International Journal of Neural Systems, Vol. 25, No. 4 (2015) 1450036 (20 pages)

DOI: 10.1142/S0129065714500361

Improving Human Activity Recognition and its Application in Early Stroke Diagnosis

José R. Villar*
Computer Science Department, University of Oviedo, ETSIMO
Oviedo, Asturias 33005, Spain
villarjose@uniovi.es

Silvia González[†] and Javier Sedano[‡]
Instituto Tecnológico de Castilla y León
c/López Bravo 70 Burgos, Burgos 09001, Spain
[†]*silvia.gonzalez@itcl.com*
[‡]*javier.sedano@itcl.com*

Camelia Chira
Computer Science Department, Tech. University of Cluj-Napoca
28 Gh. Baritiu Street, 400027 Cluj-Napoca, Romania
camelia.chira@cs.utcluj.ro

Jose M. Trejo-Gabriel-Galan
Neurology Department of the Burgos' Hospital, Burgos, Spain
jtrejogyg@gmail.com

Accepted 4 November 2014
Published Online 16 February 2015

The development of efficient stroke-detection methods is of significant importance in today's society due to the effects and impact of stroke on health and economy worldwide. This study focuses on Human Activity Recognition (HAR), which is a key component in developing an early stroke-diagnosis tool. An overview of the proposed global approach able to discriminate normal resting from stroke-related paralysis is detailed. The main contributions include an extension of the Genetic Fuzzy Finite State Machine (GFFSM) method and a new hybrid feature selection (FS) algorithm involving Principal Component Analysis (PCA) and a voting scheme putting the cross-validation results together. Experimental results show that the proposed approach is a well-performing HAR tool that can be successfully embedded in devices.

Keywords: Feature selection; genetic fuzzy finite state machine; genetic fuzzy systems; PCA; stroke.

1. Introduction

Stroke represents a huge health and economic problem: it ranks second in fatal diseases worldwide[1] and first in disability-causing diseases, entailing a total cost of $240 billion in 2010 in the USA, which is expected to double between 2012 and 2030 because of the aging of the population.[2] Stroke consists of a sudden loss of some brain functions, usually including right or left limb paralysis (also called hemiplegia) due to either a cerebral haemorrhage or a cerebral infarction, the latter being much more common (85% of strokes).

Cerebral infarction is caused by a thrombus that obstructs the blood flow in a cerebral artery and

*Corresponding author.

deprives brain cells of oxygen and glucose, which leads to cell death, and the subject will most likely suffer permanent sequelae unless the thrombus is timely dissolved by infusion of a specific drug.[3] If successful, cerebral tissue will recover and so will function, but that depends on how soon the treatment is given: In the first one and a half "golden hours", one out of three patients treated will completely recover. An estimate of 1.9 million neurons are lost[4] every minute the treatment is delayed, which means that four and a half hours after stroke onset there is no benefit left from receiving the treatment and it can even be harmful.[5]

Given these facts, the question is why only about 5% of people suffering a stroke receive this treatment. The answer is quite simple: They arrive too late at the hospital, even after informative public campaigns and in-hospital patient circuit optimization have been implemented in order to reduce the stroke onset-to-treatment interval as much as possible. All too often, the patient does not recognize the symptoms or, being asleep or alone, the paralysis does not allow him/her to seek help. The usefulness and supportive aid of a potential device able to raise an alarm in such situations is of clear and significant importance.

Hand paralysis is part of 2/3 of strokes at onset,[6] and unlike the leg, partial recovery of which frequently allows walking, hand function is not so often regained.[7] Detecting an abnormal absence of hand movement could identify 530,000 of the 795,000 strokes that happen each year in the USA.[6] In this study, we consider hand paralysis as a surrogate marker of hemiplegic strokes and thus develop a new movement detection device for stroke detection, integrating hand activity recognition and stroke onset discovery models. Unlike a simple Actigraph able to detect movement or its absence, the proposed device can discriminate normal resting from stroke-related paralysis. The new movement-detecting device consists of two modules, each mounted on one wrist of the subject, which can measure quantitative and time-related characteristics alongside the movement of each wrist, as well as the asymmetries between them. Both combined can differentiate normal hand rest from paralysis due to stroke. These two modular components should be easy to wear and cheap to produce and they have therefore been designed as small devices to be mounted in light bracelets.

The main normal hand movements and resting patterns for the different daily activities need to be included and labeled as normal in the proposed device. If unilateral absence of hand movement, significantly different from the normal pattern, is detected, an alert of a possible stroke is sent and immediate transportation to the nearest hospital for urgent treatment is implemented.

The underlying model for the development of the proposed stroke-detection device is depicted in Fig. 1.[8] The process is divided into two main steps as follows: a Human Activity Recognition (HAR) stage and a Stroke Onset Detection stage. The former is responsible for determining the current Human Activity (cHA) and detailing this topic represents the main aim of this study. The latter is devoted to evaluating the abnormality of the current movements once the cHA is known, and it is not analyzed in this work. This study addresses the HAR stage needed and a valid solution is described which is technically sound and can be embedded in wearable devices in order to be deployed.

Analyzing the current literature concerning HAR leads to the conclusion that the HAR methods are quite specific and that the required computational resources make the majority of proposals unfeasible. Furthermore, the high number of transformations from the raw acceleration values introduces a new complexity to the problem since the best subset should be found.

In this research, the well-known Genetic Fuzzy Finite State Machine (GFFSM) method[9,10] is engaged in the HAR stage due to its easy-to-embed

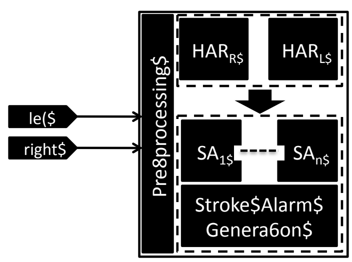

Fig. 1. The solution's block scheme. Two tri-axial accelerometers are place, one on each wrist. The data gathered from the sensors is used for (i) detecting the cHA, (ii) perform a specific pattern analysis for the cHA, and (iii) generation and submitting the corresponding alarms. Both (ii) and (iii) are part of the Stroke Onset Discovery step.

quality. The GFFSM method is extended to further learn the whole rule set in addition to learning the fuzzy partitions and rule antecedents.[11–14] Furthermore, a complete method for choosing the best feature subset based on a voting scheme together with Principal Component Analysis (PCA)[15–17] is presented. The current experimentation can be considered preliminary because no exhaustive test with high-risk population has been done yet, but the remarkable results encourage us to continue with this line of research.

The organization of this study is as follows. Section 2 deals with the basic knowledge for HAR and in Sec. 3, the complete method for HAR is fully detailed. The experimentation and the discussion of the results are included in Sec. 4. A detailed discussion on the results is included in Sec. 5. Finally, the main conclusions of this research are drawn.

2. HAR Review

Nowadays, the main research in HAR is concerned with the use of tri-axial accelerometers. After several years of study, a wide range of features calculated as transformations from the raw acceleration data have been proposed. Each HAR method chooses its own set of features according to different criteria.

Furthermore, each HAR method has its own characteristics, like the number of sensors, the transformations that are used and the recognition techniques. Consequently, one of the very first steps is to choose the best feature space among the available transformations and consider the technique to be used. Thus, this section starts by describing the main features in the literature and, afterwards, a review of the literature concerned with HAR methods is presented.

2.1. *From the acceleration data to the input feature space*

The measurements from tri-axial accelerometers, known as raw data (RD, a_i^x, a_i^y, and a_i^z; or $a_{i,j\in\{x,y,z\}}$ for the sake of brevity), can be decomposed into gravity acceleration (G) — due to each gravity, g_i^x, g_i^y, g_i^z; $g_{i,j\in\{x,y,z\}}$ — and body acceleration (BA) — which is due to the human movement, b_i^x, b_i^y, b_i^z; $b_{i,j\in\{x,y,z\}}$.

The capacity of the BA for discriminating among different human gestures is documented.[15] Nevertheless, the literature includes the use of a wide variety of transformations. Besides, all these features can potentially be computed on each of the possible signals (RD, BA, G and, in some cases, on their components), leading to a feature space with more than 190 features, which is a very challenging task indeed.

In many of the solutions, sliding windows with or without shifting are proposed; the typical window size converges to the samples within a period of 2 s. Features are typically normalized to 0-mean 1-standard deviation and/or scaled to the interval $[0, 1]$ before further pre-processing.

In order to include as much information as possible for the interested reader, an exhaustive list of the most common transformations and their mathematical expressions are included in Table 1, where w stands for the sliding window size — when needed, and subscripts $i \in \{1, \ldots, N\}$ and $j \in \{x, y, z\}$ stand for the number of the sample and the axis, respectively. Each feature will hereinafter be referred to with its name or with its corresponding transformation reference Tx.

Using frequency-derived features employing FFT or similar over long time-windows has been found to be more suitable for long duration, quasi-periodic signals like walking, cycling or brushing teeth. Otherwise, when classifying shorter duration and non-periodic activities, transitions or a short sequence of steps, the time-domain representation has been found to be better.[19]

It is interesting to mention the wide spread of transformations that have been used for specific HAR methods. Each specific case includes a number of sensors applied on different places on the body. However, there is no reason to disregard any of them because the concept they have introduced may be useful for the current HAR task. Therefore, a feature selection (FS) step must be used so that the most interesting features for the specific problem can be chosen.

2.2. *A brief review of the HAR literature*

The characterization of human movement, especially in walking, is well documented in the literature.[20] Nevertheless, stroke has a strong influence on the way patients move,[21] and particularly the patient's gait is affected. Due to this, and also because walking is one of the most sensitive parameters in human

Table 1. List of most common transformations of the acceleration data.

