***4. Title Page (with authors and addresses)**

Title:

Metric learning for Parkinsonian identification from IMU gait measurements.

Authors:

Dr Fabio Cuzzolin[1], Dr Michael Sapienza[3], Dr Patrick Esser[2], Mr Suman Saha[1]*, Miss Marloes Franssen[2], Dr Johnny Collett[2], Prof Helen Dawes[2].

1) Artificial Intelligence and Vision Group, Department of Computing and Communication Technologies, Oxford Brookes University, Wheatley campus, Oxford, OX33 1HX, UK.
2) Movement Science Group, Faculty of Health and Life Sciences, Oxford Brookes University, Gipsy Lane, Headington, Oxford, OX3 0BP, UK.
3) Torr Vision Group, Department of Engineering Science, University of Oxford, Parks Road Oxford OX1 3PJ, UK.

*** Corresponding Author:**
Mr Suman Saha
Artificial Intelligence and Vision Group
Department of Computing and Communication Technologies
Oxford Brookes University, Wheatley campus
Oxford, OX33 1HX
United Kingdom

Tel: +44 (0) 1865 48 4574
@: suman.saha-2014(at)brookes.ac.uk

Abstract word count: 183
Full text word count: 3,182
Figures: 4
Tables: 1

Acknowledgements
We would like to acknowledge all participants involved in the "Longer term exercise intervention for people with Parkinson's disease" study funded by the NIHR. In addition we would like to acknowledge the Movement Science Group for supporting this research throughout its entire duration. We would also like to acknowledge the Elizabeth Casson Trust, the National Institute for Health Research and the Physical Sciences Research Council for their partial funding throughout this project.

Author Involvement
All authors have read the manuscript and have actively been involved with the execution of the study and the preparation of this manuscript. This manuscript has not previously been published and is not under simultaneous consideration by another journal.

Dr. Fabio Cuzzolin
Tel: +44 (0) 1865 484526
@: fabio.cuzzolin@brookes.ac.uk

Dr Michael Sapienza
Tel: +44 (0) 1865 28 3144

@: michael.sapienza@eng.ox.ac.uk

Dr. Patrick Esser
Tel: +44 (0) 1865 483 833
@: pesser@brookes.ac.uk

Mr Suman Saha
Tel: +44 (0) 1865 48 4574
@: suman.saha-2014@brookes.ac.uk

Miss Marloes Franssen
Tel: +44 (0) 1865 483 272
@: marloes.franssen-2011@brookes.ac.uk

Dr. Johnny Collett
Tel: +44 (0) 1865 483 272
@: jcollett@brookes.ac.uk

Prof. Helen Dawes
Tel: +44 (0) 1865 483 293
@: hdawes@brookes.ac.uk

Postal Address for all authors:
Movement Science Group
Faculty of Health and Life Sciences
Oxford Brookes University
Gipsy Lane, Headington
Oxford, OX3 0BP
United Kingdom

Potential conflict of interest:
None of the authors have any potential conflicts of interest to declare

Funding sources for study:
This study was supported by a Wellcome Trust translation award (UK)

Metric learning for Parkinsonian identification from IMU gait measurements

Summary

Diagnosis of people with mild Parkinson's symptoms is difficult. Nevertheless, variations in gait pattern can be utilised to this purpose, when measured via Inertial Measurement Units (IMUs). Human gait, however, possesses a high degree of variability across individuals, and is subject to numerous nuisance factors. Therefore, off-the-shelf Machine Learning techniques may fail to classify it with the accuracy required in clinical trials.

In this paper we propose a novel framework in which IMU gait measurement sequences sampled during a 10 metre walk are first encoded as hidden Markov models (HMMs) to extract their dynamics and provide a fixed-length representation. Given sufficient training samples, the distance between HMMs which optimises classification performance is learned and employed in a classical Nearest Neighbour classifier. Our tests demonstrate how this technique achieves accuracy of 85.51% over a 156 people with Parkinson's with a representative range of severity and 424 typically developed adults, which is the top performance achieved so far over a cohort of such size, based on single measurement outcomes. The method displays the potential for further improvement and a wider application to distinguish other conditions.

Key Words:

Machine Learning Algorithms, Hidden Markov Models, Metric Learning, Inertial Measurement Unit, Gait, Parkinson's

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder[1]. Its clinical diagnosis, according to the UK Brain Bank criteria, is mainly based on the presence of motor symptoms (e.g. bradykinesia, rigidity, tremor)[2]. Disease progression can be monitored by analysing these motor symptoms. In established PD, the Brain Bank criteria show 90% sensitivity and specificity for the presence of midbrain Lewy bodies[2]. However, diagnosis in the community by non-experts yields a 25% error[2], supporting the need for better automated diagnostic and monitoring tools for primary care.

