Journal of Neurology (2019) 266 (Suppl 1):S108–S117
https://doi.org/10.1007/s00415-019-09458-y

ORIGINAL COMMUNICATION

Towards computerized diagnosis of neurological stance disorders: data mining and machine learning of posturography and sway

Seyed-Ahmad Ahmadi[1,2] · Gerome Vivar[1,2] · Johann Frei[1,2] · Sergej Nowoshilow[3] · Stanislav Bardins[1] · Thomas Brandt[1] · Siegbert Krafczyk[1]

Received: 11 June 2019 / Revised: 28 June 2019 / Accepted: 30 June 2019 / Published online: 8 July 2019

Abstract

We perform classification, ranking and mapping of body sway parameters from static posturography data of patients using recent machine-learning and data-mining techniques. Body sway is measured in 293 individuals with the clinical diagnoses of acute unilateral vestibulopathy (AVS, $n=49$), distal sensory polyneuropathy (PNP, $n=12$), anterior lobe cerebellar atrophy (CA, $n=48$), downbeat nystagmus syndrome (DN, $n=16$), primary orthostatic tremor (OT, $n=25$), Parkinson's disease (PD, $n=27$), phobic postural vertigo (PPV $n=59$) and healthy controls (HC, $n=57$). We classify disorders and rank sway features using supervised machine learning. We compute a continuous, human-interpretable 2D map of stance disorders using t-stochastic neighborhood embedding (t-SNE). Classification of eight diagnoses yielded 82.7% accuracy [95% CI (80.9%, 84.5%)]. Five (CA, PPV, AVS, HC, OT) were classified with a mean sensitivity and specificity of 88.4% and 97.1%, while three (PD, PNP, and DN) achieved a mean sensitivity of 53.7%. The most discriminative stance condition was ranked as "standing on foam-rubber, eyes closed". Mapping of sway path features into 2D space revealed clear clusters among CA, PPV, AVS, HC and OT subjects. We confirm previous claims that machine learning can aid in classification of clinical sway patterns measured with static posturography. Given a standardized, long-term acquisition of quantitative patient databases, modern machine learning and data analysis techniques help in visualizing, understanding and utilizing high-dimensional sensor data from clinical routine.

Keywords Neurological stance and gait disorders · Static posturography · Body sway · Machine learning · Visualization

Abbreviations

ANN	Artificial neural network
AVS	Acute unilateral vestibulopathy
CA	Anterior lobe cerebella atrophy
CI	Confidence interval
DN	Downbeat nystagmus syndrome
kNN	*k*-Nearest-neighbors
MDI	Mean decrease in impurity
HC	Healthy controls
OT	Primary orthostatic tremor
PCA	Principal component analysis
PD	Parkinson's disease
PNP	Sensory polyneuropathy
RMS	Root-mean-square
SC	Stacking classifier
SVM	Support vector machine
t-SNE	t-Stochastic neighborhood embedding

Seyed-Ahmad Ahmadi and Gerome Vivar contributed equally to this work.

This manuscript is part of a supplement sponsored by the German Federal Ministry of Education and Research within the funding initiative for integrated research and treatment centers.

✉ Seyed-Ahmad Ahmadi
aahmadi@med.lmu.de

1 German Center for Vertigo and Balance Disorders, Ludwig Maximilians Universität, Marchioninistr. 15, 81377 Munich, Germany

2 Computer Aided Medical Procedures, Technical University of Munich, 85748 Garching, Germany

3 IMP Research Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria

Introduction

Postural imbalance is an ambiguous and key symptom of various neurological disorders, including acquired loss of sensory function (visual, vestibular or somatosensory), cerebellar or extrapyramidal dysfunction due to degenerative brain disorders, or somatoform functional disorders [1-3]. Some of these conditions can be easily and reliably diagnosed by associated neurological symptoms. Others are difficult to differentiate clinically, and require further quantitative assessment of stance and gait. To this end, static posturography [4-6] records the stance and sway behavior of patients under increasingly difficult conditions, while standing on a force measurement platform. In static posturography, body sway (i.e., center-of-pressure displacements) is measured and scored based on so-called features, i.e., characteristic parameters quantifying sway path amplitude and variance, or tremors in various frequency bands using Fourier analysis. As shown in our own previous work [7], as well as the work of others [8-10], diagnostic classification can be computerized using multi-parametric statistical analysis [11] or supervised machine-learning algorithms like artificial neural networks (ANNs) [7].

Here, we attempt the application of modern machine-learning and data-mining techniques to 8 subject classes on a larger clinical dataset with 293 individuals, recruited from our neurology department, as well as our tertiary interdisciplinary outpatient dizziness unit, the Germen Center for Vertigo and Balance Disorders. The cohort includes healthy controls, and the pathological classes acute unilateral vestibulopathy (former: acute vestibular syndrome, or vestibular neuritis) [12], cerebellar disorders (anterior lobe cerebellar atrophy [13], downbeat nystagmus syndrome [14]), somatosensory deficits (sensory polyneuropathy) [8], postural tremor (orthostatic tremor) [15, 16], extrapyramidal dysfunction (Parkinson's disease) [17] and functional disorders (Persistent Postural-Perceptual Dizziness, in its sub-form phobic postural vertigo) [18].

Our analysis of posturographic sway has three objectives: (1) classification of static posturography signal patterns using modern supervised machine-learning techniques, towards computerized diagnosis for above-mentioned eight classes. Compared to many works classifying only one or two types of disorders [8, 9, 19, 20], often against healthy controls, we aim at a more general classifier distinguishing the above-mentioned eight classes. (2) We calculate the discriminative power and perform a *ranking* of the 10 posturographic stance conditions and their derived individual sway features to better assess the reasonability of the experimental paradigm. (3) We create a mapping of posturographic sway patterns of subjects by non-linear projection of high-dimensional sway patterns into a 2D map. This allows us to collectively and visually analyze the distribution of sway patterns from all subjects in our cohort. We give explanations and pointers towards improvement of the computerized diagnostic scheme in the discussion for routine clinical use.