Ref.	Transformation	Calculation	Ref.	Transformation	Calculation
T1	Mean, standard deviation, and higher momentum statistics values for the RD[34] or for the BA[18,35]	well-known statistics	T2	Intensity of the movement[39]	$\mathrm{InMo}_t^{j\in\{x,y,z\}} = \frac{1}{w}\sum_{i=0}^{w-1} \lvert a_{t-i}^v - a_{t-i-1}^v\rvert/\Delta x_t$
T3	Root Mean Square[36]	$\mathrm{RMS}_j = \sqrt{\frac{1}{w}\sum_{i=1}^{w} \lvert a_{i,j}^2\rvert}$	T4	Signal magnitude area[19]	$\mathrm{SMA}_t = \frac{1}{w}\sum_{i=1}^{w}(\lvert b_i^x\rvert + \lvert b_i^y\rvert + \lvert b_i^z\rvert)$
T5	Sum of the absolute values[37] for the BA	$\mathrm{sBA}_i = \frac{1}{w}\sum_{t=i}^{i+w}\sum_{j\in\{x,y,z\}} \lvert b_{t,j}\rvert$	T6	Amount of movement[10]	$\mathrm{AM}_i = \sum_{v=\{x,y,z\}} \lvert \max_{t=i+1}^{i+w}(b_t^v) - \min_{t=i+1}^{i+w}(b_t^v)\rvert$
T7	Correlation between axes[38] for each signal RW, G or BA	$\mathrm{corr}(x,y) = \frac{\mathrm{cov}(x,y)}{(\mathrm{std}(x)*\mathrm{std}(y))}$	T8	Shifted Delta Coefficients for the BA[19]	$\Delta b_{t+i*P}^{x,y,z} = \sum_{d=-D}^{D} d\cdot b_{t+i*P+d}^{\{x,y,z\}}/\sum_{d=-D}^{D} d^2$
T9	Mean absolute deviation[34,36]	$\mathrm{MAD}_j = \frac{1}{w}\sum_{i=1}^{w} \lvert a_{i,j} - m_j\rvert$	T10	Vibration of the sensor[35]	$\Delta_i = \frac{1}{w}\sum_{t=i}^{i+w}\sum_{j\in\{x,y,z\}} a_{t,j}^2 + g_{t,j}^2$
T11	Average energy[18,35,38]	$\mathrm{Energy} = \frac{\sum_{i=1}^{\lvert w\rvert} \lvert F_i\rvert^2}{\lvert w\rvert}$	T12	Time between peaks[34]	Different algorithms in the literature, Ref. 34 is among them
T13	Tilt of the body[10]	$\mathrm{sBA}_i = \frac{1}{w}\sum_{t=i}^{i+w} \lvert a_i^y\rvert + \lvert a_i^z\rvert$	T14	Delta coefficients for the G[19]	$\Delta g_t^{\{x,y,z\}} = \sum_{d=-D}^{D} d\cdot g_{t+d}^{\{x,y,z\}}/\sum_{d=-D}^{D} d^2$
T15	Binned distribution, relative binned distribution, and absolute binned distribution[18,34]	Statistical count of TS values within each of the window range subintervals.			

Note: TS stands for time series.

dependence, the gait and the patient's kinematics have been studied.[21,22]

Most of the studies are based on the analysis of video-images of the patient's gait or movements through well-known mechanical methods,[22–26] though there are also studies using the distributed home automation sensors.[27]

These studies are mainly focused on the rehabilitation of the patient and for developing new therapeutic training techniques, providing interesting conclusions and determining the relevant variables for characterizing the gait pattern.

Since the appearance of low cost and high performance accelerometers in the market, the recognition of human activity has gained focus. Plenty of studies have analyzed the performance of this type of device for stroke rehabilitation evaluation[28] and activity level measurement.[29] Some of these studies reported the use of different sensors and techniques, like accelerometers and electromyography[30] or accelerometers and electrocardiograph sensors.[31]

One of those studies presented a combination of accelerometers and pressure gauges within shoes for discriminating between sitting, standing, and walking of stroke patients.[32] In this study, Support Vector Machines were proposed for classification of the activity, and results of 99% of recall and 76.9% of precision were achieved. Though the classification is relatively good, it is worth noting that this approach can only discriminate between normal and abnormal walking, as the rest of the activities are mainly perceived in the upper limbs.

There are also several studies concerning the use of accelerometers as the only source of information for HAR. One of the very first studies in HAR using accelerometers is available in Ref. 33, in which several feature extraction methods were applied before modeling the classifiers of the different activities. Three different classifiers were learned: For the raw accelerometer data from the two accelerometers — one on each hip — for the PCA feature subset and for the Independent Component Analysis.

In each case, the most relevant feature subset was normalized to 0 mean and 1 standard deviation; a sliding window of 256 samples with a 64-sample shift was used. Afterwards, wavelet transformations were carried out over the windowed data. A multi-layered perceptron with back propagation was learned for the classification task of four classes: *Stop*, *Walking*, *Walking Upstairs*, *Walking Downstairs*. The classification error was used for evaluating each model, and the back propagation weight updating was based on the momentum.

A well-known contribution in HAR was proposed in Ref. 40. In this study, the authors used the "divide and conquer" strategy, detecting first static postures from dynamic activities. Specific decision trees were generated for each case, either static or dynamic. The framework was structured around a binary decision tree in which movements were divided into classes and subclasses at different hierarchical levels: General distinctions between movements were applied in the top levels, and successively more detailed subclassifications were made in the lower levels of the tree.

This framework was used to develop a classifier to identify basic movements from the signals obtained from a single, waist-mounted tri-axial accelerometer. Still, the main drawback of these methods is the complexity of the calculations, which cannot be carried out in embedded devices.

An extension of the former study was presented in Ref. 41, where a sensor in the hip was used for HAR. The movements were first divided into activity (dynamic activities) and rest (static activities) using the T4 transformation. According to the postural orientation, the current activity and orientation were then decided. Despite the results obtained from this approach and the low complexity, the main drawback of this study is that, apart from falling, no other abnormal behavior can be detected using the sensor in the hip.

In Ref. 21, a rule and heuristic based decision system is proposed for discrimination between the states of lying, standing, and walking, as well as with the transitions between states. The transitions rules are learned through a Gaussian mixture model. This work is quite interesting in the sense that the system can keep track of the current state; that is, the cHA. Nevertheless, as long as the system uses only one sensor in the hip, this is not valid for the problem faced in this study. Also, the tilt of the body and its orientation are designed for user-centered sensors, which is not the scenario for the current study. Furthermore, the complexity of the solution makes it unfeasible to transfer this to embedded systems.

Hidden Markov Models have also been proposed for automatic segmentation and classification of

HAR.[42] The underlying idea is to determine the conditions that, given the number of states and the segmented data sets, fire the set of finite state machines, making change from one state (activity) to another.

To show the benefits of this approach, the results were based on up to five sensors, four tri-axial accelerometers and one gyroscope, which were distributed among the chest, upper arms, ankle, and thigh, respectively. However, this relatively high number of sensors is the main drawback of this approach.

The GFFSM method was proposed for HAR using a sensor placed in the central part of the body.[9,10] Briefly, the GFFSM is a fuzzy finite state machine that drives the classification of cHA. The fuzzy partitions and the memberships involved in each of the transition rules are learnt in a Pittsburg style for the current subject; this learning gives the method the generalization capability. This study was found to be extremely suitable and will be further explained in the next subsection. One of the main advantages of this solution is that it can easily be transferred to embedded devices; however, the drawbacks are that the solution must be modified to manage the sensors on the wrists and that a previous training stage is needed to adapt it to the current subject.

A very interesting idea is shown in Ref. 43, where a sparse representation of the input feature domain is proposed. Briefly, this representation makes use of TS windows as a set of relevant motifs for each activity; each motif is the input features with the information from the TS window.

Whenever a new TS window is to be classified, the most suitable set of motifs is determined and thus the corresponding activity is proposed. Although this approach is conceptually quite novel and interesting, its main drawback is the computational cost, which makes it unfeasible if it is to be introduced in embedded devices.

The use of tri-axial accelerometers on the wrist has been previously documented.[15,36] In Ref. 36, up to 24 features were analyzed using dynamic linear discriminant analysis; the best ranked features in each step drive the function basis classifier update in an iterative process that ends when a suitable error value is obtained.

Interestingly, this approach allows the evolution of the activity set to be detected. The study presented in Ref. 15 details the general procedure for obtaining the set of features and, in this case, the neural classifier. Furthermore, this study proposed the Common PCA for choosing the best feature subset, although they proposed feature extraction instead of FS.

Currently, there is a trend to introduce HAR as an add-on to smartphones as long as the deployment cost is significantly reduced.[34,39,44] In Ref. 39, three sensors were placed in the dominant wrist, hip and ankle; up to six different activities were studied: resting, typing, gesticulating, walking, running, and cycling.

To classify the activity two well-known methods were analyzed: C4.5 and feed-forward neural network, although only the latter was found to be useful. Similarly, but with a different feature subset,[34] the time between peaks and a discretization of the sliding window to feed the J48 decision tree was proposed. Finally, the benefits of using the data from sensors included in smartphones were discussed in Ref. 44.

3. A Novel Approach to HAR in Embedded Devices

As stated before, hand paralysis is part of 2/3 of strokes at onset[6]; therefore, for this study the sensors have been placed on the wrists. The idea is to learn the way the subject moves to determine the cHA.

