* Manuscript

Posterior probability profiles for the automated assessment of the recovery of patients with stroke from activity of daily living tasks

1st author and corresponding author:

Name: Gert Van Dijck, PhD.

Affiliation: Katholieke Universiteit Leuven, Laboratorium voor Neurofysiologie, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

e-mail: gert.vandijck@med.kuleuven.be,

Phone: +32 16 33 04 28

Fax: +32 16 34 59 60

2nd author:

Name: Jo Van Vaerenbergh, MSc.

Affiliation: Katholieke Universiteit Leuven, Laboratorium voor Neurofysiologie, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

e-mail: jozef.vanvaerenbergh@med.kuleuven.be

3rd author:

Name: Marc M. Van Hulle, PhD., Professor

Affiliation: Katholieke Universiteit Leuven, Laboratorium voor Neurofysiologie, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

e-mail: Marc.vanhulle@med.kuleuven.be,

Phone: +32 16 34 59 61

Fax: +32 16 34 59 60

Abstract

Assessing recovery after stroke has been so far a time consuming procedure in which trained clinicians are required. This article proposes class posterior probabilities computed over time as a quantitative and statistically sound tool to assess functional recovery from isometric force and torque measurements. Force and torque measurements were obtained when patients with stroke tried to perform activity of daily living tasks. The performance of the patients was quantified by means of the posterior probability to belong to the class of normal subjects. The posterior probabilities express whether the selected features which characterize the execution of the tasks is typical for normal controls. The mechatronic platform for force and torque measurements and the class posterior probabilities enable to automate functional recovery assessment. It is shown that the class posterior probability profiles are highly correlated, $r \approx 0.8$, with the well-established Fugl-Meyer scale assessment in motor recovery. The posterior probability profiles confirm the importance of initial recovery within a few weeks after the stroke to obtain a high recovery level. These results have been obtained through careful feature subset selection procedures in order to prune the large feature set being generated. The overall approach is general and can be applied to many other health monitoring systems where different categories (diseased vs. healthy) can be identified.

Key words: Activity of daily living tasks; Classification; Feature construction; Feature subset selection, Mutual information, Stroke recovery.

1. Introduction

1.1. General background on stroke

The World Health Organization defines stroke as a syndrome consisting of the rapid onset of a focal cerebral deficit of vascular origin lasting more than 24 hours [1]. Stroke, also known as cerebrovascular accident (CVA) or 'brain attack', ranks 3rd among all causes of death behind heart diseases and cancer in the United States [2]. It is expected to become the major cause of death worldwide. Moreover, it is number 1 as a leading cause in long-term disability in the United States [3]. Patients with stroke suffer from disabilities ranging from hemiparesis, gait disturbance, incontinence, cognitive disturbance, vision disturbance, dependency in activities of daily living tasks, aphasia, numbness to depressive symptoms [4]. The symptoms are largely dependent on which part of the brain is affected by the stroke and the size of the affected part. In [5] the degree of white matter lesions (WML) was found to be correlated with global cognitive function, executive dysfunction, impaired memory functions and impaired activity of daily living.

Several factors have been shown to contribute to an increased risk for stroke: atrial fibrillation, cigarette smoking, diabetes, high blood pressure, hypercholesterolemia and obesity among others [2].

It is estimated that the direct and indirect cost related to stroke is $65.5 billion in 2008 in the U.S. [2]. Largest contributors in the acute care costs are [6]: room charges (50%), medical management (21%) and diagnostic costs (19%). For details on stroke incidence, stroke symptoms, stroke risk factors and stroke costs the reader is referred to [2] and the references therein.

1.2. Research motivation

For Europe it is expected that the proportion of the population aged ≥ 65, in which most stroke events occur, will increase from 20% in 2000 to 35% in 2050 and the median population age will rise from 37.7 years in 2000 to 47.7 years in 2050. Moreover the global number of people living in Europe will decrease from 728 million in 2000 to 705 million in 2050 resulting in fewer young people taking care of the increasing proportion of elderly people [7]. Tempering the costs of stroke care will be a tremendous challenge for future health care systems [8]. Nowadays a large number of patients with stroke benefits from comprehensive inpatient rehabilitation and to, support and quantify functional recovery robots and mechatronics technology were successfully introduced [9], [10]. It is important that rehabilitation specialists have objective tools to assess recovery, the ability to recover and the effect of therapies on the recovery process. This will help in reducing the stay of patients in hospitals and hence moderate the costs. Current techniques require a clinician to score the performance of patients in some tasks on specific scales: disability scales e.g. Barthel index (BI), Functional Independence Measure (FIM), global deficit rating scales e.g., the National Institutes of Health Stroke Scale (NIHSS), [11], [12]. These scales require trained therapists to score patients with stroke according preset rules, which can be very time consuming [11]. A demand for automated assessment techniques arises due to the increasing number of patients and the continuous growth of new treatment options [11].

Artificial intelligence (AI) techniques may be good candidates to assist clinical experts in decision making in stroke rehabilitation. They have been proven useful as a decision tool for thrombolysis [13] after stroke. However, if the patient cannot be treated within 3 hours [14] after the onset of the first symptoms of the stroke with thrombolysis damage to the brain is likely to occur and one has to consider rehabilitation, e.g. by means of physical therapy [15], [16], [17], [18].

The following research uses artificial intelligence techniques to automate the assessment of functional recovery after stroke by processing force and torque patterns exerted by patients during the performance of particular activity of daily living tasks.

For this purpose a mechatronic platform [19], [20] (section 2.2) and feature subset selection technology (section 2.4.2) were developed and the Bayesian inference mechanism (section 2.4.1) was used. These technologies together form the artificial intelligence (AI) system.

Our research is inspired by the research in [21] and [22] demonstrating the usefulness of isometric force and torque pattern analysis for stroke assessment. [22] showed characteristic forces and torques patterns for hemiparetic patients during range of motion analysis of the elbow, though the relation with functional recovery was not firmly assessed. We used six degrees of freedom (6 DOF) force/torque cells as in [21], but this time to derive functional recovery processes from the analysis of isometric force and torque patterns generated during attempts of performing daily living tasks.

The clinical evidence of the AI system for the quantification of the patients' progress is also clearly demonstrated in this article. The posterior class probabilities quantifying the recovery progress are validated by means of a correlation analysis with the Lindmark modified Fugl-Meyer scale [23], [24]. Hereto, patients were regularly monitored until 6 months after the stroke with both the mechatronic platform and the Fugl-Meyer scoring. With the aim of simplifying and downscaling the mechatronic device in the future, sensors that did not contribute to characterizing patients with stroke are eliminated by a new hybrid filter-wrapper approach that has been developed in [25]. A smaller part of the correlation analysis in section 3 has been previously presented in [26].

2. Materials and methods

2.1. Subject data

Fugl-Meyer scores and force/torque measurements were obtained from patients who were admitted from 2004 till 2005 to 3 stroke units: in Belgium (Gent), in Hungary (Budapest) and in Ireland (Dublin). Computed tomography (CT) and/or magnetic resonance imaging (MRI) scans were obtained at a mean of 6 days after stroke. Force and torque measurements were obtained at 6 days after the stroke from a total of 57 patients with stroke. Force and torque measurements from an age and sex matched control group of 57 subjects were recorded as well to obtain baseline force and torque patterns. The control group was free of any neurological or orthopaedic disorder.

16 patients out of the total group of 57 were retrospectively selected. These patients had a first-ever ischemic or haemorrhagic stroke within the middle cerebral artery (MCA) territory and completed the full six month follow up period during which they were measured at least once a week. Patients with cognitive impairments, hampering collaboration or orthopaedic comorbidity influencing the measurements were rejected as well. All measured patients reached a stable general medical condition within the first week after stroke.

A summary of the group of 16 patients is provided in table 1. Patients are between 47 and 86 years old with a mean age of 65 and a standard deviation of 11 years. Of the 16 patients, 10 are of male gender and 6 of female gender. One half of the patients has a lesion on the left side, the others on the right side.

TABLE 1 HERE.

2.2. Measurements from ADL tasks

The quality of the patients' life is strongly related to regaining the ability of performing activities of daily living (ADL) tasks. This consists of a set of frequently executed tasks such as: drinking from a glass, turning a key, grasping objects and carrying objects. These tasks are thoroughly described in textbooks for physical and occupational therapists [27], [28].

The 57 patients with stroke and 57 normal controls were asked to perform 6 very different ADL tasks: ‘drinking a glass of water’, ‘turning a key’, ‘picking up a spoon’, ‘lifting a bag’, ‘reaching for a bottle’ and ‘lifting and carrying a bottle’.

Every subject executed each task 3 times. To perform the tasks patients and normal controls were seated in a mechatronic platform [20] containing 8, 6 DOF sensors (JR3 Inc), which were closely connected to different parts of the human body in figure 1.

The sensors are located at the thumb, index, middle finger, the lower arm, below the posterior, behind the trunk, a foot and a big toe.

FIGURE 1 HERE

The mechatronic platform was designed to make it adaptable to major anthropometrical differences: small persons (1530-1625 mm), medium persons (1625-1751 mm) and large persons (1751-1870 mm). It guaranteed a same anatomical start position for all subjects [20] and consequently the intra and inter reliability of the measurements. The force measurement resolution was 0.1 N or better and signals were sampled at 100 Hz. This sampling rate is at least twice as high than the highest frequency that can be expected from involuntary movements. Tremor frequencies, e.g. resulting from maintaining a posture, are typically in a range of 8 to 12 Hz in the hands [29].

The measurement protocol started with showing to the subjects a video of a particular task performed by a normal subject. Immediately after and in response to a sound signal, the subjects attempted to perform the task. During this perfomance, force and torque signals were recorded from the sensors both in X, Y and Z direction. Figure 2 shows 3 times series $F_x(k)$, $F_y(k)$ and $F_z(k)$ during the drinking a glass task.

