Posterior Probability Profiles for the Automated Assessment of the Recovery of Stroke Patients

Gert Van Dijck(1), **Jo Van Vaerenbergh**(2), **Marc M. Van Hulle**(3)

(1), (3)Katholieke Universiteit Leuven, (2)Arteveldehogeschool
(1), (3)Laboratorium voor Neurofysiologie, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
(2)Hoogpoort 15, B-9000, Gent, Belgium
(1)Gert.vandijck@mtm.kuleuven.be, (2)jo.vanvaerenbergh@cmat.be, (3)Marc@neuro.kuleuven.be

Abstract

Assessing recovery from stroke has been so far a time consuming procedure in which highly trained clinicians are required. This paper proposes a mechatronic platform which measures low forces and torques exerted by subjects. Class posterior probabilities are used as a quantitative and statistically sound tool to assess motor recovery from these force and torque measurements. The performance of the patients is expressed in terms of the posterior probability to belong to the class of normal subjects. The mechatronic platform together with the class posterior probabilities enables to automate motor recovery assessment without the need for highly trained clinicians. 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. 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.

Introduction

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 (WHO 1988). Stroke, also known as cerebrovascular accident (CVA) or 'brain attack', ranks 3rd among all causes of death after heart diseases and cancer in the United States (Thom et al. 2006). It is expected to become soon the first cause of death worldwide. Moreover, it is number 1 as a leading cause in long-term disability in the United States (CDCP 1999). Stroke patients 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 (Kelly-Hayes et al. 2003). The symptoms are largely dependent on which part of the brain is affected by the stroke and the size of the affected part.

It is estimated that the direct and indirect cost related to stroke is \$57.9 billion in 2006 in the US (Thom et al. 2006).

Largest contributors in the acute care costs are (Diringer et al. 1999): room charges (50%), medical management (21%) and diagnostic costs (19%).

It is important for rehabilitation specialists to be able to assess recovery and the effects of different physical therapies. This will help in reducing the stay of patients in hospitals and hence has a huge potential in reducing part of all aforementioned costs. Therefore, it is important that the recovery of the patient can be monitored objectively, reliably and cost-effectively. Current techniques require a highly skilled clinician to score the performance of patients in some tasks on specific scales such as the well-established Fugl-Meyer (FM) motor recovery scale However, the scales suffer from inter-rater variability (Gladstone et al. 2002).

Measurements from ADL Tasks

Paramount in the appreciation of the quality of the patients' life is the ability to perform activities of daily living (ADL) tasks. This consists of a set of frequently executed tasks such as: drinking from a glass, turning a key, taking objects, … and so on. These tasks are thoroughly described in textbooks for physical and occupational therapists (Carr and Shepherd 1998) and (Perfetti 1997). In this research 57 stroke patients and 57 normal controls are asked to perform 6 ADL tasks: 'drinking a glass of water', 'turning a key', 'picking up a spoon', 'lifting a bag', 'reaching for a bottle' and 'bringing a bottle to the opposite side'. Every subject is required to execute each task 3 times. During these experiments patients and normal controls are seated in a mechatronic platform. This platform consists of 8 sensors which are connected to different parts of the body, see (Mazzoleni et al. 2005). Some of these body parts are fundamental in executing these tasks: thumb, index, middle finger and the lower arm. Others potentially play a secondary role e.g. with the purpose of balancing, anticipation to a task, synchronization… and so on. These sensors are located at: below the posterior, behind the trunk, a foot and a big toe. The removal of irrelevant sensors, in terms of irrelevant in discriminating stroke patients from normal controls, is addressed later in this paper by means of feature selection procedures.

Simultaneously during execution of these tasks force and torque signals are recorded from the sensors. Forces and torques are recorded both in X, Y, and Z direction. Figure 1 shows 3 times series $F_x(k)$, $F_y(k)$ and $F_z(k)$ during the drinking a glass task.

