

Feature Selection and Construction for the Discrimination of Neurodegenerative Diseases Based on Gait Analysis

Mingjing Yang¹, Huiru Zheng², Haiying Wang², Sally McClean²

Faculty of Computing and Engineering
University of Ulster
N. Ireland, UK

¹Yang-m@email.ulster.ac.uk

²{h.zheng, hy.wang, si.mcclean}@ulster.ac.uk

Abstract—Gait disorder is one symptom of neurodegenerative disease. Using wearable motion sensors to monitor the motor function of patients with neurodegenerative disease has attracted more attention. Research has shown that machine learning techniques can be applied to the classification of neurodegenerative diseases from the gait data recorded by footswitches. In order to identify the most valuable features from 10 raw temporal variables extracted from gait cycles to improve the classification performance, we examine four types of feature selection and construction methods, namely, maximum signal-to-noise ratio based feature selection method, maximum signal-to-noise ratio combined with minimum correlation based feature selection method, maximum prediction power combined with minimum correlation based feature selection method and principal component analysis. Results show that using a set of four features, a relatively high prediction performance has been achieved with classification accuracies ranging from 79.04% to 93.96%. The continual increase of the number of features does not significantly contribute to the improvement of classification performance. This is consistent with clustering-based feature analysis.

Keywords: feature selection, feature construction, classification, neurodegenerative diseases

I. INTRODUCTION

Neurodegenerative disease is a group of conditions which results from the lost of cells in brain and spinal cord leading to movement disturbance [1]. Amyotrophic Lateral Sclerosis (ALS), Parkinson's disease (PD) and Huntington's disease (HD) are progressive neurodegenerative diseases for which impacting gait are often reported. Gait impairments in patients are typically examined by measuring spatial and temporal parameters. Gait disorder is the main symptom used by a clinician to diagnosis and assesses the progress of the neurodegenerative disease. Hausdorff *et al.* [2] reported that patients with ALS, compared to healthy controls, walk more slowly, with a longer average stride time, and have a less steady and more temporally disorganized gait. Patients with HD exhibit walking with a shorter and more variable stride length, and lower cadence and greater variability in stride time, swing time and double support time [3]. Patients with PD show some important features in walking, such as shortened stride

length, reduced gait speed, increased stride-to-stride variability, shuffling gait and freezing [4].

With the recent development of wearable sensor technology, there has been a growing interest in utilising wearable motion sensors in the area of activity monitoring and motion function tests for mobility impaired person in out-lab environments [4]-[13]. For example, Moore *et al.* [11] developed a wearable system for long term monitoring of gait in PD, by which the variable stride length and the fluctuation of motor function associated with levodopa efficiency can be obtained. Salarian *et al.* [13] used wearable gyroscopes to estimate spatial-temporal parameters of gait from Parkinson's disease patients. Recent advances in wearable sensor systems have led to the accumulation of a large amount of movement data. However, the techniques that can take full advantages of a large amount of sensory data are lacking. There are many research questions deserve further investigation. For example, which information is most clinically relevant and can be applied for clinical diagnosis, monitoring and/or assessment? In this paper, we aim to identify optimum feature sets for the classification of neurodegenerative patients and healthy control subjects. A support vector machine (SVM) is employed. The rest of this paper is organized as follows: In section 2, we review the related work on machine learning and feature selection techniques in gait analysis, followed by a brief description of the data and method adopted in the study. Section 4 we present the classify results and discuss the results of feature reduction. The last Section reports the outlook and conclusion.

