Automatic, unsupervised classification of dyskinesia in patients with Parkinsons Disease.

Noël L. W. Keijsers*, Martin W. I. M. Horstink†, Stan C.A.M. Gielen†
*Dept. Biophysics, University of Nijmegen, NL 6525 EZ Nijmegen, The Netherlands
Email: {noelk, stan}@mbfys.kun.nl
†Dept. Neurology, University of Nijmegen, NL 6500 HB Nijmegen, The Netherlands
Email: M.Horstink@czzoneu.azn.nl

Abstract—One of the characteristic symptoms of patients with Parkinson Disease (PD) is a rigidity of movement. These symptoms disappear after administration of Levodopa. However, the long-term use of levodopa causes involuntary movements (dyskinesia). A proper diagnosis requires an automatic, unsupervised method for the detection and classification of levodopa induced dyskinesia. The main problem, however, is to distinguish dyskinesia from voluntary movements. The aim of this study was to train a neural network for the classification and rating of dyskinesia and to use the trained neural networks to extract parameters, which are important to distinguish between dyskinesia and normal voluntary movements. The neural network has a performance near 97% correct, which greatly improves upon previous methods (typically 75% correct).

I. INTRODUCTION

After several years of levodopa medication, many patients with Parkinsons disease suffer from levodopa induced dyskinesia (LID). Usually, the severity of LID is evaluated using self-report by the patient or using semi-quantitative rating scales during consults. However, self-assessment appears to be unreliable and motor behavior of patients during a consult is not always representative for the behavior in daily life. For these reasons an automatic, unsupervised ambulatory assessment of dyskinesia would be highly useful. Recently, several investigators successfully used an ambulatory accelerometry to monitor (abnormal) activities of patients (Veltink et al., 1996). The main challenge in automatically assessing dyskinesia is to distinguish between dyskinesia and voluntary movements. This requires information about the specific movement features, which distinguish normal voluntary movements from dyskinesia. Several studies have indicated that many movement parameters may be important for such a distinction (Keijsers et al., 2000; Hoff et al., 2001). Therefore, we have used a large number of movement parameters to train a neural network and to investigate the relevance of the various parameters. We have trained a neural network on the input data and used various techniques for data pruning. The aim of the pruning technique was to obtain insight in the various parameters, which allow the detection of LID and the distinction between LID and voluntary movements. This is important for two reasons. The first is that acceptance of a new technique will be easier if physicians do understand why the network is successful. The other reason is that insight in the movement parameters, which underlie pathological behavior, might be valuable for understanding normal motor behavior. For example, several studies have shown that angular velocities in elbow and shoulder are highly correlated in normal aiming movements of the hand (Soechting et al., 1986; Gielen et al., 1997). This has been interpreted as evidence for the existence of specific muscle synergies in human motor control. It would be interesting to investigate whether and to what extent muscle synergies are also observed in LID. In summary, the purpose of this study was to explore the behavior of neural networks in the detection and rating of dyskinesia and to describe the relevant movement parameters and their relation to the severity of LID.

II. METHODS

Thirteen patients with Parkinsons disease (8 male and 5 female; mean age 61 years (range between 48 and 71)) participated in this study. The patients were on levodopa medication for about fifteen years and all patients suffered from LID. The study was approved by the Medical Ethical Committee of the University Medical Center of the University of Nijmegen. The patients were continuously monitored for a period of approximately 2.5 hours. During the 2.5 hour monitoring session, the patients performed about 35 functional daily-life activities, like walking, putting on a coat, making coffee, preparing lunch, eating, taking off their shoes, reading a newspaper, drinking coffee and washing hands. The order of the activities was randomized between subjects. Subjects were allowed to do the activities in their own way and at their own pace.