FIGURE 2 HERE

The forces and torques were measured under isometric constraints. Under these conditions, the subject contracts his muscles according to a given pattern, e.g. a prescribed number of kg's: for a prescribed number of seconds. It is very important that the position of the subject is fixed in an exactly duplicable way over different trials [30].

When a patient tried to drink from a cup, the patient had to insert the thumb, index and middle finger within the sensors, which were attached to the cup. An isometric setting was selected because it allows to measure in a standardized way the movement initiation [31] at a moment the active motion range of the patient may be very limited. It has been shown that the isometric force and torque patterns of patients with hemiparesis are different from normal controls [21], [22], [32]. Previous research concentrated on features such as maximum force and/or torque values [21], [22] or on movement smoothness. The following research uses traditional features and constructs new features as well (in section 2.3.) with the aim to better quantify and predict functional recovery.

It has to be noticed that the recordings lead to a large amount of data per experiment: (6 ADL tasks per experiment) * (3 repetitions per ADL task) * (8 sensors) * (3 spatial directions per sensor) * (2 types of measurements: force and torque) = 864 measurements in total.

2.3. Feature construction

'Raw' time series, in this case force and torque signals, are seldom used for making predictions in pattern recognition systems. In this study the pattern corresponds with 'normal' or 'stroke'. The disadvantage of using the raw time series can be seen as follows. If one uses 3 seconds of the force/torque measurements this would imply at our sampling rate of 100 Hz that 300 samples would be used per time series. This dimensionality of 300 is already a multiple of the 57 patients and 57 normal controls available. It is well known that making predictions in such sparsely populated spaces may become inaccurate due to the 'curse of dimensionality' [33]. Secondly, one often disposes of prior knowledge about the difference in behavior between normal controls and the patients. It is common to define features based on these times series. Extracting features, defined as functions of the times series, is a first important step to dimensionality reduction and allows expressing one's prior assumptions (or hypotheses) by defining functions of the time series.

In figure 3, the force trajectories are shown by connecting the end-points of subsequent force vectors $\mathbf{F}(k-1) = [F_x(k-1), F_y(k-1), F_z(k-1)]$ and $\mathbf{F}(k) = [F_x(k), F_y(k), F_z(k)]$, where 'k' is the time index. These trajectories are shown for one patient and one normal control.

FIGURE 3 HERE

Next the construction of features is described. The features were defined on the force and the torque vectors.

2.3.1. Planning of a trajectory

Simple observation shows that normal controls better persist in their efforts maintaining a mean trajectory, e.g. when trying to bring a glass to the mouth. Hence, it can be expected that the angular deviations of **F**(k) and **T**(k) relative to the mean efforts $\mathbf{F}_m$ (average force over time) and $\mathbf{T}_m$ (average torque over time) show larger deviations and abnormalities for patients. Indeed, experiments have shown that more severely impaired subjects produce larger directional errors [34]. These deviations and abnormalities can be assessed by calculating the maximal deviation, the standard deviation, the skewness, the kurtosis of the angular deviations from the mean effort. These statistical measures however, do not take temporal aspects into account. Therefore other sets of features that are able to take the temporal aspects into account were defined as well, an autoregressive model (AR-model) was fitted to these angular deviations from the mean effort with the Bayesian information criterion (BIC) [35] to assess the time lag needed. This model allows to take linear dependencies into account over time by fitting the current angular deviations based on the previous angular deviations. Moreover, rather than building vectors $\mathbf{F}(k) = [F_x(k), F_y(k), F_z(k)]$ based on all 3 signals, similarly vectors based on 2 force and 2 torque signals were built:
$\mathbf{F}_{xy}(k) = [F_x(k), F_y(k)]$, $\mathbf{F}_{xz}(k) = [F_x(k), F_z(k)]$ and $\mathbf{F}_{yz}(k) = [F_y(k), F_z(k)]$. The same holds for the torque vectors **T**(k), $\mathbf{T}_{xy}(k) = [T_x(k), T_y(k)]$, $\mathbf{T}_{xz}(k) = [T_x(k), T_z(k)]$ and $\mathbf{T}_{yz}(k) = [T_y(k), T_z(k)]$. These new vector definitions enabled to exclude force or torque signal components that were noisy signals.

2.3.2. Continuity in voluntary movement

The ‘continuity’ of the voluntary movement can be quantified by computing the sequence of angles between subsequent force vectors, $\theta[k] = \text{angle}(\mathbf{F}(k),\mathbf{F}(k\text{-}1))$, and torque vectors, $\varphi[k] = \text{angle}(\mathbf{T}(k),\mathbf{T}(k\text{-}1))$. The difference with ‘the planning of a trajectory’, in the ‘continuity in voluntary movement’ is that angles between sequential force and torque vectors are computed instead of the angles between **F**(k) and $\mathbf{F}_m$ and the angles between **T**(k) and $\mathbf{T}_m$.
Hence, parameters of these angles such as maximal deviations from the mean, skewness, standard deviations, autoregressive coefficients quantify both the statistical and temporal aspects of the abnormalities in sequential angular deviations.

2.3.3. Velocity components

From the force and torque values one can extract linear velocity and angular velocity respectively:

$$\frac{m}{l+1}\mathbf{v}(l) = \frac{1}{l+1}\sum_{k=0}^{l}\mathbf{F}(k), \tag{1}$$

$$\frac{I}{l+1}\boldsymbol{\omega}(l)=\frac{1}{l+1}\sum_{k=0}^{l}\mathbf{T}(k). \tag{2}$$

However, it needs to be emphasized that, strictly speaking, there is no real movement, while the objects were fixed in the isometric setting. Therefore, these velocity features have no physical meaning. They represent real movements in case of freely moving objects and time independent mass (m) and moment of inertia (I). These velocities are called 'imaginary' linear velocities, $\mathbf{v}(l)$, and 'imaginary' angular (or rotational) velocities $\boldsymbol{\omega}(l)$. The statistics about the newly obtained angular and linear velocities were summarized in the same way as for the force and torque signals: using the definitions of 'planning of a trajectory' and 'continuity in voluntary movement' on the velocity vectors.

2.3.4. Coordination between body parts

The voluntary movement of objects in the ADL tasks needs a careful coordination between some body parts, which implies that forces exerted by one body part are likely to be statistically dependent on forces from other body parts. Consider e.g. the 'drinking a glass' task, it is clear that one needs a proper coordination between forces or torques exerted by the thumb, the index and the middle finger. This coordination was measured by the information theoretic measure of statistical dependency, known as mutual information [36]. The mutual information between norms of the force vector from sensor 's1', $\|\mathbf{F}_{s1}(k)\|$, and sensor 's2', $\|\mathbf{F}_{s2}(k)\|$, were computed as follows:

$$\begin{aligned} &MI(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k)\|)= \\ &\max_{a}\sum_{k} p(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k-a)\|)\ln\left(\frac{p(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k-a)\|}{p(\|\mathbf{F}_{s1}(k)\|)\,p(\|\mathbf{F}_{s2}(k-a)\|)}\right). \end{aligned} \tag{3}$$

One needs to take into account that signals belonging to different sensors can be shifted in time. Therefore, the delay 'a' between the sensors was searched for, such that the norms of the vectors became maximally dependent. Hence, this shift parameter 'a' was estimated by means of a maximization of mutual information approach which is clear from formula (3) by means of the 'max' operator. This is a new feature that has not been described yet in stroke recovery. Figure 4 shows that this feature is indeed able to describe differences between patients and normal controls. The maximal mutual information from formula (3) for all 57 normal controls and 57 patients with stroke is shown for the norms of the force vector between the thumb and the index in the 'turning a key' task.

It is clear from figure 4, that the average mutual information for normal controls is higher than for the patients with stroke. A t-test rejects the null-hypothesis of equal means with $p \approx 2.31*10^{-11}$. A similar conclusion can be drawn for the thumb and the middle finger. Hence, this provides evidence for the fact that these forces are much less dependent for the group of patients with stroke. This is a very plausible result, while 'turning a key' needs a fine coordination between the thumb-index and the thumb-middle finger. It is clear from this example that feature construction and putting forward hypotheses go hand in hand.

FIGURE 4 HERE

2.3.5. Time series fitting

As can be seen in figure 2, the force and torque signals consist of a rising part and a decaying part. This corresponds with an attempt of the subject to manipulate the object followed with the release of the attempt. This behavior was modeled by a sum of 2 exponential functions. The parameters were used as features, as well as the residuals obtained after fitting the model.

In summary, the aforementioned definitions led to a total of 59472 features. The result of applying a given feature definition to each ADL task (6 in total), each attempt (3 attempts per task) and each sensor (8 sensors) is considered as a different feature. From this point on, the data mining was not performed on the raw time series, but on the 59472 features extracted from the time series.

2.4. Class posterior probabilities and dimensionality reduction

First, the rationale behind class posterior probabilities is discussed, subsequently a combined feature filter and feature wrapper selection approach is applied to decrease the dimensionality prior to density estimation.

2.4.1. Rationale of class posterior probabilities

Given the set of normal controls and patients, the differences in the features (F_1, F_2, …, F_n) between patients with stroke and normal controls are described by means of their class conditional distributions:

$$p(F_1, F_2, ..., F_n \mid normal),$$

$$p(F_1, F_2, ..., F_n \mid stroke).$$

These probabilities need to be estimated from measurements on a group of normal controls and patients with stroke. For the estimation of $p(F_1, F_2, ..., F_n \mid stroke)$ the features (F_1, F_2, …, F_n) were computed when patients were in the acute phase, less than 6 days after the stroke. This is the starting point from which we want to monitor changes in functional recovery.