II. RELATED WORK

There have been several attempts to use supervised or unsupervised machine learning approaches for gait analysis and to use different feature selection methods to identify the most valuable features. For example, Kamruzzaman *et al.* [14] used the three neural networks, namely, Standard Backpropagation; Scaled Conjugate Gradient and Backpropagation with Bayesian Regularization, to automated classification of young/elderly gait patterns. 24 features were extracted from raw gait data recorded by optical motion analysis system. With the help of a forward selection algorithm, Backpropagation with Bayesian Regularization attained 100% classification accuracy when

trained with 3 selected features. Pazit Levinger *et al.* [15] applied SVM to classify gait patterns of patients with knee osteoarthritis before and after knee replacement surgery using 8 spatial-temporal gait parameters as input features. When the hill climbing feature selection algorithm is applied, the SVM can identify the osteoarthritis gait from the healthy ones with the highest accuracy of 97.1% using two symmetry index features. In this study, the gait data were recorded in a laboratory environment using a six-camera Vicon System. Shyamal Patel *et al.* [16] collected the daily movement of PD patient using triaxial accelerometers placed on both side of upper arm, forearm and shin. They ranked six features the intensity, modulation, rate, periodicity and coordination of upper and lower limbs movement based on clustering ability, they also used classifiers to assess the impact of ranked features on predicting accuracy in bradykinesia and dyskinesia. However in this study, they only considered the impact of a single feature, and did not consider the impact of combining features. Zheng *et al.* [17] investigated classification performance of three classifiers (SVM, Random Forest and KStar) used in gait data recorded from neurodegenerative diseases patients and control subjects by a pair of footswitches, and applied a power prediction feature ranking method combined with feature selection based on minimum correlation to select the maximum-relevant and minimum-redundancy feature set. Their results showed that SVM classifier with a selected four features out of ten can achieve the best performance in the classification of neurodegenerative diseases and control subjects. However, no other feature selection technique were examined or compared.

In this paper, we investigate three feature selection techniques and one feature construction method, namely the maximum signal-to-noise ratio (SNR) based feature selection method (MSNR), maximum signal-to-noise ratio (SNR) combined with minimum correlation based feature selection method (MSNR&MC), maximum prediction power combined with minimum correlation based feature selection method (MPP&MC) and principal component analysis (PCA), for the analysis of neurodegenerative patients gait datasets. A SVM-based classifier is used to assess the prediction power of feature combination.

III. METHODOLOGIES

A. Datasets under study

In this study, Neuro-Degenerative Disease Gait Dynamic Database (NDDGD) was used, which was published by Physionet [18]. The dataset contains stride interval time series extracted from the vertical ground reaction force records of subjects as they walked along a 77m hallway for 5 minutes. The data includes records from 64 individuals, i.e. 15 Parkinson disease subjects (age ranging from 44 to 77), 20 Huntington disease subjects (age ranging from 29 to 71), 13 Amyotrophic Lateral Sclerosis subjects (age ranging from 36 to 70) and 16 healthy control subjects (age ranging from 22 to 74). Readers are referred to [2] for more details.

B. Feature extraction

Human walking is a quasi-period movement consisting of gait cycles. In this study, one gait cycle (stride) is used as an

instance. Ten temporal features are extracted from patient gait cycles, namely, Right Stride (rs), Right Swing (rw), Right Stance (rt), Right Swing (% of stride) (rwp), Left Stride (ls), Left Swing (lw), Left Stance (lt), Left Swing (% of stride) (lwp), Double Support (d), Double Support (% stride) (dp). As illustrated in Fig 1, rs is the period of time from right heel contacting with the ground to the same heel contacting with the ground again; rt is the period of time for right foot contact with the ground (from right heel contact to right toe off); rw is the period time for right foot leave form the ground (from right toe off to right heel contact again). Double Support (dp) is the period of time for two feet contact with the ground in a cycle.

C. SVM classifier

SVMs have attracted a lot of attention since they were first presented by Vapnik in 1992 [19]. SVMs nonlinearly map the original data to a sufficiently high dimension space in which a maximum margin hyperplane is identified to separate the data from different classes. It has shown that SVMs can achieve good performance in classifying nonlinear separate data found in many areas such as handwritten digit recognition, speaker identification etc..