Walking has been signalled as a sensitive indicator for the progression of PD[3], as individuals present an altered gait pattern with increased cadence and reduced stride lengths[4]. Inertial Measurement Units (IMUs) can be used to gather gait measurements inexpensively, quickly and easily in clinical environments[4]. However, basic temporal (steptime/cadence) and spatial (stride-length and walking speed) parameters cannot be used as discriminative function, as they lack disease specificity[5, 6]. Alternate Centre of Mass (CoM) excursion in conjunction with sophisticated classification methodologies has been relatively successful as disease discriminative functions over short distances[6].

Motor symptoms are useful for distinguishing different forms of Parkinson's and for determining severity progression[2] for example, postural instability and gait disability versus tremor dominant phenotypes and stages of motor decline in line with functional mobility.

Machine learning (ML) techniques can utilise gait data uniquely, providing a non-intrusive means of monitoring the development and onset of neurodegenerative conditions. Artificial Neural Networks have been employed to distinguish gait pattern between typically developed adults (TDA) and subjects with pathological conditions with an accuracy of 95%[7], or those with lower limbs arthritis with 80% accuracy[8]. They have also been applied for detecting and classifying walking pattern changes due to ageing, achieving a maximum generalisation performance of 83.3%[9].

Machine learning has been successfully used for the diagnosis of individual forms of dementia related Parkinson's[10], but also early Alzheimer's[11]. ML disease progression approaches have also been explored to rate the severity[12] in PD (based on the UPDRS scale), for example via postural sway analysis employing support vector machine (SVM) classification[13] or via longitudinal measurements combined with random forest[3] regression. These methods differ from the clinicians' own UPDRS estimates by a range between ±5 and ±10 UPDRS points. More effective methods applying feature selection methods achieve a 2 UPDRS points difference from clinicians' estimates[14].

[31] and [32] have proposed to diagnose PD using ground reaction forces (as gait signals), captured using force-sensitive sensors placed underneath the subject's feet. Features are extracted either by computing statistics (e.g. min, max, mean and standard deviation) of the force signals [31], or by applying Fourier transform to these signals[32]. Features are then selected using genetic algorithms [31], or their histogram is computed[32]. Subsequently, either radial basis[31] or chi-square distance kernels[32] are used to train SVMs for classification. Hidden Markov models (HMMs) with Gaussian

Mixtures have also been used for classifying PD[33]. Factorial HMMs have been employed for distinguishing amyotrophic lateral sclerosis[34] patients from healthy subjects.

All these works used relatively smaller cohort of subjects[15], ranging between 29 to 166 subjects. In opposition, we consider here an increased clinical sample, covering a wide range of severities and phenotypes of PD (including lesser affected people) in addition to a large age-matched cohort of TDA. As soon as a much bigger share of the population is analysed, issues with the generalisation power of ML methods arise[16], signalling the need for novel paradigms. In response, whereas others have used standard off-the-shelf classifiers[17,18,31,32], we propose a tailored classification method which applies to time-series of gait measurements represented as dynamical models. Unlike [33,34], we use HMMs only to encode measurement series. Motivated by recent ML advances, we then construct an optimal classifier for the problem at hand from the available training data via metric learning techniques, achieving promising results in classifying human action image sequences[19] belonging to tens of different classes. This study explores whether this novel optimal metric learning-based classifier can: firstly, automatically distinguish those with and without PD (including people with mild symptoms), during a clinically standardised 10-metre walk test, within a large cohort; and secondly, determine disease severity.

Methodology

Classification approach

The problem of automatically determining whether a person has PD and its severity from IMU data can be formalised within Machine Learning as follows. Given a 'training set' $D = \{(G_1,Y_1), \ldots, (G_n,Y_n)\}$ of *n* gait motions G_k, each associated with a 'class label' Y_k (e.g. normal versus PD), we want to learn an appropriate machinery (a 'classifier') which, given as input a new, unlabelled gait motion, produces the class label of the new sequence, therefore deciding whether the subject performing the motion is affected by Parkinson or not. Solving a classification problem involves:

i) Finding a suitable representation for the input data;

ii) Designing the most appropriate classifier for the problem.

Here, each instance of gait motion is represented by a time series of IMU. For each time instant, a vector of 9 components is formed by collecting the X,Y,Z values produced by the device's accelerometer, magnetometer and gyroscope. The IMU's X, Y and Z axes were aligned to the longitudinal, transverse and frontal axis, respectively, whereby positive values were measured as up, right, and forwards movements.