The class posterior probability is then obtained by means of Bayes' theorem:

$$P(normal \mid F_1, ..., F_n) = \frac{P(F_1, ..., F_n \mid normal)\, p(normal)}{P(F_1, ..., F_n \mid normal).p(normal) + P(F_1, ..., F_n \mid stroke).p(stroke)},$$
$$and\ P(stroke \mid F_1, ..., F_n) = 1 - P(normal \mid F_1, ..., F_n), \tag{4}$$
$$p(stroke) + p(normal) = 1.$$

If the features of a patient with stroke are obtained at a certain time instant, say 'd' days after the stroke, (f_1(d), f_2(d), …f_n(d)), then $P(normal \mid F_1(d), ..., F_n(d))$ provides a quantitative measure for 'being normal' by means of the class posterior probability to belong to the group of normal controls. The class posterior probability has the advantage that it takes explicitly both the normal controls and patients with stroke into account. This is a clear difference with the hypothesized ideal performance of the traditional scaling methods discussed in the introduction. Several normal controls perform ADL tasks in potentially a different manner; therefore it is natural to take the statistical distribution for the normal controls into account. Moreover, the class posterior probability is an easily interpretable measure, bounded by two extreme values: 1 as a measure for complete 'normality' and 0 as measure for complete 'disability'. A third advantage of the approach is that is Bayesian: the prior knowledge of an clinical expert about the 'normality' of a patient is represented by p(normal). This serves as an interface between the clinical expert and the Bayesian inference mechanism. In the validation section, the non-informative prior probability, p(normal) = ½, was used. This prior probability was set to this value in order to attribute the result to the trained system, rather than the prior knowledge of the expert.

2.4.2. Feature subset selection

It is well known that probability densities can not be estimated accurately in high dimensional feature spaces, using well-established statistical techniques such as kernel-density estimation. In general, the accuracy is a function of both the number of data points and the dimensionality of the feature space, such as e.g. in the AMISE (Asymptotic Mean Integrated Square Error) expression in [37] for kernel density estimation:

$$AMISE(h_{opt}) \sim N^{-\frac{4}{4+D}}. \tag{5}$$

Here, ‘h_{opt}’ is the optimal bandwidth of the kernel, ‘N’ the number of samples used in the density estimate and D the dimensionality of the feature space. Hence, in this case, the accuracy decreases exponentially with increasing dimensionality ‘D’ of the feature space for a fixed number of data points.

A general observation of the decrease in accuracy with increasing dimensionality is the so-called ‘curse of dimensionality’. A second important reason for feature subset selection is to achieve a reduction in the number of sensors with the aim to downscale the mechatronic device to make it smaller and cheaper. This is a goal that in general cannot be achieved by means of feature extraction. Feature extraction extracts new features which are combinations of the original features and therefore exist of combinations of the sensors. Feature subset selection allows attributing the selected features to particular sensors. There are 2 main paradigms to perform feature subset selection [38]: the filter-based feature subset selection and the wrapper-based feature subset selection. In the filter-based selection the induction algorithm, see section 3, is not included in the search, whereas in the wrapper search the induction algorithm is used to obtain classification performances of already selected subsets. Filter approaches find feature subsets with a lower computational cost, but the obtained feature subsets result in lower classification accuracies as those obtained from a wrapper approach [39]. Very recently it was shown in [25] that preceding a wrapper search with a filter results in subsets with the same accuracy as obtained from wrapper searches, but with a reduced computational cost. The resulting feature subset selection approach is called a hybrid filter-wrapper approach. The used filter and wrapper approaches are described in following 2 sections.

2.4.2.1. Filter based feature selection

Because quantitative analysis by means of posterior probabilities is central in this article, one should be able to reduce the high-dimensional feature set to lower dimensions, without the risk of losing important features and without actually the need for calculating the probabilities in high-dimensional spaces. As a first step, the features are preprocessed with a filter that relies on following theorem [25]:

$$\begin{aligned} &KL(p(C \mid F_1,...,F_n) \,\|\, p(C \mid F_{r_1},...,F_{r_{n1}})) = \\ &\sum_{i=1}^{n2} MI(F_{s_i};C) \qquad\qquad\qquad . \\ &with,\ \{r_1,r_2,...,r_{n1}\} \cup \{s_1,...,s_{n2}\} = \{1,2,...n\} \end{aligned} \tag{6}$$

Here we denote with 'C' the class variable that will be either the class of patients or the class of normal controls, $\{F_1, \ldots, F_n\}$ is the set of all features, $\{Fr_1, \ldots, Fr_{n1}\}$ is a selected subset and $\{Fs_1, \ldots, Fs_{n2}\}$ is the set of omitted features. This theorem states that the Kullback-Leibler distance [36] between the full class posterior probability $p(C \mid F_1,...,F_n)$ and the class posterior probability of a selected subset $p(C \mid F_{r_1},...,F_{r_{n1}})$ is equal to a sum of marginal mutual information contributions ($MI(F_{s_i};C)$) between the class variable and the omitted features. This theorem holds when features are class conditional independent and independent, i.e. $p(f_1,...,f_n \mid c) = p(f_1 \mid c).p(f_2 \mid c)...p(f_n \mid c)$ and $p(f_1,...,f_n) = p(f_1)...p(f_n)$ respectively for all c, f_1, …, f_n. Then discarding those features $\{F_{S1}, F_{S2}, \ldots F_{Sn2}\}$ for which $MI(F_{Si};C) = 0$ guarantees that $KL(p(C \mid F_1,...,F_n) \,\|\, p(C \mid F_{r_1},...,F_{r_{n1}})) = 0$ and hence no information is lost w.r.t. to the full class posterior probability $p(C \mid F_1,...,F_n)$. Strictly speaking, the conditions of class conditional feature independence and feature independence may not be fulfilled. Nevertheless, extensive simulations on gene expression data sets in [25] have shown that preceding a sequential forward [40] wrapper search with this filtering step, i.e. removing features for which $MI(F_{S1};C) = 0$, speeds up the wrapper search without a decrease in classification accuracy. The theorem establishes a connection between two filter-based feature selection frameworks: the conditional relative entropy framework in [41] and the mutual information feature selection framework [42].

This theorem is optimally suited for the problem considered here: one wants to remove features Fs_1, …, Fs_{n2} such that the original posterior probabilities $p(normal \mid F_1,...,F_n)$ and $p(stroke \mid F_1,...,F_n)$ are not changed.

The mutual information was estimated by means of the entropies:

$$\begin{aligned} MI(F_i;C) &= H(F_i) - H(F_i \mid C) \\ &= H(F_i) - \sum_{j=1}^{2} H(F_i \mid c_j).p(c_j). \end{aligned} \tag{7}$$

Where the entropies were estimated by equation (20) in [43]:

$$\hat{H}(F_i) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N}\sum_{i=1}^{N}\log\varepsilon(i)\,. \tag{8}$$

Here ψ(.) is the psi-function, 'N' the number of data points (57 for the class of patients with stroke and 57 for the normal controls), ε(i) is twice the distance from the i'th data point to its k'th neighbor, 'd' the dimensionality and 'c_d' the volume of the d-dimensional unit ball. For the Euclidean norm $c_d = \pi^{d/2}/\Gamma(1 + d/2)$, with Γ(.) the gamma-function. Because features are considered individually in the theorem, 'd' is equal to 1.

It has to be noted that due to the finite sample estimation, $MI(F_i; C)$ is different from 0, even if the feature 'F_i' is statistically independent from the class variable 'C'. Statistical relevance can be easily tested, under the null hypothesis of irrelevance, by random permutation of the class labels. A total of 1000 permutations were used to assess the statistical significance of feature relevance. Features for which the MI exceeds a certain significance level (α was taken equal to 0.01 in this case) can be considered as statistically relevant. In figure 5, an example is shown when the permutation test, using 1000 permutations, was applied to the appealing feature of the mutual information between the norm of the force exerted by the index and the norm of the force exerted by the thumb in the turning a key task which (which was shown in figure 4). The actual MI is higher than those obtained under permutations.

FIGURE 5 HERE

This procedure allowed to reduce the original set of 59472 features to a set of 2637 most relevant features. Note that the set of 2637 features is larger than can be expected than when every feature would be irrelevant. This would lead to approximately 59472*0.01 ≈ 600 features. Hence, this indicates that the feature construction lead to some discriminative features.

2.4.2.2. Wrapper based feature selection

The set of 2637 features is still too high to estimate the posterior probabilities accurately for the 57 normal controls and the 57 patients. The set of features is further reduced by taking the induction algorithm into account in the search [38] and retaining those features for which the best classification could be obtained in terms of classification accuracy with a leave-one-out validation. Strictly speaking, statistical optimality such as in the filter procedure cannot be guaranteed anymore. Three induction algorithms were compared to search for the best subset of features, in terms of discriminating normal controls from patients with stroke: k-nearest neighbor (KNN), least squares support vector machines (LSSVM) [44] and a Bayesian classifier with kernel density estimation (KDE), for KDE see [45]. The number of nearest neighbors 'K' was set equal to 1.

In case of the KDE, the class conditional probability for normal subjects was estimated as:

$$P(F \mid normal) = \frac{1}{n.\left(h_{normal}\right)^d} \sum_{i=1}^{n} K\left(\frac{F - F_{i,normal}}{h_{normal}}\right). \tag{9}$$

Where index 'i' runs over all training subjects from the normal class, 'K' is a Gaussian kernel, h_{normal} the kernel bandwidth, 'd' the dimensionality of feature vectors $F_{i,normal}$. A similar expression is obtained for the class of patients with stroke.

Gaussian kernels were used with the maximum likelihood cross-validation method for kernel bandwidth estimation in KDE.

The hyper-parameters of the LSSVM were set by hand such that initial experiments with a leave-one-out validation provided good results: regularization parameter 'γ' = 29.6 and the square of the kernel-width 'σ^2' = 1.15. The LSSVM hyperparameters were then further tuned manually for the best subset found in order to increase the performance.

The sequential forward search (SFS) [40] procedure was used in the wrapper search for feature subsets. The SFS procedure gradually adds the next best feature to an existing subset, starting from the empty set.

