a tri-axial gyroscope located proximal to the wrist. In a training phase, movements are initially performed in a controlled environment which are represented by a ranked set of 30 time-domain features. Using a sequential forward selection technique, for each set of feature combinations three clusters are formed using k-means clustering followed by 10 runs of 10-fold cross validation on the training data to determine the best feature combinations. For the testing phase, movements performed during the ADL are associated with each cluster label using a minimum distance classifier in a multi-dimensional feature space, comprised of the best ranked features, using Euclidean or Mahalonobis distance as the metric. Experiments were performed with four healthy subjects and four stroke survivors and our results show that the proposed methodology can detect the three movements performed during the ADL with an overall average accuracy of 88% using the accelerometer data and 83% using the gyroscope data across all healthy subjects and arm movement types. The average accuracy across all stroke survivors was 70% using accelerometer data and 66% using gyroscope data. We also use a Linear Discriminant Analysis (LDA) classifier and a Support Vector Machine (SVM) classifier in association with the same set of features to detect the three arm movements and compare the results to demonstrate the effectiveness of our proposed methodology.

Keywords: Activities of daily living, activity recognition, inertial sensors, k-means clustering, minimum distance classifier, remote telehealth monitoring.

1. Introduction

Tracking of human body movement has been performed across various applications such as home-based remote health monitoring [1-2], human computer interaction [3-4] and sports coaching [5-6] using a wide range of sensing technologies, including: mechanical tracking, optical systems, radio-frequency identification (RFID) [7], low-cost body-worn inertial sensors [8-9], and fusion of vision-based and inertial sensor based approaches [10]. Most of these approaches are however, primarily restricted to indoor activities within a defined region and require an un-hindered surveillance of the vision system [11]. Furthermore, the use of high complexity image processing algorithms can result in slower analysis which can be particularly challenging if real-time information analysis is required [10]. Low cost body-worn inertial sensors provide an effective mechanism when the sensor data is used to train an activity model which is used later to recognize the activities from new observation data [12]. Recognition of activities using sensor data in an unconstrained daily living scenario is particularly challenging owing to the considerable amount of variability inherent in movement patterns of each subject without any manual intervention or a-priori knowledge [13].

In principle there are three steps for activity recognition using inertial sensors: (1) data capture by appropriate sensor; (2) segmentation of the captured data to identify the beginning and end of an activity; and (3) recognition of the activity using appropriate classification techniques [11]. Continuous monitoring of activities in an unconstrained scenario involves data segmentation and activity recognition which are in practice interrelated but are individually two separate research problems owing to the possible qualitative non-uniqueness of an activity pattern exhibited by an individual subject and due to inter-person variability. In the research reported here we concentrate only on the activity recognition

part as a proof-of-concept methodology.

We perform a systematic exploration to recognize three fundamental movements of the upper limb which are generally associated with activities of daily living, using data collected from a wrist-worn, wireless tri-axial accelerometer and tri-axial rate gyroscope. The motivation behind this work was to detect the occurrence of these specific arm movements in an out-of-laboratory condition (i.e. real world) using the minimal number of sensors, as opposed to detecting gross static or dynamic activities and postures like standing, sitting, running, cycling, brushing teeth [1, 6, 9, 14], etc. The detection and classification of particular arm movements (e.g. prescribed exercises) during daily activities can over time provide a measure of arm rehabilitation progress in remote health monitoring applications, especially in neurodegenerative pathologies such as stroke or cerebral palsy. Our objective is to detect the use of the impaired arm of a stroke patient during prescribed exercises and normal daily activities and to classify the type of movements performed as a means to assess their rehabilitation. Enumerating occurrences of these movements over time can act as an indicator of rehabilitation progress since the frequency of these movements is more likely to increase as the motor functionality of the patient improves. The specific arm movements considered were: (1) reach out and retrieve object, (2) lift cup to mouth and (3) performing pouring/(un)locking action, all of which involve rotations of the forearm about various axes.

The fundamental concept is to first form a set of 3 clusters in multi-dimensional feature space (selected from a ranked set of 30 features), with each cluster representing a particular type of movement, using sets of features generated from person-centric data collected in a constrained *training* phase (e.g. in the laboratory). Subsequent data collected during an unconstrained *testing* phase (e.g. real world) is tested for its proximity to each of the clusters when the same features are extracted by using a pattern recognition technique. We used the regularized Mahalonobis distance based k-means clustering technique to form the clusters on the *training* data and use 10 runs of a 10-fold cross validation technique to determine the best combination of cluster forming features. A minimum distance classifier based on Euclidean and Mahalonobis distance was used for matching the test data to the formed clusters. We adopted a personalized recognition approach, thereby formulating the clusters from person-centric *training* data. This approach was adopted in view of the large degree of inter-person variability expected amongst the sample population. Furthermore, this would be beneficial when applied to monitoring individual patients who demonstrate differences in levels of impairment depending on their stage of rehabilitation. Experiments performed with four healthy subjects and four stroke patients show that the proposed method can recognize the three movements with an overall average accuracy of 88% using just accelerometer data and 83% when using only gyroscope data across all healthy subjects and an average accuracy of 70% using accelerometer data and 66% using gyroscope data across all stroke patients and arm movement types. The minimum accuracy for detecting each individual arm movement was 80% for healthy subjects and 60% for stroke patients if more than one sensor is used. The movements were further classified using the learning algorithms LDA and SVM employing a radial basis function (RBF) kernel. A comparison of the achieved sensitivity and overall accuracy values demonstrates the effectiveness of our proposed methodology.

The remainder of the paper is structured as follows. In Section II we provide some background on the state-of-the-art techniques for activity recognition, relating prior work to our methodology, and in Section III we describe the experiment protocol. The data processing techniques and the results are analyzed in Sections IV and V respectively. Finally a discussion is presented in Section VI.

2. Background

Activity recognition of gross movements and postures using inertial sensors is generally analyzed using statistical signal processing involving the primary steps of feature extraction, feature selection and classification [1, 6, 8, 9, 15].

Different machine learning techniques have been used for such movement recognition, e.g. Support Vector Machines (SVM) [16-17], Decision Trees (DT) [8, 15], Naive Bayes (NB) [15], Multi-Layer Perceptron (MLP) [18], Artificial Neural Networks (ANN) [8], Hidden Markov Models (HMM) [19], or a combination of these techniques [10]. Instance-based classification algorithms have also been used successfully to classify data from inertial sensors but suffer from high memory usage and long processing times [20].

By comparison, very little has been reported in terms of recognizing fine grain activities e.g., upper limb movement in out-of-laboratory settings, which is an important aspect for assessing rehabilitation of impaired limb functionality such as in stroke. Furthermore, to the best of our knowledge, very little has been reported on clustering techniques for upper limb activity recognition. Nevertheless, clustering techniques have been successfully used in diverse fields such as medicine (EEG, Functional MRI), geography or marketing and can be conveniently deployed with limited resources (memory and CPU) [21]. We extend the application area further to activity recognition involving the upper limb by using the widely popular *k*-means partitional clustering algorithm. A major advantage of the *k*-means algorithm is its computational simplicity making it an attractive choice for a wide variety of applications [22]. It is a well-perceived fact in the research community that cluster analysis is primarily used for unsupervised learning where the class labels for the training data are not available. However, the *k*-means algorithm can also be used for supervised learning where the class labels of the training data are known a priori [23], [24]. In our proposed methodology, we are aware of the class labels for the training data pertaining to the three movements performed in a constrained *training* phase. This helps to have a definitive estimate of the underlying cluster structure to be formed on the data (three clusters), thereby facilitating a faster convergence during cluster formation for reduced time complexity [22]. To verify the robustness of our proposed methodology we use two conventional supervised learning algorithms and compare the achieved results. We use the LDA classifier and the SVM classifier with a RBF kernel, which is trained on the data collected in the *training* phase and evaluated to predict the movements performed during the ADL phase (*testing*) using the same best features used by the minimum distance classifier in our proposed methodology. LDA is considered to be computationally less complex [25] and SVM is a very popular algorithm in the machine learning community and generally produces high accuracy rates with moderate computational complexity (depending on the number of support vectors used) [15], [18]. In principle, SVM is a binary classifier but has been extended to handle multiple classes using the 'one versus all' or the 'one versus one' approach [26]. However, both of these methods can be computationally intensive depending on the number of target classes. Hence, we have used the toolbox LIBSVM which is a library for SVM and provides an efficient platform for multi-class classification [27].