Methods

Patient cohorts

For this study, we included a cohort of 293 adults with patients ($n = 236$) and healthy controls ($n = 57$), who performed posturographic examination under the same examination protocol. Group statistics are detailed in Table 1. Presumptive clinical diagnoses were determined by expert physicians from a neurological clinic and policlinic (Dept. of Neurology, University of Munich, Germany) and a tertiary interdisciplinary outpatient dizziness unit (German Center for Vertigo and Balance Disorders, Munich, Germany). Patients in the primary orthostatic tremor (OT) group were diagnosed according to criteria defined in [21, 22], with a focus on a high-frequency body sway (11–19 Hz) present already while standing with eyes open on firm ground. Downbeat nystagmus (DN) was diagnosed with the aid of quantitative video-oculography according to criteria defined in [23]. Acute unilateral vestibulopathy (AVS) was diagnosed following guidelines in [24]. Parkinson's disease (PD) patients were clinically diagnosed [25] and selected to only include patients with a tremor-dominant type (4–8 Hz). Anterior lobe cerebellar atrophy (CA) was based on cerebellar symptoms during neurological examination and confirmation with MR imaging (atrophy in the vermal and anterior lobe region of the cerebellum). Polyneuropathy (PNP) patients presented a distal, symmetric sensorimotor loss, according to guidelines in [26]. Patients belonging to the

Table 1 Details of patient cohorts

Diagnosis	Anterior lobe cerebellar atrophy	Downbeat nystagmus	Acute unilateral vestibular syndrome	Poly-neuropathy	Phobic postural vertigo	Parkinson's disease	Primary orthostatic tremor	Healthy control
Diagnosis abbreviation	CA	DN	AVS	PNP	PPV	PD	OT	HC
N	48	16	49	12	59	27	25	57
Age mean (s.d.)	63.0 (10.5)	62.3 (16.1)	54.5 (12.9)	64.7 (9.7)	50.1 (15.7)	61.8 (12.7)	62.8 (10.1)	31.0 (10.8)
Male/female	37/11	8/8	33/16	8/4	22/37	15/12	14/11	31/26

postural phobic vertigo (PPV) group were recruited earlier than 2017, i.e., before the official consensus criteria for Persistent Postural-Perceptual Dizziness (PPPV) were defined [3], hence the diagnostic process followed guidelines in [27]. For this study, neither disease severity nor comorbidities could be considered, instead we represented patients with respect to their primary diagnosis. All patients participated under written consent, the study was authorized by the ethics committee of the medical faculty, Ludwig-Maximilians-University, Munich, and conducted in conformity with the Declaration of Helsinki.

Instrumentation

Posturographic examination

We followed the posturographic examination protocol of [7]. Subjects were examined while standing on a stabilometer platform (Type 9261 A; Kistler, Winterthur, Switzerland). The protocol consisted of ten trials during which subjects had to maintain balance in upright stance, with arms hanging, and under conditions with increasing difficulty. Posturographic stance conditions were: (1) eyes open; (2) eye closed; (3) eyes open, head in neck; (4) eyes closed, head in neck; (5) eyes open, standing on foam-rubber block; (6) eyes closed, foam-rubber block; (7) eyes open, head in neck, foam-rubber block; (8) eyes closed, head in neck, foam-rubber block; (9) eyes open, foam-rubber block, tandem stance; (10) eyes closed, foam-rubber block, tandem stance.

Signal processing and feature extraction

The posturography sensor hardware records raw data of fore-aft (y) and lateral (x) body sway and body weight (z), at a sampling frequency of 40 Hz. Each trial was recorded for a duration of 30 s. Machine-learning classification was performed on a set of discriminative sway features. For each trial, recorded as a 3-dimensional time series of $x/y/z$ raw data, 18 features were computed [7], which are summarized in Table 2. A full examination with 10 trials thus yields a 180-dimensional feature vector (10 trials with 18 features each) for each subject. The features represent accumulated sway path in x, y and z directions [28], root-mean square values in x, y and z directions [29], and spectral energy magnitudes [7] in different frequency bands, which were computed after application of a Hamming window, followed by discrete Fourier analysis (MATLAB, MathWorks Inc., USA). Feature abbreviations, descriptions, and formulae are summarized in Table 2.

Classification using supervised machine learning

Our procedure for supervised machine learning consists of data pre-processing, cross-validation and classification. Pre-processing is performed by normalization of the data range for each sway feature. Normalization applied a transformation to zero-mean ($\mu = 0$) and unit-variance ($\sigma^2 = 1$) distribution for each sway feature. For assessment of classification robustness, we utilized a stratified k-fold cross-validation, which provides training- and test-set splits that compensate for the class imbalance of our dataset. To obtain robust cross-validation statistics, we utilized $k = 50$ randomized and stratified splits at 90% training vs. 10% test data for each class.