3. Results

We evaluate our new methodology of section 2 in different manners. It is verified whether the selected features can be used to make a good distinction between the patients with stroke and normal controls when using different classification algorithms. This distinction helped us in finding the important features and sensors, but is on its own not sufficient. The posterior probabilities should also be informative about the degree of recovery. This is verified by means of a correlation study with different subscores of the Fugl-Meyer scoring. Moreover, we verify whether the posterior probabilities lead to plausible recovery patterns.

3.1. Distinguishing patients from normal controls

Table 2 shows that the best performance using a leave-one-out validation is obtained for the Bayesian classifier with KDE 98.25% (shown in the 4th column) with a subset size of 6 features. The first column represents the number of features selected, abbreviated as feature subset size (FS Size). The K-NN approach (shown in the 2nd column) reaches its maximal accuracy of 94.74% for feature subset sizes of 5 and 6 features. The LSSVM approach (shown in the third column) reaches a maximal accuracy of 97.37% for a feature subset size of 6 features.

TABLE 2 HERE

In fact in case of the KDE, only 2 patients with stroke were erroneously assigned to the class of normal controls and 0 normal controls to the class of patients with stroke. Hence, this leads to an overall classification accuracy of (112/114)*100 = 98.25%. The sensitivity, which is the number of true positives / (number of true positives + number of false negatives), is equal to (55 / (55 + 2))*100 = 96.5%. The specificity, which is the number of true negatives / (number of true negatives + number of false positives), is equal to (57 / (57 + 0))*100 = 100%.
Performing experiments with more than 6 features for all classifiers until the number of features was equal to 10, did not increase the classification performance. In case of the KDE, the resulting set consists of 6 features in which only 4 sensors appeared: the thumb, index, middle finger and the seat. This suggests that the platform can be reduced from 8 to 4 sensors. The most discriminative tasks obtained from the KDE feature selection are: ‘drinking a glass’, ‘lifting a bag’ and ‘lifting and carrying a bottle’. The sensors and ADL tasks that were selected in case of the KDE feature selection are included in table 2.
The most discriminative features obtained from the KDE feature selection are: ‘angular velocity’, angular deviations from ‘continuity in voluntary movement’ and the residual from the ‘time series fitting’. We use the 6 selected features to calculate the evolution of the posterior probability from (4) over time.

3.2. Correlation with Fugl-Meyer scoring

It needs to be shown that the whole procedure of defining features (section 2.3), reducing the feature set by means of the hybrid filter-wrapper approach (section 2.4.2), followed by the calculation of the posterior probabilities (formula 4) on the reduced feature set leads to results that satisfy the initial goals of monitoring the patient’s functional recovery over time with the posterior probability.

The validation can be achieved by performing a correlation study between the class posterior probabilities and the widely accepted modified Lindmark version [23], [24] of the Fugl-Meyer (FM) scoring [46], [47].

Figure 6 shows the class posterior probability for the range 6 to 180 days after stroke for one patient who recovered fast. Also shown are the normalized global FM score (normalization is obtained by adding up all items and dividing by the maximum), the normalized sum of the upper extremity of part A (ability to perform active movements), the normalized sum of Part B (ability to perform rapid movement changes). The normalization results in a score between 0 and 1. This normalization makes a comparison between de Fugl-Meyer scale and the posterior probabilities possible, the latter also having values between 0 and 1. In order to quantify the correspondence between posterior probabilities and the subscores of the Fugl-Meyer scale the Pearson correlations are computed.

FIGURE 6 HERE

Table 3 gives an overview of the average correlation of the 16 selected patients from section 2.1 from the total of 57 patients.

TABLE 3 HERE

The posterior probability profiles are obtained with KDE on the 6 selected features to compute the results in table 3. The correlations are computed between the posterior class probabilities and Lindmark modified Fugl-Meyer scoring.

The Lindmark modified Fugl-Meyer scale is divided into 7 domains: ability to perform active movements, ability to perform rapid movement changes, mobility, balance, sensation, joint pain and joint motion. Different physical tests are performed which focus on different aspects of motor recovery and on different parts of the body:

- Total Fugl-Meyer score: this is the sum of all subscores,
- Part A: ability to perform active movements, in case of the ‘upper extremity’ subsection, 3 parts of the body are tested:
    - Upper extremity part of the body:
        - Arm,
        - Wrist,
        - Hand function,

    - Lower extremity part of the body,
- Part B: ability to perform rapid movement changes,
- Part C: mobility,
- Part D: balance,
- Part E: sensation,
    - Light touch,
    - Joint position,
- Part F: joint pain,
- Part G: joint motion:
    - Upper extremity part,
    - Lower extremity part.

The exercises and body parts that have the highest degree of correlation with the class posterior probability are shown in Table 4 (Part A).

TABLE 4 HERE

In each of the parts (A, B, D, E, F, G) a patient has to perform some small physical tests. The performance of the patient is supervised by a clinician who assigns a score to the tests. The score is typically a discrete value: e.g. in 'Part A', 'Part B', 'Part C' and 'Part D' most tests are given a score from 0 to 3, while in 'Part E', 'Part F' and 'Part G' most tests are given a score from 0 to 2: 0 = cannot perform, 1 = performs partially, 2 = performs fully.
The average correlation scores have been obtained by averaging the correlations of 6, 31, 56, 81, 106, 131, 156 and 180 days after stroke, between the posterior class probabilities (obtained from the top 6 features) and the scoring by a clinician.
It should be noted that in particular the average correlations with the global FM score and the 'upper extremity' of part A (ability to perform active movements) score are high.

3.3. Recovery patterns

Longitudinal studies using repeated measurements over time indicate that recovery after stroke shows a non-linear pattern as a function of time and is determined by certain not yet fully understood biological processes, leaving aside why damage in certain brain areas offers better recovery perspectives than others [48]. This means that the outcome of patients with stroke is heterogeneous by nature and individual recovery patterns differ. Nevertheless, some regularities have been found in patients' functional recovery. Curve-fitting of time series of the Barthel index (BI) with a logistic model recorded in patients with a first-ever MCA stroke revealed that patients having greater improvements within the first weeks post-stroke reached higher plateaus at six months than those with later BI improvements [49]. These results are very similar to an older study [50], also using the BI, demonstrating the importance of an early rapid phase of recovery for the final functional outcome and in line with the high correlation between the admission motor–FIM (Functional Independence Measure) and discharge FIM [51]. Our results are comparable with the former references in that the time curve of the class posterior probability evenly shows the importance of a fast recovery in the first weeks for obtaining a high recovery plateau. In figure 6 it can be seen that the recovery measured by means of the posterior probability in the first few weeks is very effective and that a plateau with high normality is achieved after approximately two months. A similar observation is made for the global FM score and the 2 subscores shown in figure 6. Figure 7 also confirms the need of an initial recovery in the first few weeks to obtain a high recovery plateau.

FIGURE 7 HERE

It shows the class posterior probability over time of a patient having no initial recovery, only showing some changes after 2 months. However, the level of functionality reached at month 6 is very low. A similar pattern is observed for several subscores of the FM assessment in figure 7. Given the strong correlation of the class posterior probability with the FM, stages in its time course might predict the future outcome of patients in analogy with [49]. In the previous reference the authors showed that, based on the FM scores of the flaccid arm, prediction of arm function outcome at 6 months could be made within 4 weeks after onset. We admit that prediction of outcome in functional recovery based on the posterior probabilities is still somewhat speculative given the limited number of involved patients, but the strong correlations with several FM subscores are promising. Whether month 6 is a breaking point or not in recovery and what can still be expected in the subsequent months is an ongoing scientific debate without clear conclusions yet [48].

4. Discussion

4.1. Correlation studies

In [11] references are made to previous correlation studies between FM and other scales. The correlation between FM score and ADL capacity has been studied in [47]. The correlation between FM motor scale and the ADL total score was 75% six months after discharge from the hospital. In [52] the correlation between FM and the BI index (Barthel index disability measure) was studied. The correlation scores that have been reported there between the Fugl-Meyer upper extremity motor subscore and the BI is 75% in the acute stage and 82% after 5 weeks. The average correlation result of 80.29% in our experiments between the same subscore and the posterior probability is approximately as high as those correlations in [52] and hence provide in that sense a satisfying degree of correlation. However, it has to be noted that Fugl-Meyer and Barthel index scoring need to be performed by trained clinicians. The Fugl-Meyer scoring easily takes 30 minutes to administer which may limit its widespread applicability in clinical practice [52]. Installing patients in the platform, executing the measurement protocol and computing the posterior takes 5 to 10 minutes. The posterior probability profiles hence allow to assess motor recovery for some limited subscores from the Lindmark modified Fugl-Meyer score with an equal performance (degree of correlation) as would be achieved with some other established supervised scoring techniques.

The high correlation between the posterior probability profile and the ‘upper part A’ can be expected: the selected ADL tasks and sensors concentrated mainly on the use of the hand, the upper arm and the wrist. As opposed to the Fugl-Meyer scoring, the ADL tasks intend to measure more directly the performance of a patient in some tasks which are important in daily living. A low Fugl-Meyer score on ‘Part E’ sensation does not necessarily exclude whether a patient can still perform a task. However, in order to make any quantitative assessment possible a correlation with existing quantitative techniques needed to be established.

In [53] a correlation study between several movement smoothness parameters and the change in global Fugl-Meyer score has been performed. The movement smoothness parameters were computed from the movement of a robotic arm that was displaced by patients with stroke. This robot arm restricted patient’s arm movement to a horizontal plane. Such a system could in principle also be used to automate assessment of functional recovery. The largest correlations the authors found was with the jerk parameter, which had a correlation of -0.48 with the FM, and the speed parameter of movement which had a correlation of 0.40 with the FM. As opposed to that research, our research considered isometric measurements with six degrees of freedom per sensor. Rather than restricting ourselves to a single parameter, we integrated the six most discriminative features in the computation of the posterior probabilities. Estimation of functional recovery is thus obtained in a multi-dimensional space.

This may explain the higher correlation between the posterior probability and the global FM score. It is remarkable that we obtain such high correlations in an automated system: we do not require any scoring of a clinician in the computation of the posterior probabilities.