The accuracy of any movement recognition technique is dependent on the system components and requirements, covering areas such as: type of activities, number of activities, type of sensors, number of sensors, placement of sensors [18], level of data fusion, and most importantly the classification methodology adopted. Further, there is a need for subject specific training especially for tracking activities that are susceptible to individual and temporal variation [28]. Recognition strategies generally follow one of three themes. Firstly, using only data collected under controlled conditions (e.g. in the laboratory) for training as well as testing, which results in high accuracies [8]. Secondly, using both controlled and un-controlled data (e.g. out-of-laboratory) for both training and testing, which results in reasonably high accuracies [6, 28]. Finally, using controlled data for training and only un-controlled data for testing, which generally results in lower accuracies but is more realistic of real-world applications [6, 28].

In this work we opted to use controlled data for training and un-controlled data for testing in order to explore the levels of recognition accuracy for a robust classification mechanism applicable in the field of home based rehabilitation. Here, a subject is instructed to follow a particular exercise regime involving the impaired arm in a controlled environment (clinic or home) and is later monitored to track occurrences of these specific exercises while they perform

daily activities that involve the impaired arm, facilitating a measure of rehabilitation progress. The highlights of our exploration can be summarised as: (1) we demonstrate a completely personalized approach of detecting elementary arm movements that accommodates different levels of impairment and/or rehabilitation status; (2) we analyse kinematic data of healthy subjects and stroke patients to test the robustness of our methodology; and (3) we demonstrate a system that does not require periodic training and can therefore be realistically implemented for real-time detection of arm movements in a resource constrained environment of a wireless sensor node (WSN) [29].

3. Experimental Protocol

A Shimmer 9DoF wireless kinematic sensor module comprising mutually orthogonal tri-axial accelerometers, rate gyroscopes and magnetometers, was used as the sensing platform [30]. For our experiments we use the tri-axial accelerometer (range ±1.5 g) and gyroscope (range ± 500 °/s) and exclude the magnetometer since it can be affected by the presence of ferromagnetic materials which are expected to be present in the home environment [31]. The dorsal side of the forearm proximal to the wrist on the dominant arm for healthy subjects or impaired arm for stroke patients was chosen as the sensing position. The dorsal side was in contact with the XY plane of the sensor with the X-axis pointing towards the hand and the Z-axis pointing away from the dorsal aspect. Sensor data was collected at a rate of 50 Hz, deemed sufficient for assessing habitual limb movement, which is on the high side when compared to assessing holistic activity as in [15, 18]. We focused on three elementary types of arm movement (*actions*) [11]:

- *Action A* – Reach and retrieve an object (extension and flexion of the forearm).
- *Action B* – Lift cup to mouth (rotation of the forearm about the elbow).
- *Action C* – Perform pouring or (un)locking action (rotation of the wrist about long axis of forearm).

In principle, these elementary movements constitute a significant proportion of the complex movements performed with the upper limb in daily life and also resemble three of the tasks in the standard Wolf Motor Function Test (WMFT); an established clinical assessment method for quantifying upper extremity motor ability [32-34].

In this investigation, experiments were performed at the University of Southampton (UoS) with four healthy subjects (age range 24 to 40, male, all right arm dominant) and at the Brandenburg Klinik (BBK) with four stroke impaired patients (age range 45 to 73, both sexes, both left and right arm dominant). Experiments were performed within an open laboratory with an attached kitchen at UoS and within a treatment centre at BBK under the supervision of the expert physiotherapist members of the research team, using the same set of equipment. To generate the *training* phase data for the target cluster formation all four healthy participants performed 240 trials of *Action A*, 120 trials of *Action B* and 120 trials of *Action C*, separated into groups of five repetitions, with each group of trials being separated by approximately three minutes. This was done to avoid unrepresentative data due to fatigue and to minimize the effects of unconscious self-learning of the activities. The four stroke patients performed – 80 trials of *A* and 40 trials each of *B* and *C* (patients 1 and 4) and 40 trials of *A* and 20 trials each of *B* and *C* (patients 2 and 3) separated similarly into groups of five repetitions for each action. This data collection produces a large *training* set offering greater potential for accurate recognition [28] since the cluster formulation on the *training* data inherently captures the person-centric nature of movement patterns. There were more number of trials for *Action A* since it is a generic movement performed more frequently in our daily lives as is further evident from Table I.

We designed an activity-list (cf. Table 1) which emulated the process of '*making-a-cup-of-tea*', a common activity performed in daily life, having repeated occurrences of the three elementary types of arm movement (*actions*) [11]. The activity list in our experiment protocol comprises 20 individual activities including 10 occurrences of *Action A*, and 5 each of *Action B* and *Action C*. There were no restrictions on the various physical factors of the experiment such as the

seating position or standing position with respect to the kitchen surface or the time required to complete the actions. The experiment was unconstrained in this manner to ensure a wider range of variability in the data paving the way for a robust arm movement recognition system which will produce acceptable levels of accuracy in a real world application.

	Activity	Action
1.	Fetch cup from desk	A
2.	Place cup on kitchen surface	A
3.	Fetch kettle	A
4.	Pour out extra water from kettle	C
5.	Put kettle onto charging point	A
6.	Reach out for the power switch on the wall	A
7.	Drink a glass of water while waiting for kettle to boil	B
8.	Reach out to switch off the kettle	A
9.	Pour hot water from the kettle in to cup	C
10.	Fetch milk from the shelf	A
11.	Pour milk into cup	C
12.	Put the bottle of milk back on shelf	A
13.	Fetch cup from kitchen surface	A
14.	Have a sip and taste the drink	B
15.	Have another sip while walking back to desk	B
16.	Unlock drawer	C
17.	Retrieve biscuits from drawer	A
18.	Eat a biscuit	B
19.	Lock drawer	C
20.	Have a drink	B

Table1: Use case activity list - 'Making-a-cup-of-tea'

On a separate day, the healthy participants were recalled and requested to perform four repetitions of the activity-list in Table I, at a comfortable speed in a kitchen, with a 10 minute rest period between repetitions. Similarly, the stroke patients were requested to perform 2 repetitions of the same activity-list. The disparity in the number of trials performed by healthy subjects and stroke patients was due to the fact that the latter tend to tire quickly and were asked to perform the tasks only whilst they felt comfortable to do so. The activity-list was prepared to facilitate the evaluation of the recognition methodology under semi-naturalistic conditions [28]. Dense sensing-based activity monitoring has been used to detect such holistic activities as 'making tea or coffee' by the use of ambient and wearable sensors to detect user-object interactions within the paradigm of ambient assisted living or smart homes [11]. In this work our focus is on detecting the three elementary movements used during the archetypal activity of '*making-a-cup-of-tea*'. The start and stop time of the activities were noted down by the researcher observing them as they performed the designated tasks. The corresponding data collected was segmented using the annotations from the researcher and used for the *testing* phase. Since our focus was primarily on recognizing the occurrence of these elementary movements, we did not implement an automated segmentation method.