This proposal is focused on generating a valid HAR solution that can be embedded in cheap electronic devices, which is a preliminary step before developing the whole solution described in the introduction section. Besides, it is known that the older the subject the smaller the arm characteristic movements, thus, further study will be needed in order to validate this approach with a group of subjects from the focus population.

To this end, the device placed on each wrist should be responsible for HAR. Clearly, the HAR method should be as light as possible if we consider current embedded devices' computational resources for reduced power consumption.

Additionally, taking into account the focus population, the activities to be identified are bound to a greatly reduced set. In this study, we have considered only three activities — Walking, Resting, and Standing, which represent the main set of activities carried out by the subjects with high stroke risk.

Considering these restrictions and constraints, we have chosen what seems to be the most promising technique. In this sense, one of the valid techniques, if we consider the computational limitations, is the GFFSM. This technique might not be valid for a medium-high size set of activities, but it has been found valid for the considered problem.[8] This approach has made use of only one sensor placed on the back of the subject, so adaptation of the method is required: This is the most interesting feature set for using the GFFSM when the sensors are placed on the wrists. Furthermore, the GFFSM requires a learning stage for adapting the model to the current subject.

In preliminary studies, the GFFSM technique was adapted to the problem.[37] Furthermore, a FS based on filtering according to the Information Correlation Coefficient measurement and then a wrapper method has allowed us to optimize the adapted GFFSM.[45] However, in this latter study it was concluded that further analysis was needed with methods like the PCA. Besides, it was also found that the learning method of the GFFSM would be enhanced with a Michigan approach.[46] Consequently, the current study is concerned with two main issues:

- A two-stage FS devoted to (a) evaluating and filtering the set of features using a novel voting approach based on PCA and (b) using the wrapper FS presented in Refs. 45 and 47 with the best ranked features from step (a).
- Developing a Michigan approach for the GFFSM that makes use of a boosting meta-heuristic in order to obtain the most suitable set of rules to drive the HAR.

The remaining contents of the section are organized as follows. In Sec. 3.1, the FS method is explained and detailed, including the voting scheme proposed based on the PCA transformation. Section 3.2 is devoted to the description of the original GFFSM method, its adaptation and enhancements, together with the novel Michigan approach using boosting. It is interesting to highlight that the whole approach is a complex solution based on the memetic computing principles of design.[48]

3.1. *A two-step FS method*

The FS includes two stages: First filtering using a voting scheme based on PCA ranking and a second stage using a GFFSM wrapper FS with the most interesting feature subset found in the filtering FS.

The reason for this hybridization comes from the experience in the preliminary studies with the different runs of the experiments for the same subject. It was found that running the same experiment for the same subject and performing the same filtering strategy — including PCA — could lead to different feature subsets. And if all the data sets for the same subject are joined and PCA is performed, another different feature subset is also obtained.

The question that arises is which of the feature subsets is the best one. The answer to this question can be found by introducing a second wrapper type FS stage that considers the conjunction of all the feature subsets found.

Introducing this second stage is a good solution found in the literature so far. But the problem is that the conjunction of the feature subsets increases the dimensionality of the feature space for the second FS stage, which eventually decreases the performance of the FS and even introduces high computational costs.

Consequently, there is a need for a method that joins the results from all the PCA analysis over each of the folds and over the sequencing of all the data sets. The filtering stage that is proposed in the next subsection tackles this problem, while the wrapper FS stage is detailed in Sec. 3.1.2.

3.1.1. *The filtering FS stage*

The underlying idea is to take advantage of carrying out several runs of the same rehabilitation test for a subject[25]; each test generates a TS of accelerations and its transformations that are segmented as stated in Ref. 10. This study proposes a feature filtering method as follows (see Fig. 2).

Two main ballots are managed: The first one considers each of the TS independently, performing a poll to obtain a feature subset and then merging the results from all the TS; the second one considers all the TS in a single TS. Finally, the results from both ballots are merged and the output feature subset is obtained.

The same *voting scheme* is performed for all the polls. This voting scheme will be used for transferring the PCA rankings to the original feature space. The voting scheme is defined by the following

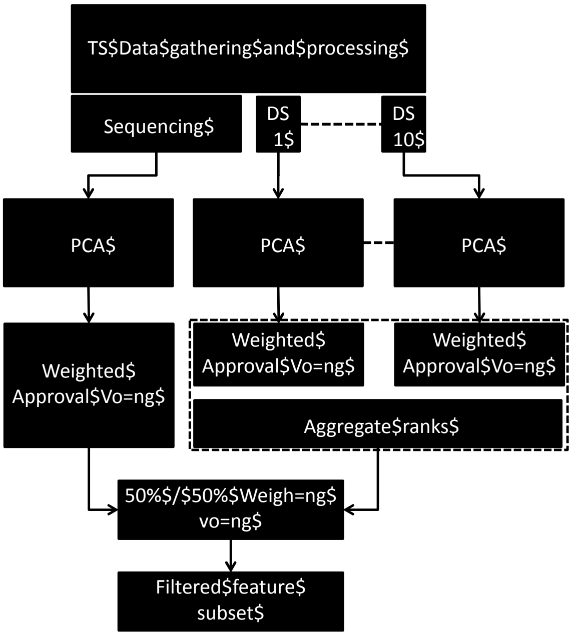

Fig. 2. Overview of the voting scheme. Two ballots are carried out, one for the set of TS and a second one for the sequence of TS. The weighted approval voting method is used for the aggregation of the translated PCA rankings. Both ballots have the same weight in the final results. DS stands for data set.

rules, introducing a weighted *Approval Voting* scheme[49]:

- The TS is normalized using the mean and standard deviation of the sample.
- PCA chooses the transformations to reach 95% of representation. Let $\boldsymbol{M}$ be the number of transformations at that representation level.
- The transformations are ranked in ascending percentage of representation, that is, the best ranked transformation is assigned with the rank $\boldsymbol{M}/\boldsymbol{M}$, the second best with the $(\boldsymbol{M}-\mathbf{1})/\boldsymbol{M}$, etc. All the values are in the interval $(0, 1]$.
- Each transformation votes for as many features as are involved in its computation.
- Voting means assigning the rank to each of the candidate features.
- Features from the original space accumulate the ranked votes.
- The higher the sum of ranked votes, the higher the relevance of the feature.

As mentioned before, a ballot is carried out to poll for a feature subset from each TS. For each poll, the PCA ranking is transferred using the voting scheme described above. In order to accumulate the results from each poll, the mean rank of each feature is calculated; the voted features are sorted according to this mean value. This rank is the outcome of the first ballot.

The second ballot considers the samples from all the TS, joining them in one data set. After applying PCA to this data set, the ranking is transferred to the original feature space using the same voting scheme. The outcome of this second ballot is the rank of the features in the original space.

Finally, we merge the results from each ballot using a 50/50% weighted voting. Sorting the features in the original space according to the descending order of votes allows the input feature domain to be filtered.

The selection of the weight for the final rank (50/50%) has been chosen as a compromise since there is no certain knowledge of which of the approaches may be better. Perhaps further analysis is required, but as far as the research team envisages, the results will be rather problem-specific and no general proposal of weights will eventually be found.

3.1.2. *The wrapper FS stage*

A second FS stage is needed in order to choose the best three features for the GFFSM. In this study, the well-known Steady State Genetic Algorithm FS method,[50] as stated in Ref. 47, is adapted and applied to the problem as in Ref. 45.

In short, this wrapper FS selects a subset of features of a given certain dimension by means of a genetic algorithm (GA).[51] Each individual represents a subset of features from the input domain and includes a model. This model is learnt during the individual's fitness evaluation, and the error measurement obtained is assigned to optimize the individual fitness. Thus, this FS method not only selects the feature subset but also generates the desired model. Consequently, the computational cost of this method is relatively high.

For this approach, the feature subset is set to a size of 3. To evaluate an individual, a GFFSM model is learnt for the individual's feature subset and the classification error generated from the GFFSM's validation is set as the individual's fitness.

A 5×2 cross-validation scheme is used for training and testing the GFFSMs: From the collections of 10 TS gathered for the subject, a random half are used for training and the remainder is used for validation. A relaxed set of parameters must be used to decrease the computational cost of the GFFSMs training.

The outcome of this stage is the reduced feature subset of the most promising three features for the HAR method. However, as a relaxed set of parameters was used within the wrapper, the obtained GFFSM model can still be improved, so a complete learning stage should come after the FS.

3.2. *The HAR method*

This subsection is devoted to the thorough explanation of the GFFSM, from the original proposal to current developments. Thus, it will deal with the following aspects: (i) the original GFFSM contribution, (ii) the adaptation and enhancements found so far, (iii) a brief on a boosting method for learning fuzzy rule sets, and (iv) the novel boosting meta-heuristic to learn the Michigan approach GFFSM.

3.2.1. *The original GFFSM*

The GFFSM was detailed in Ref. 10: A GA evolves the Fuzzy Finite State Machine for HAR. The GFFSM is defined as the tuple $\{Q, U, f, Y, g\}$, where Q is the state of the system, U is the input vector, f is the transition function which calculates the state of the system, Y is the output vector, and g is the output function which calculates the output vector.

The state of the system is defined as a linguistic variable, with $\{q_1, \ldots, q_n\}$ as labels, and one label per state. For each time step the system has a state activation $S[t] = (s_1[t], \ldots, s_n[t])$, with $s_i[t]$ in $[0, 1]$ and the sum of the activation levels is always 1. The transitions are the fuzzy rules that are allowed, which are defined *a priori*, as depicted in Fig. 3.