The movements and postures were measured using 3-D accelerometers (at both upper arms, at both upper legs, at the wrist of the most dyskinetic side, and at the top of the sternum) and a portable data recorder. The accelerometer signals were digitally stored on a recorder for off-line analysis. The data were also used to rate the severity of LID on the modified AIMS-scale (a five-point scale with a value between 0 (no dyskinesia) and 4 (extreme dyskinesia)) by two experienced physicians, independently. The neural network was trained using these variables as input and the rating scores given by the physicians as output. An overview of the parameters is shown in Table 1. The first 9 variables in the table were calculated digitally stored on a recorder for off-line analysis. The data were also used to rate the severity of LID on the modified AIMS-scale (a five-point scale with a value between 0 (no dyskinesia) and 4 (extreme dyskinesia)) by two experienced physicians, independently. The neural network was trained using these variables as input and the rating scores given by the physicians as output. An overview of the parameters is shown in Table 1. The first 9 variables in the table were calculated
The performance of the network was evaluated using the mean square error (MSE) between the neural network output and the score given by the physicians. In addition, the percentage of correctly classified signals by the neural network had one hidden unit and required 12 input parameters to reach a correct classification performance up to about 75% correct. Figure 1 shows typical results of the performance by the neural network on a test-set for one patient in a 81 minute interval. On the five-point AIMS-scale the neural network closely follows the ratings by the physicians.

For movements of the trunk, the best performing neural network architecture was cross-correlation parameters between movements of the 6 body segments added another 36 input parameters, adding up to a total of 92 input parameters. The neural network used in this study was a MultiLayer Perceptron (MLP) using backpropagation. The inputs for the MLP are shown in Table 1. There was one output unit for each body segment, the value of which reflects the severity of LID of that body segment. This segment could be the most dyskinetic arm, the trunk, or the most dyskinetic leg. The output of the units in the hidden layer was given by a hyperbolic tangent sigmoid transfer function that gives a value between 1 and +1. The output of the unit in the output layer was given by a linear transfer function and had a value in the range between 0 and 4 reflecting the AIMS score.

The performance of the network was evaluated using the mean square error (MSE) between the neural network output and the score given by the physicians. In addition, the percentage of correctly classified signals by the neural network was used as a second criterion to evaluate the performance of the network. Since physicians rate dyskinesia by integers in the range between zero and four, the neural network classification was seen as correct when the difference between the neural network output and the score given by the physicians was smaller than 0.5. The generalization performance of the network was tested by training the network with 80% of the data-set and testing the network with the remaining 20% of the data. This was done 50 times for randomly selected sets of training and test-sets. The optimal architecture of the network was seen as the network, which gave on average the smallest mean square error (MSE) on the test-set for the 50 randomly selected sets. The performance was investigated for various neural network architectures. For each number of hidden units the procedure of forward selection (Laar et al., 1999) was used to find the most valuable input variables to the neural network. Forward selection means, that we started with an empty set of variables, and add, one after another, the variable which causes the largest reduction of the mean square error between the neural network output and the score given by the physicians. After each step we look for the next most important variable, etc. If the performance by a neural network is measured by means of the squared error \( E = \frac{1}{N} \sum_{i=1}^{P} (T^\mu - O^\mu)^2 \), with \( T^\mu \) and \( O^\mu \) the target and output of the network for pattern \( \mu \), \( P \) the number of input patterns and with \( T^\mu - O^\mu \) normally distributed with a standard deviation \( \sigma \), the hypothesis that the performance of two networks is identical has to be rejected with significance \( \alpha \) if

\[
p(Z < \frac{\sqrt{P}E_1 - E_2}{\sigma}) < \alpha
\]

The neural network correctly classified dyskinesia or the absence of dyskinesia in 93.7, 99.7 and 97.0% of the time for the arm, trunk and leg, respectively. This performance was considerably better than that of previous studies, which reached a classification performance up to about 75% correct.