The correlations with the different exercises of the FM scoring, examples to test the arm, wrist and hand function are given in table 4, were also calculated. The highest possible correlations between the posterior probabilities and single exercises were found for the ‘hook grasp’, the description is given in exercise 4 of the hand function in table 4, and the ‘lateral grasp’, exercise 5 of the hand function in table 4. The correlations for the ‘hook grasp’ and the ‘lateral grasp’ are respectively 83.46% ± 8.95 and 82.86% ± 8.84. These two exercises form together with five other exercises part A of the hand function, table 4. These correlations are higher than the correlation between the posterior probabilities and part A of the hand function which was equal to 78.0% ± 9.65 and given in table 3.

4.2. Discussion recovery patterns

Fast recovery patterns are rather reflections of strong spontaneous neurological improvement and might comprise the ‘milder’ strokes. Having this information shortly after hospitalization is important. Patients having capabilities to recover without special physiotherapeutic intervention can be early discharged. Several mechanisms have been found to underlie functional recovery. Indeed, the spontaneous recovery in the first few days after stroke has been attributed to edema resolution and the reperfusion of ischemic penumbra, while recovery afterwards has been attributed to brain plasticity [48], [54]. Several mechanisms are likely to be involved in brain plasticity and might be reflected through the variable slope followed by a plateau phase, expressing the way how meaningful brain signals are again combined and reconnected [55]. A time interval without clear changes initially after the stroke, as in figure 7, may reflect ‘intrinsic cerebral damage’ and is an important predictor of poor outcome [49].

In our research we discovered several important features, all contributing to the ‘process’ of recovery. Though some of them such as ‘trajectory planning’, ‘continuity in voluntary movement’ and ‘velocity components’ might encompass the discrete building blocks of human movement described in [56], it is impossible until now to link them in a consistent way to the interplay between different brain areas. This means that a thorough understanding of recovery curves from a neurophysiologic point of view needs research that correlates data, similar to those described in this paper, with advanced measurement techniques such as functional magnetic resonance imaging (fMRI), transcranial magnetic stimulation (TMS) and computed electroencephalography EEG.

4.3. Future research

A question that has not been treated in our research so far is whether the mechatronic platform could be used to train patients in order to order to enhance their functional recovery. This would require the design of a protocol with an appropriate visual feedback to the patients of their performance [57]. An isometric torque measurement protocol in [57] has been shown successful in reducing abnormal torque patterns in patients with stroke. However, it has not been shown so far whether this would lead to increased performance in activity of daily living tasks. This still remains a topic for future research.

5. Conclusion

Posterior probability profiles are proposed as a new quantitative and statistically sound measure for the automation of the assessment of functional recovery of patients with stroke in activity of daily living tasks. The posterior probabilities take the performance of both normal controls and patients with stroke into account to select the discriminative features and sensors. It is shown that the posterior probability profiles have a high correlation of 76.6% and 80.29% with the global score of the Lindmark modified Fugl-Meyer scale and 'Part A' (upper extremity subscore) respectively. This degree of correlation is as high as obtained with supervised scoring techniques such as the Barthel index. This allows for an automated assessment of the functional recovery in a more time-effective way than by means of the traditional manual scoring by trained clinicians.

An important prerequisite is to reduce the number of features. This is solved by means of a hybrid filter-wrapper approach. A statistically optimal filter allows to reduce the feature set from 59472 features to 2637 features. The wrapper search on this smaller set reduces the feature set to a total of 6 features.

The wrapper part of the feature selection approach simplifies the experimental set-up from 8 to 4 sensors (ignoring the non-discriminative sensors) hence, reducing the cost of the set-up.

Comparison between different induction algorithms reveals that the best results in discriminating normal controls from patients with stroke is obtained by kernel density estimation, achieving a recognition rate of 98.25%.

The posterior probability profiles confirm recovery patterns that have been observed with human supervised assessment techniques. A significant recovery within the first few weeks is necessary to obtain a high level of normality after 6 months. A lack of recovery within the first few weeks on the other hand results in poor recovery after 6 months, observed as a low level of normality.

Acknowledgments

This work was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

The first author is supported by the CREA Financing (CREA/07/027) program of the K.U.Leuven. The second author was funded by the European Commission (IST-2002-507424). The third author is supported by research grants received from the Excellence Financing (EF 2005) program of the K.U.Leuven, the Belgian Fund for Scientific Research -- Flanders (G.0248.03 and G.0234.04), the Interuniversity Attraction Poles Programme -- Belgian Science Policy (IUAP P5/04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), and the European Commission (NEST-2003-012963, STREP-2002-016276, IST-2004-027017, and IST-2007-217077).

The mechatronic platform was developed by Mazzoleni et al. [19] in the ALLADIN project. We are grateful to these authors for their collaboration. The ALLADIN project is funded by the European Commission under the 6th Framework Programme, IST Programme, Priority 2.3.1.11 – eHealth, IST Contract No.: IST-2002-507424.

The experiments comply with the national laws and local ethical committees.

References

[1] WHO MONICA project principal investigators. The World Health Organization MONICA project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. J Clin Epidemiol 1988;41:105–14.

[2] Rosamond W, Flegal K, Furie K, Go A, Greenland K, Haase N, et al. Heart disease and stroke statistics - 2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2008;117:e25–146.

[3] Centers for Disease Control and Prevention. Prevalence of disabilities and associated health conditions among adults: United States, 1999. MMWR 2001;50:120–5.

[4] Kelly-Hayes M, Beiser A, Kase CS, Scaramucci A, D'Agostino RB, Wolf PA. The influence of gender and age on disability following ischemic stroke: the Framingham study. J Stroke Cerebrovasc Dis 2003;12:119–26.

[5] Pohjasvaara TI, Jokinen H, Ylikoski R, Kalska H, Mäntylä R, Kaste M, Erkinjuntti T. White matter lesions are related to impaired instrumental activities of daily living poststroke. J Stroke Cerebrovasc Dis 2007;16:251–258.

[6] Diringer MN, Edwards DF, Mattson DT, Akins PT, Sheedy CW, Hsu CY, et al. Predictors of acute hospital costs for treatment of ischemic stroke in an academic center. Stroke 1999;30:724–8.

[7] Truelsen T, Piechowski-Jóźwiak B, Bonita R, Mathers C, Bogousslavsky J, Boysen G. Stroke incidence and prevalence in Europe: a review of available data. Eur J Neurol 2006;13:581–98.

[8] Truelsen T, Ekman M, Boysen G. Cost of stroke in Europe. Eur J Neurol 2005;12:78–84.

[9] Erol D, Sarkar N. Coordinated control of assistive robotic devices for activities of daily living tasks. IEEE Trans Neural Syst Rehabil Eng 2008;16:278–85.

[10] Volpe BT, Lynch D, Rykman-Berland A, Ferraro M, Galgano M, Hogan N, Krebs HI. Intensive sensorimotor arm training mediated by therapist or robot improves hemiparesis in patients with chronic stroke. Neurorehab Neural Re 2008;22:305–10.

[11] Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehab Neural Re 2002;16:232–40.

[12] Kasner, SE. Clinical interpretation and use of stroke scales. Lancet Neurol 2006;5:603–12.

[13] Gottrup C, Thomsen K, Locht P, Wu O, Sorensen AG, Koroshetz WJ, Østergaard L. Applying instance-based techniques to prediction of final outcome in acute stroke. Artif Intell Med 2005;33:223–36.

[14] Grond M, Stenzel C, Schmülling S, Rudolf J, Neveling M, Lechleuthner A, et al. Early intravenous thrombolysis for acute ischemic stroke in a community-based approach. Stroke 1998;29:1544–9.

[15] Taub E, Miller NE, Novack TA, Cook EWD, Fleming WC, Nepomuceno CS, et al. Technique to improve chronic motor deficit after stroke. Arch Phys Med Rehabil 1993;74:347–54.

[16] Taub E, Uswatte G, Pidikiti RD. Constrained-induced movement therapy: a new family of techniques with broad application to physical rehabilitation: a clinical review. J Rehab Res Dev 1999;36:237–51.

[17] Miltner WH, Bauder H, Sommer M, Dettmers C, Taub E. Effects of constrained-induced movement therapy on patients with chronic motor deficits after stroke: a replication. Stroke 1999;30:586–92.

[18] Stewart KC, Cauraugh JH, Summers JJ. Bilateral movement training and stroke rehabilitation: a systematic review and meta-analysis. J Neurol Sci 2006;244:89–95.

[19] Mazzoleni S, Van Vaerenbergh J, Toth A, Munih M, Guglielmelli E, Dario P. ALLADIN: a novel mechatronic platform for assessing post-stroke functional recovery. In: Proceedings of the 9th international conference on rehabilitation robotics. Chicago, Illinois: IEEE; 2005. p. 156–9.

[20] Van Vaerenbergh J, Mazzoleni S, Toth A, Guglielmelli E, Munih M, Stokes E, et al. Assessment of recovery at stroke patients by whole-body isometric force-torque measurements of functional tasks I: mechanical design of the device. In: Proceedings of the 3rd European medical and biological engineering conference. Prague, Czech Republic: IFMBE; 2005.

[21] Dewald JP, Beer RF. Abnormal joint torque patterns in the paretic upper limb of subjects with hemiparesis. Muscle Nerve 2001;24:273–83.

[22] Koo TK, Mak AF, Hung LK, Dewald JP. Joint position dependence of weakness during maximum isometric voluntary contractions in subjects with hemiparesis. Arch Phys Med Rehabil 2003;84:1380–6.

[23] Lindmark B, Hamrin E. Evaluation of functional capacity after stroke as basis for active intervention. Presentation of a modified chart for motor capacity assessment and its reliability. Scand J Rehabil Med 1988;20:103–9.

[24] Lindmark B, Hamrin E. Evaluation of functional capacity after stroke as a basis for active intervention. Validation of a modified chart for motor capacity assessment. Scand J Rehabil Med 1988;20:111–5.