A change in the state is considered if any of the rules has a firing strength higher than 0. The output of the system is the set of new state activation levels, which induces the new state.

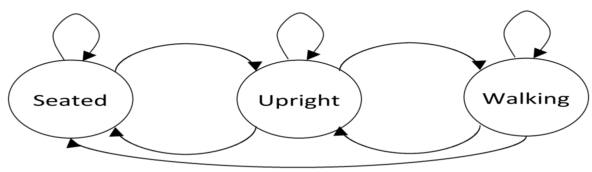

Fig. 3. The original GFFSM *a priori* defined states and transitions.

Three input variables are used: The dorsoventral acceleration a_i^z, the amount of movement (T6) and the tilt of the body (T13), each of them considered linguistic variables with three labels each. Ruspini trapezoid membership functions are used, so we need to learn up to four parameters for each input variable. A GA evolves the partitions and the rules in a Michigan style: Up to 72 binary genes coding the rules and 12 real-coded genes for the membership function parameters.

The fitness function is the Mean Absolute Error (MAE), which is calculated using Eq. (1). In this equation, T is the number of examples in the data set, and $s_i[t]$ and $s_i^*[t]$ are the degree of activation and the expected degree of activation, respectively, of state q_i at time step $t = j$.

$$\mathrm{MAE} = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=0}^{T}|s_i[t] - s_i^*[t]|. \tag{1}$$

The GA is completely defined with a binary tournament, generational replacement with elitism, two-point crossover for the rule base and BLX-α crossover for the real-coded genes applied twice for obtaining two new pairs of chromosomes and a classical bitwise mutation for the rule base and uniform mutation for the real-coded part. Termination occurs whenever one of the following conditions applies: MAE reaches 0 value, or the expected number of generations is reached or there are a number of generations without MAE changes.

3.2.2. *Previous works for adapting the GFFSM*

As in this study the sensors are placed on the wrists, this approach cannot be directly applied: When the tilt of the body, though easy to compute, loses its meaning, the dorsoventral acceleration is no longer available.

In Ref. 37, a different set of variables was chosen as the most conceptually similar to the original ones: The signal-magnitude area (T4) and the sensor vibration (T10) have been used instead, while the amount of movement (T6) was retained. Furthermore, the genetic parameters were slightly modified to 100 individuals and 200 generations. This model will henceforward be denoted by aGFFSM.[37]

Moreover, a higher resource consuming method was proposed and evaluated (hereinafter, gGFFSM) with 300 generations, 100 individuals and considering the crossover operator as being able to combine genes from more than one rule antecedent.

The Information Correlation Coefficient measurement has been used for filtering the feature domain together with the wrapper FS stage detailed in the previous subsection[38]; the 20 most suitable features were chosen from the filtering stage; then the wrapper FS was carried out. The most representative features found were the a^x, g^y and T15(a^y, 10) for the right hand and T13(a^x, a^z), T6 and T15(a^z, 10) for the left hand. The GFFSM model trained using this feature subset is denoted by wGFFSM, using the same parameter setting as gGFFSM.

Actually, gGFFSM and wGFFSM triggered similar results while both were found to outperform the adapted aGFFSM when using 10-fold cross-validation, and the two approaches were comparable (see Fig. 4). The wGFFSM performed with less deviation and a slightly higher median MAE value of the best individual found in each run during the cross-validation process.

However, some repetitive errors have provided evidence that perhaps the current rule set could be optimized, so a change in the learning paradigm was required. The next two subsections deal with the new GFFSM proposal.

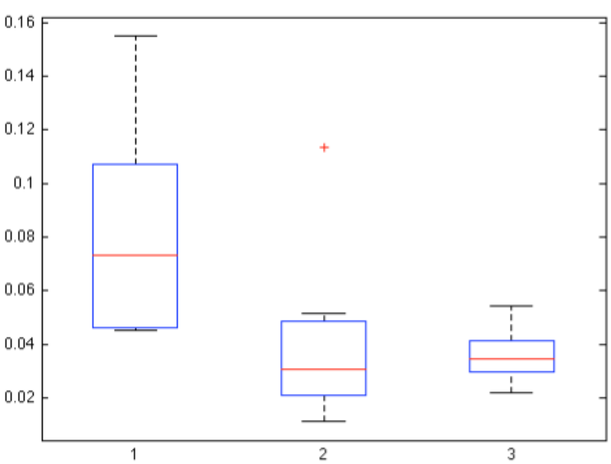

Fig. 4. Comparison of MAE classification errors of the aGFFSM (1), gGFFSM (2), and wGFFSM (3): The two latter are comparable and clearly outperform the former.

3.2.3. *Learning fuzzy rules through boosting*

Fuzzy classifiers can be learnt using boosting, that is, by learning an incremental rule set that best suits the training data. In this study, we made use of the proposal for learning fuzzy classifiers through boosting by means of the single-winner inference.[52] Thus, the current subsection only gives an outline of the algorithm; readers interested in further details should refer to the original contribution.

Let p be the number of classes in the data set of size m samples, n the number of features, and N the number of rules. A fuzzy rule set is represented as the relationship $Ax\{1,\ldots,p\}$, with $A = \{Aj\}_{j=1,\ldots,N}$ being the antecedent of a fuzzy rule. Typically, the antecedents are expressed as Cartesian products of fuzzy sets.

A fuzzy classifier is that model that infers a class k to an example x by means of $\mathbf{arg\,max}_k \vee_j A^j(x) \wedge s^j_k$, with s^j_k being the membership degree of rule j to class k.

The single-winner boosting fuzzy rule learning algorithm provides a rule for each iteration based on the number of explained samples. The algorithm first finds a feasible rule antecedent A_j and then estimates the rule-class memberships.

The full algorithm is outlined in Figs. 5 and 6. The former includes a GA algorithm for finding valid fuzzy rule antecedents, while the latter deals with the evaluation of each individual within the population and the single winner inference.

3.2.4. *A boosting extension to the GFFSM*

It is proposed to use the Michigan style to tackle the learning of the fuzzy rule base of the GFFSM, instead

Procedure: Boosting Fuzzy Rules
Input:
  a data set of size m; the number of rules to learn N
Output:
  a rule base of size N
R=[]
for each rule r=1,…,N
  run a GA
    call *AddOneRule* for each individual
  add the rule of minimum fitness value to R
for j=1,…,N
  make all s_k^j=0 but the maximum one $s_{q(j)}^j$
return R

Fig. 5. The Boosting Fuzzy Rules learning algorithm with the single winner inference.

Procedure: AddOneRule
Input:
  a rule base of size N; a fuzzy rule antecedent A^{N+1}
Output:
  a rule base of size N+1 and a numerical value of fitness
$S_{kj}=s_k^j$, $\forall k=1,\ldots,p$ $j=1,\ldots,N$
Initialize S_{kN+1} randomly
do
  $I'=I$
  for $j=1,\ldots,N+1$, $i=1,\ldots,m$
    $I_{ij}=1$ if rule j wins in example x_i, 0 otherwise
  $F_{ji}=A^i(x_i)\cdot I_{ij}$ $\forall j=1,\ldots,N+1$ $i=1,\ldots,m$
  $Z_{ki}=4(y_k(x_i)-0.5)$ $\forall k=1,\ldots,p$ $i=1,\ldots,m$
  $S'=S$
  $S=Z\cdot F^t\cdot(F\cdot F^t)^{-1}$
  $S=\alpha\cdot S+(1-\alpha)\cdot S'$
while $\|S-S'\|<\varepsilon$
return S and fitness$=\|Z-S\cdot F\|$

Fig. 6. The *AddOneRule* algorithm for learning one new rule, including the single winner inference.

of pre-defining it. Hereinafter, this model is referred to as boosting GFFSM (bGFFSM). The Michigan learning style represents each rule as an individual of the population, and the final fuzzy rule set is obtained as the set of the best individuals found up to that point in the evolution.

However, the method for learning the fuzzy rules plays an important role in the final model. In this study, we have chosen to learn fuzzy rules using boosting and single-winner inference, as proposed in Ref. 51, although some adaptations were needed. For the sake of simplicity, this section only highlights the main idea of the chosen algorithm, referring interested readers to the original contribution for further details. The required adaptations will be detailed afterwards.

As explained in Ref. 52, each rule added to the rule base increases the number of explained examples in the data set (see the algorithm in Fig. 5). The point is that while Adaboost makes use of the max operator as t-conorm, boosting makes use of the sum operator: The candidate to be added to the rule set is the one that maximizes the updated number of explained samples. Thus, instead of having the capacity of a fuzzy state variable as in the original GFFSM contribution, using boosting ends up choosing one single label as a consequence, with its corresponding weight.

Moreover, the authors proposed an algorithm *AddOneRule* that makes use of a GA to find a rule that maximizes the number of explained examples (see algorithm depicted in Fig. 6). Consequently, the computational cost of this learning scheme is higher than that of the Pittsburg approach.

Some adaptations to the method are needed in order to learn GFFSM. First, the rules should reflect the Fuzzy Finite State Machine with a single state as a consequence. This means that, although some other linguistic labels can be learned for each rule candidate, only the label that maximizes the number of explained samples receives the award: No feedback concerning the rest of the state labels is generated. It is therefore better to include more than one single label as the possible outcome of a rule.

The linguistic labels are provided with three uniformly distributed trapezoidal membership functions; thus four real parameters should be learnt for each fuzzy variable and 12 real parameters should be learnt for the partitions. To do so, it is proposed to learn the partitions with the original GFFSM Pittsburg approach; once the partition is learnt, then the boosting learning takes place.