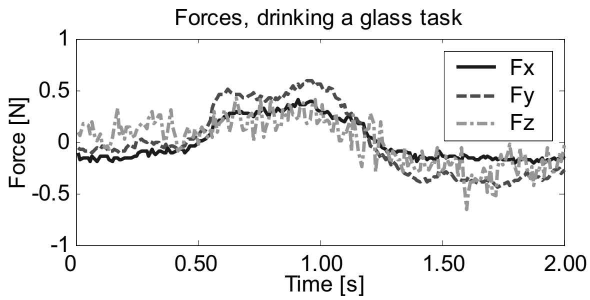

Figure 1: example recording of the three forces $F_x(k)$, $F_y(k)$ and $F_z(k)$ from the thumb during the 'drinking a glass' task. The onset of the effort can be seen at approximately 0.5 s, as an increase in the forces.

The forces and torques are being measured under isometric constraints. The isometric approach implies that objects which are handled by the subjects within the platform cannot be moved. When a patient tries to drink from a cup, the patient has to insert the thumb, index and middle finger within the sensors, which are attached to the cup. The cup itself cannot be moved and is attached to a frame. Hence, the reaction forces will be recorded. The reason for this approach is two-fold: firstly for the ease of recording and secondly to take into account that stroke patients are often still capable of manipulating objects in a manner which is very different from normal subjects. This ability alone, however, is not indicative for the assessment of how 'close' stroke patients perform the task compared to normal subjects. The isometric setting enforces the stroke patients to manipulate the objects in the same way as normal subjects would be required. 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) = 864 measurements in total. Signals are recorded at a sampling rate of 100 Hz. Hence, this implies that recording of 1 s per experiment already leads to 86400 time samples.

Feature Construction

'Raw' time series, in this case force and torque signals, are seldom used for making predictions in pattern recognition systems. It is common to define features based on these times series. Extracting features, defined as functions of the times series, are a first important step to dimensionality reduction. Suppose that we would limit ourselves to fixed windows of 100 samples, using raw time series and retaining only 1 time series out of 864 times series would imply that one would use at least a 100 dimensional space. Secondly, one often disposes of prior knowledge about the difference in behavior between normal controls and the patients. Feature extraction then allows to express one's prior assumptions (or hypotheses) by defining functions of the time series.

It should be noticed that it is natural to consider combinations of X, Y and Z components to construct force vectors in vector space. In figure 2, we show the force trajectories by connecting the end-points of subsequent force vectors **F**(k-1) = [F_x(k-1), F_y(k-1), F_z(k-1)] and **F**(k) = [F_x(k), F_y(k), F_z(k)].

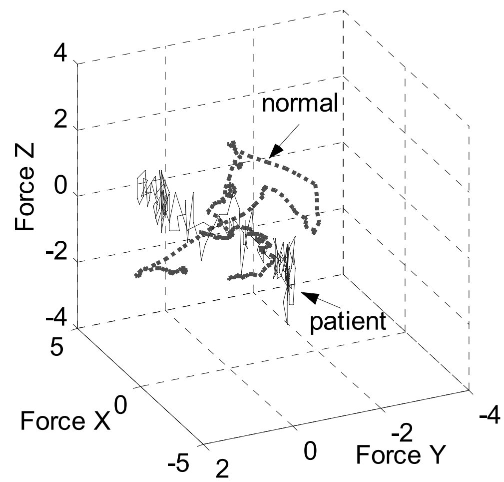

Figure 2: force trajectories over time for the 'drinking a glass' task. The trajectory is obtained by linking consecutive end-points of the force vectors. The patient trajectory and normal control trajectory are discriminative by their smoothness: the normal control force trajectory seems smoother, while the patient's trajectory is less predictive.

Next we describe the construction of features which are determined by the force and the torque vectors.