Classification with supervised machine learning and ensemble models

As classifiers, we used different machine-learning algorithms which are often applied in high-dimensional data analysis [30]. Additionally, we compare the selected algorithms' performance on our new dataset to the previously

Table 2 List of spatial and spectral features extracted from posturographic raw data

Feature abbreviation	Description	Formula (if applicable)
swaypath_x	Accumulated sway path distance in x-direction	$\mathrm{sp}_x = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2}$
swaypath_y	Accumulated sway path distance in y-direction	$\mathrm{sp}_y = \sum_{i=1}^{n-1} \sqrt{(y_{i+1} - y_i)^2}$
swaypath_z	Accumulated sway path distance in z-direction	$\mathrm{sp}_z = \sum_{i=1}^{n-1} \sqrt{(z_{i+1} - z_i)^2}$
rms_x	Root-mean-square in x-direction	$\mathrm{RMS}_x = \sqrt{1/n \sum_{i=1}^{n} x_i^2}$
rms_y	Root-mean-square in y-direction	$\mathrm{RMS}_y = \sqrt{1/n \sum_{i=1}^{n} y_i^2}$
rms_z	Root-mean-square in z-direction	$\mathrm{RMS}_z = \sqrt{1/n \sum_{i=1}^{n} z_i^2}$
fft_$x/y/z$_01_24	Fourier spectral energy integrals in frequency band: 0.1–2.4 Hz	FFT of $x/y/z$ series
fft_$x/y/z$_24_35	2.4–3.5 Hz	FFT of $x/y/z$ series
fft_$x/y/z$_35_8	3.5–8.0 Hz	FFT of $x/y/z$ series
fft_$x/y/z$_11_19	11.0–19.0 Hz	FFT of $x/y/z$ series

proposed artificial neural network (ANN) configuration from [7]. We compare logistic regression (LogRegr) [31], *k*-nearest-neighbors (kNN) [32], a single-hidden-layer perceptron ANN (ANNsingle [7]), a deeper multi-layer perceptron ANN with three hidden layers (ANNmulti) [33], support vector machines (SVM) [34], random forests (RandForest [35]), and extra-randomized trees (ExtraForest [36]). For our experiments, we used implementations provided in the scikit-learn framework for pre-processing, cross-validation and classification [37]. In addition to regular classifiers, we employed an ensemble-learning technique called Stacking Classifier (SC [38]), which combines probabilistic classifications of all above-mentioned classifiers into a stronger and more robust meta-classifier, with logistic regression as the meta-learning algorithm.

Evaluation scores

Cross-validation results are evaluated according to accuracy, which summarizes the multi-class classification in a single scalar according to the following formula:

$$\mathrm{Acc}(c_{\mathrm{true}}, c_{\mathrm{pred}}) = \frac{1}{n_{\mathrm{samples}}} \sum_{i=1}^{n_{\mathrm{samples}}} 1(c_{i,\mathrm{true}} = c_{i,\mathrm{pred}}),$$

where c_{true} is the true class of a sample, c_{pred} is the predicted class, and $1(x)$ is the indicator function. Furthermore, we report confusion matrices with detailed (mis-)classification rates, as well as sensitivities and specificities in results.

Ranking of stance conditions and sway feature importance

Analyses on feature importance and embedding were performed using the ExtraForest algorithm [36]. Feature importance was calculated according to Mean Decrease in Impurity (MDI [39]). In MDI, each feature importance is computed as the number of splits that involve the feature, summed over all decision trees in the ensemble, and weighted by the number of samples that were split by the feature. We rank the discriminative power of the ten stance conditions and of the extracted sway features by computing relative MDI coefficients, scaled by the most important feature in each group.

Mapping disorders of stance using posturography

We calculate a 2D map of the sway pattern distribution in our cohort using t-Stochastic Neighborhood Embedding (t-SNE [40]) a dimensionality reduction technique designed for visualization of high-dimensional datasets. Unlike, e.g., Principal Component Analysis (PCA), t-SNE performs a non-linear projection of the 180-dimensional feature vectors of each subject into the 2D plane for visualization. The projection preserves the stochastic distribution of data points in high-dimensional space into the low-dimensional (2D) space, by aligning their respective distributions. Alignment is achieved by maximization of Kullback–Leibler divergence [40]. The 2D map allows for simultaneous assessment of the relative sway similarity of all subjects in our dataset, along with the distribution of their diagnostic classes. Two main parameters, dimensions and perplexity, affect the layout of the resulting 2D map. As discussed in the original paper, t-SNE is relatively robust to these parameters, we hence use a standard parametrization (see "Results").

Statistical analysis

Data were analyzed using the Python programming language and open-source modules for scientific computing, in particular scipy-stats [41] for statistical testing and scikit-learn [37] for machine-learning experiments. Comparison between the best- and second-best performing classifiers (SC vs. ExtraForest) was performed using Wilcoxon signed-rank test, the level of significance was set at $p < 0.05$. Sensitivities and specificities of diagnostic classification were calculated based on 50-fold randomized, stratified cross-validation, the best-performing algorithm's accuracy is reported with 95% confidence intervals.

Results

Supervised machine learning for classification and feature ranking

Classification results

Classification accuracy is evaluated with 50 randomized folds of stratified cross-validation (90% training vs. 10% test data). Mean classification accuracies for single classifiers range from 64.5% (kNN) to 80.7% (ExtraForest). Across the 50 cross-validation runs per algorithm, the maximum accuracy ranges from 76.7% (kNN) to 93.3% (ANNMulti, RandForest, ExtraForest).

The application of a Stacked Classifier (SC) increases the confidence and robustness of the classification further, raising the mean classification accuracy from 80.7% (ExtraForest) to 82.7% (95% CI: [80.9%, 84.5%]). The difference of paired accuracies between the SC and the second-best performing method (ExtraForest) is statistically significant ($p = 0.015$, Wilcoxon signed-rank test). A box-plot of classification accuracies is depicted in Fig. 1.