4. Data Processing

The primary steps involved in our data processing for each subject's data is illustrated in Figure 1 and described in detail in the following sections.
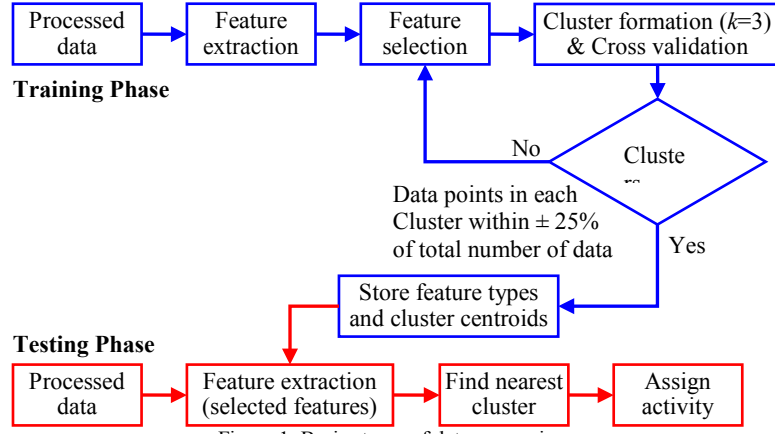
Figure 1. Basic stages of data processing.

4.1. Acquisition and Pre-processing

The tri-axial accelerometer and gyroscope located on the wrist transmit data along with a time stamp to a host computer using the Bluetooth wireless transmission protocol. The raw sensor data is band-pass filtered with a 3rd order Butterworth filter having cut-off frequencies of 0.1 Hz and 12 Hz to respectively attenuate the low frequency artefacts and high frequency noise components introduced in the data due to physical effects such as drift [15].

4.2. Feature Extraction

Typical feature sets for human activity recognition include statistical functions, time and/or frequency domain features, as well as heuristic features [9]. Each accelerometer and gyroscope data stream (x, y and z) exhibit signal patterns that are distinctive for each of the arm movements, which is characterized by a set of features extracted from the signals [15]. In this investigation, we consider 10 time-domain features, extracted from the data from each of the three accelerometer axes and from each of the three gyroscope axes as follows: 1) standard deviation – measure of the variability from the mean of the signal, 2) root mean square (rms) –measure of the signal energy normalized by the number of samples, 3) information entropy - measure of the randomness of a signal [35], 4) jerk metric - rms value of the second derivative of the data normalized with respect to the maximum value of the first derivative [36], 5) peak number - obtained from gradient analysis of the signal, 6) maximum peak amplitude - measure of the amplitude of the peaks obtained after gradient analysis, 7) absolute difference - absolute difference between the maximum and the minimum value of a signal, 8) index of dispersion - ratio of variance to the mean, 9) kurtosis - measure of the 'peakedness' of a signal assuming a non-Gaussian distribution in the data, 10) skewness - measure of the symmetry of the data assuming a non-Gaussian distribution in the data [37].

Although the last two features are usually associated with defining the shape of a probability distribution, they can still be used as classifying features if they routinely return values that distinguish one pattern of data from another. Moreover, features such as jerk metric, peak number and maximum peak amplitude are reflective of the movement fluidity associated with assessing motor functionality in human limb [38]. Hence, we compute 10 one-dimensional features on each individual accelerometer (*acc_x*, *acc_y*, *acc_z*) and gyroscope (*gyro_x*, *gyro_y*, *gyro_z*) data segment for each movement trial of each subject. The subsequent process of feature selection and cluster formation is performed on the sensor specific feature space (comprising of 30 features), represented as:

$$FS_{sensor} = [f_1\_x \cdots f_{10}\_x, f_1\_y \cdots f_{10}\_y, f_1\_z \cdots f_{10}\_z] \tag{1}$$

where FS_{sensor}, represents the respective feature space for each sensor type (accelerometer or gyroscope) formed by considering the 10 individual features ($f_1 \ldots f_{10}$) computed on each tri-axial data segment together. The suffix (x, y or z) represents the sensor axis on which the respective feature was computed. As mentioned in section 3, the movements

performed by the subjects were segmented using the annotations from the researcher, therefore each feature was computed on data segments of varying lengths representative of the time taken to complete each movement trial. This is further reflected in Figures 2 and 3 where we illustrate the typical variations in accelerometer and gyroscope data recorded during a single example of each action for a healthy and a stroke patient respectively. It clearly highlights the difference in movement profiles among the two groups and the longer time taken by the stroke survivors to complete the actions with less smoothness of movement. The features are linearly normalized and the best features for each subject are selected as discussed in section 4.3.
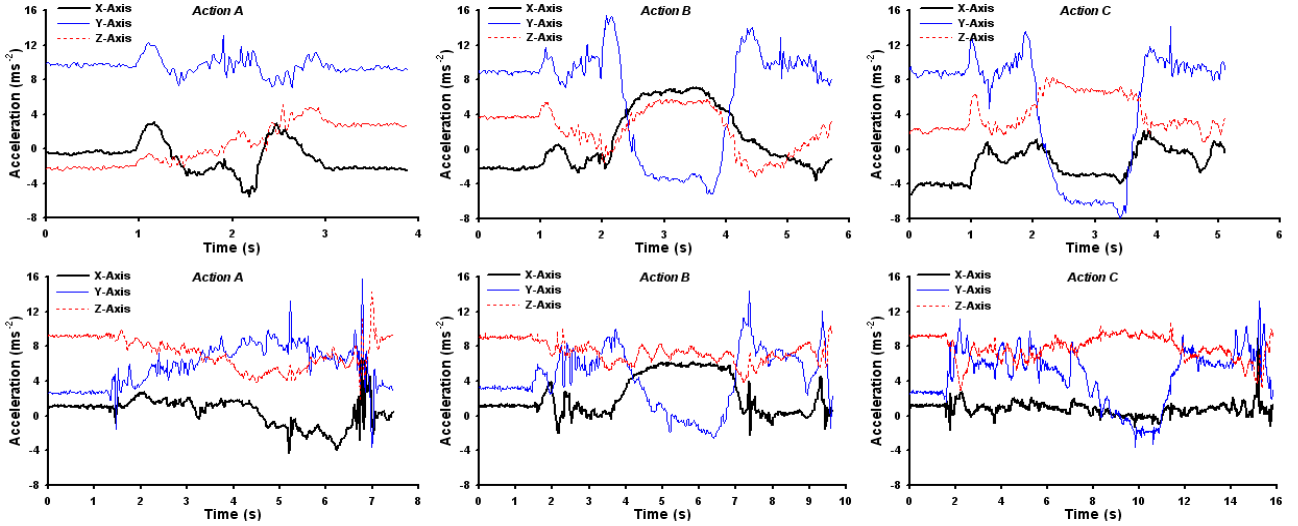