The antecedents of all the rules in the classifier have been defined according to the boosting method (see algorithms in Figs. 5 and 6)[52]: $A = \{A_f^j\}$, with $f = 1, \ldots, n$ referring to variables, $j = 1, \ldots, M$ referring to rules, n is the number of input variables (in this case, $n = 3$) and M is the number of rules to learn, which is given as a parameter. As with the original GFFSM contribution, the rules include any label combinations of the three antecedent linguistic variables (Low, Medium, and High), so A_f^j can be any combination of these labels. The rule part of the antecedent is composed of one initial state label plus 9 linguistic terms (3 per input variable) for a total of 10 binary-coded genes.

T is the number of examples in the data set, and S_k^j is the matrix of sample explanations used to determine which of the samples are explained or not and which is the rule that best explains each sample. Each candidate's rule is given a rank according to the number of samples it explains: The state that best ranks a candidate rule is the one chosen as its consequence.

For computing the antecedent of the firing strength of a rule, the min and max operators are used as the t-norm and the t-conorm, respectively.

Finally, only rules that introduce a change between the initial and final states are learnt;

otherwise, the boosting method can get stuck. To do so, the candidates with different initial and consequent states are awarded.

4. Experiments and Discussion of the Results

The different tests carried out aim to validate each of the hypotheses of this study, regarding (i) the performance of the boosting-based GFFSM, (ii) the FS step outcome endorsement, and (iii) the final experimentation with the best results found so far. This section deals with each of these steps as follows: The next subsection gives details concerning data gathering; Sec. 4.2 focuses on the pre-processing and the cross-validation, while the last subsections are devoted to testing and discussing the above-mentioned points.

It is important to point out that the experimentation to be included is not a medical study of patients from the focus population, but proof-of-concept experiments designed from the computer science perspective to obtain a valid tool that can henceforward be tested in a medical study: We do expect to carry out a medical study once the whole approach is finished. Consequently, readers should not expect to find here a detailed protocol or an exhaustive description of the subjects under study.

4.1. *Materials and methods*

For this study, a pair of data-logging bracelets with a sampling frequency 16 Hz, to be worn on the wrists, has been developed (see Fig. 7). Each bracelet, which is identified for the left or right hand, includes a triaxial accelerometer with a predefined range up to 3 G and a USB port. These bracelets are able to store data from several experiments lasting up to 40 min. The batteries last about 10 days, so there should not be any problems with their autonomy.

In order to systematically evaluate the proposal, a test should be defined. To do so, several test beds were analyzed from the literature of rehabilitation studies.[53] The different studies establish patterns to be carried out by the subject, i.e. making the subject walk 10 m discarding 2 m at each end.[54] The question is which of the tests would allow us to validate the HAR proposal and the answer needs to consider the current problem with the set of activities to detect it.

Fig. 7. The developed bracelets, each one with the mark for the left (Izda.) and right (Dcha.) hands.

After analysis of the different approaches, the stroke rehabilitation test[25] was finally selected as the test to be carried out by the subjects, though small variations were introduced. In the original test, the subject starts from sitting for T1 seconds, then stands up and stays still for T2 seconds. Afterwards, the subject has to walk a distance of 3 m, turning 45° and walking a further 3 m; then the subject retraces the path to the original position, and there stands still for T2 seconds. Finally, the subject sits and stays in that position for another T1 seconds.

Some small modifications were introduced. First, the 45° turn, which was introduced to test the stroke patients' ability to change direction, was not included. The walking distance was extended to 10 m in order to include enough walking cycles. Finally, the periods T1 and T2 were set at 10 and 5 s, respectively. This modified rehabilitation test is hereinafter referred to as SRTEST.

To test the approach proposed in this paper to determine its validity, one male and two female members of the research team were chosen as test subjects. All these subjects were in a good health condition, and a normal fit state. The age of the subjects varied from 28 to 46 years old, which is clearly outside the focus population. However, as long as the current problem is actually to evaluate if the proposal is suitable for HAR purposes, this is considered adequate for a proof-of-concept experiment.

For the purpose of this experiment, each subject carried out 12 runs of the SRTEST and the first two were discarded: They were done simply to get used to the test, and to avoid moving differently from usual when sitting, standing or walking. In addition, these repetitions of the test would allow us to obtain statistical data from the experiments.

4.2. *Data pre-processing and cross-validation*

All the data is manually segmented and classified according to the activity that the subject has to do, in the same way as proposed for the original GFFSM work.[10] For computing the features, a sliding window of 10 samples wide and one sample shift are used. In this way, known states are classified with complete certainty, while transitions between actions are assigned with imprecise data, e.g. 0.7/SEATED+0.3/STANDING. So up to 10 TS are gathered and segmented for one subject.

Finally, the segmented data for each run is considered a data set, which means that up to 10 different data sets are available for training and testing. Furthermore, one more data set has been generated containing the segmented data from all the runs.

Typically, HAR studies in the literature make use of 10-fold cross-validation. The cross-validation scheme used in this study is a kind of 5×2 cv in which five random TS data sets are chosen for training and the five remaining sets are kept for testing. This is not the usual way of working with TS, but by using this scheme a better resemblance of the real deployment of the device and models is obtained, as many completely unknown TS are given to the model. However, higher classification errors than those that can be obtained for the usual 10-fold cross-validation schemes in different related studies are also expected.

4.3. *Comparison of GFFSM HAR methods*

Four different GFFSM approaches are to be compared within this subsection: The aGFFSM, gGFFSM, and wGFFSM presented in Sec. 3.2.1 and the bGFFSM learned with the single winner inference described in Sec. 3.2.4.

The parameters for learning each of the methods are shown in Table 2. Both the number of generations and the convergence stop condition for the bGFFSM vary according to the rules that have been added.

Results are shown in Tables 3 and 4 for the right and left hand, respectively. The data shown comes from the best individual in each run with the cross-validation method. For the bGFFSM, two columns are reported. When learning the model with six rules with the Boosting approach, we need to learn the model with five rules first, so results with five and six rules are included.

As can be seen, there is a totally different scenario with regard to that presented in Ref. 45: The performance of the different methods is completely

Table 2. Experiment parameters for each of the GFFSM approaches.

Parameter	aGFFSM	gGFFSM wGFFSM	bGFFSM
Generations	200	300	300+10/rule
Population size	100	100	50
Elite pop. Size	1	1	40
Subpop. size	—	—	10
Crossover prob.	0.8	0.8	1
Mutation prob.	0.02	0.02	0.1
Interchange prob.	—	—	0.01
Crossover BLXα	0.3	0.3	0.3
No. of rules	8	8	Up to 6
Generations w/o progress	50	50	100+10/rule
Logitboost α	—	—	0.75
Maximum number of iterations	—	—	4

Table 3. Comparison of MAE classification errors results after the 5×2 cross-validation for the *right hand*. Mn, Md, and Std stand for mean, median, and standard deviation, respectively.

Fold	aGFFSM	gGFFSM	wGFFSM	bGFFSM five rules	bGFFSM six rules
1	0.0780	0.0757	0.0321	0.0283	0.0248
2	0.0552	0.1920	0.0196	0.0365	0.0370
3	0.1293	0.0645	0.1158	0.0426	0.0473
4	0.1489	0.0848	0.0397	0.0119	0.0218
5	0.0439	0.0216	0.0175	0.0753	0.0573
6	0.1362	0.0540	0.0088	0.0248	0.0248
7	0.2550	0.0525	0.0194	0.0355	0.0355
8	0.0680	0.0826	0.0041	0.0251	0.0235
9	0.0936	0.0341	0.0139	0.0244	0.0705
10	0.0208	0.0540	0.0156	0.0462	0.0611
Mn	0.1029	0.0716	0.0287	0.0349	0.0404
Md	0.0858	0.0593	0.0185	0.0319	0.0362
Std	0.0679	0.0468	0.0323	0.0169	0.0177

Table 4. Comparison of MAE classification errors results after the 5×2 cross-validation for the *left hand.* Mn, Md and Std stand for mean, median, and standard deviation, respectively.

Fold	aGFFSM	gGFFSM	wGFFSM	bGFFSM five rules	bGFFSM six rules
1	0.0842	0.1474	0.0020	0.0008	0.0877
2	0.1971	0.1161	0.0016	0.0011	0.0012
3	0.1084	0.0642	0.1250	0.0008	0.0007
4	0.2401	0.0247	0.0350	0.0015	0.0015
5	0.1768	0.1848	0.0015	0.0008	0.0001
6	0.0683	0.1763	0.0394	0.0002	0.0029
7	0.3659	0.1406	0.0834	0.0012	0.0015
8	0.1310	0.0913	0.0666	0.0010	0.0322
9	0.2257	0.0566	0.0072	0.0008	0.0001
10	0.1135	0.0339	0.0011	0.0002	0.0029
Mn	0.1711	0.1036	0.0363	0.0009	0.0131
Md	0.1539	0.1037	0.0211	0.0008	0.0015
Std	0.0903	0.0580	0.0431	0.0004	0.0280

different as a consequence of the different cross-validation methods.

This study, with bigger validation data sets, benefits from the cross-validation method representing a more general solution. However, as these methods were learnt with smaller data sets, they perform with a slightly higher error measurement and the results could perhaps still be enhanced if more TS were available for training and testing.

Furthermore, the boxplots of the results for the results shown above are depicted in Figs. 8 and 9 for the right and the left hand, respectively. Both methods wGFFSM and bGFFSM are found to be the most suitable methods, although the gGFFSM's behavior changes remarkably with the hand that is considered.