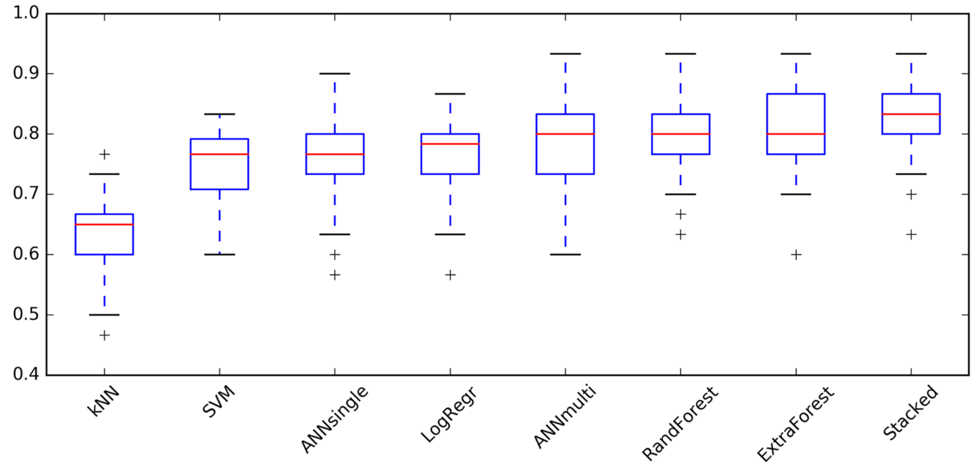

Fig. 1 Cross-validated results from different supervised classification algorithms, evaluated on the entire dataset. Classification accuracies (denoted in %/100) are sorted by mean value in ascending order, with box plots indicating confidence intervals over 50 stratified random-split cross-validation with 10% hold-out test data. The most accurate classifier (Stacked Classifier, SC) shows significantly better performance than the second-best algorithm RandForest (p = 0.015, Wilcoxon signed-rank test). Detailed classification results of SC on individual diagnostic classes are depicted in Fig. 2

Confusion matrices

To better analyze class separation, we compute normalized confusion matrices, highlighting true labels vs. predicted labels and the respective percentages. While the mean accuracy of the Stacked Classifier (SC) in the 8-class setup yields an overall accuracy of 82.7% (95% CI: [80.9%, 84.5%]), confusion matrices reveal three classes with true positive rates below 70%, namely DN, PD and PNP (cf. Fig. 1, left panel).

We additionally compare the Stacking Classifier, our best-performing method, to the classification scenario in [7], using five classes CA/AVS/NC/PPV/OT only. Figure 2 depicts the confusion matrix. Given five classes only, SC classification yields a comparable overall specificity of 97.4%, compared to 98.4% in [7]. Sensitivity, however, fairs slightly lower at 89.5%, compared to 93.4% in [7], mainly due to lower sensitivity for OT (80.0%, compared to 100%).

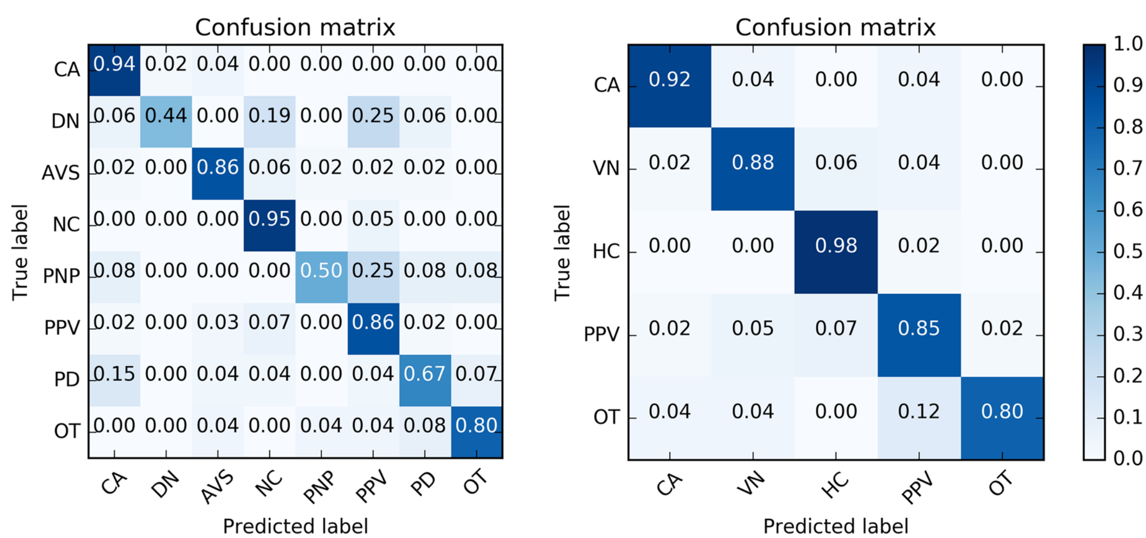

Fig. 2 Confusion matrices (true label vs. predicted label) of the Stacking Classifier algorithm for the 8-class, (left panel) and 5-class (right panel) Stacking Classifier (see Table 1 for abbreviations of disorders). Classes CA/VN/NC/PPV/OT show consistently high classification accuracy (see main diagonal of confusion matrices) with sensitivities above 80%, while classes DN/PNP/PD are more challenging to classify (left panel: classification accuracies below 70%)

Ranking of stance conditions and feature importance

Altogether, 18 features from 10 trials (180 features total) are ranked by their discriminative power derived from the MDI criterion in the ExtraForest classifier. We marginalize their importance by stance condition, and by feature type, and visualize their relative feature importance in Fig. 3.

As seen in Fig. 3, left panel, trial 6 (eyes closed, frontal stance on foam-rubber block) is evaluated with highest accumulated importance. The next two most important experiments are trials 5 (eyes open, foam-rubber block) and trial 7 (eyes open, backward head extension, foam-rubber block). Furthermore, features from two of the presumably most difficult trials 8 (eyes closed, backward head extension, foam-rubber block) and 10 (eyes closed, foam-rubber block, tandem stance) are ranked comparatively low (< 50% of importance of trial 6).

As seen in Fig. 3, right panel, the highest ranked features are accumulated swaypath lengths in x/y/z direction. The next important feature is rms_z, i.e., vertical (weight) variation. Among spectral features, horizontal x/y energies in the bands 0.1–2.4 Hz and 2.4–3.5 Hz are most influential, along with vertical tremor energy in the band 3.5–8 Hz. In the high-frequency band 11–19 Hz, lateral (x) and vertical (z) tremor energies are more central to classification than fore-aft sway tremor energy.