III. Results

The performance of the network was evaluated using the mean square error (MSE) between the neural network output and the score given by the physicians. In addition, the percentage of correctly classified signals by the neural network was used as a second criterion to evaluate the performance of the network. Since physicians rate dyskinesia by integers in the range between zero and four, the neural network classification was seen as correct when the difference between the neural network output and the score given by the physicians was smaller than 0.5. The generalization performance of the network was tested by training the network with 80% of the data-set and testing the network with the remaining 20% of the data. This was done 50 times for randomly selected sets of training and test-sets. The optimal architecture of the network was seen as the network, which gave on average the smallest mean square error (MSE) on the test-set for the 50 randomly selected sets. The performance was investigated for various neural network architectures. For each number of hidden units the procedure of forward selection (Laar et al., 1999) was used to find the most valuable input variables to the neural network. Forward selection means, that we started with an empty set of variables, and add, one after another, the variable which causes the largest reduction of the mean square error between the neural network output and the score given by the physicians. After each step we look for the next most important variable, etc. If the performance by a neural network is measured by means of the squared error \( E = \frac{1}{N} \sum_{i=1}^{P} (T^\mu - O^\mu)^2 \), with \( T^\mu \) and \( O^\mu \) the target and output of the network for pattern \( \mu \), \( P \) the number of input patterns and with \( T^\mu - O^\mu \) normally distributed with a standard deviation \( \sigma \), the hypothesis that the performance of two networks is identical has to be rejected with significance \( \alpha \) if

\[
p(Z < \frac{\sqrt{P}E_1 - E_2}{\sigma}) < \alpha
\]

The neural network correctly classified dyskinesia or the absence of dyskinesia in 93.7, 99.7 and 97.0% of the time for the arm, trunk and leg, respectively. This performance was considerably better than that of previous studies, which reached a classification performance up to about 75% correct. Figure 1 shows typical results of the performance by the neural network on a test-set for one patient in a 81 minute interval. On the five-point AIMS-scale the neural network closely follows the ratings by the physicians.

For movements of the trunk, the best performing neural network had one hidden unit and required 12 input parameters to reach a correct classification performance of over 97%, indicating that a linear classification was sufficient to assess the severity of dyskinesia for the trunk. The most important parameter for the classification of movements appears to be the percentage of time that the trunk is moving in a one-minute interval (%V_{trunk}). This parameter adds 32.4% to the correct performance of the neural network. Parameter %V_{trunk} appeared to have the largest correlation with the neural network output (0.61), which explains why this parameter appears as the most important parameter to rate dyskinesia. The second and third most important parameters are the standard deviation of the velocity of the less affected leg (SD(V) leg) and the power of the velocity signals in the range below 3 Hz (\( \bar{V}_{<3Hz,leg} \)), which add another 22.9% and 10.5%, respectively, to the performance. The contribution of the other nine parameters becomes gradually smaller, but is significant and explains an extra 9.6% to the correct performance of the neural network.

Analysis of the weights of the input units to the hidden units reveals that patients moving the trunk for a large fraction of time (%V_{trunk}) and having a small value of the standard deviation of the segment velocity of the less affected leg (SD(V) leg) are most likely to have dyskinesia. Moreover, patients tend to suffer more from dyskinesia when the trunk movements with frequency components below 3 Hz (\( \bar{V}_{<3Hz,leg} \)) are

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{V}_{segment} )</td>
<td>Mean segment velocity</td>
</tr>
<tr>
<td>( \bar{V}_{&lt;3Hz,segment} )</td>
<td>Mean segment velocity for frequencies below 3 Hz</td>
</tr>
<tr>
<td>( \bar{V}_{&gt;3Hz,segment} )</td>
<td>Mean segment velocity for frequencies above 3 Hz</td>
</tr>
<tr>
<td>( \bar{V}_{cross} )</td>
<td>Ratio between ( \bar{V}<em>{&lt;3Hz,segment} ) segment and ( \bar{V}</em>{&gt;3Hz,segment} ) segment</td>
</tr>
<tr>
<td>SD(V) segment</td>
<td>Standard deviation of segment velocity</td>
</tr>
<tr>
<td>% V( \theta ) segment</td>
<td>Percentage of time that a segment is moving at velocity ( &gt; 0.05 ) m/s.</td>
</tr>
<tr>
<td>( \bar{V}_{segment} )</td>
<td>Mean segment velocity when the segment is moving at a velocity ( &gt; 0.05 ) m/s.</td>
</tr>
<tr>
<td>( P_{1-3Hz,segment} )</td>
<td>Power of frequencies in the range between 1 - 3 Hz</td>
</tr>
<tr>
<td>( P_{&gt;3Hz,segment} )</td>
<td>Power of frequencies above 3 Hz</td>
</tr>
<tr>
<td>max( p_{segment-segment} )</td>
<td>Maximum of normalized crosscorrelation between velocities of two segments</td>
</tr>
<tr>
<td>% sitting</td>
<td>Percentage of time that a patient is sitting</td>
</tr>
<tr>
<td>% upright</td>
<td>Percentage of time that the trunk is upright</td>
</tr>
</tbody>
</table>
large relative to the standard deviation of the segment velocity of the leg (SD(V) I_leg).