In addition, it is remarkable that the bGFFSM got worse with six rules than with five rules. This might be due to different factors; for instance, the reduced number of samples for learning the sixth rule may induce rules that introduce noise instead of enhancing the model. Besides, it might also be due to the different number of samples in the data sets for each class.

4.4. *Evaluation of the FS method*

In this stage, the results from the PCA sorting and filtering on the original feature space are presented. As stated in Sec. 3.1.1, the most remarkable PCA transformations are used for the voting and ranking of the features.

Afterwards, the 20 features with the highest ranks will be chosen for a second FS stage, the wrapper FS. The exhaustive lists of the best features according to the PCA voting scheme for each hand are presented in Table 5.

The wrapper FS gathers these reduced feature subsets and learns wGFFSM models with the relaxed set of parameters. Two sets of parameters are needed: Those related with the GA for evolving the feature subset and those related with the wGFFSM's learning GA.

The parameters for the GA devoted to FS are 30 generations with 26 individuals, using the one point crossover operator with a probability of 0.8. The mutation operator is also flipping one of the three-selected features among the available candidates; the mutation probability is set to 0.02. The

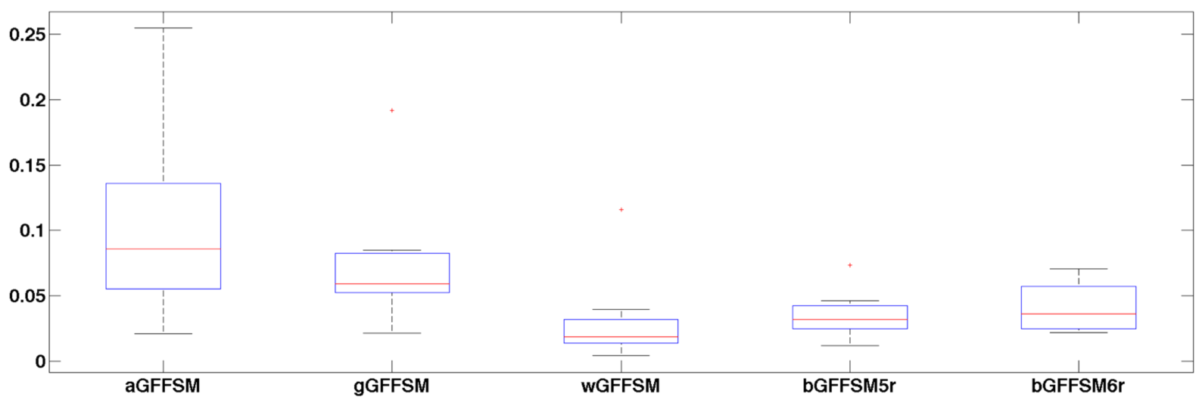

Fig. 8. MAE classification errors boxplot of the best individuals from the results of each run for the right hand. The best performance is obtained for the wGFFSM, but it is not statistically better than the bGFFSM.

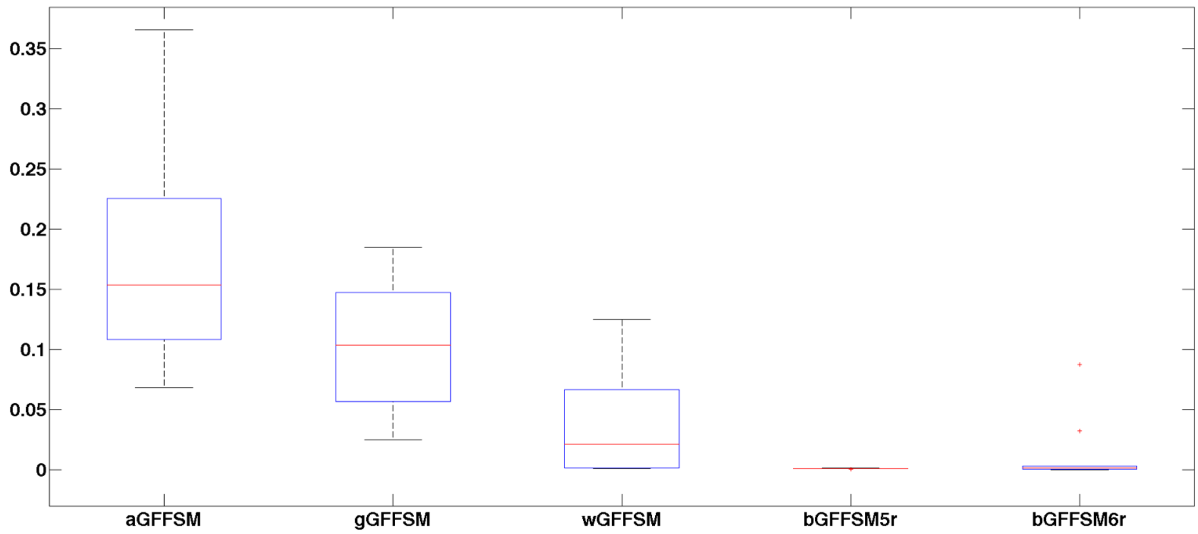

Fig. 9. MAE classification errors comparison boxplot for the best individual left hand for each run: The bGFFSM clearly beats all the other methods: Using five rules is enough.

Table 5. The 20 most relevant features after the PCA voting scheme. This reduced feature subset is the input space for the wrapper FS.

Left hand		Right hand	
Feature	Votes	Feature	Votes
Tl Kurtosis A	4585	T15(a^z, 10)	4225
T13(a^x, a^z)	4145	Tl Kurtosis b^z	4180
T3(a^x)	4105	Tl Skewness b^x	4075
g^x	4095	T15(a^x, 10)	3900
T15(a^z, 6)	4090	T13(a^x, a^z)	3900
a^x	4085	Tl Kurtosis b^y	3880
T15(a^x, 7)	4010	a^x	3845
Tl Kurtosis b^y	3835	T3(a^x)	3830
T15(a^y, 8)	3790	g^x	3795
Tl Skewness b^y	3780	Tl Skewness b^z	3775
Tl Kurtosis b^z	3555	T15(a^z, 6)	3520
T13(a^x, a^y)	3535	Tl Kurtosis b^x	3515
T15(a^x, 10)	3255	T7(a^y, a^z)	3385
T15(a^y, 7)	3170	T13(a^x, a^y)	3385
T7(a^x, a^y)	3155	T15(a^z, l)	3375
T15(a^y, 10)	3145	T15(a^y, 2)	3330
Tl Skewness b^y	3080	T13(a^y, a^z)	3240
T15(a^x, l)	3070	Tl Skewness b^y	3225
T13(a^x, a^z)	3020	g^y	3190
T15(a^z, 6)	2985	a^y	3165

fitness of each of the individuals is calculated as the MAE of the GFFSM model learnt from the feature subset the individual has chosen.

Learning the GFFSM within the wrapper is bound by the computational costs: In this case, the GA runs 50 generations with 76 individuals, and the α-crossover operator probability is set to 0.8 for an α value of 0.3. The mutation operator is the classical bitwise mutation for the rule base and the uniform mutation for the real coded part; the mutation probability is set to 0.02.

The following GA early stop conditions are defined by: (i) the convergence measured as 25 generations without changes in the MAE of the best individual and (ii) reaching an MAE fitness lower than 0.02 at any generation.

Table 6. Right hand MAE classification error test results obtained from the best model of each fold using the best feature subset after the PCA voting FS.

Fold	wGFFSM	bGFFSM five rules	bGFFSM six rules
1	0.0073	0.0107	0.1355
2	0.0062	0.0896	0.0185
3	0.0148	0.0273	0.0476
4	0.0072	0.0109	0.0109
5	0.0139	0.0215	0.0114
6	0.0051	0.0359	0.0178
7	0.0325	0.0268	0.0337
8	0.0045	0.0135	0.0136
9	0.0187	0.0127	0.0127
10	0.0096	0.0965	0.0219
Mn	0.0120	0.0345	0.0323
Md	0.0085	0.0242	0.0181
Std	0.0086	0.0320	0.0380

Table 7. Left hand MAE classification error test results obtained from the best model of each fold using the best feature subset after the PCA voting FS.

Fold	wGFFSM	bGFFSM five rules	bGFFSM six rules
1	0.0341	0.0176	0.0156
2	0.0428	0.0233	0.0207
3	0.0146	0.0261	0.0220
4	0.0093	0.0242	0.0278
5	0.0095	0.0554	0.0355
6	0.0100	0.0213	0.0213
7	0.0033	0.0235	0.0231
8	0.0071	0.0352	0.0265
9	0.0044	0.0287	0.0278
10	0.0115	0.0388	0.0408
Mn	0.0147	0.0294	0.0261
Md	0.0097	0.0252	0.0248
Std	0.0131	0.0111	0.0074

Finally, the best subset of features found for each hand were $\{g^x, \mathrm{T15}(a^x, 1), \mathrm{T15}(a^x, 10)\}$ for the left hand and $\{a^x, g^x, \mathrm{T15}(a^z, 10)\}$ for the right hand. The most interesting issue from these feature subsets is that the chosen features are basic transformations or even the raw data.

4.5. *The final GFFSM modeling*

Once the most relevant features have been chosen, we can evaluate the best models found so far with the complete set of parameters — those which were shown in Table 2: The wGFFSM and the bGFFSM. Tables 6 and 7 include the results obtained from each hand.