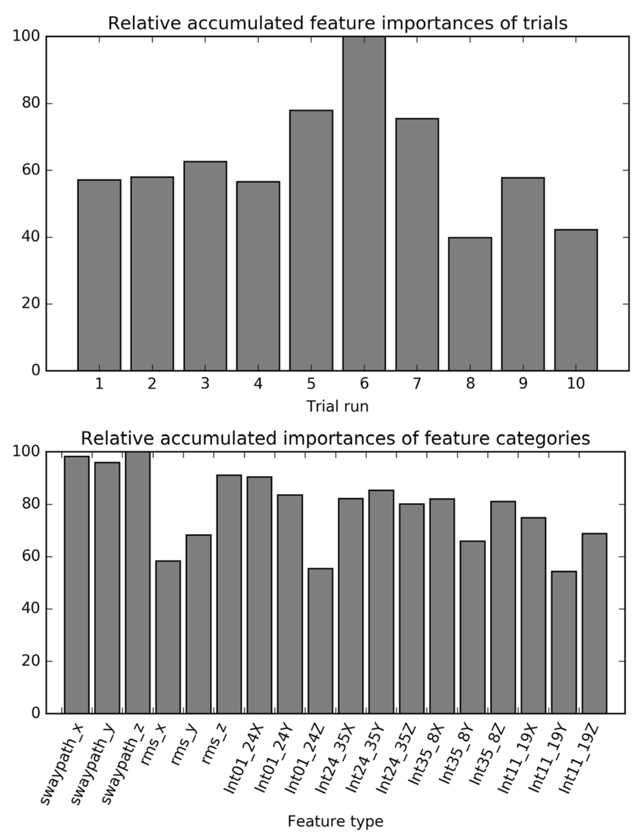

Fig. 3 Relative accumulated importance of stance conditions 1–10 (top panel) and of posturographic feature types (bottom panel)

Mapping sway path features using t-SNE

Computation of t-SNE dimensionality reduction and 2D mapping [40] is performed with a configuration initial dimension $i=20$ and perplexity $p=25$. The resulting 2D map of posturographic sway behavior can be seen in Fig. 4 in form of one comprehensive plot with color coded class distributions, as well as one-vs-all colorings for each diagnostic class.

As seen in the colored visualization, and highlighted in the one-vs-all colored subpanels, the well-classified patient groups of CA, AVS, NC and PPV form well-defined clusters in the 2D embedding space. Compared to these, classes DN, PNP and PD are not grouped together and sparsely scattered across the 2D space. Implications of these observation are discussed below.

Discussion

Classification results of various stance disorders using supervised machine learning

In the separation of five diagnostic classes, supervised classification with Stacked Classifier meta-learning yielded consistently good classification of CA, AVS, NC and PPV and OT. A mean sensitivity and specificity of 88.4% and 97.1% for all five conditions was obtained, which is comparable with previously published classification accuracies [7] and on the order of inter-observer variability across expert raters (0.86) [20]. Given that these results stem from a three times larger dataset than before and a statistically more robust 50-fold stratified random cross-validation [7], we see this study as a solid confirmation that supervised learning techniques can be used to “differentiate postural sway patterns typical of several distinct clinical balance disorders with sufficiently high sensitivity and specificity for clinical use” [7].

In comparison, regarding the classification into eight diagnostic groups, the three classes DN, PNP and PD have significantly lower true-positive classification rates at 44%, 50% and 67%, respectively (cf. Fig. 2 , left). We conclude that the choices of classifiers and their parametrization are not the decisive factor for low classification accuracy of classes DN, PD or PNP. Instead, we argue that while the extracted sway features are sufficient to discriminate the main classes, they still fail to provide enough saliency to distinguish classes such as DN, PD or PNP.