Planning of a Trajectory. From experience it is clear that normal controls are better capable of 'planning' a trajectory in their efforts, e.g. in trying to bring a glass to the mouth. Hence, it can be hypothesized that the angular deviations from **F**(k) and **T**(k) relative to the mean efforts $\mathbf{F}_m$ and $\mathbf{T}_m$ show larger deviations and abnormalities for patients. These deviations and abnormalities can be assessed e.g. by calculating the maximal deviation, the standard deviation, the skewness, the kurtosis,… of the angular deviations. These statistical measures however, do not take temporal aspects into account. In order to do so, we fit an autoregressive model (AR-model) to these angular deviations with the Bayesian information criterion (BIC) (Lütkepohl 2005) to assess the time lag needed.

This model allows to take linear dependencies into account over time by fitting the current angle deviations based on the previous angle deviations. Moreover, rather than building vectors $\mathbf{F}(k) = [F_x(k), F_y(k), F_z(k)]$ on all 3 signals, similarly vectors based on 2 force and 2 torque signals are 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 $\mathbf{T}(k)$, $\mathbf{T}_{xy}(k) = [T_x(k), T_y(k)]$, $\mathbf{T}_{xz}(k) = [T_x(k), T_z(k)]$ and $\mathbf{T}_{yz}(k) = [T_y(k), T_z(k)]$. These new vector definitions enable us to exclude force or torque signal components that are only noisy signals and are not part of voluntary movements.

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, $\phi[k] = \text{angle}(\mathbf{T}(k),\mathbf{T}(k\text{-}1))$. Due to the partial destruction of feedback loops, the patients' mental representation of the next state, is likely to need bigger corrections when confronted with the sensory feedback (Miall 1993). 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.

Velocity Components. From the force and torque values we can also extract linear velocity and angular velocity respectively:

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

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

However, we need to emphasize that strictly speaking, there is no real movement, while the objects are fixed in the isometric setting. Therefore, these velocity features have no physical meaning. Formula's (1) and (2) would be correct in case of freely moving objects and time independent mass (m) and moment of inertia (I). We call these velocities 'imaginary' linear and 'imaginary' angular velocities. The statistics about the newly obtained angular and linear velocities are summarized in the same way as for the force and torque signals.

Synchronization between Body Parts. The voluntary movement of objects in the ADL tasks needs a careful synchronization, 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 synchronization between forces and torques exerted by the thumb, the index and the middle finger. We compute this synchronization by the information theoretic measure of statistical dependency, known as mutual information (Cover and Thomas 1991). The mutual information can be computed as follows:

$$\begin{aligned}&\text{MI}(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k)\|) = \\ &\max_a \sum_k p(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k\text{-}a)\|) \\ &\ln\left(\frac{p(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k\text{-}a)\|}{p(\|\mathbf{F}_{s1}(k)\|)p(\|\mathbf{F}_{s2}(k\text{-}a)\|)}\right)\end{aligned} \qquad (3)$$

Hence, the mutual information between the forces of sensor s1 and sensor s2 is computed between the magnitudes of these forces. We need to take into account that signals belonging to different sensors can be shifted in time. Therefore, the delay 'a' between the sensors is searched for, such that the magnitudes of the vectors become maximally dependent. Hence, this shift parameter is estimated by means of a maximization of mutual information approach. In figure 3, the maximal mutual information for all 57 normal controls and 57 stroke patients is shown for the norms of the force vector between the thumb and the index in the 'turning a key' task.

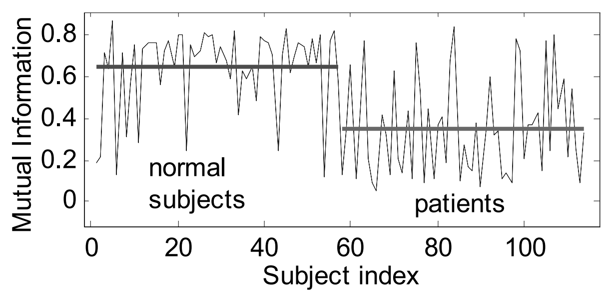

Figure 3: mutual information feature between the thumb and index in the 'turning a key' task. The first 57 subjects are normal controls, the next 57 subjects are stroke patients. The average of the features for the controls is equal to 0.64 (left horizontal line), the average of the patients is equal to 0.35 (right horizontal line).