IMU sequences may be of different lengths: we then need to find a constant-size representation for them ('time warping'[20]). Furthermore, studies in gesture and gait classification indicate that modelling time series dynamics can greatly help with their classification[20]. Researchers have employed linear, nonlinear[21] and even chaotic[22] dynamical systems to encode time series. Hidden Markov models[23] (HMMs), in particular, address the time warping issue while efficiently encoding motion dynamics[20, 24].

HMM representation of IMU sequences

An HMM is a finite-state stochastic model whose *N* states form a Markov chain. Transitions between states are governed by a *N*×*N* transition matrix $A=[a_{ij}]$, where a_{ij} specifies the probability of passing from state *i* to state *j*, for each pair of states (Fig. 2(b) left).

Although HMM states are 'hidden' (they cannot be observed directly), the measurement vector *y* (here a 9-dimensional IMU vector) they generate can instead be observed. For each state *i*, a Gaussian distribution with mean C_i describes the likelihood of a state *i* generating an observation *y*. In Fig. 1 each state *i* is associated with a specific region of the IMU signal.

Given a sequence of IMU vectors associated with a walking gait, its best HMM description can be identified via the Expectation-Maximisation (EM) algorithm[23, 25]. Each IMU sequence, regardless its length, can then be represented by a HMM $H=\{A,C\}$ with the same number of states (a parameter of EM), where *A* is the transition matrix and $C=[C_1,\ldots,C_N]$ is the matrix whose columns are the means of the *N* Gaussian output densities. N=3-state automata have been demonstrated to represent simple actions effectively[19].

Classifying HMMs

Disease diagnosis reduces then to the binary classification of walking gaits of unknown test subjects represented as hidden Markov models, learnt from the associated series of IMU measurements. HMMs are typically classified by: 1. learning a new model H=(A,C) for each test sequence; 2. computing its distance (appropriately measured) from each training model in $D' = \{(H_1,Y_1), \ldots,(H_n,Y_n)\}$, and: 3. assigning to H the label of the closest training model.

Various distance functions for dynamical systems[26] and HMMs[25] have been proposed. None can suit every classification problem, as the same models can be endowed with different labels. A widely supported approach[27], consists of learning the most appropriate distance function for each specific classification problem, e.g. by maximising the classification performance achieved on the available training data.

Learning an optimal HMM metric

Two of the authors have proposed in a very recent paper[19] a principled framework for learning such an optimal distance function for a training set of models. This framework can be applied here once IMU gait sequences are encoded as HMMs, yielding the disease recognition pipeline of Fig. 2(a).

Firstly, each IMU gait sequence is encoded by a HMM via Expectation-Maximisation (stage 1). The optimal distance function for a given training set of models can then be learned in a 'pullback metric' framework[28] (stage 2), in which the space of HMMs is stretched via a differentiable deformation and the classification performance on the training data of the resulting 'pullback' distance in the deformed space is assessed (Fig. 2(b)). The maximal-performance pullback distance (stage 3) is finally passed to an off-the-shelf classifier (for instance a Nearest Neighbour (1-NN) classifier, stage 4).

In this work test HMMs encoding IMU sequences to classify are therefore assigned the label of the closest training HMM, with respect to the selected optimal metric.

More technical details on the pullback metric framework can be found in a recent paper[19].

Severity estimation

The 36 item short-form (SF-36) was designed to obtain self-perceived information on 8 health domains, namely: limitations in physical or social activities, limitations due to physical health or to emotional problems, bodily pain, general mental health, vitality and general health perception[29]. Training gait sequences in our dataset are assigned a physical functioning severity score in the range 0 to 100 (higher scores representing more favourable health states) from SF-36.

We can then estimate the severity level of each new test IMU sequence (Fig. 3) by locating for each test HMM (denoted by "?") its K=5 nearest training HMMs (according to the optimal pullback distance learned), and averaging those severity levels associated with PwP (circled).

Protocol and inclusion criteria

Participants were included if between the ages of 39 and 80, whose condition had been stable (in terms no relapse or exacerbation, causing a significant change in their condition), who could walk at least 10 metres independently with or without their walking aid(s) and whom were in their ON-state

when visiting the research centre. Participants were excluded if they were pregnant, allergic to adhesive materials or had a condition that precludes safe participation in assessment (as indicated by referring clinician) or were unable to give consent.