The optimal neural network for rating the severity of dyskinesia for the arm was a neural network with two hidden units and six input parameters. The three most important parameters were \( \frac{\text{V}_{\text{V}}}{V <3Hz} \), \( \tilde{\rho}_{\text{wrist-trunk}} \) and \( \%V_{\text{m}1} \text{mleg} \), adding 23.1%, 18.0% and 7.5% to the correct performance of the neural network. The other 3 parameters, added in the forward selection procedure, provided an increase in the performance of the neural network by another 11.6%. Analysis of the projections of the input units to the hidden units revealed a nonlinear interaction between the ratio between low and high frequencies of the most affected leg (\( \frac{\text{V}_{\text{V}}}{V <3Hz} \)), the cross-correlation between wrist and trunk (\( \tilde{\rho}_{\text{wrist-trunk}} \)), and the parameters % sitting, \( V_{\text{wrist}} \) and \( \tilde{\rho}_{\text{wrist-arm}} \). Dyskinesia is found most likely for the arm when the movements of the most affected leg are predominantly lower, rather than at higher frequencies (large value for parameter \( \frac{\text{V}_{\text{V}}}{V <3Hz} \)) and for relatively larger cross-correlation values between wrist and trunk (\( \tilde{\rho}_{\text{wrist-trunk}} \)) (see Fig. 2A). A larger cross-correlation value between movements of the wrist and the trunk (\( \tilde{\rho}_{\text{wrist-trunk}} \)) than between movements of the wrist and least affected arm (\( \tilde{\rho}_{\text{wrist-arm}} \)) resulted in a higher probability that hidden unit 1 will contribute to a rating of dyskinesia (see Fig. 2B). The percentage of the time that the most affected leg is moving (\( \%V_{\text{m}1} \text{mleg} \)) contributed to rating movements as dyskinesia when movements are predominantly at lower, rather than at higher frequencies (large value for parameter \( \frac{\text{V}_{\text{V}}}{V <3Hz} \)) and when the movements between wrist and trunk are uncoordinated (small value for parameter \( \tilde{\rho}_{\text{wrist-trunk}} \)) (see Fig. 3).

The optimal neural network for rating the severity of dyskinesia for the leg was a neural network with three hidden units and seven input parameters. The parameters SD(V) I_leg and \( \%V_{\text{m}1} \text{mleg} \) were the most important parameters and explained together 72.1% of the performance for rating one-minute intervals. The other 5 parameters added in the forward selection procedure, provided an increase of 13.4% to the performance. The neural network for the leg used three hidden units. The combinations of parameters, which determined the output of these 3 hidden units were SD(V) I_leg, \( \%V_{\text{m}1} \text{mleg} \) and \( P_{\text{1-3Hztrunk}} \) (see Fig. 4), which shows the probability, that one of the hidden units contributes to a rating of dyskinesia as a function of the two most valuable parameters (SD(V) I_leg and \( \%V_{\text{m}1} \text{mleg} \), 4A) and as a function of the parameters SD(V) I_leg and \( P_{\text{1-3Hztrunk}} \), 4B). Dyskinesia will occur when the standard deviation of the less dyskinetic leg has a small value and when the most dyskinetic leg is moving for a large fraction of time and for a higher power for frequencies in the range between 1 and 3Hz of the trunk. The other hidden unit contributes to a rating of dyskinesia when the mean velocity of the dyskinetic leg during moving is relatively small and when the power for frequencies in the range between 1 and 3Hz of the trunk is large. The third hidden unit contributes to
Fig. 4. The relation between the three most valuable parameters and the output of hidden unit 1 of the network for assessing the severity of dyskinesia for the leg. The gray scale indicates the probability that hidden unit 1 contributes to a rating of dyskinesia (see gray bar on the right). Panel A shows the probability that hidden unit 1 contributes to a rating of dyskinesia as a function of the two most valuable parameters ( and ). Panel B shows the probability that hidden unit 1 contributes to a rating of dyskinesia as a function of the most important parameter ( ) and third valuable parameter ( ).