As indicated in figure 3, the average mutual information for normal controls is higher than for the stroke patients. A t test rejects the null-hypothesis of equal means with $p \approx 2.31 \cdot 10^{-11}$. 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 (less synchronized) for the group of stroke patients. 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 extraction and putting forward hypotheses go hand in hand.

Time Series Fitting. As can be seen in figure 1, the force and torque signals consist of a rising part and a decaying part. This behavior has been modeled by a sum of 2 exponential functions. The parameters are used as features, as well as the residuals obtained after fitting the model.

In summary, the aforementioned definitions lead to a total of 59472 features, and for which we consider the result of applying a given feature definition to each ADL task (6 in total), attempt (3 attempts / task) and sensor (8 sensors) as different features.

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.

Rationale of Class Posterior Probabilities

Given the set of normal controls and patients, one can obtain the conditional probabilities of the features (F_1, F_2, … F_n) for the normal controls and the stroke patients:

$$p(F_1,F_2,...,F_n|\text{normal}) \tag{4}$$

$$p(F_1,F_2,...,F_n|\text{stroke}) \tag{5}$$

These probabilities need to be estimated from measurements on a group of normal controls and stroke patients. Stroke patients should be selected in the phase where they are clearly not fully recovered, otherwise they should be considered as a subject with normal performance.

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

$$p(\text{normal}|F_1,...,F_n) = \frac{p(F_1,...,F_n|\text{normal})p(\text{normal})}{p(F_1,...,F_n)} \tag{6}$$

$$\text{and } p(\text{stroke}|F_1,...,F_n) = 1 - p(\text{normal}|F_1,...,F_n)$$

If the features of a stroke patient are obtained at a certain time instant, say d days after the stroke, ($f_1(d)$, $f_2(d)$, …$f_n(d)$), then (6) 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 stroke patients into account. This is a clear difference with the hypothesized ideal performance of the traditional scaling methods discussed in the introduction. Normal controls perform ADL tasks in a potentially 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 effectively serves as an interface between the clinical expert and the inference engine. In the validation part of this paper, we used the non-informative prior probability: p(normal) = 1/2. This value was set in order to attribute the result to the trained system, rather than the expert.

Feature Subset Selection

It is well known that densities can not be estimated accurately in high dimensional feature spaces. In general, the accuracy is a function of both the number of data points and the dimensionality of the feature space, e.g. see the AMISE (Asymptotic Mean Integrated Square Error) expression in (Härdle et al. 2004) for kernel density estimation. 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 that we want to reduce the number of sensors, hence, to decrease of the cost of the mechatronic device. This is a goal that in general cannot be achieved by feature extraction, because it makes combinations of the original features.

Filter Based Feature Selection. Because quantitative analysis by means of posterior probabilities is central in this paper, we should be able to reduce the high-dimensional feature set to lower dimensions, without the risk of losing important features and without actually calculating the probabilities in high-dimensional spaces, which cannot be performed accurately.

One can prune the feature set in a statistically optimal way by means of the Kullback-Leibler framework for feature subset selection that has been proposed in (Koller and Sahami 1996).

It can be proven that if the following assumptions hold: 1) the features are independent and 2) independent when conditioned on the class label (class conditional independence), then the Kullback-Leibler divergence between the original feature set and a subset thereof is equivalent to the sum of the mutual information contributions from the omitted features. This can be written as follows:

$$\text{if } \forall\, F_1,...,F_n : p(F_1,...,F_n) = \prod_{i=1}^{n} p(F_i)$$

$$\text{and } \forall\, F_1,...,F_n, C : p(F_1,...,F_n|C) = \prod_{i=1}^{n} p(F_i|C)$$

$$\text{then: } KL(p(C|F_1,...,F_n)\,||\,p(C|F_{r_1},...,F_{r_{n1}})) = \sum_{i=1}^{n2} MI(F_{s_i};C) \tag{7}$$

$$\text{with } \{r_1,r_2,...,r_{n1}\} \cup \{s_1,...,s_{n2}\} = \{1,2,...,n\}$$

Here, we represent the class variable by means of variable C, which can e.g. take values -1 and +1 or 0 and +1.