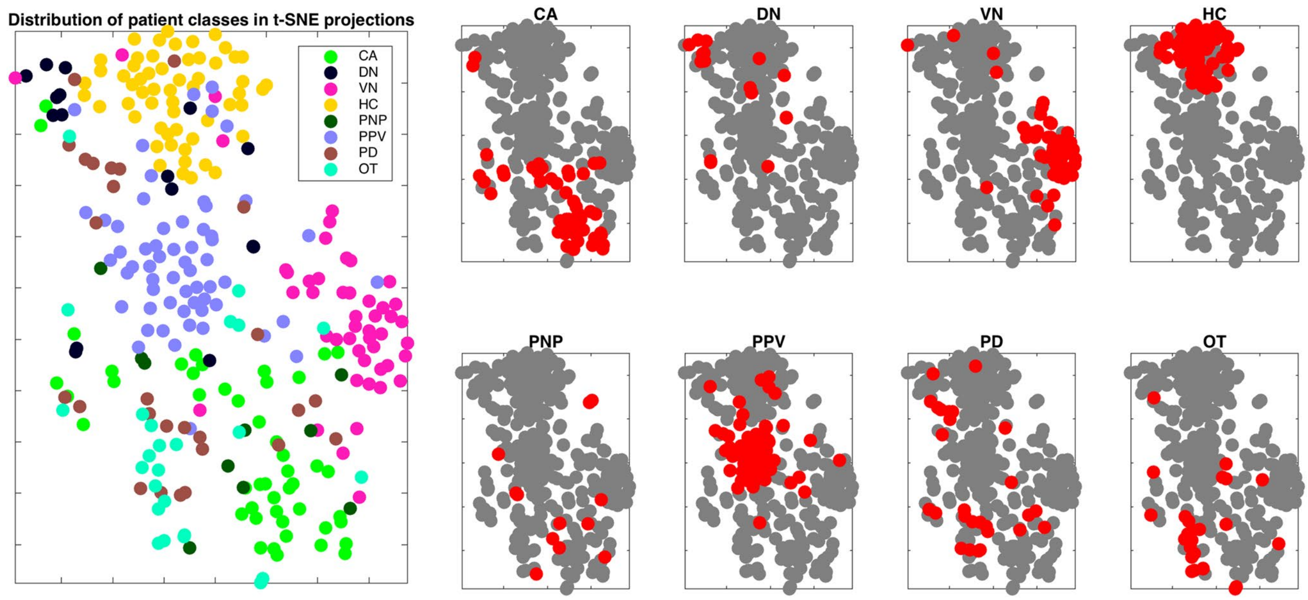

Fig. 4 Distribution of high-dimensional sway features for all study subjects, projected into 2D space via t-SNE (x-/y-axes represent relative similarity between subjects by proximity, and are unit-less). Left panel: color coding of eight diseases reveal clear clustering of disorders CA, AVS, NC, PPV. Right subpanels: one-vs-all color coding for all eight classes highlight weak clustering in disorders DN, PD and PNP

Ranking of stance conditions and feature importance

In our study, each patient is represented as a 180-dimensional feature vector. While all these are derived from extensive expert knowledge [7, 8, 10], they do not carry the same discriminative power. Figure 3 showed relative feature importance of the ten stance conditions and 18 feature categories (see Table 2). Concerning features, swaypath_*x*/*y*/*z* and spectral energies in lower bands, in particular 2.4–3.5 Hz are most relevant. Concerning stance conditions, the five most important conditions are 6, 5, 7, 3 and 9. Four of these are acquired during trials with the condition "eyes open", or while standing on foam-rubber block. Despite their lower importance, trials 8 and 10 still contribute to classification performance, and should only be left out of the diagnostic protocol if examination time needs to be strictly reduced.

Mapping of sway patterns using t-SNE for visualization of disease specific data clustering

The non-linear projection of 180-dimensional features into a 2D map via t-SNE allows us for the first time to assess the distribution of posturographic sway patterns for close to 300 subjects simultaneously in one plot. The resulting 2D map reveals naturally occurring clusters of stance disorders in our cohort. It should be noted that t-SNE achieves this mapping in an entirely unsupervised fashion, i.e., without a-priori knowledge of specific symptoms or disorders of the tested individuals.

Class distributions in t-SNE space, visualized in Fig. 4, reveal clusters for the well-detectable disorders NC, CA, PPV, AVS and a subset of OT. Disease classes get grouped together solely based on the stochastic similarity of sway patterns, which confirms their high discriminative power. In contrast, classes DN, PNP and PD are sparsely scattered across the 2D map, indicating a low similarity of patients with respect to their sway path feature representation. In high-dimensional feature space, patients from these classes are similarly scattered, which is the reason why supervised classification algorithms fail to find sufficiently separating boundaries between diagnoses, regardless of the concrete classifier type.

An interesting special case is class OT, where classification sensitivity dropped from 100 to 80% compared to [7]. In Fig. 4, half of subjects in class OT ($n=13$) are well clustered in the lower left of the data plane, while remaining OT subjects ($n=12$) are more scattered. This coincides with findings in [16], where up to 50% of OT subjects were identified to lack characteristic discharges at OT-typical peak frequencies of 13 Hz [15]. Furthermore, most of the off-cluster OT subjects reside close to subjects from class CA. Interestingly, a recent study [42] found evidence for cerebellar origins of orthostatic tremor, which is further discussed and explained in [16]. Multimodal analysis based on PET imaging, electromyography and posturography revealed pathological ponto-cerebello-thalamo-cortical activations in primary orthostatic tremor during lying and stance. Similarity of posturographic sway patterns between subjects in classes OT and CA may, therefore, be due to comorbidity, which was not considered in our diagnostic categorization.

Limitations and clinical implications

One limitation of this study is the large age difference between the HC group and the seven disease groups. The high classification accuracy of HC is thus partly explainable by age difference, as aging effects can cause a decline in biomechanical or sensory function, and as such have been shown to considerably affect static postural function [43, 44]. Therefore, a control group with a higher age average would probably lead to a slight decrease in classification accuracy. From a clinical perspective, however, the main classification goal is the differentiation of disease groups, and a relatively high separation of HC to the rest has little consequence on the inter-class separability of the remaining groups. As such, it is safe to assume that the main conclusions of this study are not affected by the age difference. In future work, demographic differences such as age, gender, as well as basic clinical scores could be incorporated into the classification to analyze their effect on differential diagnosis across disease groups.

Apart from HC, our results indicate that not all disorders of stance and gait can be classified with an equally high sensitivity and specificity. We see several potential reasons for this.

First, Fig. 4 showed that while certain disorders are well-clustered (CA/AVS/NC/PPV), the less distinguishable diagnoses (DN, PNP, PD) are distributed sparsely and interleaved with other classes. We hypothesize that there is room for classification improvement, if a more specific feature representation or additional stance conditions (e.g., leaning forward or backward; balancing on one foot) can be found for these diagnoses. Simply put, the number and type of features and conditions might not yet be sufficient to optimally distinguish certain disorders. Spatial sway features (swaypath, RMS) and spectral coefficients (FFT energy integrals) considered in this and previous work are global scalars computed over the entire 30 s examination window for each trial, while the temporal dynamics of motion might be of relevance, such as characteristic sway dynamics under proprioceptive blocks during the foam-rubber block condition. Further features might be required to model external factors, such as the number of holds of a patient by the examiner during examination, to avoid falls, which are particularly frequent in elderly patients and patients with PD [9, 17]. The number of holds, if not well documented by examination staff, could be extracted from the z-amplitude, since a fall-avoiding hold causes a dip in body weight.