It is worth noting that the wGFFSM has been really enhanced while the bGFFSM has not: This is totally the opposite of what was expected. The

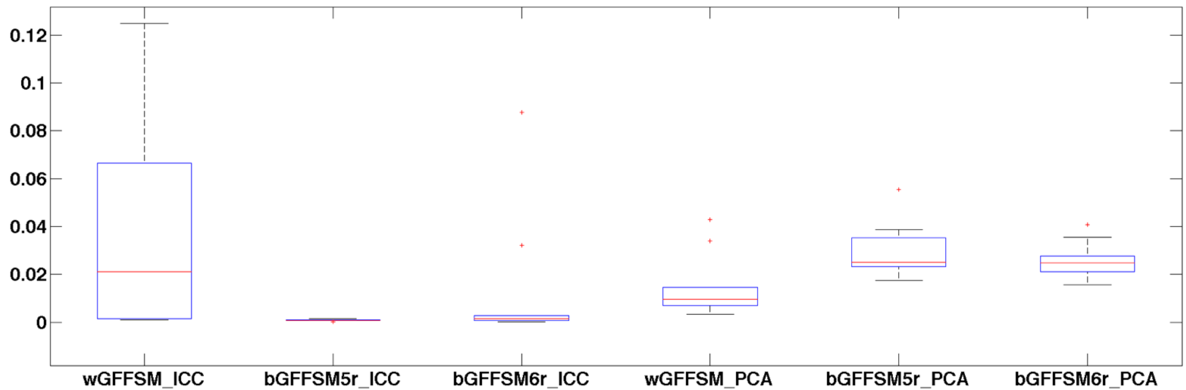

Fig. 10. Left hand comparison MAE classification error boxplot. The bGFFSM did not improve with the PCA FS. The ICC refers to the feature subset using the Information Correlation Coefficient obtained from the previous studies.

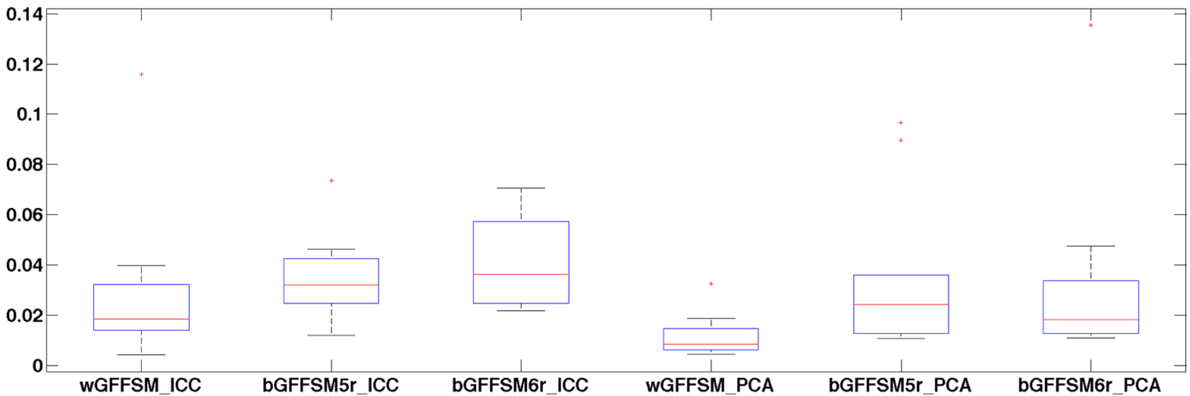

Fig. 11. Right hand comparison MAE classification error boxplot. The feature subset found after the PCA filtering FS slightly enhances the different methods, but does not clearly outperform the results found using the Information Correlation Coefficient.

Table 8. Left hand aggregated confusion matrix for the gFFSM6r when using the ICC/PCA feature subset. The errors for the STANDING state have increased significantly.

	Resting	Standing	Walking
Resting	23,655/23,862	220/4	15/24
Standing	675/10	11,990/11,545	45/1155
Walking	84/118	21/765	16,265/15,487

boxplots can potentially help us to understand what is happening in this process.

Figures 10 and 11 depict the boxplots for the left and right hand, respectively. As mentioned before, the bGFFSM got worse for both cases of 5 and 6 rules. This fact can be easily explained if we observe the aggregated confusion matrix (see Table 8) for all the folds. In fact, the number of samples labeled with WALKING is significantly higher than the sum of the rest of the classes.

Consequently, PCA may have driven the FS to those features that better reflect the most common class. Therefore, the results for the class STANDING were worse, which in turn affects the overall results in a negative way. One of the main conclusions that is drawn from this experimentation is that the SRTEST should be more carefully designed, considering an increase of the times T1 and T2.

5. Discussion of the Results

Some comments can be drawn from the experience detailed above. First, the hybrid FS method including the PCA might represent a good compromise when dealing with TS. In general, the deviation of the results is greatly reduced, although in some cases the model performs worse than when carrying out the filtering FS with the Information Correlation Coefficient and the same wrapper FS. It is clear that the experiments should be very carefully designed, but there is no doubt either that the results obtained have improved. The latter leads to our second point: The SRTEST in its current formulation is not completely valid and the different scheduled states should last about the same time.

Besides, the GFFSM models perform satisfactorily and the results are reasonable. Moreover, this model can be easily deployed in embedded systems. Only the fuzzy characteristics concerning the membership functions and the table for looking up the rules need updating from one subject to other. The results for all the subjects were quite similar, so the models are robust among the testing population. However, the tests were run on a population that is different from the target users in the overall research.

This experimentation was aimed at finding suitable models to use and to test if their outcome was accurate or not. To validate this HAR model in stroke patients, a specific study in this population should be undertaken.

One interesting question is what would happen if it were found that more than the presented activities are required. This model was developed with the premise that only three states were required and that the GFFSM will perform well with them, but the performance will certainly decrease with the number of activities.

From this, two main possibilities arise. The first one is to expand this model to five activities or introduce the “divide and conquer” principle to group the activities, including a classifier for each group: This would eventually require more computation in the bracelet, which might increase its cost. The second one includes evaluating alternative solutions that might be easily transferred to embedded devices. In this case, for instance, totally different approaches will be analyzed, because, as indicated in the state-of-the-art, most solutions are too complex to be embedded.

Furthermore, the same approach for HAR can be used for different topics, for instance, trying to detect the onset of epilepsy. This is also a very interesting topic in which the cHA is needed, and good discrimination between some types of activities and the onset characteristics that might resemble a similar response in the accelerometers is required.

6. Conclusions

This study is devoted to HAR in the context of stroke alarm generation. A proposal for stroke alarm generation is given and the HAR block has been fully analyzed, while stroke onset detection is left as future work. In this scenario, the focus population must perform the activities considered: Only three main activities were to be detected.

In order to develop an embedded solution, the GFFSM model has been adapted and extended with new algorithms, including the use of a boosting heuristic for learning the fuzzy system. Furthermore, a novel voting FS method based on rolling back the PCA rankings has enhanced the performance of the method.

The step for HAR included in our solution for developing a stroke onset alarm generation can be considered solved, as the MAE results are really impressive. Both the wGFFSM and the bGFFSM approach seem to be good options, although the latter is preferable due to the reduced number of rules required: This reduction makes its integration in embedded solutions easier.

However, the SRTEST test has been found to lack class balance so a better design would be needed. When testing the solution with subjects within the focus population, the balance of the classes within the data sets obtained from the SRTEST should be guaranteed.

Acknowledgments

This research has been partially supported through the projects of the Spanish Ministry of Science and Innovation PID 560300-2009-11 and TIN2011-24302, Fundación Universidad de Oviedo project FUO-EM-340-13, Junta de Castilla y León CCTT/10/BU/0002 and Agencia de Inversiones y Servicios de Castilla y León (record CCTT/10/LE/0001).

References

1. V. L. Feigin, M. H. Forouzanfar, R. Krishnamurthi, G. A. Mensah, M. Connor, D. A. Bennett *et al.*, Global and regional burden of stroke during 1990–2010: Findings from the Global Burden of Disease Study 2010, *Lancet* **383:9913** (2014) 245–254.
2. B. Ovbiagele, L. B. Goldstein, R. T. Higashida, V. J. Howard, S. C. Johnston, O. A. Khavjou *et al.*, American Heart Association Advocacy Coordinating Committee and Stroke Council. Forecasting the future of stroke in the United States: A policy statement from the American Heart Association and American Stroke Association, *Stroke* **44**(8) (2013) 2361–2375.
3. G. Adams, M. J. Del Zoppo, D. L. Alberts, L. Bhatt, A. Brass, R. L. Furlan *et al.*, Guidelines for the early management of adults with ischemic stroke, *Stroke* **38** (2007) 1655–1711.
4. J. L. Saver, Time is brain — quantified, *Stroke* **37**(1) (2006) 263–266.
5. W. Hacke, M. Kaste, E. Bluhmki, M. Brozman, A. Dávalos, D. Guidetti *et al.*, Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke, *The New England J. Med.* **359** (2008) 1317–1329.
6. R. L. Sacco, S. E. Kasner, J. P. Broderick, L. R. Caplan, J. J. Connors, A. Culebras *et al.*, An updated definition of stroke for the 21st century. A statement for healthcare professionals from the American Heart Association, *Stroke* **44** (2013) 2064–2089.
7. T. S. Olsen, Arm and leg paresis as outcome predictors in stroke rehabilitation, *Stroke* **21** (1990) 247–251.
8. J. R. Villar, S. González, J. Sedano, C. Chira and J. M. Trejo, Early diagnosis of stroke: Bridging the gap through wearable sensors and computational models, in *Proc. 9th Int. Conf. Applied Mathematics ICAM9* (2013), pp. 122–125.
9. A. Álvarez-Álvarez, G. Triviño and O. Cordón, Human gait modeling using a genetic fuzzy finite state machine, *IEEE Trans. Fuzzy Syst.* **20**(2) (2012) 205–223.
10. A. Ávarez-Álvarez, G. Triviño and O. Cordón, Body posture recognition by means of a genetic fuzzy finite state machine, in *IEEE 5th Int. Workshop on Genetic and Evolutionary Fuzzy Systems (GEFs)* (2011), pp. 60–65.
11. H. Adeli and S. L. Hung, *Machine Learning — Neural Networks, Genetic Algorithms, and Fuzzy Sets* (John Wiley and Sons, New York, 1995).
12. H. Adeli and K. Sarma, *Cost Optimization of Structures — Fuzzy Logic, Genetic Algorithms, and Parallel Computing* (John Wiley and Sons, West Sussex, United Kingdom, 2006).
13. N. Siddique and H. Adeli, *Computational Intelligence — Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing* (John Wiley, West Sussex, United Kingdom, 2013).
14. K. Sarma and H. Adeli, Fuzzy genetic algorithm for optimization of steel structures, *J. Struct. Eng. ASCE* **126**(5) (2000) 596–604.
15. A. Weingessel and K. Hornik, A robust subspace algorithm for principal component analysis, *Int. J. Neur. Syst.* **13**(307) (2003). Doi: 10.1142/S0129065703001650.
16. A. Meraoumia, S. Chitroub and A. Bouridane, 2D and 3D palmprint information, PCA and HMM for an improved person recognition performance, *Integ. Comput. Aided Eng.* **20**(3) (2013) 303–319.
17. S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, *IEEE Trans. Biomed. Eng.* **55**(2) (2008) 512–518.
18. J. Y. Yang, J. S. Wang and Y. P. Chen, Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural networks, *Patt. Recog. Lett.* **29** (2008) 2213–2220.
19. F. R. Allen, E. Ambikairajah, N. H. Lovell and B. G. Celler, Classification of a known sequence of motions