Each participant's date of birth, time since diagnosis, and leg length was recorded. Gait measurements were collected via an IMU attached to the lower spine (Lumbar4 region) by double-sided adhesive tape. Participants in both studies were instructed to walk over a 10-metre walkway free of obstacles at their self-selected walking speed. IMU data was transferred via Bluetooth protocol to a laptop where the data was stored and processed accordingly post assessment. Walking speed was derived from IMU data by using well established algorithms[13]. In detail descriptions of the instrumented walking protocol can be found in Esser et.al., 2011 [4]. The studies involved in data collection were approved by the University Ethics committee and participants consented according to the Declaration of Helsinki.

Data analysis

An experiment was set up to determine how much better our optimal metric learning classifier is at predicting disease labels for both TDA and PwP as compared to a machine randomly assigning a label to each test subject (random guessing). The disease's degree of severity was also estimated as in Fig. 3.

For each IMU sequence an HMM with n=3 states was learned via EM. Since the latter suffers from local minima, the algorithm was applied 10 times to each sequence, retaining the model parameters yielding the highest likelihood. That yielded a dataset of hidden Markov models, each associated with the whole IMU gait sequence captured for a given individual.

Classification of PwP versus TDA

We quantified the performance of our classification algorithm as follows. An optimal pullback distance function is learned by maximising its classification performance on a "training set" of HMMs by cross validation. Then, the Nearest-Neighbour classifier associated with the learned optimal distance is evaluated on a "testing set". In order to produce a robust evaluation result, we randomly generated 25 distinct splits between training and testing sets, and reported the mean performance over the 25 evaluation runs. Each train/test split of the HMM dataset was obtained by randomly sampling two-thirds of the dataset for training and holding the remaining third for testing.

As base distance between two HMMs, $H_1=\{A_1,C_1\}$ and $H_2=\{A_2,C_2\}$, we used the Frobenius norm $|A_1-A_2|_F+|C_1-C_2|_F$, where $|M|_F=\sqrt{Tr(M^T M)}$. No gait cycle from the same individual appeared in both training and testing sets at any time.

In both training and testing each unlabelled HMM was assigned the class of the nearest model in the training set (according to the learned optimal distance).

Severity estimation

Disease severity for PwP was estimated for each test HMM by finding the 5 closest neighbouring HMMs in the training data and averaging the severity levels for those among them with PwP (Fig. 3). In the Results section, good performance is associated with a low Root-Mean-Square-Deviation

(RMSD) of the estimate; an RMSD score of zero signifies that ground truth and predicted severity scores are equal.

Results

Experimental setup

Subject demographics

Gait data from TDA (n=424, mean age 51.9±10.0yrs, range 39-80years) and age matched PwP(n=156, 67.2±8.0yrs, range 39-80) was analysed. Height distribution was found to be 1.70±0.09m for PwP and 1.71±0.10m for TDA. Weight distribution was 76.7±15.2kg for PwP and 76.3±15.5kg for TDA. Those with PD scored a median of 70(range 20-97) on the complete SF-36 with a median Hoehn&Yahr rating of 1(range 0-4). PwP were assessed by the MDS-UPDRS scale on which on average they scored 17 (range 0-57) on the motor section part 3. Furthermore, PwP were found to score an average of 75 (range 0-100) on the physical functioning section of the SF-36.

IMU-derived walking speed was found to be 1.12±0.18ms^{-1} (range 0.59-1.70ms^{-1}) for PwP, and 1.39±0.18ms^{-1} (range 0.86-1.96ms^{-1}) for TDA. Fig. 4 shows the associated normal distributions of speed for the two groups. Their significant overlap shows that simple discrimination based on speed is inadequate to classify mild PD, supporting the need for the more sophisticated metric learning approach proposed.

Results of disease classification and severity prediction

We applied the metric-learning methodology described to the above data.

Classification results

Classification results are here expressed as a 'confusion matrix', which compares predicted (by the classifier) and actual classes of the test samples (Table 1). Results presented are the average over the 25 repeated runs of the classification procedure.

A false positive (fp) occurs when a person is predicted with PD but does not actually have PD. A false negative (fn) occurs when a person is classified as TDA when they actually have PD. The notations (tp) and (tn) denote the numbers of true positive and true negative cases, respectively.

The following measures are typically used to assess classification performance: 'Recall' = tp/(tp+fn); 'Precision' = tp/(tp+fp); 'Accuracy' = (tp+tn)/(tp+tn+fp+fn); and F1 score (harmonic mean of precision and sensitivity):

F1 = 2*tp/(2*tp +fp+fn).