[25] Van Dijck G, Van Hulle MM. Speeding up feature subset selection through mutual information relevance filtering. In: Kok JN, Koronacki J, López de Mántaras R, Matwin S, Mladenic D, Skowron A, editors. In: Proceedings of the 11th European conference on principles and practice of knowledge discovery in databases. Warsaw, Poland: Springer-Verlag; 2007. p. 277–87.

[26] Van Dijck G, Van Vaerenbergh J, Van Hulle MM. Posterior probability profiles for the automated assessment of the recovery of stroke patients. In: Proceedings of the 22nd AAAI conference on artificial intelligence. Vancouver, Canada: AAAI Press; 2007. p. 347–53.

[27] Perfetti C. Der hemiplegische patient. Kognitiv-therapeutische übungen. München, Germany: Richard Pflaum Verlag GmbH & Co; 1997.

[28] Carr JH, Shepherd RB. Neurological rehabilitation: optimizing motor performance. Oxford, UK: Butterworth-Heinemann; 1998.

[29] Charles PD, Esper GJ, Davis TL, Maciunas RJ, Robertson D. Classification of tremor and update on treatment. Am Fam Physician 1999;59:1565–72.

[30] Hettinger T, Muller EA. Progress of increase of muscle power after a single maximum training stimulus. Int Z Angew Physiol 1956;16:184–91.

[31] Van Dijck G, Van Hulle MM, Van Vaerenbergh J. A novel criterion for onset detection: differential information redundancy with application to human movement initiation. In: Proceedings of the 7th IEEE international conference on data mining. Omaha, Nebraska: IEEE Computer Society; 2007. p. 673–8.

[32] Dewald JP, Pope PS, Given JD, Buchanan TS, Rymer WZ. Abnormal muscle coactivation patterns during isometric torque generation at the elbow and shoulder in hemiparetic subjects. Brain 1995;118:495–510.

[33] Duda RO, Hart PE, Stork DG. Pattern Classification, second edition. New York, USA: John Wiley & Sons; 2001.

[34] Lum PS, Burgar CG, Kenney DE, Van der Loos HFM. Quantification of force abnormalities during passive and active-assisted upper-limb reaching movements in post-stroke hemiparesis. IEEE Trans Biomed Eng 1999;46:652–662.

[35] Lütkepohl H. New Introduction to Multiple Time Series Analysis. Berlin, Germany: Springer-Verlag; 2005.

[36] Cover TM, Thomas JA. Elements of Information Theory, second edition. Palo Alto, USA: John Wiley & Sons; 2006.

[37] Härdle W, Müller M, Sperlich S, Werwatz A. Nonparametric and Semiparametric Models. Berlin, Germany: Springer-Verlag; 2004.

[38] Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell 1997;97:273–324.

[39] Inza I, Larrañaga P, Blanco R, Cerrolaza AJ. Filter versus wrapper gene selection approaches in DNA microarray domains. Artif Intell Med 2004;31:91–103.

[40] Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. Pattern Recogn Lett 1994;15:1119–25.

[41] Koller D, Sahami M. Toward optimal feature selection. In: Saitta L, editors. Proceedings of the 13th international conference on machine learning. Bari, Italy: Morgan-Kaufmann; 1996. p. 284–92.

[42] Battiti R. Using mutual information for selecting features in supervised neural net learning. IEEE T Neural Netw 1994;5:537–50.

[43] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E 2004;96:066138.1–16.

[44] Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. Least Squares Support Vector Machines. Singapore: World Scientific; 2002.

[45] Devroye L, Lugosi G. Combinatorial Methods in Density Estimation. New-York, USA: Springer-Verlag; 2001.

[46] Fugl-Meyer A, Jaasko L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. Scand J Rehabil Med 1975;7:13–31.

[47] Fugl-Meyer A, Jaasko L. Post-stroke hemiplegia and ADL performance. Scand J Rehabil Med Suppl 1980;7:140□52.

[48] Kwakkel G, Kollen BJ, Lindeman E. Understanding the pattern of functional recovery after stroke: facts and theories. Restor Neurol Neurosci 2004;22:281□99.

[49] Kwakkel G, Kollen BJ, van der Grond J, Prevo AJH. Probability of regaining dexterity in the flaccid upper limb. Impact of severity of paresis and time since onset in acute stroke. Stroke 2003;34:2181□6.

[50] Skilbeck CE, Wade DT, Hewer RL, Wood VA. Recovery after stroke. J Neurol Neurosurg Psychiatry 1983;46:5□8.

[51] Sonoda S, Saitoh E, Nagai S, Okuyama Y, Suzuki T, Suzuki M. Stroke outcome prediction using reciprocal number of initial activities of daily living status. J Stroke Cerebrovasc Dis 2005;14:8□11.

[52] Wood-Dauphinee S, Williams J, Shapiro S. Examining outcome measures in a clinical study of stroke. Stroke 1990;21:731□739.

[53] Rohrer B, Fasoli S, Krebs HI, Hughes R, Volpe B, Frontera WR, Stein J, Hogan N. Movement smoothness changes during stroke recovery. J Neurosci 2002;22:8297□304.

[54] Hurtado O, Pradillo JM, Alonso-Escolano D, Lorenzo P, Sobrino T, Castillo J, et al. Neurorepair versus neuroprotection in stroke. Cerebrovasc Dis 2006;21:54□63.

[55] Gerloff C, Bushara K, Sailer A, Wassermann EM, Chen R, Matsuoka T, et al. Multimodal imaging of brain reorganization in motor areas of the contralesional hemisphere of well recovered patients after capsular stroke. Brain 2006;129:791□808.

[56] Dipietro L, Krebs HI, Fasoli SE, Volpe BT, Hogan N. Submovement changes characterize generalization of motor recovery after stroke. Cortex 2008; in press.

[57] Ellis MD, Holubar B, Acosta A, Beer RF, Dewald JP. Modifiability of abnormal isometric elbow and shoulder joint torque coupling after stroke. Muscle Nerve 2005;32:170-8.

Tables

Table 1. Patient data of the group of 16 patients. Country code: B = Belgium, H = Hungary and I = Ireland. Gender code: F = female and M = male. Lesion code: L = left and R = right.

Patient Nr (country)	Gender	Age	Lesion side
1 (H)	F	74	R
2 (H)	M	70	L
3 (I)	M	50	R
4 (I)	F	67	L
5 (H)	M	51	L
6 (H)	M	63	L
7 (B)	M	72	R
8 (I)	M	65	R
9 (I)	M	69	R
10 (I)	M	54	L
11 (H)	F	60	L
12 (B)	F	47	R
13 (B)	F	67	L
14 (I)	M	66	R
15 (B)	F	86	L
16 (B)	M	82	R

Table 2. Results of the comparison of: k-NN, LSSVM and KDE, using leave-one-out validation.

FS Size	k-NN	LSSVM	KDE	Sensor	ADL Task
1	80.70%	82.46%	82.46%	Middle Finger	Drinking
2	88.60%	85.96%	87.82%	Thumb	Lifting and carrying a bottle
3	92.11%	90.35%	92.11%	Seat	Drinking
4	93.86%	93.86%	96.49%	Thumb	Lifting and carrying a bottle
5	94.74%	96.49%	97.37%	Thumb	Drinking
6	94.74%	97.37%	98.25%	Index	Lifting a bag

Table 3. Pearson correlation between Fugl-Meyer subscores and the class posterior probability profile.

Total FM-score	Part A	Upper Part A 80.29% ± 8.56			Part B	Part C	Part D	Part E	Part F	Part G
		Arm	Wrist	Hand						
76.6% ± 4.63	**75.7% ± 6.20**	**78.9% ± 8.51**	**79.6% ± 5.24**	**78.0% ± 9.65**	**72.4% ± 9.55**	**65.4% ± 1.87**	**67.5% ± 6.15**	**52.5% ± 4.84**	**46.4% ±11.09**	**43.0% ± 3.52**

Table 4. Items of upper extremity of part A. Scores are omitted and replaced by '#'.

The Modified motor assessment chart according to Lindmark			
P = paretic side; NP = non paretic side; p = points			
Part A: Ability to perform active movements.			
	Maximum score Part A	**90**	**90**
		Score 0-3	
Upper extremity	Maximum score upper extremity part A	**54**	**54**
Arm	Maximum score Arm	24	24
Sitting on the edge of the bed or on a chair		P	NP
1	Bring the hand to the mouth by bending the elbow and supinating the forearm and touch the lips with the fingers	#	#
2	Bring the hand to the back of the neck by abducting the shoulder, bending the elbow and pronating the forearm	#	#
3	Flexion of the arm to 180 degrees with extended elbow	#	#
4	Abduction of the arm to 180 degrees with extended elbow	#	#
5	Bring the hand to the lateral side of the opposite knee by adduction and inward rotation of the arm, extension of the elbow and pronation of the forearm	#	#
6	Supination of the forearm. For 1-2 p the elbow is flexed 90 degrees. For 3 p the elbow must be extended and the shoulder joint flexed about 45 degrees	#	#
7	Pronation of the forearm. For scoring see 6	#	#
8	Bring the arm around the body and put the back of the hand against the waist	#	#
	Subtotal score	#	#
Wrist	Maximum score Wrist	9	9
For 1-2 p the elbow is supported, for 3p the elbow is unsupported and extended			
1	Dorsiflexion (extension)	#	#
2	Volarflexion (flexion)	#	#
3	Circumduction	#	#
	Subtotal score	#	#

Hand Function	Maximum score Hand	21	21
1	Flexion of all fingers	#	#
2	Extension of all fingers	#	#
3	Opposition of thumb against the tip of the second finger	#	#
4	Hook grasp. Hold around a stick with the metacarpophalangial joint extended and interphalangeal joint flexed	#	#
5	Lateral grasp. A paper is held between the thumb and the lateral side of the second finger. The thumb must be extended and abducted	#	#
6	Pinch grasp. A pen is held between the thumb and the second finger	#	#
7	Cylindrical grasp. A drinking glass is held with the thumb and the second finger opposing each other	#	#
	Subtotal score	#	#
	Maximum score upper extremity		

Figure captions

Figure 1. Mechatronic platform to acquire force and torque signals from different tasks. The positions of the different force and torque sensors are shown.