Given some training data which is written as:

$$D = \left\{ (x_i, c_i) \mid x_i \in R^k, c_i \in \{-1, 1\} \right\}_{i=1}^n \quad (1)$$

where x_i is a training example consisting of a k dimensional vector, c_i is class label either 1 or -1. The hyperplanes which separate two classes can be described as:

$$f(x) = W \bullet \phi(x) + b \quad (2)$$

where W is the normal vector of hyperplane, $\phi(x)$ is the function which define the nonlinear mapping from input space to a higher dimensional space, and b is the hyperplane bias from the origin. The main idea of SVM is to choose the W and b which maximize the perpendicular distance between the class boundary and the hyperplane.

In this study, SVM was applied and the radial basis function kernel was selected. 10-folder-cross-validation was performed to estimate the performance of the classifier. All the models were implemented within the framework provided by publicly available package RapidMiner [20].

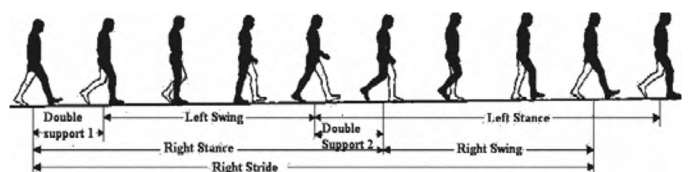


Figure 1. The illustration of a gait cycle

D. Features selection and construction

It has been generally recognized that high dimensionality of data will lead to high complexity of the classifier. Feature selection and construction technology are often used to reduce the data dimensionality. There are two common methods for dimensionality reduction, one is the selection of an appropriate subset among the existing features and the other is

transformation of the original features to construct a smaller set of new features [20]. In this study, we proposed 3 feature selection methods: MSNR, MSNR&MC, and MPP&MC. MSNR is a feature ranking based selection method, while SNR&C and MPP&C are feature selection methods following the maximum-relevant-minimum-redundant strategy [22]. PCA, one of feature construction methods, was also investigated.

1) *Maximum signal-to-noise ratio (SNR) based feature selection method (MSNR)*: SNR defines the separation power of a feature between two classes by the ratio of the difference of feature mean values between two classes and the feature variance within a class [23]. For each feature, the SNR is calculated using (3)

$$SNR = (\mu_{class1} - \mu_{class2}) / (\sigma_{class1} + \sigma_{class2}) \quad (3)$$

Where μ_{class1} and μ_{class2} are the means of the features in class1 and class2 respectively; σ_{class1} and σ_{class2} are standard deviations of the features in class1 and class2 respectively. A higher ratio of SNR means a stronger prediction power.

MSNR consists of two steps. Firstly, ranking features was performed on the strength of SNR. A feature with minimum SNR was removed from the feature set to obtain a smaller feature set. The process was repeated until only one feature left in the feature subset. The algorithm is shown as (4):

$$F_{n-1} = F_n - f_{\min-SNR} \quad (4)$$

Where n is the number of features, F_n is the features set with the number of features is n , $f_{\min-SNR}$ is the feature with the minimum SNR in F_n . If n change from 10 to 2, we can get 9 feature subsets.

2) *Maximum signal-to-noise ratio (SNR) combined with minimum correlation based feature selection method (MSNR&MC)*: The method includes three steps. The first step is ranking the features based on the SNR which is the same as in the method 1. In the second step, the pairwise correlations of each pair of features were calculated and formed a correlation matrix. After that the feature with lowest SNR in the feature pair with the highest correlation value was removed to obtain a new feature subset. The process was repeated until there was only one feature left in feature subset. The algorithm is summarized in algorithm 1.

Algorithm 1:

1. Rank all the features (F) in terms of SNR. The SNR of $f_i >$ the SNR of f_j , when $i < j$, where f_i is the i_{th} feature.

2. Calculate the pairwise correlation of feature pairs to obtain the correlation matrix R .

For $n=10$ to 2, (n is the size of feature set)

3. Search the maximum element r_{ij} in R , where $i < j$

4. Remove f_j from F_n to obtain a new feature set F_{n-1}

5. Remove the i_{th} row and the j_{th} column from R

End

3) *Maximum prediction power combined with minimum correlation based feature selection method(MPP&MC)*: The method proposed by Zheng *et al.* [17] comprised two steps. The first step is classifying two groups using one feature, and ranking features based on prediction power (accuracy).