The Accuracy of the proposed classification approach in determining PwP from TDA, averaged over the 25 repeated runs, was (85.51±4.73%), compared to (49.62±3.43%) obtained when assigning TDA/PwP labels at random (when indeed a 50% accuracy is expected). We achieved a mean F1 score (a more reliable performance measure, given the imbalance in the number of TDA and PwP samples) of (81.54±5.92%), compared to the (46.43±3.49%) of random guessing.

Results on severity estimation

The average estimation error (or 'RMSD') of the SF-36 predicted motor severity score was found in our approach to be 27.81±3.07 points on the 0-100 range (i.e., an estimate of 50 could refer to a real score between 22 and 78). In comparison, random assignment produced an RMSD of 39.53±3.84 scale points.

Discussion

This study indicates that our classifier was able to both correctly identify Parkinsonian gait within a large subset of typically developed adults (TDA), and discriminate low from high motor severity scores.

PD discrimination

Our method compares favourably with existing competitors on PD discrimination.

Shukla et al.[13] for instance, studied the SVM classification of postural balance test data by evaluating 24 PwP, without a control group. Their results with respect to medication condition (before and after medication test results) show an accuracy of 64.5%, i.e. 21% lower than ours.

Others have adopted, for example, LS-SVM for classifying PD movements acquired via optoelectronic cameras[15]. These are not mobile sensors and are relatively more expensive than IMU devices. The authors' experimental setup is relatively complicated, while we follow the standard UPDRS rating scale. Finally, they analysed a much smaller cohort compared to ours (14 PwP and 14 TDA).

Cancela et al.[18] collected data from 3-axis accelerometers located on limbs and trunk of 20 PwP. Statistical features were extracted from the collected signals, and off-the-shelf classifiers employed to discriminate PD. The authors achieved classification accuracy in the range 70.83%-75% when analysing walking gaits, and 86.48% accuracy for hand movements. Our approach, instead, is not limited to specific action classes and exhibits an accuracy of 85.51+-4.73% on 156 PwP. Patel et al.[3] analysed 5 PwP, focussing on "heel tapping" and "alternate hand movements" tasks and manually selecting feature measurements based on the action class, which severely limits applicability.

Very significantly, a very recent work by Zhan, et al[30]. conducting a similar large scale PD monitoring from smartphone data (121 PwP and 105 controls) has achieved a 71.0% accuracy.

Severity estimation

Relatively few studies currently employ MLA for the classification of disease severity. Barth et al.[17], for instance, use six different ML classifiers to automatically detect the severity of walking-derived bradykinesia on the UPDRS scale. They use a SHIMMER sensor with integrated gyroscope and accelerometers, and combine multiple gait features. Compared to them, we achieve comparable recall and significantly better specificity (90.35% versus 86%[17]), while covering a significantly larger cohort (156 versus 27). A model has also been proposed to estimate average PD progression using speech signals[14]. However, this work focuses on PD telemonitoring, arguably less challenging than PD diagnosis, does not incorporate healthy controls and is tested on just 42 PwP.

Conclusions

Importantly we have included people with a range of motor severity and found that our framework can cope with larger cohorts of subjects with differing presentations, offering greater ecological validity, while yielding state-of-the-art accuracy, demonstrating a significantly higher generalisation capability than existing methods. Simply our method offers the opportunity to more sensitively classify people with Parkinson's and to offer a methodology for more sensitive monitoring change in

motor symptoms across classifications that could be simply implemented in clinical practice and allow clinicians to monitor the effect of medications and other therapies. Furthermore, as opposed to other works[3, 14, 15], in our empirical validation we used performance measures widely considered more complete and reliable (precision, recall and F1 score). Further supporting this clinical application, we use simple gait measures that only take two minutes to implement, for instance in primary care pathways to support general practitioners, with an accuracy level that even at this early stage is very competitive." Our methods may have application across a number of movement disorders offering a means for clinicians to classify their clients and monitor change across classification levels. The methodology is open to further improvements in all areas, including the use of models with a greater number of states, the adoption of more sophisticated generative models (rather than HMMs), the design of a richer search space of distances to optimize upon[19], all elements that may significantly improve performance further. Collecting additional PwP samples will also lower the imprecision of motor severity level estimates.

References

[1] Horsfall, L., et al., *Time trends in incidence of Parkinson's disease diagnosis in UK primary care.* J Neurol, 2013. **260**(5): p. 1351-7.

[2] Brooks, D.J., *Parkinson's disease: diagnosis.* Parkinsonism Relat Disord, 2012. **18 Suppl 1**: p. S31-3.

[3] Patel, S., et al., *Longitudinal monitoring of patients with Parkinson's disease via wearable sensor technology in the home setting.* Conf Proc IEEE Eng Med Biol Soc, 2011. **2011**: p. 1552-5.