Second, as mentioned in the patient cohort description, we neither considered disease severity nor comorbidity in this study. In particular, for those disease groups with low classification accuracy (i.e., DN, PD, and OT), clinically known heterogeneity might explain why they do not cluster as well as the other classes. For example, OT might be caused by a cerebellar pathology and thus linked to CA, as evidenced by multi-modal data with nuclear imaging [42] and longitudinal posturographic examination [45]. Further, there is evidence for a deficit in lower limb somatosensation in OT [46, 47]. Dichotomous classification, as applied in this study, might thus be too simple an assumption for future studies and for application in the clinic. Instead, a guideline for incorporation of machine learning into clinical routine would be to rely on probabilistic outputs of classifiers rather than hard labels and investigation of multi-class, probabilistic prediction for modeling of cases with comorbidity.

Third, a straightforward limitation is that supervised classification in this study is solely determined by a single modality, i.e., features extracted purely from static posturography. As discussed above, balance maintenance in humans is dependent on an interplay of highly multi-modal body sensors and brain functions, which should be matched by a multi-modal examination paradigm. Dynamic posturography [19] for example considers force platform stabilometry, under simultaneous monitor-based visual stimuli and external foot support actuation for balance manipulation. Further modalities can be thought of, such as demographics, questionnaires, video-oculography [48] and medical imaging [49]. Combination of static and dynamic posturography with ANN-based gait analysis [50] is most promising, since patients with postural imbalance also suffer from impairment of gait. As demonstrated in this work, as soon as discriminative feature extraction schemes are in place for each modality, modern supervised machine-learning algorithms such as randomized forests and meta-learners are capable of identifying discriminative patterns in high-dimensional clinical sensor data, for breakthrough performance in computerized diagnosis.

Conclusion

In this study, we performed supervised classification experiments on posturographic sway features from a large clinical cohort of patients with various vertigo, balance and movement disorders. For a subset of classes considered in this study, namely cerebellar ataxia, functional phobic postural vertigo, acute unilateral vestibulopathy, orthostatic tremor and healthy controls, we consider quantitative static posturography as a useful tool for computer-aided diagnosis in clinical routine. Other stance disorders remain a challenge, which may be addressed by extraction of more meaningful features, incorporation of further stance conditions, and inclusion of, e.g., video oculography, gait analysis, questionnaire data, demographics or medical imaging into a multi-modal examination paradigm. Importantly, modern methods from the fields of machine learning and data mining have reached a high level of maturity. A mapping of high-dimensional clinical data into a low-dimensional 2D-space can provide an informative visualization approach for the analyzed data. In particular, it allows to clinically interpret shortcomings of the automated classification routines with respect to, e.g., heterogeneous disease groups. Overall, machine learning and data science can help in visualizing, understanding, and utilizing high-dimensional clinical data, towards a computerized and more "objective" evaluation of balance disorders.

Acknowledgements The study was supported by the German Federal Ministry of Education and Research (BMBF) in connection with the foundation of the German Center for Vertigo and Balance Disorders (DSGZ) (Grant number 01 EO 0901).

Compliance with ethical standards

Conflicts of interest None of the authors have potential conflicts of interest to be disclosed.