Figure 2. Data from a tri-axial accelerometer located on the wrist collected while performing arm actions *A*, *B* and *C* from a healthy subject (upper) and a stroke survivor (lower).
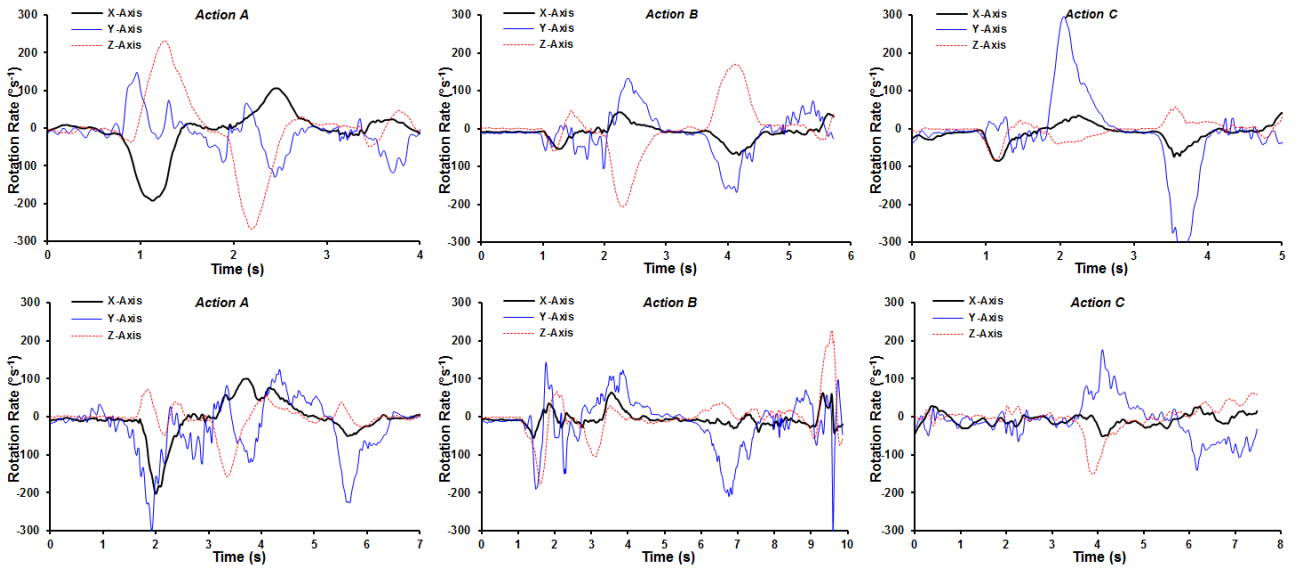

Figure 3. Data from a tri-axial rate gyroscope located on the wrist collected while performing arm actions *A*, *B* and *C* from a healthy subject (upper) and a stroke survivor (lower).

4.3. Feature Selection

The RELIEF algorithm [36], Clamping technique [18] and Principal Component Analysis (PCA) [12] are the most commonly used ranking/selection algorithms in the field of human activity recognition but are computationally intensive. We use the low-complexity class-separability measure based on scatter matrices to rank the 30 features for each sensor/movement combination. The scatter matrices quantify the scatter of feature vectors in the feature space. The rank of each individual feature for a multiple-class scenario is determined by the R value as shown below [22]:

8

$$R = \frac{S_w + S_b}{S_w} \tag{2}$$

where S_w and S_b are the within-class and between-class scatter matrices respectively. A high value of R represents a small within-class variance and a large between-class distance among the data points in the respective feature space [22]. The ranked features are sorted in descending order with respect to their R values. We employ a sequential forward selection (*sfs*) technique, selecting the first i features of the ranked feature set in each iteration ($i = 2,…,30$) and check if the data from the *training* phase can be correctly clustered in a multi-dimensional feature space as described in the next section.

4.4. Cluster Formation

The fundamental concept of cluster analysis is to form groups of similar objects as a means of distinguishing them from each other and can be applied in any discipline involving multivariate data [21]. With a given dataset $X = \{x_i\}$, $i = 1,...,n$ to be clustered into a set of k clusters, the k-means algorithm iterates to minimize the squared error between the empirical mean of a cluster and the individual data points, defined as the cost function, J:

$$J(\theta,u) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij} \left\| x_i - \theta_j \right\|^2 \tag{3}$$

where θ_j is the cluster center and $u_{ij} = 1$ if x_i lies close to θ_j, or 0 if otherwise [39]. Initially k centroids are defined and the data vectors are assigned to a cluster label depending on how close they are to each centroid. The k centroids are recalculated from the newly defined clusters and the process of reassignment of each data vector to each new centroid is repeated. The algorithm iterates over this loop until the data vectors from the dataset X form clusters and the cost function J is minimized [22].

The Euclidean distance used to compute the squared distance between the vectors x_i and the mean of each cluster θ_j has an undesirable effect of splitting large and elongated clusters, since most real datasets do not have a well-defined, isolated and spherical underlying cluster structure. By comparison, the use of the Mahalonobis distance which involves computing the covariance matrix of the data vector causes a large cluster to absorb nearby smaller clusters, leading to the creation of unusually large or small clusters. Hence we use the regularized Mahalonobis distance as mentioned in [39] which prevents the clustering algorithm from producing unusually large or small clusters.

$$J(x_i,\theta_j) = (x_i - \theta_j)^T [(1-\lambda)(\textstyle\sum_j + \varepsilon I)^{-1} + \lambda I](x_i - \theta_j) \tag{4}$$

where Σ_j is the covariance matrix of the k-th cluster and I is the $d \times d$ identity matrix, d is the input dimensionality (no. of feature vectors representing the data vector) and ε (10^{-6}) is the regularization parameter. The value of λ can be used as a parameter to control the choice of distance measure to be used, with $\lambda=0$, J is the squared Mahalonobis distance and when $\lambda=1$, J is the squared Euclidean distance [39]. In our exploration, we start with an initial value of $\lambda=1$ and after 3 iterations change it to $\lambda=0$. The cluster formation and cross validation steps are illustrated in Figure 4.

We run our algorithm on the feature vectors computed from the accelerometer and gyroscope data which characterizes the movement trials of the *training* phase (480 trials for each healthy subject, 160 trials for patients 1 and 4 and 80 trials for patients 2 and 3) to form three clusters representing the three arm movements. The algorithm runs in conjunction with the *sfs* algorithm sequentially selecting a combination of 2 to 30 ranked features in each step (i). For a particular set of feature vectors selected (i), 10 runs (n) of 10-fold cross validation are carried out whereby 10 segments of the *training* data are created. In each run of the stipulated 10 runs, one segment is used as the *test* dataset while the rest of the 9 segments are used as the *training* dataset. In each run (n), the k-means clustering runs on the *training* dataset.
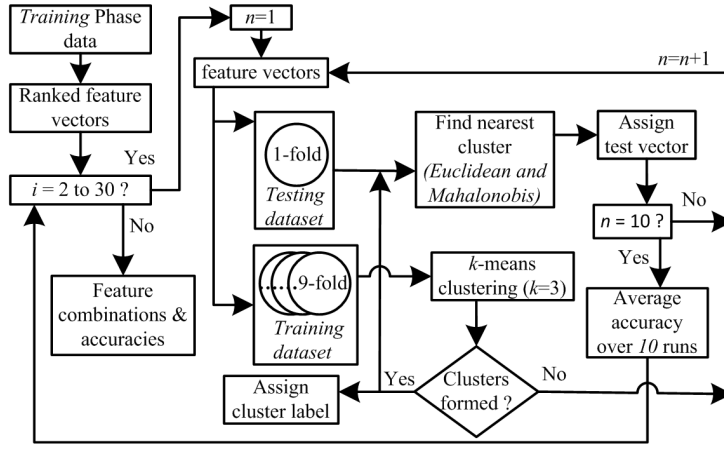
9

Figure 4. Overview of cluster formation and cross validation on the *Training* phase data collected from each sensor and each subject.

We set a threshold of 25% of the expected number of data points for each of the three clusters formed (i.e. for healthy subjects: 240 ± 60 for *Action A* and 120 ± 30 for *Action B* and *Action C*, for patients 1 and 4: 80 ± 20 for *Action A*, 40 ± 10 for *Action B* and *Action C*, for patients 2 and 3: 40 ± 10 for *Action A*, 20 ± 5 for *Action B* and *Action C*). This threshold value was experimentally selected since it produced the best results. If the number of data points in each cluster is within the threshold, we consider it as correctly clustered for that particular combination of features selected (i = 2,...,30). We compute the distance of the mean of the *training* dataset for each class label from the cluster centroids and thereby assign each cluster with the class label that has its closest proximity to that particular class of the *training* dataset.