[4] Esser, P., et al., *Assessment of spatio-temporal gait parameters using inertial measurement units in neurological populations.* Gait Posture, 2011. **34**(4): p. 558-60.

[5] Beavers, K.M., et al., *Associations between body composition and gait-speed decline: results from the Health, Aging, and Body Composition study.* Am J Clin Nutr, 2013.

[6] Hamer, M., et al., *Walking speed and subclinical atherosclerosis in healthy older adults: the Whitehall II study.* Heart, 2010. **96**(5): p. 380-4.

[7] Holzreiter, S. and M. Kohle, *Assessment of gait patterns using neural networks.* J Biomech, 2004. **26**(6): p. 645-651.

[8] Lafuente, R., et al., *Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis.* Clin Biomech (Bristol, Avon), 1998. **13**(3): p. 216-229.

[9] Begg, R. and J. Kamruzzaman, *Neural networks for detection and classification of walking pattern changes due to ageing.* Australas Phys Eng Sci Med, 2006. **29**(2): p. 188-95.

[10] Ahlrichs, C. and M. Lawo, *Parkinson's disease motor symptoms in machine learning: a review.* Health Informatics, 2013. **2**(4).

[11] Woon, W.L., et al., *Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings.* Physiol Meas, 2007. **28**(4): p. 335-47.

[12] Joshi, S., et al., *Classification and treatment of different stages of alzheimer's disease using various machine learning methods.* International Journal of Bioinformatics Research, 2010. **2**(1): p. 44-52.

[13] Shukla, A., et al., *Classification of Postural Response in Parkinson's Patients Using Support Vector Machines.* 2013: p. V002T22A004.

[14] Tsanas, A., et al., *Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity.* J R Soc Interface, 2011. **8**(59): p. 842-55.

[15] Chan, E.Y. Leung, and H. Poizner, *Correlation among joint motions allows classification of Parkinsonian versus normal 3-D reaching.* Neural Systems and Rehabilitation Engineering, IEEE Transactions, 2010. **18**(2): p. 142-149.

[16] Maja, S., *Towards less supervision for scalable recognition of daily activities*. 2010, Technische Universitat Darmstadt.

[17] Barth, J., et al., *Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease.* Conf Proc IEEE Eng Med Biol Soc, 2011. **2011**: p. 868-71.

[18] Cancela, J., et al., *A comprehensive motor symptom monitoring and management system: the bradykinesia case.* Conf Proc IEEE Eng Med Biol Soc, 2010. **2010**: p. 1008-11.

[19] Cuzzolin, F. and M. Sapienza, *Learning pullback HMM distances.* IEEE TRANS. ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2014. **36**(7): p. 1483-1489.

[20] Shi, Q., L. Wang, and L. Cheng, *Discriminative human action segmentation and recognition using semi-Markov model.* Proc. CVPR'08, 2008: p. 1-8.

[21] Chaundhry, R., et al., *Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions.* Proc. CVPR'09, 2009: p. 1932-1939.
[22] Ali, S., A. Basharat, and M. Shah, *Chaotic Invariants for Human Action Recognition.* Proc. ICCV'07, 2007.
[23] Elliott, R., L. Aggoun, and J. Moore, *Hidden Markov models: estimation and control*. Vol. 29. 1995: Springer.
[24] Lee, Y.-S. and S.-B. Cho, *Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer.* Hybrid Artificial Intelligent Systems, 2011. **6678**: p. 460-467.
[25] Do, M., *Fast Approximation of Kullback–Leibler Distance for Dependence Trees and Hidden Markov Models.* IEEE SIGNAL PROCESSING LETTERS, 2003. **10**(4): p. 115-118.
[26] Mensink, T., et al., *Metric Learning for Large Scale Image Classification: Generalizing to New Classes at Near-Zero Cost.* Computer Vicion - ECCV, 2012: p. 488-501.
[27] Shental, N., et al., *Adjustment learning and relevant component analysis.* Proc. ECCV'02, 2002: p. 776–792.
[28] Lebanon, G., *Metric Learning for Text Document.* IEEE Trans Pattern Analysis & Machine Intelligence, 2006. **28**(04): p. 497-508.
[29] Ware, J.E., Jr. and C.D. Sherbourne, *The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection.* Med Care, 1992. **30**(6): p. 473-83.
[30] Zhan, A., et al., *High Frequency Remote Monitoring of Parkinson's Disease via Smartphone: Platform Overview and Medication Response Detection.* 2016.