References

1. Dieterich M, Brandt T (2015) The bilateral central vestibular system: Its pathways, functions, and disorders. Ann N Y Acad Sci. 1343:10–26
2. Brandt T, Dieterich M (2017) The dizzy patient: don't forget disorders of the central vestibular system. Nat Rev Neurol. 13:352–362
3. Staab JP, Eckhardt-Henn A, Horii A, Jacob R, Strupp M, Brandt T et al (2017) Diagnostic criteria for persistent postural-perceptual dizziness (PPPD): consensus document of the committee for the classification of vestibular disorders of the Bárány Society. J Vestib Res. 27:191–208
4. Black FO (2001) What can posturography tell us about vestibular function? Ann N Y Acad Sci. 942:446–464
5. de Waele C. Vestibular tests: static and dynamic posturography BT—encyclopedia of neuroscience. In: Binder MD, Hirokawa N, Windhorst U (eds) Springer, Berlin. 2009;4217–8.
6. Monsell EM, Furman JM, Herdman SJ, Konrad HR, Shepard NT (1997) Computerized dynamic platform posturography. Otolaryngol Head Neck Surg 117:394–398
7. Krafczyk S, Tietze S, Swoboda W, Valkovi P, Brandt T (2006) Artificial neural network: a new diagnostic posturographic tool for disorders of stance. Clin Neurophysiol. 117:1692–1698
8. Fioretti S, Scocco M, Ladislao L, Ghetti G, Rabini RA (2010) Identification of peripheral neuropathy in type-2 diabetic subjects by static posturography and linear discriminant analysis. Gait Posture. 32:317–320
9. Howcroft J, Lemaire ED, Kofman J, McIlroy WE (2017) Elderly fall risk prediction using static posturography. PLoS One 12:1–13
10. Saripalle SK, Paiva GC, Derakhshani RR, King GW, Lovelace CT (2014) Classification of body movements based on posturographic data. Hum Mov Sci. 33:238–250
11. Vonk J, Horlings CGC, Allum JHJ (2010) Differentiating malingering balance disorder patients from healthy controls, compensated unilateral vestibular loss, and whiplash patients using stance and gait posturography. Audiol Neurotol. 15:261–272
12. Redfern MS, Furman JM (1994) Postural sway of patients with vestibular disorders during optic flow. J Vestib Res. 4:221–230
13. Diener HC, Dichgans J, Bacher M, Gompf B (1984) Quantification of postural sway in normals and patients with cerebellar diseases. Clin Neurophysiol. Elsevier 57:134–142
14. Helmchen C, Kirchhoff J-B, Göttlich M, Sprenger A (2017) Postural ataxia in cerebellar downbeat nystagmus: its relation to visual, proprioceptive and vestibular signals and cerebellar atrophy. PLoS One 12:e0168808
15. Rigby HB, Rigby MH, Caviness JN (2015) Orthostatic tremor: a spectrum of fast and slow frequencies or distinct entities? Tremor Other Hyperkinet Mov (N Y). 5:324
16. Schöberl F, Feil K, Xiong G, Bartenstein P, la Fougére C, Jahn K et al (2017) Pathological ponto-cerebello-thalamo-cortical activations in primary orthostatic tremor during lying and stance. Brain 140:83–97
17. Ickenstein GW, Ambach H, Klöditz A, Koch H, Isenmann S, Reichmann H et al (2012) Static posturography in aging and Parkinson's disease. Front Aging Neurosci. 4:20
18. Brandt T, Strupp M, Novozhilov S, Krafczyk S. Artificial neural network posturography detects the transition of vestibular neuritis to phobic postural vertigo. J. Neurol. 2012. p. 182–4.
19. Allum JHJ, Bloem BR, Carpenter MG, Honegger F (2001) Differential diagnosis of proprioceptive and vestibular deficits using dynamic support-surface posturography. Gait Posture. 14:217–226
20. Schwesig R, Fischer D, Becker S, Lauenroth A (2014) Intraobserver reliability of posturography in patients with vestibular neuritis. Somatosens Mot Res. 31:28–34
21. Bronstein AM, Guerraz M (1999) Visual-vestibular control of posture and gait: physiological mechanisms and disorders. Curr Opin Neurol. 12:5–11
22. Deuschl G, Bain P, Brin M (1998) Consensus statement of the movement disorder society on tremor. Ad Hoc Scientific Committee. Mov Disord. 13:2–23
23. Strupp M, Hufner K, Sandmann R, Zwergal A, Dieterich M, Jahn K et al (2011) Central oculomotor disturbances and nystagmus: a window into the brainstem and cerebellum. Dtsch Arztebl Int. 108:197–204
24. Strupp M, Brandt T (1999) Vestibular neuritis. Adv Otorhinolaryngol. 55:111–136
25. Gelb DJ, Oliver E, Gilman S (1999) Diagnostic criteria for Parkinson disease. Arch Neurol. 56:33–39
26. Richardson JK, Thies S, Ashton-Miller JA (2008) An exploration of step time variability on smooth and irregular surfaces in older persons with neuropathy. Clin Biomech. 23:349–356
27. Brandt T (1996) Phobic postural vertigo. Neurology. 46:1515–1519
28. Hufschmidt A, Dichgans J, Mauritz K-H, Hufschmidt M (1980) Some methods and parameters of body sway quantification and their neurological applications. Arch Psychiatr Nervenkr. 228:135–150
29. Brandt T, Krafczyk S, Malsbenden I (1981) Postural imbalance with head extension: improvement by training as a model for ataxia therapy. Ann N Y Acad Sci. 374:636–649
30. Caruana R, Karampatziakis N, Yessenalina A. An empirical evaluation of supervised learning in high dimensions. Proc 25th Int Conf Mach Learn-ICML '08. 2008. p. 96–103.
31. Yu H-F, Huang F-L, Lin C-J (2011) Dual coordinate descent methods for logistic regression and maximum entropy models. Mach Learn. 85:41–75
32. Altman NS (1992) An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am Stat. 46:175–185

33. Hinton GE (1989) Connectionist learning procedures. Artif Intell. 40:185–234
34. Smola AJ, Sch B, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput. 14:199–222
35. Criminisi, A, Konukoglu E, Shotton J (2011) Decision forests for classification, regression, density estimation manifold learning and semi-supervised learning microsoft technical report
36. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn. 63:3–42
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res. 12:2825–2830
38. Wolpert DH (1992) Stacked generalization. Neural Netw. 5:241–259
39. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. ed. 1. Wadsworth Stat. Ser. Chapman and Hall/CRC; 1984.
40. van der Maaten L, Hinton GE (2008) Visualizing high-dimensional data using t-SNE. J Mach Learn Res. 9:2579–2605
41. Jones E, Oliphant T, Peterson P, others. SciPy: Open source scientific tools for Python [Internet]. 2001. https://www.scipy.org/
42. Gallea C, Popa T, García-Lorenzo D, Valabregue R, Legrand AP, Apartis E et al (2016) Orthostatic tremor: a cerebellar pathology? Brain 139:2182–2197
43. Hageman PA, Leibowitz JM, Blanke D (1995) Age and gender effects on postural control measures. Arch Phys Med Rehabil. 76:961–965
44. Gill J, Allum JHJ, Carpenter MG, Held-Ziolkowska M, Adkin AL, Honegger F et al (2001) Trunk sway measures of postural stability during clinical balance tests: effects of age. J Gerontol Ser A Biol Sci Med Sci. 56:438–447
45. Feil K, Böttcher N, Guri F, Krafczyk S, Schöberl F, Zwergal A et al (2015) Long-term course of orthostatic tremor in serial posturographic measurement. Park Relat Disord. 21:905–910
46. Fung VSC, Sauner D, Day BL (2001) A dissociation between subjective and objective unsteadiness in primary orthostatic tremor. Brain 124:322–330
47. Wuehr M, Schlick C, Möhwald K, Schniepp R (2018) Walking in orthostatic tremor modulates tremor features and is characterized by impaired gait stability. Sci Rep. 8:14152
48. Kumar N, Kohlbecher S, Schneider E. A novel approach to video-based pupil tracking. Conf Proc—IEEE Int Conf Syst Man Cybern. 2009. pp 1255–62.
49. Kirsch V, Keeser D, Hergenroeder T, Erat O, Ertl-Wagner B, Brandt T et al (2016) Structural and functional connectivity mapping of the vestibular circuitry from human brainstem to cortex. Brain Struct Funct. 221:1291–1308
50. Pradhan C, Wuehr M, Akrami F, Neuhaeusser M, Huth S, Brandt T et al (2015) Automated classification of neurological disorders of gait using spatio-temporal gait parameters. J Electromyogr Kinesiol. 25:413–422