Figure 2. Example recording of the three forces $F_x(k)$, $F_y(k)$ and $F_z(k)$ from the thumb during the 'drinking a glass' task. The onset of the effort can be seen at approximately 0.5 s, as an increase in the forces.

Figure 3. Force trajectories over time for the 'drinking a glass' task. The trajectory is obtained by linking consecutive end-points of the force vectors. The patient trajectory and normal control trajectory can be distinguished by their smoothness: the normal control force trajectory seems smoother, while the patient's trajectory is less smooth.

Figure 4. Mutual information feature between the thumb and index in the 'turning a key' task. The first 57 subjects are normal controls, the next 57 subjects are patients with stroke. The average value of the feature for the controls is equal to 0.64 (left horizontal line), the average of the patients is equal to 0.35 (right horizontal line).

Figure 5. Histogram of the mutual information test statistic for the mutual information feature between the norm of the thumb and the norm of the index force in 'turning a key' ADL task. The histogram is obtained by performing 1000 permutations of the class label. The vertical bar at approximately 0.12 indicates the 99% percentile. The bar at the right is the actual value of the mutual information test statistic, without permutations. The feature is clearly significant for the mutual information relevance filter.

Figure 6. Class posterior probability profile and subscores of Fugl-Meyer assessment for a subject with fast recovery.

Figure 7. Class posterior probability profile and subscores of Fugl-Meyer assessment for a subject with poor recovery.

Summary:

Objective: Assessing recovery after stroke has been so far a time consuming procedure in which trained clinicians are required. A demand for automated assessment techniques arises due to the increasing number of patients with stroke and the continuous growth of new treatment options. In this study, we investigate the applicability of isometric force and torque measurements in activity of daily living tasks to assess the functional recovery after stroke in an automated way.

Methods and materials: A new hybrid filter-wrapper feature subset technology was developed for a new mechatronic platform with the aim to identify the most important features and sensors that can distinguish normal controls from patients with stroke. We compared three different classification algorithms to make the distinction: k-nearest neighbors, kernel density estimation and least-squares support vector machines. Based on isometric force and torque measurements obtained from 16 patients with a first-ever ischemic or haemorrhagic stroke within the middle cerebral artery (MCA) territory, we computed for each subject the probability to belong to the class of normal subjects. These probabilities were computed during a period of 6 months post-stroke to quantify the level of recovery during this period. The posterior probabilities were validated by means of a correlation study with the Lindmark modified Fugl-Meyer assessment.

Results: Patients with stroke and normal controls could be distinguished with an accuracy of 98.25% by means of kernel density estimation. The posterior probability profiles had a correlation of 76.6% and 80.29% with the global score of the Lindmark modified Fugl-Meyer scale and ‘Part A’, the upper extremity subscore, respectively. This degree of correlation was as high as obtained with supervised scoring techniques such as the Barthel index.

Conclusion: This study shows that the assessment of recovery after stroke can be automated by means of posterior probability profiles due to their high correlation with the Fugl-Meyer assessment. The posterior probability profiles confirm the importance of a recovery within the first weeks after stroke to obtain a higher recovery plateau compared to later changes in recovery.