[31] Daliri, Mohammad Reza. *Automatic diagnosis of neuro-degenerative diseases using gait dynamics*. Measurement 45.7 2012. p. 1729-1734.

[32] Daliri, Mohammad Reza. *Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease*. Biomedical Signal Processing and Control 8.1 2013. p. 66-70.

[33] Khorasani, Abed, and Mohammad Reza Daliri. *HMM for classification of Parkinson's disease based on the raw gait data*. Journal of medical systems 38.12 2014. p. 1-6.

[34] Khorasani, Abed, Mohammad Reza Daliri, and Mohammad Pooyan. Recognition of amyotrophic lateral sclerosis disease using factorial hidden Markov model. Biomedical Engineering/Biomedizinische Technik 61.1 2016. P. 119-126.

Metric learning for Parkinsonian identification from IMU gait measurements

Tables

Table 1: Average confusion matrices for classifying PwP vs. TDA.

HMM Metric Learning		**Predictions**	
		PwP	*TDA*
True labels	*PwP*	tp=37.7	fn=14.3
	TDA	fp=13.6	tn=127.4

Random		**Predictions**	
		PwP	*TDA*
True labels	*PwP*	tp=24.6	fn=27.4
	TDA	fp=69.8	tn=71.2

Metric learning for Parkinsonian identification from IMU gait measurements

Figures:

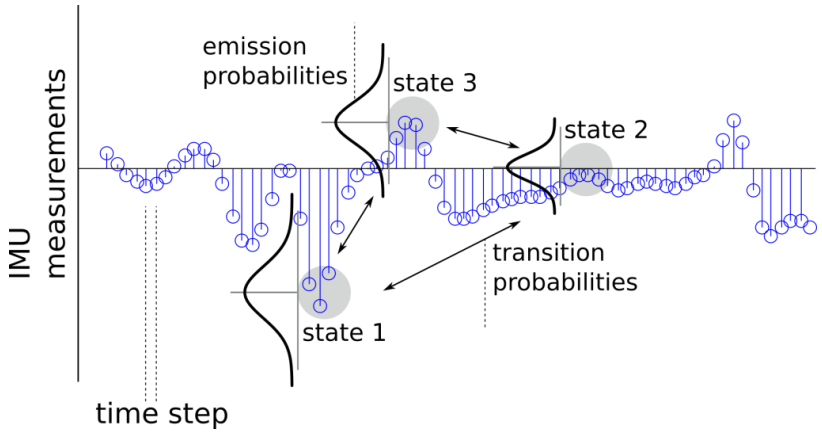

Fig. 1. Pictorial representation of an HMM encoding an IMU sequence.

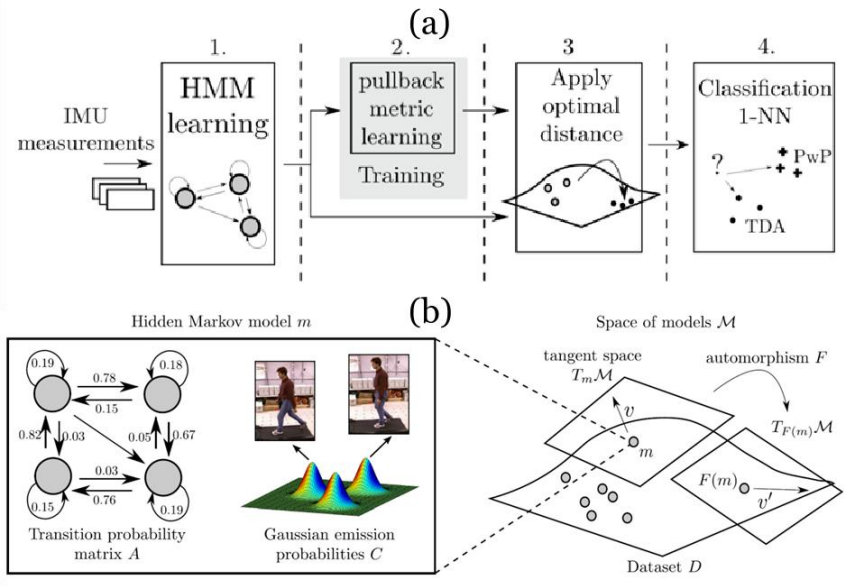

Fig. 2. **(a)** Overview of the metric learning algorithm proposed in [19] for time-series classification; **(b)** in pullback metric learning each training HMM (left) is a point in the space of models M (right). Given a 'base' distance on M, any differential stretching F of M generates a 'pullback' distance there. Any parameterised family of such stretchings induces a family of distances on M, among which we can select that achieving maximal classification performance on the training set.