We then use a minimum distance classifier [22] to compute the distance of the test vector (*test* dataset) from the centroid of each cluster in a multi-dimensional feature space (considering the feature combination of the current step, i) based upon: a). Euclidean distance and b). Mahalonobis distance. The Mahalonobis distance is used to measure the distance of a point from a data distribution. The data distribution is characterized by the mean and the covariance matrix which defines the shape of how the data is distributed in the feature space and is generally hypothesized as a multivariate Gaussian distribution. Here, the Mahalonobis distance takes into consideration the covariance of the clusters along with their mean for the maximum likelihood estimation of the covariance matrix and hence is effective for clusters with larger variance along one or many directions and in general having an ellipsoidal shape. The *test* dataset is assigned to a particular cluster depending on the minimum distance computed for each of the two measures. The predicted label is verified with respect to the known annotations thereby ascertaining the accuracy of the prediction for a single run. The accuracy of prediction for a particular feature combination is determined by averaging the results produced over the 10 runs. This process is repeated for each of the sequentially selected feature combinations. Therefore at the end (i = 30), we have a detailed list for each subject and each sensor of the feature combinations that resulted in a successful cluster formation and the corresponding accuracies achieved both with Euclidean and Mahalonobis distance measures.

The data from the ADL phase (*testing*) (80 test vectors for each healthy subject [(10*A* + 5*B* + 5*C*) × 4 trials] and 40 test vectors for each stroke patient [(10*A* + 5*B* + 5*C*) × 2 trials]) is pre-processed for each type of sensor, and only those features are extracted from each test vector which resulted in the best accuracy in the cross validation of the *training* data. We use a Euclidean and Mahalonobis distance based classifier to compute the distance of each test vector (represented by the extracted features) from the centroid of each cluster in a multi-dimensional feature space. The test vector is assigned to a particular cluster depending on the minimum distance computed for each of the two measures. The predicted label is verified with respect to the annotations in the activity-list of Table I.

5. Results and Analysis

Overall correct classification or accuracy is generally a viable measure to estimate the performance of a binary classifier (considering two classes). For multi-class classification problems, with class imbalance (differing number of test patterns associated with each class) and having possible dissimilar classification rates of different classes, it is also essential to measure the classification level for each class, also known as the sensitivity [40]. Hence the sensitivity (S) of a given class i (where, $i = A, B, C$) is represented by the number of patterns correctly predicted to be in class i with respect to the total number of patterns in class i. The overall accuracy of movement detection for each subject is represented by the sum of individual class sensitivities (success rate of individual movements) with respect to the total number of test patterns to be classified. This can be further illustrated with an example confusion matrix, as shown in Figure 5. The diagonal and the off-diagonal elements of the confusion matrix correspond to correctly classified and misclassified patterns respectively. This example shows near perfect classification since all diagonal elements approach unity and all off-diagonal elements approach zero. Therefore, the sensitivity of class A (expressed as a percentage), can be computed as $S_A = \dfrac{0.95}{(0.95 + 0.05)} = 95\%$ and the overall accuracy (expressed as a percentage) can be computed as

$$Accuracy = \frac{(0.95 + 0.9 + 0.98)}{3} = 94\%.$$

	Predicted 'A' $j = 1$	Predicted 'B' $j = 2$	Predicted 'C' $j = 3$
Actual 'A', $i = 1$	0.95	0.05	0
Actual 'B', $i = 2$	0.1	0.9	0
Actual 'C', $i = 3$	0.02	0	0.98

Figure 5. Sample confusion matrix for three classes (A, B, C)

We had in total 80 movement trials (*actions*) to be recognized (40 of *A*, 20 of *B*, 20 of *C*) for each healthy subject and 40 movement trials to be recognized (20 of *A*, 10 of *B*, 10 of *C*) for each stroke patient. The sensitivity of recognizing each action performed in the ADL phase and the overall accuracy for the healthy subjects using accelerometer and gyroscope data are presented in Tables 2 and 3 respectively. The tables also show the minimum number of features that were required to successfully form the three clusters for each subject. The number of features has been determined by 10 runs of 10-fold cross validation on the *training* phase data as discussed in Section 4. The results in general show that each subject required a different minimum number of features to successfully form 3 separate clusters from the *training* data, reflecting the variability in arm movement patterns between individuals. The minimum distance classifier used for recognizing the arm movements has also been shown in each table. The right hand column in Tables 2, 3, 5 and 6 show the overall detection accuracy (total number of recognized actions expressed as a percentage of the total number of actions performed) for each subject.

5.1. Healthy Subjects

The overall accuracy covers the range 61% to 100% (average of 88%) using accelerometer data and 60% to 94% using gyroscope data (average of 83%) for all healthy subjects. In general, these recognition accuracies are quite favourable considering the elementary nature of the arm movements being detected and using only controlled data collected in the laboratory for *training*. The obvious exceptions are the detection of *Action B* for Subject 2 (sensitivity of 5%) using accelerometer and *Action A* for Subject 4 (sensitivity of 30%) using gyroscope data respectively. It is

worth mentioning that Subject 2 required the smallest number of features (2) to form clusters from the *training* data. This is somewhat counter-intuitive – fewer features imply sufficient differences in arm movement patterns to make unique cluster formation easier. Whilst this may be the case, however, the low detection accuracy for *Action B* could be accounted for by poor repeatability by the subject in this particular arm movement. Similarly, for Subject 4, using 20 features and a Euclidean distance classifier the low sensitivity for *Action A* may be attributed to poor repeatability. However, for these specific cases, for Subject 2, the sensitivity of *Action B* with gyroscope data is 80% (27 features) and for Subject 4, the sensitivity of detecting *Action A* with accelerometer data is 95% (with 23 features). Therefore, although the overall recognition accuracies using both accelerometer and gyroscope data are nearly similar, for specific cases considering more than one sensor type can improve the overall accuracy of detection.

Subject	Features	Minimum Distance Classifier	Sensitivities (%)			Overall accuracy (%)
			A	*B*	*C*	
Subject1	11	Euclidean	100	100	100	100
Subject2	2	Euclidean	80	5	80	61
Subject3	7	Euclidean	95	100	90	95
Subject4	23	Euclidean	95	100	85	94

Table 2: Recognition sensitivities for each arm movement and overall accuracies with accelerometer data for healthy subjects.

Subject	Features	Minimum Distance Classifier	Sensitivities (%)			Overall accuracy (%)
			A	*B*	*C*	
Subject1	10	Euclidean	93	90	100	94
Subject2	27	Euclidean	100	80	60	85
Subject3	18	Mahalonobis	90	90	100	93
Subject4	20	Euclidean	30	95	85	60

Table 3: Recognition sensitivities for each arm movement and overall accuracies with gyroscope data for healthy subjects.

We present a list of features selected from a ranked list (sorted in descending order) in Table 4, specific to each subject and each sensor, illustrating the difference in the number of features required to form the clusters. Although for each subject the ranked order of features is different (reflecting the different ways in which they perform a movement) there is a strong commonality in the top ranked features across all subjects for both the sensors. Specifically, the features *stddev_y* and *rms_y* are the top two features extracted on the accelerometer data across all subjects except for Subject 3, where it lies within the best five features used. By comparison, the features *rms_x*, *stddev_x* and *diff_x* are routinely amongst the best five features extracted from the gyroscope data for each subject. The recognition accuracies for Subject 2 with accelerometer data using additional features are illustrated in Figure 6, which reveals that increasing the number of features beyond 2 does not yield successful cluster formations (blank spaces) or improved accuracy.