Figure 1
Click here to download high resolution image

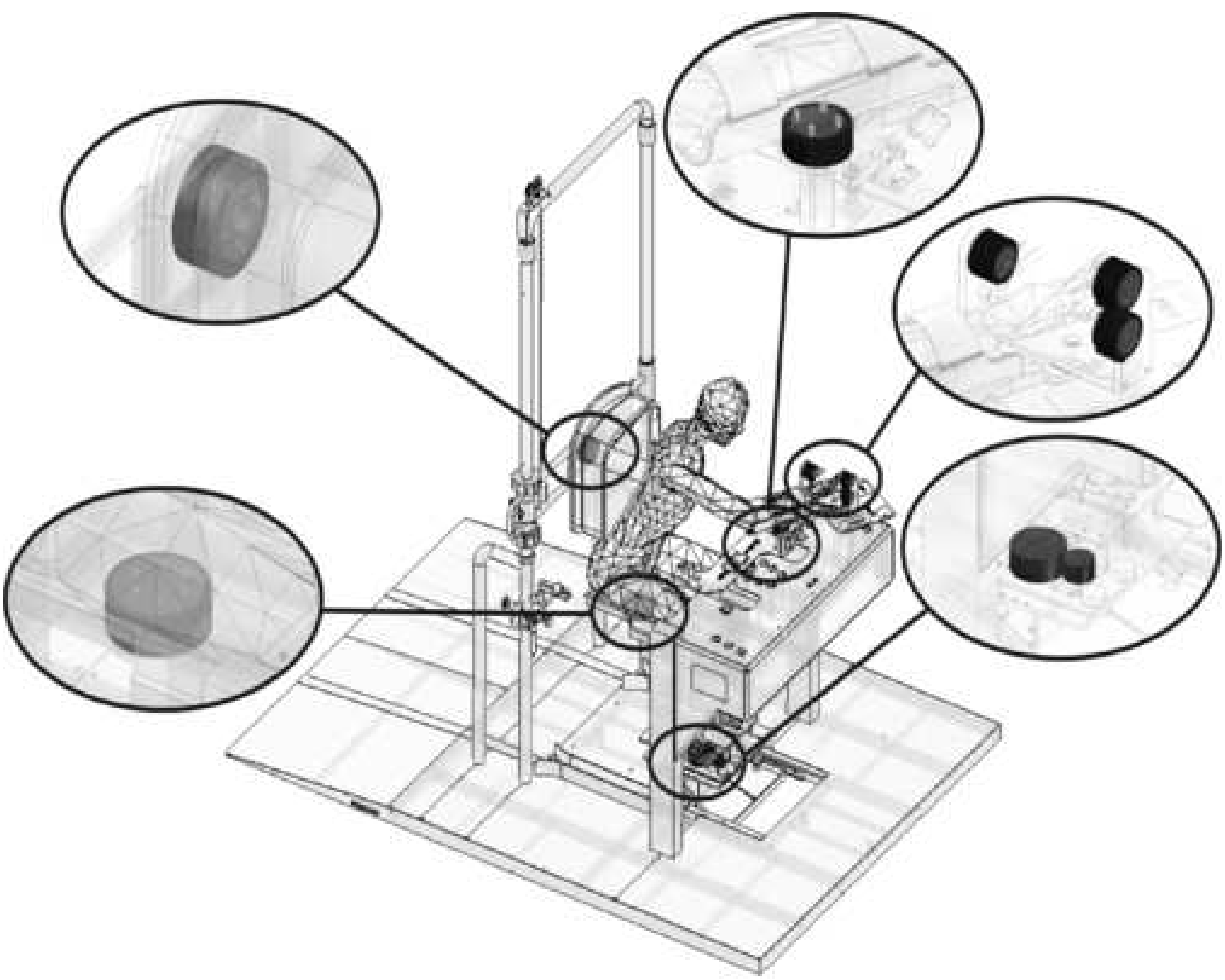

Figure 2
Click here to download high resolution image

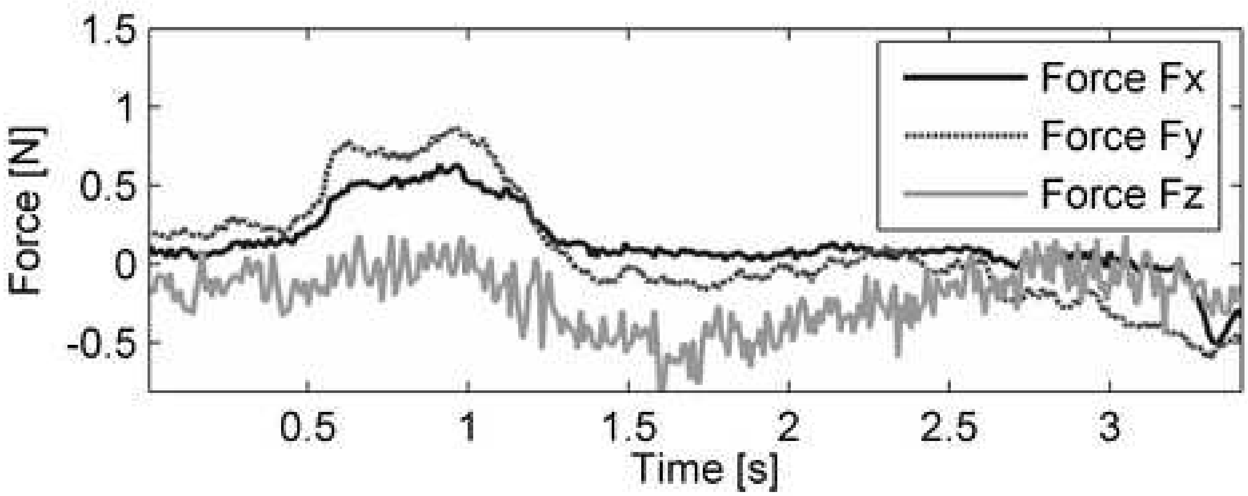

Figure 3
Click here to download high resolution image

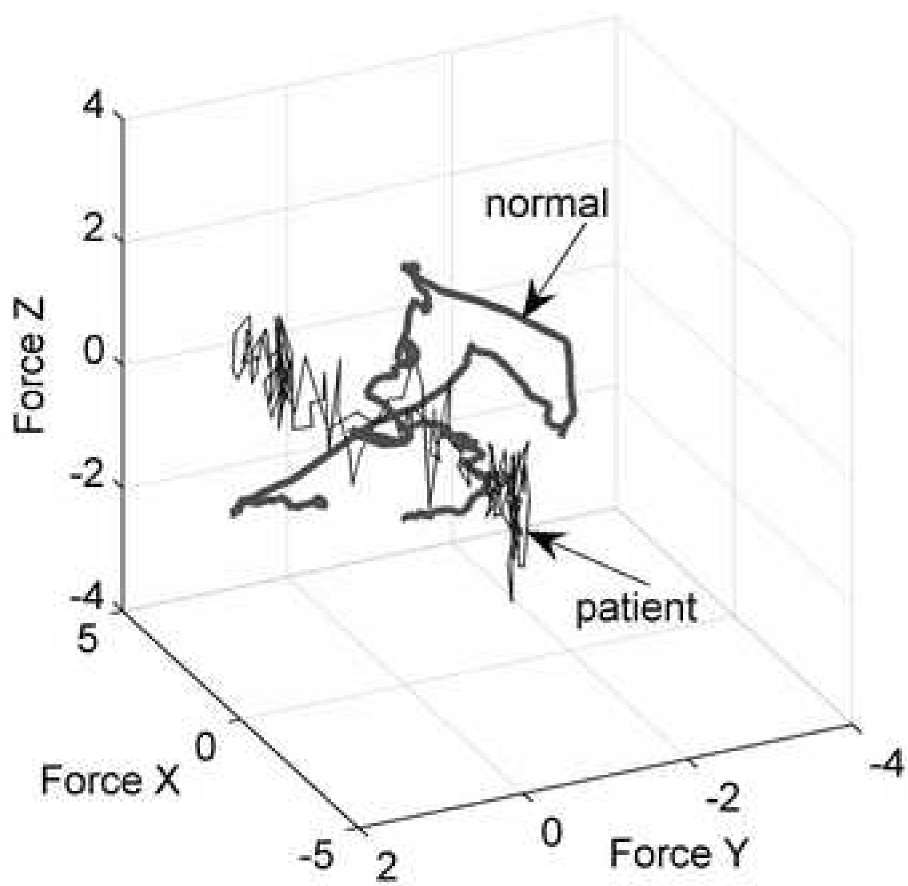

Figure 4
Click here to download high resolution image

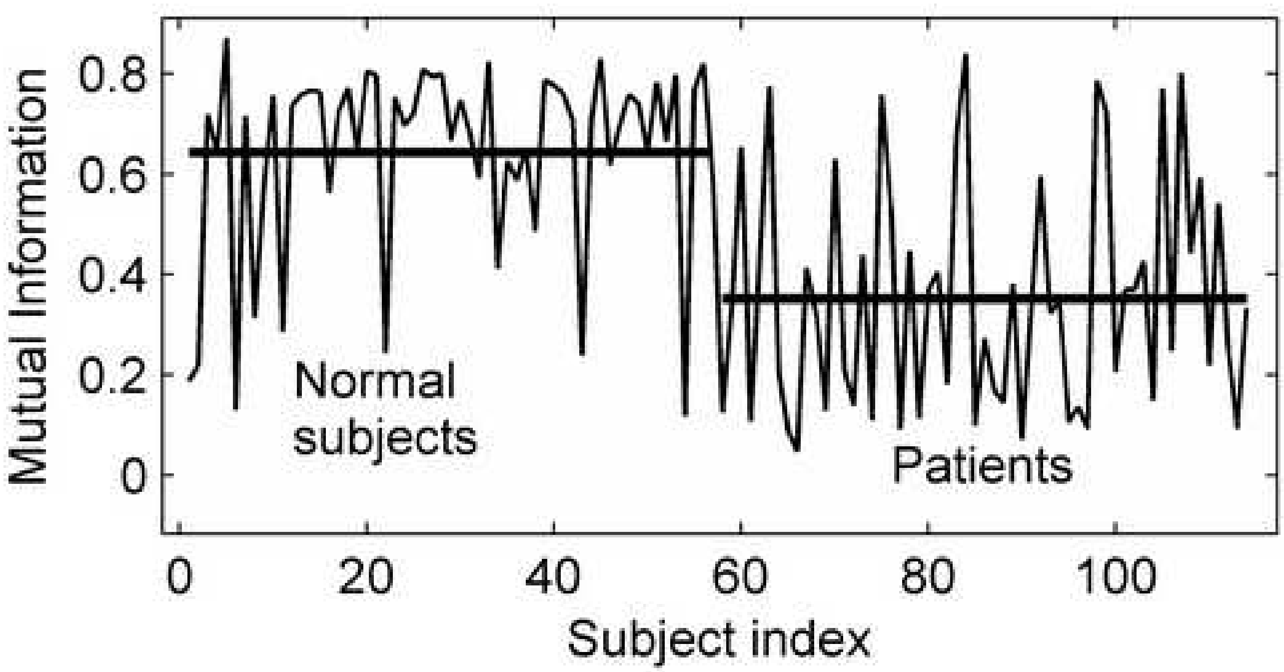

Figure 5
Click here to download high resolution image

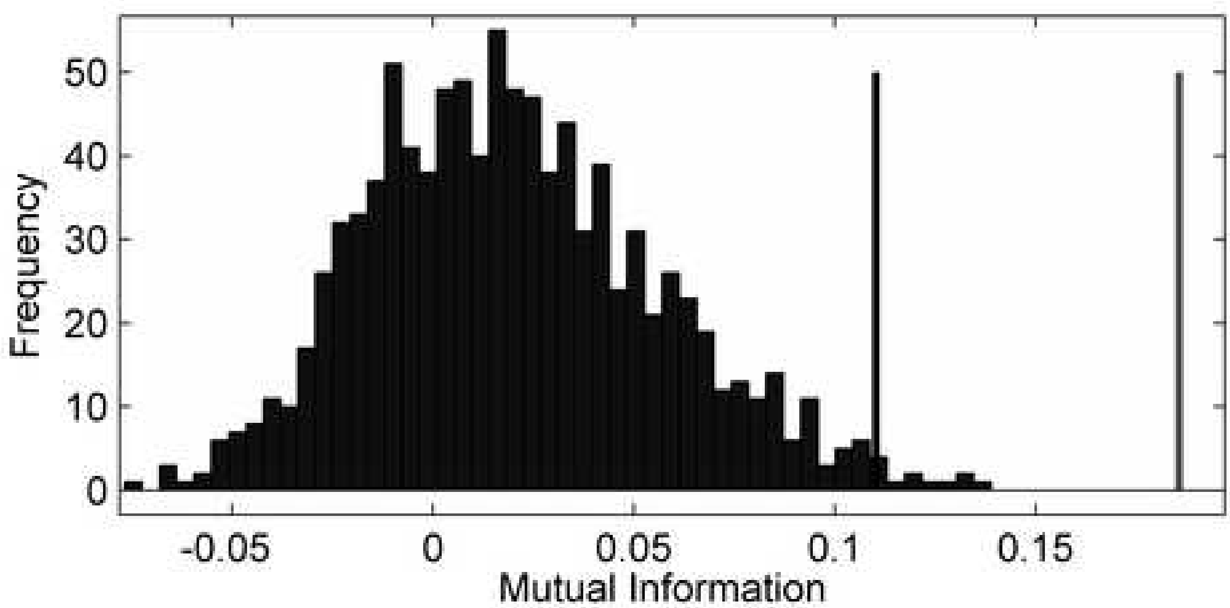

Figure 6
Click here to download high resolution image

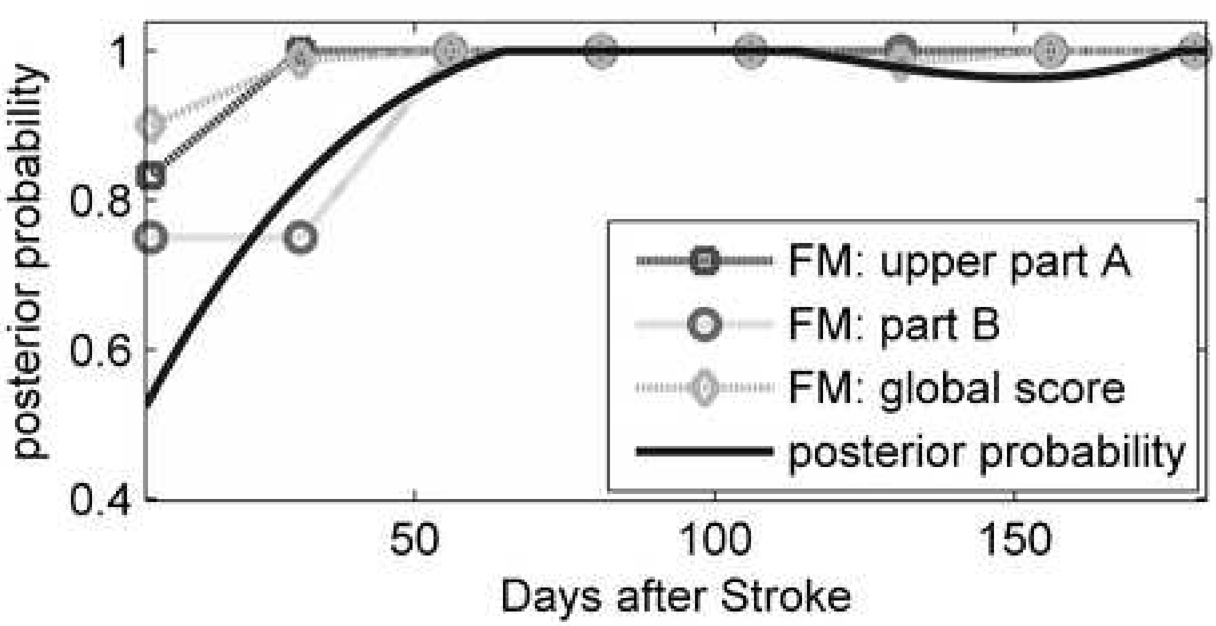

Figure 7
Click here to download high resolution image