The second step is to calculate the correlation matrix of features, and to remove the smaller prediction power feature in the pair of features with the highest correlation. The process was repeated until there was only one feature left. Further details can be found in [17].

4) *Principal component analysis (PCA)*: The method is a classical approach for finding optimal linear transformation which can best represent the data in a least square sense. PCA transforms a data set into a new coordinate system which makes the largest variance of the data to lie on the first coordinate (called principal component), the second largest variance on the second coordinate, and so on [24]. The dimension reduction is performed by keeping the first few principal components which retain the most variance of the data.

Here, first, a Z-transformation was performed on the original feature set to make each feature have mean 0 and variance 1. After that PCA was performed. Last, we can obtain the new feature subset 1 which contained the first principal component and the new feature subset 2 which contain the first 2 principal components, and so on.

IV. RESULTS AND DISCUSSION

A. Classification analysis

We are interested in classifying subjects into four groups: ALS, PD, HD and healthy controls using SVM with the temporal features extracted from gait cycles as input. We considered 7 binary classification problems, namely:

ALS vs. Control group	(AC)
PD vs. Control group	(PC)
HD vs. Control Group	(HC)
ALS vs. PD	(AP)
ALS vs. HD	(AH)
PD vs. HD	(PH)
ALS + PD + HD vs. Control Group	(CD)

The classification results for each problem with 10 features using 10-fold-cross-validation are shown in table I [17]. The results show that ALS can be most easily distinguished from PD (accuracy 85.47%), HD (accuracy 86.52%) and the healthy control group (accuracy 93.96%). PD and HD can be distinguished from the healthy control group with accuracy 86.43% and 84.17%. However, the classification accuracy of PD vs. HD is only 79.04% which is less than other classification problems. This may due to the common cause of PD and HD which is the impairment of basal ganglia. Their gait patterns have similar characteristics. All in all, it is feasible to use machine learning methods in classification neurodegenerative patients with healthy controls. Readers are referred to [17] for more details.

Table I. THE CLASSIFICATION RESULT OF SVM WITH 10 FEATURES ACROSS 7 BINARY CLASSIFICATION PROBLEM

Classification Problem	Ac (%)	AUC	Positive Class			Negative Class		
			Pr (%)	Se (%)	Sp (%)	Pr (%)	Se (%)	Sp (%)
ALS vs. Control group	93.96	0.98	90.76	93.80	94.06	96.07	94.06	93.80
PD vs. Control group	86.43	0.92	87.64	83.11	89.43	85.44	89.43	83.11
HD vs. Control group	84.17	0.91	89.03	80.01	88.86	79.74	88.86	80.01
ALS vs. PD	85.47	0.92	88.15	87.19	82.98	81.69	82.98	87.19
ALS vs. HD	86.52	0.93	91.76	86.9	85.83	78.29	85.83	86.9
PD vs. HD	79.04	0.85	80.45	82.32	74.95	77.19	74.95	82.32
ALS+PD+HD vs. Control group	86.85	0.91	91.45	90.33	77.6	75.17	77.6	90.33

B. The impact of feature selection and construction methods on the classification performance

Feature selection technologies are important in machine learning for reasons such as generalization performance, improved learning speed and building more robust models. In some cases, feature selection technologies can help us to obtain a better understanding of data by finding out which features are most important and what is the relationship between features. In this paper, four feature reduction and selection methods (i.e. PCA, MSNR, MSNR&MC, and MPP&MC) were examined using the classifier SVM. The prediction performance was measured by the area under the receiver operating characteristic curve (AUC). The results are shown in Fig 2. Where the x axis indicates the number of features and the y axis represents the AUC values obtained.