A rating of dyskinesia when the leg was moving in at least 91 % of the time and when the standard deviation of velocities of the less affected leg (SD(V) lleg ) is small. The probability of dyskinesia increases when the cross-correlation between the less affected leg and the trunk (θtleg–trunk ) is relatively small for the large number of movements.

IV. Discussion

The main results of this study are two-fold. The first is that the neural network could correctly classify dyskinesia or the absence of dyskinesia in 93.7, 99.7 and 97.0% for the arm, the trunk and the leg, respectively. This is a huge improvement upon previous approaches, which reached a performance close to 75%. The main reason for the huge improvement is the ability to find suitable non-linear combinations of parameters, which play a role in the detection of dyskinesia. The second major result is the rule-extraction, which revealed new insights in the characteristics of dyskinesia. A major advantage of using neural networks for the detection and rating of LID with the forward selection procedure to find the most relevant parameters is that this procedure searches for the most relevant parameters without any prior information and restriction.

Our results showed that the most important parameters ( , %V1trunk and SD(V) lleg for arm, trunk and leg respectively) were the best parameters for all segments, whatever the search algorithm. We also found that sometimes one parameter could be replaced by another parameter without large consequences for the performance of the neural network. This was usually related to the fact that some parameters were highly correlated. For example, parameter gives almost the same performance as parameter .

Previous studies in assessing dyskinesia focussed mainly on parameters in the frequency domain (Manson et al., 2000; Hoff et al., 2001). The results of these studies showed that dyskinetic movements were represented in the lower frequency bands (between 1 and 4Hz, refs). In the present study, parameters \( \bar{V}_{<3Hz,\text{trunk}} \), \( \bar{V}_{<3Hz,\text{lleg}} \) and \( P_{1-3Hz,\text{trunk}} \) showed relatively larger values for patients suffering from dyskinesia. Therefore, these results support the results of previous studies that dyskinesia is most dominant for movements in the lower frequency range. The cross-correlation parameter played an important role in assessing the severity of dyskinesia but its role is somewhat complicated. Subjects showing small values of the mean cross correlation (below 0.2) or large mean cross correlation values (above 0.38) were not suffering from dyskinesia while patients showing mean cross-correlation values between 0.2 and 0.38 tend to reveal a larger probability for dyskinesia. Values of the mean cross-correlation below 0.2 are usually a result of minor motor activity, while values of the mean cross-correlation above 0.38 are a result of a large number of well correlated voluntary movements. The large mean value of the cross-correlation for normal movements corresponds to the observation by Soechting et al. (1986) that joint velocities in elbow and shoulder covary during reaching and pointing movements to targets in 3D space. When the mean cross-correlation has a value between 0.2 and 0.38, parameter \( \bar{V}_{<3Hz,\text{lleg}} \) appears to be an important parameter to indicate whether a subject is dyskinetic. Patients with mean cross-correlation values between 0.2 and 0.38 are most likely dyskinetic in conditions when there are a lot movements ( %V1 large) when movements are predominantly at lower, rather than at higher frequencies. The role of the cross-correlation suggests that movements of body segments are not well coordinated in dyskinesia.

A. Acknowledgment

This study was financially supported by the Prinses Beatrix Fonds (project-number: MAR 00-104) and the Parkinson Patienten Vereniging.

References