Accelerometer	
Subject	**Ranked Features**
Subject1	stddev_y, rms_y, rms_z, stddev_z, rms_x, diff_y, stddev_x, diff_z, max_mag_y, diff_x, max_mag_z
Subject2	stddev_y, rms_y
Subject3	rms_z, rms_x, stddev_y, stddev_x, rms_y, entropy_z, stddev_z
Subject4	stddev_y, rms_y, stddev_x, rms_x, diff_y, max_mag_y, diff_x, max_mag_x, kurtosis_x, kurtosis_z, skewness_z, entropy_y, diff_z, max_mag_z, kurtosis_y, stddev_z, entropy_x, skewness_x, peaks_y, skewness_y, entropy_z, rms_z, peaks_x
Gyroscope	
Subject	**Ranked Features**
Subject1	rms_x, stddev_x, rms_z, diff_x, diff_z, rms_y, stddev_z, max_mag_x, max_mag_z, stddev_y
Subject2	rms_x, stddev_x, diff_x, max_mag_x, stddev_y, rms_y, diff_y, max_mag_y, rms_z, stddev_z, entropy_y, skewness_z, diff_z, entropy_z, entropy_x, skewness_x, skewness_y, kurtosis_y, kurtosis_x, kurtosis_z, max_mag_z, jerk_y, peaks_x, jerk_z, peaks_z, jerk_x, peaks_y
Subject3	rms_z, rms_x, stddev_z, stddev_x, diff_x, diff_z, max_mag_x, max_mag_z, max_mag_y, entropy_y, diff_y, entropy_z, rms_y, entropy_x, stddev_y, skewness_y, skewness_z, peaks_x
Subject4	rms_x, stddev_x, diff_x, max_mag_x, rms_z, max_mag_y, stddev_z, stddev_y, rms_y, diff_y, skewness_y, skewness_z, diff_z, entropy_x, kurtosis_y, entropy_z, jerk_y, max_mag_z, entropy_y, kurtosis_z

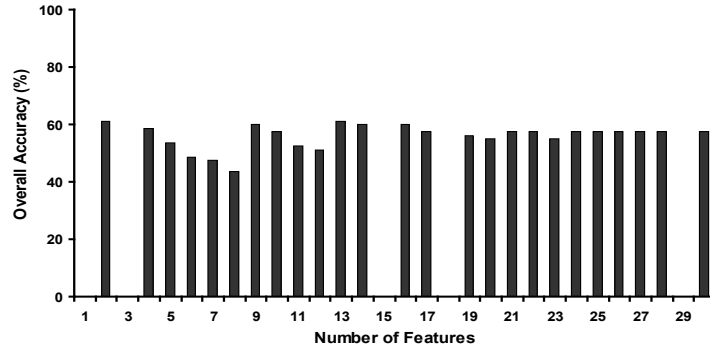Table 4: Selected features in ranked order for healthy subjects

Figure 6. Change in accuracy with the number of features selected for healthy Subject 2 with accelerometer data using Euclidean distance.

5.2. Stroke Patients

For the stroke patients, the overall accuracy is in the range of 40% to 88% (average of 70%) using accelerometer data and 40% to 83% (average of 66%) using gyroscope data as shown in Tables 5 and 6 respectively. Under closer examination, we observe that for Subject 1 the sensitivity for individual actions with acceleration data are above 80%, but for Subjects 2 and 3, although the overall accuracy is above 70%, the sensitivities for *Action B* and *Action C* are quite low (20%). When we consider the gyroscope data, these particular action/patient combinations are improved considerably: 100% sensitivity for *Action B* with Subject 2 and 80% sensitivity for *Action C* with Subject 3. For Subject 4, the overall results with both sensors are not high, although we can see that *Action A* can be recognized by 60% (gyroscope), *Action B* by 80% and *Action C* by 60% (accelerometer). The low overall accuracy can be attributed to the fact that Subject 4 was at an early stage of rehabilitation and the impaired arm being tested was not the naturally dominant arm thereby resulting in poor repeatability. This further emphasizes how detection accuracies may be improved by considering more than one sensor type for specific cases. Moreover, it is worth noting that for stroke patients the Mahalonobis distance based classifier is more effective than the Euclidean distance. This is further reflective of the fact that there is a high degree of variability in their movement profile resulting in clusters having a larger variance along one or many directions, under which conditions the Mahalonobis distance acts as a more effective distance measure.

| Subject | Features | Minimum Distance Classifier | Sensitivities (%) | | | Overall accuracy (%) |
			A	B	C	
Subject1	19	Mahalonobis	80	90	100	88
Subject2	19	Mahalonobis	90	20	100	75
Subject3	21	Mahalonobis	95	100	20	78
Subject4	8	Euclidean	10	80	60	40

Table 5: Recognition sensitivities for each arm movement and overall accuracies with accelerometer data for stroke patients.

| Subject | Features | Minimum Distance Classifier | Sensitivities (%) | | | Overall accuracy (%) |
			A	B	C	
Subject1	8	Euclidean	90	50	100	83
Subject2	10	Euclidean	60	100	60	70
Subject3	24	Mahalonobis	85	30	80	70
Subject4	30	Mahalonobis	60	40	0	40

Table 6: Recognition sensitivities for each arm movement and overall accuracies with gyroscope data for stroke patients.

The ranked list of features selected for each stroke patient using each sensor is presented in Table 7. In comparison to healthy subjects, the features selected for stroke patient are quite different, exhibiting a minimal degree of commonality among the top ranked features. This is primarily due to the fact that each of the four subjects had different levels of functional ability (as assessed by the therapists). Hence, there was a great degree of intra-subject and inter-subject variability in the movement patterns.

Accelerometer	
Subject	**Ranked Features**
Subject1	stddev_y, stddev_x, rms_x, rms_y, entropy_y, diff_x, entropy_z, max_mag_x, stddev_z, rms_z, peaks_z, entropy_x, diff_y, peaks_x, diff_z, kurtosis_x, max_mag_z, peaks_y, kurtosis_y
Subject2	diff_x, stddev_y, skewness_x, entropy_y, rms_y, stddev_x, rms_x, peaks_x, entropy_z, entropy_x, max_mag_x, peaks_z, kurtosis_y, diff_y, max_mag_y, peaks_y, kurtosis_x, skewness_y, max_mag_z
Subject3	entropy_z, entropy_y, entropy_x, stddev_z, rms_z, rms_y, stddev_y, skewness_z, skewness_x, kurtosis_z, diff_z, kurtosis_x, max_mag_z, peaks_z, kurtosis_y, peaks_x, peaks_y, skewness_y, diff_x, rms_x
Subject4	entropy_y, entropy_z, stddev_y, peaks_x, entropy_x, rms_x, stddev_x, rms_y
Gyroscope	
Subject	**Ranked Features**
Subject1	stddev_x, rms_x, diff_x, max_mag_x, stddev_z, rms_z, entropy_z, entropy_y
Subject2	diff_z, stddev_z, rms_z, max_mag_z, diff_x, stddev_x, rms_x, entropy_y, max_mag_x, stddev_y
Subject3	diff_z, stddev_z, rms_z, max_mag_z, entropy_y, entropy_z, entropy_x, stddev_x, rms_x, skewness_x, diff_x, max_mag_x, jerk_x, peaks_z, peaks_x, peaks_y, skewness_y, stddev_y, kurtosis_z, jerk_y, kurtosis_y, rms_y, kurtosis_x, diff_y
Subject4	skewness_y, stddev_y, entropy_y, entropy_z, stddev_x, rms_x, peaks_z, max_mag_x, diff_x, peaks_x, entropy_x, peaks_y, rms_y, diff_y, stddev_z, rms_z, skewness_x, diff_z, jerk_y, max_mag_y, kurtosis_y, max_mag_z, kurtosis_z, kewness_z, jerk_x, kurtosis_x, jerk_z, disp_y, disp_z, disp_x.