The class conditional independence condition should not come as a surprise: the naive Bayesian (NB) classifier similarly assumes class conditional independence. Although class conditional independence seems a very strong condition, it has been shown that the naive Bayesian classifier easily outperforms many other classifiers (Domingos and Pazzani 1997).

It also has to be remarked that features which are selected based on the mutual information criterion will not necessarily be class conditional independent. This can be easily seen as follows. Suppose that we add a copy of a feature to the feature set. If the original feature is selected based on its high mutual information with the class variable, then the copy will be selected as well, because the mutual information is equal.

From (7), it is clear that eliminating those features for which $MI(F_{si};C) = 0$ leads to the KL divergence = 0, due to the fact that the mutual information is larger or equal to zero. Hence, there is no loss in information by removing those features. We estimated the mutual information by entropies and conditional entropies:

$$\begin{aligned} MI(F_i;C) &= H(F_i) - H(F_i|C) \\ &= H(F_i) - \sum_{j=1}^{2} H(F_i|c_j)p(c_j) \end{aligned} \quad (8)$$

The entropy estimate was obtained by a recently proposed nearest-neighbor approach to entropy estimation, formula (20) in (Kraskov et al. 2004). It has to be noticed that due to the finite sample estimation, $MI(F_i; C)$ is different from 0, even if the feature is statistically irrelevant. Statistical relevance can be easily tested, under the null hypothesis of irrelevance, by random permutation of the class label. 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.

This procedure, which is more thoroughly explained in a regression context (Van Dijck and Van Hulle 2006), allowed us to reduce the original set of 59472 features to a set of 2637 features.

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 (Kohavi et al. 1997) 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. We compare 3 induction algorithms to search for the best subset of features, in terms of discriminating normal controls from stroke patients: k-nearest neighbor, least squares support vector machines (LSSVM, Suykens et al. 2002) and a Bayesian classifier with kernel density estimation (KDE), for KDE see (Devroye et al. 2001). We use Gaussian kernels and use the maximum likelihood cross-validation method for kernel bandwidth estimation in KDE. The hyperparameters of the LSSVM were set to standard values. The LSSVM hyperparameters were then further tuned manually for the best subset in order to increase the performance.

We used the sequential forward search (SFS) procedure, which gradually adds the next best feature to an existing subset, starting from the empty set. From a recent comparison (Oh et al. 2004) among search procedures it can be seen that the difference with more advanced search procedures such as genetic algorithms is rather small, and maximally a few percent, see table 3 in (Oh et al. 2004). Moreover, the more advanced algorithms typically test much more subsets as compared to the SFS, this increases the risk of finding spurious results, due to the larger number of hypotheses being tested.

Comparison from table 1 shows that the best performances are obtained for the Bayesian classifier with KDE (98.25 %). The first column represents the number of features selected, abbreviated as feature subset (FS) size.

FS Size	k-NN	LS-SVM	KDE
1	80.70 %	82.46 %	82.46 %
2	88.60 %	85.96 %	87.72 %
3	92.11 %	90.35 %	92.11 %
4	93.86 %	93.86 %	96.49 %
5	94.74 %	96.49 %	97.37 %
6	94.74 %	97.37 %	**98.25 %**

Table 1: results of the comparison of: k-NN, LS-SVM and KDE, using leave-one-out validation.

In fact, 2 stroke patients were erroneously assigned to the class of normal controls. Performing experiments with more than 6 features for all classifiers until d = 10, did not increase the performance. The resulting set consists of 6 features in which only 4 sensors play a role: thumb, index, middle finger and the seat. This effectively reduces the cost of the platform from 8 to 4 sensors.