A close examination of the results presented in Fig 2 reveals that:

- In all four feature selection and construction methods, when the number of features increases from 1 to 4, the curves rise sharply indicating that the prediction performance improved significantly with an increase in feature number.
- In using MSNR&C, MPP&C or PCA, when the feature number was more than 4, adding new features did not improve the AUC significantly.
- The MSNR-based feature selection achieved the best performance when using more than four features in the most of classification problems (PC, HC, AP, PH, and CD). In the other problem (AC, AH), MSNR achieved its best performance using 4 features.
- When the feature number is greater than 4, three curves (respectively representing MSNR&C, MPP&C or PCA) nearly overlap except for PCA which is smaller at feature number 9 and 10.
- As shown in the Fig 2, the AUC curves become flat when the number of features is more than 4, suggesting the continual increase of features did not make significant contribution to the increase of the prediction performance.

C. The element of optimal feature subset

The results showed in Fig 2 indicate that the prediction performance of SVM is sufficient with 4 features as input. To find the composition of the feature subset with size 4 (subset 4), the elements of feature subset selected by MSNR&MC which have better performance across 7 classification problem were examined. Table II shows the feature ranking result of 7 classification problems and the elements of feature subset which are formed by MSNR&MC. The elements of subset 4 are different, but in most problems except for Control vs. Disease, we can find that the four elements come from four groups (rt, rs), (lw, lwp), (rw, rwp) and (d, dp). This is consistent with the results obtained from hierarchical clustering analysis of all the input samples. As shown in Fig 3 [17], features ls and lt, d and dp, lw and lwp, rw and rwp are grouped together in the hierarchical tree of 10 features. In order to obtain an insight into the relationships between features grouped in the same cluster, we replaced one feature in the feature subset with the other one in the same group, for example replace lw with lwp, and examine the impact of new feature set on prediction power. In the problem Control vs. Disease, the lwp was replaced by dp and lw was replaced by d. The classification performance of the new feature set was measured by AUC and a t-test was applied to assess the significant difference between the performance of the original feature set and the new feature set. As shown in Table III, the AUC of the new feature sets have no significant difference ($p > 0.05$) compared with the original features selected by MSNR&MC. The feature sets in the last two columns of Table III were formed by an element (rt, rs) replaced by one element from the group (lt, ls) which is nearest to the group (rs,rt) in the hierarchical tree. In the classification of AC, PC, AP, no significant difference was observed in terms of AUC values. Interestingly, the AUC values were found lower in the classification problem of HC, AH, and PH. This may be due to the unbalance of gait found in HD patients, which make different left strides and right strides.

From Table II, it can be found that the elements of feature subset 4 are different across 7 classification problem. We wonder if a feature subset selected for a classification problem is suitable for other problems. In order to test it, the feature set (rt,d,rw,lw) were examined using SVM classifier in every classification problem. The prediction performances are shown in Table IV. Compared with the AUC obtained with the original feature subset 4 which is obtained by MSNR&MC in

each binary classification problem, the two results are nearly same.