Table 7: Selected features in ranked order for stroke patients.

5.3. Evaluation with LDA and SVM

For comparing the performance of our proposed methodology we use two well-known supervised learning algorithms LDA and SVM for classifying the three arm movements performed during the ADL phase. We use the data collected in the *training* phase to train the classifiers and evaluate them to detect the movements performed in the ADL (*testing*) phase. In this investigation we used the same set of features as used in our proposed clustering and minimum distance classifier based methodology (cf. Tables 4 and 7) for both the healthy subjects and the stroke patients. The classification results are presented in Tables 8 – 11. Using LDA, for the healthy subjects, the overall accuracy is in the range of 25% to 57% (average of 45%) using accelerometer data and 42% to 60% (average of 53%) using gyroscope data as shown in Table 8. For the stroke patients, the overall accuracy was in the range of 32% - 60% (average of 49%) using accelerometer data and a range of 35% - 67% (average of 46%) using gyroscope data as shown in Table 9.

Subject	Accelerometer				Gyroscope			
	Sensitivities (%)			Overall accuracy (%)	Sensitivities (%)			Overall accuracy (%)
	A	*B*	*C*		*A*	*B*	*C*	
Subject1	45	80	0	42	60	0	95	54
Subject2	70	0	85	56	30	15	95	42
Subject3	60	95	15	57	100	15	10	56
Subject4	0	100	0	25	100	0	40	60

Table 8: Sensitivities for each arm movement and overall accuracies for healthy subjects using LDA.

Subject	Accelerometer				Gyroscope			
	Sensitivities (%)			Overall accuracy (%)	Sensitivities (%)			Overall accuracy (%)
	A	*B*	*C*		*A*	*B*	*C*	
Subject1	35	100	40	52	40	90	100	67
Subject2	15	0	100	32	20	10	100	37
Subject3	60	100	20	60	35	0	70	35
Subject4	95	0	20	52	60	0	60	45

Table 9: Sensitivities for each arm movement and overall accuracies for stroke patients using LDA.

Using SVM, for the healthy subjects, the overall accuracy is in the range of 48% - 61% (average of 54%) using accelerometer data and 67% - 70% (average of 68%) using gyroscope data as shown in Table 10. For the stroke patients, the overall accuracy was in the range of 35% to 70% (average of 55%) using accelerometer data and 25% to 85% (average of 50%) using gyroscope data as shown in Table 11.

Subject	Accelerometer				Gyroscope			
	Sensitivities (%)			Overall accuracy (%)	Sensitivities (%)			Overall accuracy (%)
	A	*B*	*C*		*A*	*B*	*C*	
Subject1	50	95	0	48	85	80	20	67
Subject2	80	5	80	61	87	75	25	68
Subject3	45	95	10	49	90	75	20	69
Subject4	65	95	5	57	95	70	20	70

Table 10: Sensitivities for each arm movement and overall accuracies for healthy subjects using SVM.

Subject	Accelerometer				Gyroscope			
	Sensitivities (%)			Overall accuracy (%)	Sensitivities (%)			Overall accuracy (%)
	A	*B*	*C*		*A*	*B*	*C*	
Subject1	30	70	100	57	95	60	90	85
Subject2	55	100	70	70	100	0	0	50
Subject3	65	20	90	60	30	0	100	40
Subject4	10	50	70	35	5	0	90	25

Table 11: Sensitivities for each arm movement and overall accuracies for stroke patients using SVM.

The overall accuracies using LDA and SVM are comparatively on the lower side when compared to the results achieved using our proposed methodology (cf. Tables 2, 3, 5 and 6). The sensitivities for individual actions achieved using LDA and SVM are high for some cases, though none of the subjects have all three movements classified with a sensitivity of greater than 60%. This therefore proves the effectiveness of our proposed methodology. We believe our proposed methodology can provide a gross measure of the impaired arm usage and help in assessing the rehabilitation progress of the patients over time outside the clinical environment.

6. Discussion and Future Work

In this paper we describe a proof-of-concept methodology to recognize three fundamental movements of the upper limb that are typically used in activities of daily living, using data collected by a wrist-worn, wireless tri-axial accelerometer and a tri-axial rate gyroscope in conjunction with partitional *k*-means clustering. The clusters are formed using a ranked list of features (extracted from the *training* data), whereby we sequentially select a combination of 2 to 30 features in each iteration and perform 10 runs of 10-fold cross validation on the selected feature vectors. In each run, the training data (comprising of the selected feature vector combinations) are divided into 10 segments out of which one segment acts as the *test* dataset and the remaining 9 segments act as the *training* dataset. The *training* dataset is used to form three clusters using the regularized Mahalonobis distance based *k*-means clustering (*k*=3) and use a minimum distance classifier based upon Euclidean and Mahalonobis distance to associate the *test dataset* with the formed clusters. This cross validation process helps in producing a specific set of feature combinations which is used for associating the *testing* phase data ('*making-a-cup-of-tea*') with Euclidean or Mahalonobis distance metric to the clusters pre-computed on the laboratory setup *training* phase data.

Our results show that we can detect all three arm movements performed in an out-of-laboratory situation with an accuracy of 61-100% (average of 88%) using between 2-23 time domain features extracted from the accelerometer data and a Euclidean distance based classifier. Similarly, the levels of accuracy lie between 60% - 94% (average of 83%) using 10-27 features extracted from the gyroscope data using Euclidean and Mahalonobis distance for all the healthy subjects. For stroke patients, the accuracy lies in the range of 40% to 88% (average of 70%) using accelerometer data and 40% to 83% (average of 66%) using gyroscope data. Furthermore, our results suggest that the accelerometer and the gyroscope can be used to complement each other for cases where the sensitivity of an individual action using data from just one of these types of sensor is low.

We also demonstrate that the number of features used to achieve these high levels of accuracy is subject-specific, reflecting the variability inherent in human movement. There was some degree of commonality amongst the top ranked

features across the four healthy subjects, whereas for stroke patients each subject had a different set of features in the ranked list. It is important to note that various physical factors like age, stage of rehabilitation and affected arm play a significant part in achievable detection accuracies since they affect the level of repeatability of individual movements as well as introducing a high degree of temporal variation. This is evident from the differences in the results when healthy subjects are compared with the stroke impaired patients. A further comparison of our achieved results against the classification results obtained using a LDA and a SVM classifier clearly reflect the effectiveness of our proposed methodology in detecting the three investigated arm movements.

We intend to extend this study further to observe the effects of sensor fusion and other attachment locations with a larger sample population. We believe that this methodology can be implemented for online detection of arm movements in a resource constrained environment of body-worn wireless sensor networks (WSN) to detect, classify and enumerate the various occurrences of prescribed exercises (i.e. arm movements) during normal daily activities, which over time can provide an indication of rehabilitation progress. The *training* phase, cluster formation and feature selection (being relatively time and memory intensive) need only be done in an offline mode when requested by the clinician, depending on the rehabilitation progress of the patient. This will take into account the improved motor functionalities of patients over time. Further, an online detection module can be used to compute only the required features and the distance to the pre-computed cluster centroids in near real-time, thereby providing an energy efficient solution towards WSN operation for long durations. This is particularly relevant in many telehealth systems and applications where trade-offs exist between useable operating lifetime, available power and power source mass [29].