and postures from accelerometry data using adapted Gaussian mixture models, *Physiolo. Measurement* **27** (2006) 935–951.
20. M. P. Murray, A. B. Drought and R. C. Kory, Walking patterns of normal men, *J. Bone Joint Surgery* **46**(2) (1964) 335–360.
21. M. Ahmed and S. Ahmed, Kinetics and kinematics of loading response in stroke patients (a review article), *Ann. King Edward Medical University* **14**(4) (2008) 143–147.
22. I. de Quervain, S. Simon, S. Leurgans, W. Pease and D. McAllister, Gait pattern in the early recovery period after stroke, *J. Bone Joint Surgery. American* **78**(10) (1996) 1506–1514.
23. K. Subramanian and S. Suresh, Human action recognition using meta-cognitive neuro-fuzzy inference system, *Int. J. Neural Syst.* **22**(6) (2012) 1250028-15.
24. H. P. von Schroeder, R. D. Coutts, P. D. Lyden, E. Billings Jr. and V. Nickel, Gait parameters following stroke: A practical assessment, *J. Rehab. Res. Develop.* **32**(1) (1995) 25–31.
25. K. Hollands, Whole body coordination during turning while walking in stroke survivors, Ph.D. Thesis, School of Health and Population Sciences, University of Birmingham (2010).
26. S. R. Ke, H. L. U. Thuc, Y. J. Lee, J. N. Hwang, J. H. Yoo and K. H. Choi, A review on video-based human activity recognition, *Computers* **2** (2013) 88–131.
27. J. A. Iglesias, P. Angelov, A. Ledezma and A. Sanchis, Human activity recognition based on evolving fuzzy systems, *Int. J. Neural Syst.* **20**(5) (2010) 355–364.
28. B. Knorr, R. Hughes, D. Sherrill, J. Stein, M. Akay and P. Bonato, Quantitative measures of functional upper limb movement in persons after stroke, in *Proc. 2nd Int. IEEE Conf. Neural Engineering EMBS* (2005), pp. 252–255.
29. D. Rand, J. J. Eng, P. F. Tang, J. S. Jeng and C. Hung, How active are people with stroke? Use of accelerometers to assess physical activity, *Stroke* **40** (2009) 163–168.
30. S. H. Roy, M. S. Cheng, S. S. Chang, J. Moore, G. D. Luca, S. H. Nawab and C. J. D. Luca, A combined SEMG and accelerometer system for monitoring functional activity in stroke, *IEEE Trans. Neural Syst. Rehab. Eng.* **17**(6) (2009) 585–594.
31. T. Fujimoto, H. Nakajima, N. Tsuchiya, H. Marukawa, K. Kuramoto, S. Kobashi and Y. Hata, Wearable human activity recognition by electrocardiograph and accelerometer, in *43rd Int. IEEE Symp. Multiple-Valued Logic* (2013).
32. G. D. Fulk and E. Sazonov, Using sensors to measure activity in people with stroke, *Top Stroke Rehab.* **18**(6) (2011) 746–757.
33. J. Mantyla and T. Himberg, Recognizing human motion with multiple acceleration sensors, in *IEEE Int. Conf. Syst. Man Cybernet.* **3494** (2001) 747–752.
34. J. R. Kwapisz, G. M. Weiss and S. A. Moore, Activity recognition using cell phone accelerometers, *ACM SIGKDD Explorations Newslett.* **2** (2010) 74–82.
35. S. Wang, J. Yang, N. Chen, X. Chen and Q. Zhang, Human activity recognition with user-free accelerometers in the sensor networks, in *Proc. Int. Conf. Neural Networks and Brain ICNN&B'05*, Vol. 2, IEEE Conference Publications (2005), pp. 1212–1217.
36. Y. P. Chen, J. Y. Yang, S. N. Liou, G. Y. Lee and J. S. Wang, Online classifier construction algorithm for human activity detection using a tri-axial accelerometer, *Appl. Math. Comput.* **205**(2) (2008) 849–860.
37. S. González, J. R. Villar, J. Sedano and C. Chira, A preliminary study on early diagnosis of illnesses based on activity disturbances, in *Proc. 10th Int. Conf. Distributed Computing and Artificial Intelligence* (2013).
38. L. Bao and S. S. Intille, Activity recognition from user-annotated acceleration data, in *Proc. Second Int. Conf. Pervasive Computing, PERVASIVE 2004*, LNCS 3001 (Springer Verlag, Heidelberg, 2004), pp. 1–17.
39. N. Györbiro, A. Fábían and G. Hományi, An activity recognition system for mobile phones, *Mobile Networks and Applications* **14** (2009) 82–91.
40. M. Mathie, B. Celler, N. Lovell and A. Coster, Classification of basic daily movements using a triaxial accelerometer, *Med. Biol. Eng. Comput.* **42**(5) (2004) 679–687.
41. D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell and B. G. Celler, Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, *IEEE Trans. Inform. Technol. Biomed.* **10**(1) (2006) 156–167.
42. F. Chamroukhi, S. Mohammed, D. Trabelsi, L. Oukhellou and Y. Amirat, Joint segmentation of multivariate time series with hidden process regression for human activity recognition, *Neurocomputing* **120**(23) (2013) 633–644.
43. M. Zhang and A. A. Sawchuk, Human daily activity recognition with sparse representation using wearable sensors, *IEEE J. Biomed. Health Inform.* **17**(3) (2013) 553–560.
44. L. G. Mojica, S. Raghuraman, A. Balasubramanian and B. Prabhakaran, Exploring unconstrained mobile sensor based human activity recognition, *3rd Int. Workshop on Mobile Sensing* (2013).
45. J. R. Villar, S. González, J. Sedano, C. Chira and J. M. Trejo, Human activity recognition and feature selection for stroke early diagnosis, in *8th Int. Conf. Hybrid Artificial Intelligent System HAIS*

2013, LNCS 8073 (Springer Verlag, Berlin Heidelberg, 2013), pp. 659–668.

46. F. Herrera, Genetic fuzzy systems: Taxonomy, current research trends and prospects, *Evol. Intell.* **1** (2008) 27–46.
47. J. R. Villar, S. González, J. Sedano, E. Corchado, L. Puigpinós and J. de Ciurana, Meta-heuristic improvements applied for steel sheet incremental cold shaping, *Memetic Comput.* **4**(4) (2012) 249–261.
48. F. Neri and C. Cotta, Memetic algorithms and memetic computing optimization: A literature review, *Swarm and Evol. Comput.* **2** (2012) 1–14.
49. S. J. Brams and P. C. Fishburn, Going from theory to practice: The mixed success of approval voting, 2003, in *Handbook on Approval Voting Studies in Choice and Welfare* (Springer Verlag, 2010), pp. 19–37.
50. J. Casillas, O. Cordón, M. J. del Jesus and F. Herrera, Genetic feature selection in a fuzzy rule-based classification system learning process, *Inform. Sci.* **136**(1–4) (2001) 135–157.
51. T. Weise, *Global Optimization Algorithms — Theory and Application*. Available at http://www.it-weise.de/projects/book.pdf, accessed 1st December 2014.
52. L. Sánchez and J. Otero, Boosting fuzzy rules in classification problems under single-winner inference, *Int. J. Intelligent Syst.* **22**(9) (2007) 1021–1035.
53. B. Langhammer, Physical therapy tests in stroke rehabilitation, International Encyclopedia of Rehabilitation. Available at http://cirrie.buffalo.edu/encyclopedia/en/article/116. Accessed on October 2014.
54. S. Ahmed, N. E. Mayo, J. Higgins, N. M. Salbach, L. Finch and S. L. Wood-Dauphinee, The Stroke Rehabilitation Assessment of Movement (STREAM): A comparison with other measures used to evaluate effects of stroke and rehabilitation, *Phys. Therapy J.* **87** (2003) 617–630.