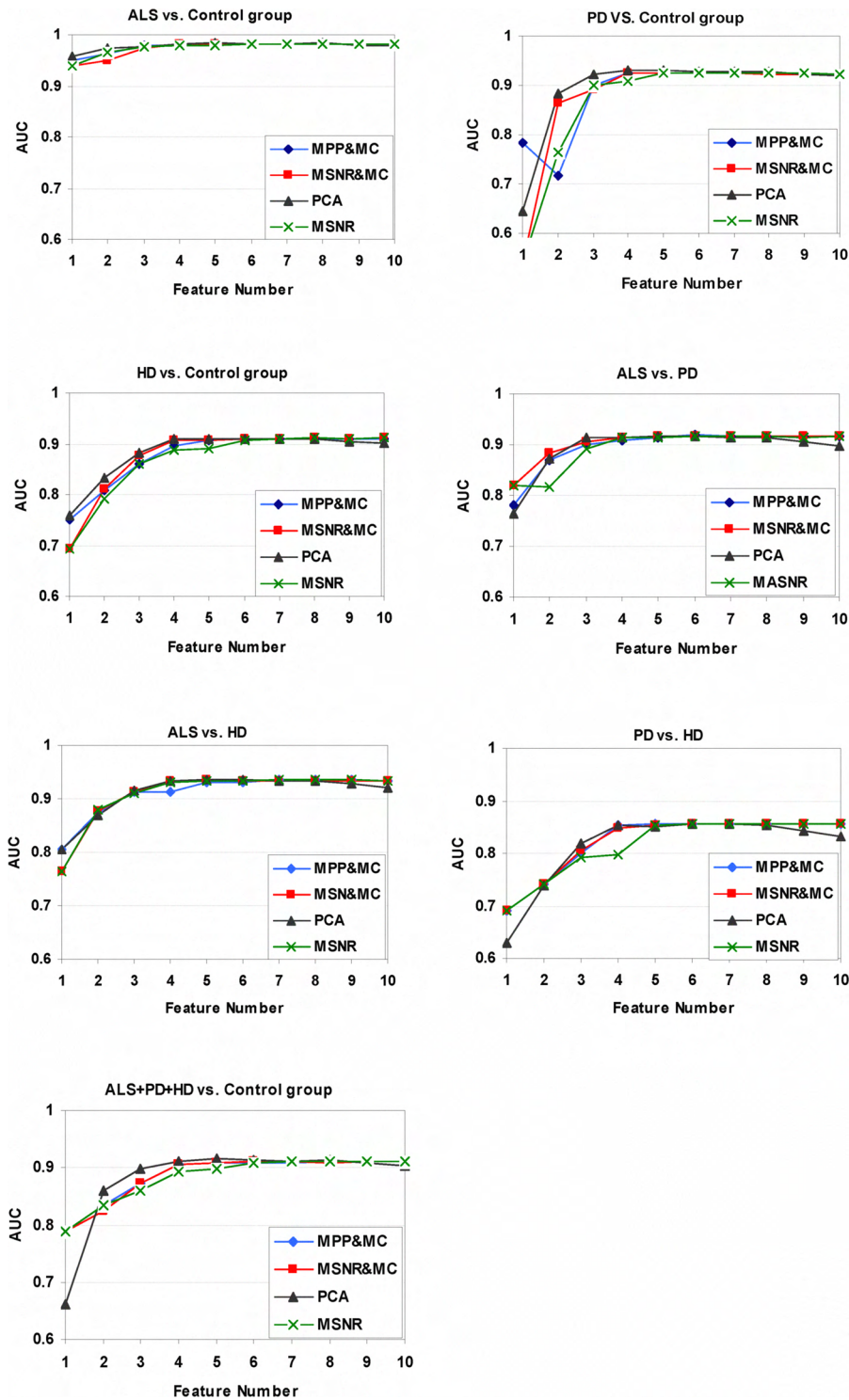


Figure 2. Impact of number of features on the predictive power (AUC) for seven classification problems based on three features reduction methods (PCA, MSNR, MSNR&MC, MPP&MC)

Table II. THE FEATURE RANKING BASED SIGNAL NOISE RATE AND THE LAST 4 FEATURES SELECTED BY CORRELATION BASED REMOVE METHOD

Classification problem	Feature ranking based SNR	4 features selected by MSNR&MC
ALS vs. Control group	rt,rs,lt,ls,lwp,rwp,d,dp,rw,lw	rt,d,rw,lw
PD vs. Control group	dp,lwp,d,rwp,rt,lt,lw,rw,rs,ls	dp,rwp,rt,lw
HD vs. Control group	dp,d,rwp,rt,lt,lwp,rs,ls,lw,rw	dp,rt,lw,rw
ALS vs. PD	rs,ls,rt,lw,lt,rw,d,dp,lwp,rwp	rs,d,lwp,rwp
ALS vs. HD	rs,ls,lt,rt,d,rw,dp,lw,lwp,rwp	rs,d,lw,rwp,
PD vs. HD	lwp,rs,lt,ls,rw,rt,d,dp,rwp,lw	lwp,rs,rw,d
ALS+PD+HD vs. Control group	rt,d,dp,rwp,lt,lwp,rs,ls,rw,lw	rt,rwp,lwp,lw