Acknowledgments

Conflict of Interest: The authors report no conflict of interest.

References

[1] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C.J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly," *IEEE Trans. Biomed. Eng.*, vol. 50, no.6, pp. 711-723, Jun. 2003.

[2] A. Hadjidj, A. Bouabdallah, and Y. Challal, "Rehabilitation supervision using wireless sensor networks," *in Proc. IEEE WoWMoM*, Luca, Italy, pp.1-3, Jun. 2011.

[3] M.S. Raisinghania, A. Benoit, and J. Ding, "Ambient intelligence: Changing forms of human-computer interaction and their social implications," *Texas Digital Library*, vol. 5, no.4, 2004.

[4] D. Merrill, J. Kalanithi, and P. Maes, "Siftables: Towards sensor network user interfaces," *in Proc 1st Int. Conf. on Tangible and Embedded Interaction*, Baton Rouge, LA, pp. 75-78, Feb. 2007.

[5] S. Armstrong, "Wireless connectivity for health and sports monitoring: a review," *B. J. Sports Medicine*, vol. 41, pp. 285-289, Jan. 2007.

[6] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no.1, pp. 20-26, Jan. 2008.

[7] F.E. Martınez-Pérez, J.A. González-Fraga, J.C. Cuevas-Tello, and M.D. Rodrıguez, "Activity inference for ambient intelligence through handling artifacts in a healthcare environment," *Sensors*, vol. 12, no.1, pp. 1072–1099, Jan. 2012.

[8] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol.10, no.1, pp.119-128, Jan. 2006.

[9] K. Altun, B. Barshan, O. Tuncel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors", *Pattern Recognition*, vol. 43, no. 10, pp. 3605-3620, Oct. 2010.

[10] C. Zhu and W. Sheng, "Motion- and location-based online human daily activity recognition," *Pervasive and Mobile Computing*, vol. 7, no. 2, pp. 256-269, Apr. 2011.

[11] L. Chen, J. Hoey, C.D. Nugent, D.J. Cook, and Z. Yu, "Sensor-Based activity recognition," *IEEE Trans. Syst. Man And Cybernetics*, vol. 42, no. 6, pp. 790-808, Nov. 2012.

[12] L.Wang, T. Gu, X. Tao, J. Lu, "A hierarchical approach to real-time activity recognition in body sensor networks", *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 115-130, Feb. 2012.

[13] P. Turaga, R. Chellappa, V.S. Subrahmanian, and O. Udrea, "Machine recognition of human activities", *IEEE Trans. Circ. Syst. Video Technol.*, vol. 18, no. 11, pp. 1473-1488, Nov. 2008.

[14] J.J. Guiry, P.V.D. Ven, J. Nelson, L. Warmerdam and H. Riper, "Activity recognition with smartphone support", *Medical Engineering & Physics*, vol. 36, no. 6, pp. 670-675, Jun. 2014.

[15] O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas, "Daily living activity recognition based on statistical feature quality group selection," *Expert Systems with Applications*, vol. 39, no. 9, pp. 8013-8021, Jul. 2012.

[16] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results," *IEEE Trans. Inf. Technol. Biomed.*, vol.14, pp. 274-283, Mar. 2010.

[17] D. Fuentes, L. Gonzalez-Abril, C. Angulo, and J.A. Ortega, "Online motion recognition using an accelerometer in a mobile device," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2461-2465, Feb. 2012.

[18] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1662-1674, Apr. 2013.

[19] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognition*, vol. 41, no. 6, pp. 2010-2024, Jun. 2008.

[20] N. Bicocchi, M. Mamei, F. Zambonelli, "Detecting activities from body-worn acclerometers via instance-based algorithms", *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 482-495, Aug. 2010.

[21] T.Warren Liao, "Clustering of time series data-a survey", *Pattern Recognition*, vol. 38, no. 11, pp. 1857-1874, Nov. 2005.

[22] S. Theodoridis and K. Koutroumbas, "Pattern Recognition", 4th ed., Elsevier, pp. 30-31 and pp. 280-288, 2008.

[23] J.L. Semmlow, "Biomedical and Medical Image Processing", 2nd ed., CRC Press, pp. 361-398, 2008.

[24] T. Hastie, R. Tibshirani and J. Friedman, "The Elements of Statistical Learning", 2nd ed., Springer, pp. 459-484, 2009.

[25] T. Chen, E. B. Mazomenos, K. Maharatna, S. Dasmahapatra and M. Niranjan, "On the trade-off of accuracy and computational complexity for classifying normal and abnormal ECG in remote CVD monitoring systems," in *Proc. IEEE Workshop Signal Processing Systems (SIPS)*, Quebec, pp. 37–42, Oct. 2012.

[26] C.-H. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. on Neural Networks*, vol. 13, no. 2, pp. 415-425, Mar. 2002.

[27] LIBSVM - A Library for Support Vector Machines. 2013. [Online] Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[28] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data", *in Proc. 2nd Int. Conf. Pervasive Comput.*, 2004, pp. 1-17.

[29] K. Maharatna, E. B. Mazomenos, J. Morgan, and S. Bonfiglio, "Towards the development of next-generation remote healthcare system: Some practical considerations," *in Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Seoul, pp. 1–4, May. 2012.

[30] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "Shimmer – A wireless sensor platform for noninvasive biomedical research," *IEEE Sensors Journal*, vol. 10, no. 9, pp.1527–1534, Sep. 2010.

[31] C. Kendell and E.D. Lemaire, "Effect of mobility devices on orientation sensors that contain magnetometers," *J. Rehab. Res. Dev.*, vol. 46, no. 7, pp. 957-962, 2009.

[32] S. L. Wolf, H. Newton, D. Maddy, S. Blanton, Q. Zhang, C. J. Winstein, D. M. Morris, and K. Light, "The Excite trial: relationship of intensity of constraint induced movement therapy to improvement in the wolf motor function test," *Restorative Neurology and Neuroscience*, vol. 25, no. 5, pp. 549–562, Mar. 2007.

[33] S. L. Wolf, J. P. McJunkin, M. L. Swanson, and P. S. Weiss, "Pilot normative database for the wolf motor function test," *Archives of Physical Medicine and Rehabilitation*, vol. 87, no. 3, pp. 443–445, Mar. 2006.

[34] D. M. Morris, G. Uswatte, J. E. Crago, E. W. Cook III, and E. Taub, "The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 6, pp. 750–755, Jun. 2001.

[35] I. Kim, S. Im, E. Hong, S.C. Ahn, and H.-G. Kim, "ADL classification using triaxial accelerometers and RFID," *in Proc. First Int. Conf. Ubiquitous Computing Convergence Technology*, 2007.

[36] 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, Mar. 2010.

[37] S.A. Lowe, G. Olaighin, "Monitoirng human health behaviour in one's living environment: a technological review," *Medical Engineering & Physics*, vol. 36, no. 2, pp. 147-168, Feb. 2014.

[38] B. Rohrer, S. Fasoli, et al, "Movement smoothness changes during stroke recovery," *The Journal of Neuroscience*, vol. 22, pp. 8297–8304, Sep. 2002.

[39] J. Mao, A.K. Jain, "A self-organizing network for hyperellipsoidal clustering (HEC)", *IEEE Trans. Neural Networks*, vol. 7, no.1, pp. 16-29, Jan. 1996.

[40] J. C. F. Caballero, F. J. Martinez, C. Hervas and P. A. Gutierrez, "Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks," *IEEE Trans. On Neural Networks*, vol. 21, no. 5, pp. 750-770, May 2010.