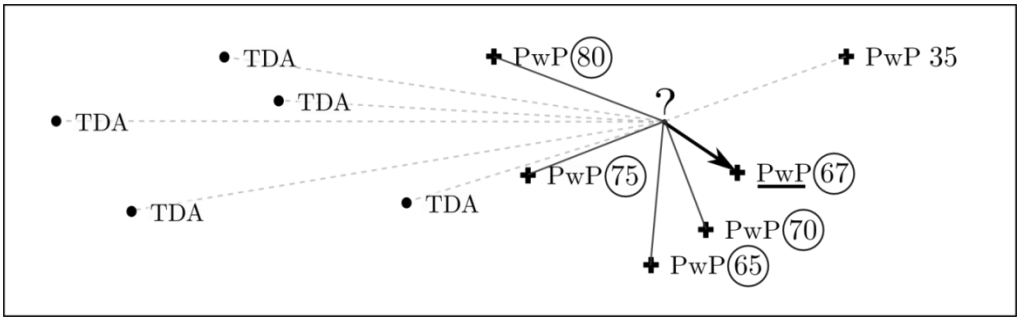

Fig. 3. Disease severity estimation.

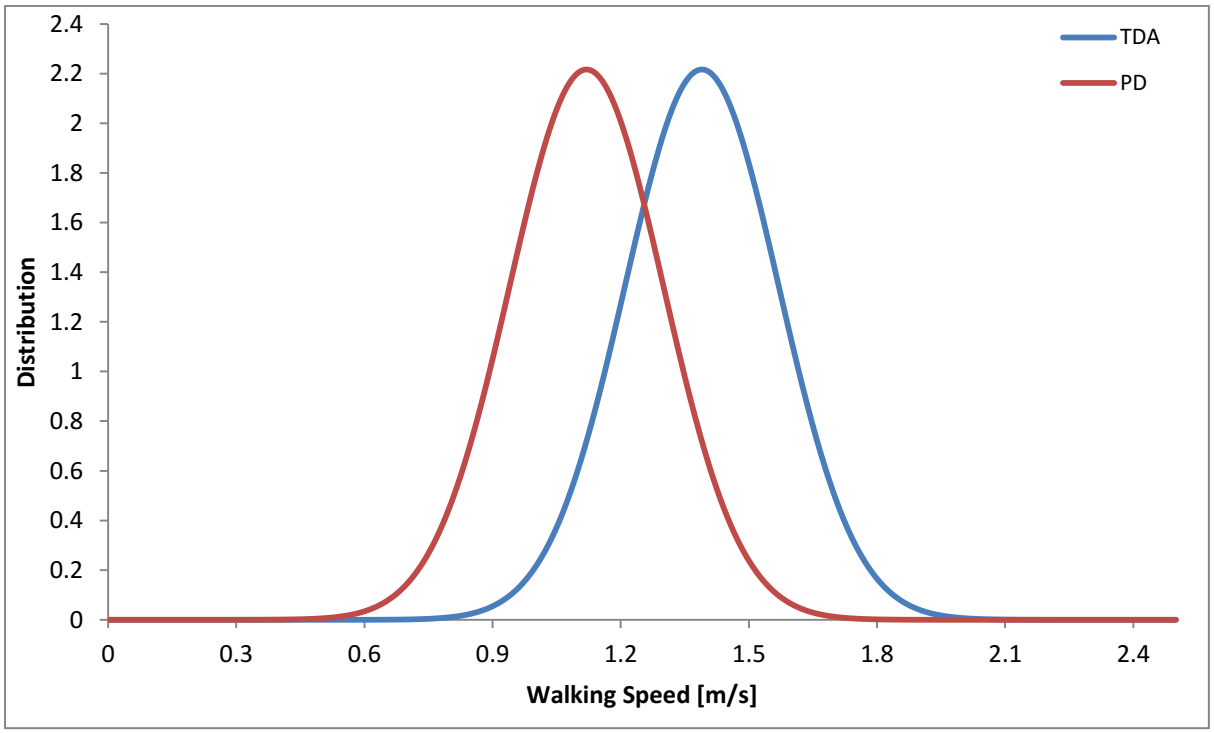

Figure 4. Empirical normal distribution for walking speed of Parkinson's and TDAs in our tests.

Highlights

- A metric learning approach to identify Parkinsonian gait from IMU data is proposed.
- The approach learns the best classification strategy for the given training data.
- Consequently, it can cope with larger cohorts with better generalisation power.
- We achieve 85.51% accuracy over 580 subjects, the best yet over such large cohort.