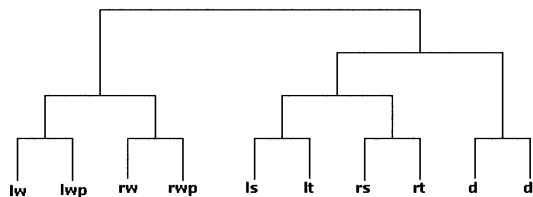


Figure 3. A hierarchical tree of 10 features based in the analysis of all the samples

Table III. THE AUC OF ORIGINAL FEATURE SUBSET 4 AND NEW FEATURE SET, THE VALUE IN BALCKETS ARE T-TEST VALUE

Classification Problem	Origin set	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
ALS vs. Control group	0.982	0.982 (0.945)	0.982 (0.974)	0.981 (0.89)	0.981 (0.753)	0.978 (0.945)	0.978 (0.309)
PD vs. Control group	0.925	0.926 (0.96)	0.926 (0.792)	0.926 (0.808)	0.923 (0.618)	0.921 (0.561)	0.92 (0.449)
HD vs. Control group	0.908	0.908 (0.903)	0.904 (0.468)	0.901 (0.324)	0.908 (0.545)	0.878 (0.002)	0.879 (0.002)
ALS vs. PD	0.913	0.911 (0.957)	0.91 (0.74)	0.917 (0.359)	0.912 (0.977)	0.912 (0.719)	0.907 (0.334)
ALS vs. HD	0.934	0.933 (0.786)	0.934 (0.876)	0.931 (0.556)	0.932 (0.587)	0.918 (0.01)	0.921 (0.007)
PD vs. HD	0.848	0.854 (0.350)	0.851 (0.323)	0.851 (0.480)	0.838 (0.260)	0.815 (0)	0.811 (0)
ALS+PD+HD vs. Control group	0.907	0.91 (0.482)	0.906 (0.014)	0.907 (0.022)	0.903 (0.038)	0.897 (0.013)	0.898 (0.038)

Table IV. THE AUC OF ORIGINAL FEATURE SUBSET 4 AND THE NEW FEATURE SET (RT,D,RW,LW)

Classification problem	Original feature subset 4	New feature set (rt,d,rw,lw)
AC	0.982	0.982
PC	0.925	0.927
HC	0.908	0.908
AP	0.913	0.918
AH	0.934	0.934
PH	0.848	0.855
CD	0.907	0.905

V. CONCLUSIONS AND FUTURE WORK

Gait analysis is an important tool for clinicians to assess the state of neurodegenerative disease to support therapy. In this study, we used the SVM to classify the ALS, PD, HD and control condition achieving good performance (accuracy ranging from 79.04% to 93.96). In order to find the most valuable temporal gait features, we examined 4 feature selection and construction methods. The result shows that using a feature set including 4 features are sufficient to achieve relatively high classification performance. The continual increase of the number of feature does not significantly contribute to the improvement of prediction performance. These 4 features respectively come from four groups, namely (lw, lwp), (rw, rwp), (d, dp) and (rt, rs), which is consistent with the result obtained from hierarchical clustering analysis. In this paper we only considered temporal gait features. We will examine the contribution of other features, such as spatial features, balance features, and

features derived from the gait data, in neurodegenerative disease gait pattern classification in the future.