The most important tasks are: 'drinking a glass', 'lifting a bag' and 'lifting a bottle'. The most important features are: 'angular velocity', angular deviations from 'continuity in voluntary movement' and the residual from the 'time series fitting'.

Validation

We need to show that the process of defining features, reducing the feature set, followed by the definition of the posterior probabilities on the reduced feature set leads to results that satisfy our goals: the posterior probability should be able to monitor the patient's motor recovery over time. We can achieve this by performing a correlation study between the class posterior probabilities and the broadly accepted, Fugl-Meyer (FM) score.

Figure 4, shows the class posterior probability for the range 6 to 180 days after stroke for one subject that recovered fast after the occurrence of the stroke. Also shown are the normalized global FM score (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). In order to quantify the correspondence between posterior probabilities and the subscores of the Fugl-Meyer scale we compute the Pearson correlations.

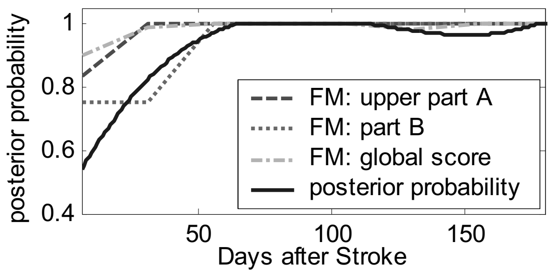

Figure 4: class posterior probability and subsets of Fugl-Meyer scores for one subject.

Table 2 gives an overview of the average correlation for 16 of the 57 patients which were selected based on the constraints of a first-ever ischemic or haemorrhagic stroke and a lesion within the middle cerebral artery (MCA) territory proven by a CT and/or MRI scan.

Global FM score	Upper part A	Part B	Part C	Part D
76.6 % ± 4.63	80.29 % ± 8.56	72.43 % ± 9.55	65.41 % ± 1.87	67.50 % ± 6.15

Table 2: average Pearson correlation scores between posterior probabilities and subscores of the Fugl-Meyer scale.

Patients are between 47 and 87 years old with a mean age of 65 and a standard deviation of 11 years. The posterior probability profiles are obtained with KDE on the 6 selected features to compute the results in table 2. The average correlation scores have been obtained by averaging the correlations after 6, 31, 56, 81, 106, 131, 156 and 180 days after stroke. From table 2 we note that especially the average correlations with the global FM score and the upper extremity of part A (ability to perform active movements) score are high. Part B, C and D are respectively: the ability to perform rapid movement changes, mobility and balance.

In (Gladstone et al. 2002) many references are made to previous correlation studies between FM and other scales e.g. the Barthel index (BI). 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. Our average correlation result of 80.29 % between the same subscore and the posterior probability is approximately as high as those correlations. However, it has to be noticed that Fugl-Meyer and Barthel index scoring need to be performed by trained clinicians and easily take a few hours. Installing patients in the platform, executing the measurement protocol and computing the posterior takes 5 to 10 minutes.

Acknowledgments

The first author is supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

The second author is supported by the Excellence Financing program of the K.U.Leuven (EF 2005), the Belgian Fund for Scientific Research Flanders (G.0248.03, G.0234.04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), the Belgian Science Policy (IUAP P5/04), and the European Commission (NEST-2003-012963, STREP-2002-016276, and IST-2004-027017).

The mechatronic platform is developed within the ALLADIN project.

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.

This work made use of the HPC (High Performance Computing) infrastructure of the K. U. Leuven.

References

Carr, J.; and Shepherd, R. 1998. *Neurological Rehabilitation: Optimizing Motor Performance*. Oxford: Butterworth-Heinemann.

Centers for Disease Control and Prevention (CDCP) 1999. Prevalence of Disabilities and Associated Health Conditions among Adults: United States. *Morbidity and Mortality Weekly Report* 50: 120-125.