REFERENCES

- [1] Serge Przedborski, Miquel Vila, and Vernice Jackson-Lewis, "Series Introduction: Neurodegeneration: What is it and where are we?" *J Clin Invest.*, vol. 111(1), pp: 3-10. 2003
- [2] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol.*, vol. 88, pp. 2045-2053, 2000.
- [3] M. Banaie, Y. Sarbaz, S. Gharibzadeh and F. Towhidkha, "Huntington's disease: Modeling the gait disorder and proposing novel treatments," *J. Theor. Biol.*, vol. 254, pp. 361-367, 2008.
- [4] J. I. Hoff, A. A. vander Plas, E. A. H. Wagemans and J. J. van Hilten, "Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease," *Movement Disorders*, vol. 16, pp. 58-61, 2001.
- [5] P. Bonato, D. M. Sherrill, D. G. Standaert, S. S. Salles and M. Akay, "Data mining techniques to detect motor fluctuations in Parkinson's disease." *Conf Proc IEEE Eng Med Biol Soc*, vol. 7, pp. 4766-9, 2004.
- [6] R. J. W. Dunnewold, J. I. Hoff, H. C. J. van Pelt, P. Q. Fredrikze, E. A. H. Wagemans and B. J. J. van Hilten, "Ambulatory quantitative assessment of body position, bradykinesia, and hypokinesia in Parkinson's disease," *Journal of Clinical Neurophysiology*, vol. 15, pp. 235-242, 1998.
- [7] N. Keijsers, M. Horstink and S. Gielen, "Ambulatory automatic assessment of motor fluctuations in Parkinson's disease," *Movement Disorders*, vol. 20, pp. S73-S73, 2005.
- [8] K. Lorincz, B. Kuris, S. M. Ayer, S. Patel, P. Bonato and M. Welsh, "Wearable wireless sensor network to assess clinical status in patients with neurological disorders," *Proceedings of the Sixth International*

- [9]. Symposium on Information Processing in Sensor Networks, pp. 563-564, 2007.
- [10]. N. L. W. Keijsers, M. W. I. M. Horstink and S. C. A. M. Gielen, "Ambulatory motor assessment in Parkinson's disease," *Movement Disorders*, vol. 21, pp. 34-44, 2006.
- [11]. S. T. Moore, H. G. MacDougall and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *J. Neurosci. Methods*, vol. 167, pp. 340-348, 2008.
- [12]. S. T. Moore, H. G. MacDougall and W. G. Ondo, "Ambulatory monitoring of motor fluctuations and freezing of gait: Objective assessment of the efficacy of pharmacological treatments in Parkinson's disease," *Movement Disorders*, vol. 23, pp. S216-S216, 2008.
- [13]. Salarian, H. Russmann, F. J. G. Vingerhoets, P. R. Burkhard and K. Aminian, "Ambulatory monitoring of physical activities in patients with Parkinson's disease," *Ieee Transactions on Biomedical Engineering*, vol. 54, pp. 2296-2299, 2007.
- [14]. R. Begg and J. Kamruzzaman, "Neural networks for detection and classification of walking pattern changes due to ageing," *Australas. Phys. Eng. Sci. Med.*, vol. 29, pp. 188-195, 2006 Jun. 2006.
- [15]. P. Levinger, D. T. H. Lai, K. Webster, R. K. Begg and J. Feller, "Support Vector Machines for detecting recovery from knee replacement surgery using quantitative gait measures." *Conf Proc IEEE Eng Med Biol Soc*, vol. 2007, pp. 4875-4878, 2007. 2007.
- [16]. S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. H. Growdon, M. Welsh and P. Bonato, "Analysis of Feature Space for Monitoring Persons with Parkinson's Disease With Application to a Wireless Wearable Sensor System," *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 6290-6293, 2007.
- [17]. H. R. Zheng, M. J. Yang, H. Y. Wang, S. McClean, "Machine Learning and Statistical Approaches to Support the Discrimination of Neuro-Degenerative Diseases Based on Gait Analysis", Accepted by the post-conference Springer special volume on Intelligent Patient Management.
- [18]. A.L. Goldberger, LAN Amaral, L. Glass, J.M. Hausdorff, PCh Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000.
- [19]. V. N. Vapnik, *The nature of statistical learning*, 2nd ed. New York: Springer, 2000.
- [20]. M. Mierswa, R. Wurst, M. Scholz, E. Timm, "YALE: Rapid Prototyping for Complex Data Mining Tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD-06)(2006)
- [21]. J. Han, M. Kamber, *Data Mining Concepts and Techniques*, 2nd ed. China Machine Press, 2006.
- [22]. Peng, H.C., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(2005) 1226-1238
- [23]. G. Liang, Q. Song, N. Kasabov, "A Novel Feature Selection Method to Improve Classification of Gene Expression Data", 2nd Asia-Pacific Bioinformatics Conference
- [24]. I.T. Jolliffe *Principal Component Analysis*, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illu.