Cover, T. M.; and Thomas, J. A. 1991. *Elements of Information Theory*. New York: John Wiley & Sons.

Devroye, L.; and Lugosi, G. 2001. Combinatorial Methods in Density Estimation. New York: Springer-Verlag.

Diringer, M. N.; Edwards, D. F.; Mattson, D. T.; et al. 1999. Predictors of Acute Hospital Costs for Treatment of Ischemic Stroke in an Academic Center. *Stroke* 30(4): 724–728.

Domingos, P.; and Pazzani, M. 1997. On the Optimality of the Simple Bayesian Classifier. *Machine Learning* 29(2-3): 103-130.

Gladstone, D. J.; Danells, C. J.; and Black, S. E. 2002. The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties. *Neurorehabilitation and Neural Repair* 16(3): 232-240.

Härdle, W.; Müller, M.; Sperlich, S.; and Werwatz, A. 2004. Nonparametric and Semiparametric models. Berlin: Springer-Verlag.

Kelly-Hayes, M.; Beiser, A.; Kase, C. S.; et al. 2003. The Influence of Gender and Age on Disability Following Ischemic Stroke: The Framingham Study, *Journal of Stroke and Cerebrovascular Diseases* 12(3): 119-126.

Kohavi, R.; and John, G. H. 1997. Wrappers for Feature Subset Selection. *Artificial Intelligence* 97(1-2): 273 – 324.

Koller, D.; and Sahami, M. 1996. Toward Optimal Feature Selection. In *Proceedings of the Thirteenth International Conference on Machine learning*, 284-292. Bari, Italy: Morgan-Kaufmann.

Kraskov, A.; Stögbauer, H.; and Grassberger, P. 2004. Estimating Mutual Information. *Physical Review E* 69(066138): 1-16.

Lindmark, B.; and Harmin, E. 1988. Evaluation of Functional Capacity after Stroke as Basis for Active Intervention. Presentation of a Modified Chart for Motor Capacity Assessment and its Reliability. *Scandinavian Journal of Rehabilitation Medicine* 20: 103-109

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg: Springer-Verlag.

Mazzoleni, S.; Van Vaerenbergh, J.; Toth, A.; Munih, M.; Guglielmelli, E.; and Dario, P. 2005. ALLADIN: A Novel Mechatronic Platform for Assessing Post-stroke Functional Recovery. In *Proceedings of the 9th International Conference on Rehabilitation Robotics*, 156-159. Chicago, USA: IEEE Press.

Miall, R. C.; Weir, D. J.; Wolpert, D. M.; and Stein, J. F. 1993. Is the Cerebellum a Smith Predictor? *Journal of Motor Behavior* 25(3): 203-216.

Oh, I.-S.; Lee, J.-S.; and Moon, B.-R. 2004. Hybrid Genetic Algorithms for Feature Selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(11): 1424-1437.

Perfetti, C. 1997. *Der Hemiplegische Patient. Kognitiv-Therapeutische Übungen.* München: Richard Pflaum Verlag GmbH & Co.

Suykens, J. A. K.; Van Gestel, T.; De Brabanter, J., De Moor, B.; and Vandewalle, J. 2002. Least Squares Support Vector Machines. Singapore: World Scientific.

Thom, T. et al. 2006. Heart Disease and Stroke Statistics - 2006 Update: A Report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 113(6): e85 – e151.

Van Dijck, G.; and Van Hulle, M. M. 2006. Speeding up the Wrapper Feature Subset Selection in Regression by Mutual Information Relevance and Redundancy Analysis. In *Proceedings of the 16th International Conference on Artificial Neural Networks*, 31-40. Athens, Greece: Springer.

WHO MONICA Project Principal Investigators 1988. The World Health Organization MONICA Project (Monitoring Trends and Determinants in Cardiovascular Disease): a Major International Collaboration. *Journal of Clinical Epidemiology* 41(2): 105-114.