Multi-Complexity Ensemble Measures for Gait Time Series Analysis: Application to Diagnostics, Monitoring and Biometrics

Valeriy Gavrishchaka, Olga Senyukova and Kristina Davis

Abstract Previously, we have proposed to use complementary complexity measures discovered by boosting-like ensemble learning for the enhancement of quantitative indicators dealing with necessarily short physiological time series. We have confirmed robustness of such multi-complexity measures for heart rate variability analysis with the emphasis on detection of emerging and intermittent cardiac abnormalities. Recently, we presented preliminary results suggesting that such ensemble-based approach could be also effective in discovering universal meta-indicators for early detection and convenient monitoring of neurological abnormalities using gait time series. Here, we argue and demonstrate that these multi-complexity ensemble measures for gait time series analysis could have significantly wider application scope ranging from diagnostics and early detection of physiological regime change to gait-based biometrics applications.

1 Introduction

Development of new technologies has provided inexpensive and unobtrusive means of collecting multi-scale physiological data and led to continuous improvements in clinical instrumentation. Sophisticated portable and wearable systems for real-time collection of physiological data have also become affordable for routine individual

Valeriy Gavrishchaka
Department of Physics, West Virginia University, Morgantown, WV 26506, USA, e-mail: gavrishchaka@gmail.com

Olga Senyukova
Department of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskie Gory, 119991 Moscow, Russia, e-mail: osenyukova@graphics.cs.msu.ru

Kristina Davis
Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA, e-mail: kristinadavismd@gmail.com

use. Increased availability of high-resolution data provides new opportunities for quantitative diagnostics, early abnormality detection, and convenient monitoring.

Analysis techniques compatible with necessarily short time series are essential for many applications. However, it is challenging to construct universal measures or indicators for robust quantification of physiological states from short time series. In express diagnostics, preventive monitoring, and personalization of medical treatment, it is important to find and correctly interpret quantitative measures capable of detecting emerging and transient abnormalities and other subtle regime changes.

Variability analysis of physiological time series provides a generic framework for robust discrimination between normal and abnormal states [1–5]. The well known application of this methodology is heart rate variability (HRV) analysis approved as one of the modalities for cardiac diagnostics [1–4]. Compared to traditional electrocardiography (ECG) analysis the method is more robust to noise because it relies only on the interbeat interval signal (RR data) which is very important for analysis of the data from portable and wearable devices. Moreover it is able to detect cardiac and non-cardiac (e.g., emotional) abnormalities lacking well-defined ECG form patterns.

Variability analysis is usually based on nonlinear dynamics (NLD) complexity measures and advanced linear indicators. Unfortunately, the accuracy and stability of such variability measures tend to decrease significantly when the analysis is performed on shorter data segments [1–6]. This limitation diminishes the predictive capability of these measures for early detection of both short-lived precursors of emerging physiological regimes and abnormalities with transient patterns.

Recently we have demonstrated that performance of HRV indicators dealing with short time series could be significantly improved through optimal combination of complementary complexity measures, using boosting-like ensemble learning [6, 7]. Such an approach is especially important for early detection of emerging abnormalities and other regime changes where other techniques could often fail.

ECG time series is an important diagnostic and monitoring modality for cardiac abnormalities and related state changes. However, alternative physiological data channels could be informative for other abnormalities. For example, due to advances in technology, gait time series can be easily collected with wearable clinical equipment as well as with general-purpose portable devices such as smartphones where built-in accelerometers are now part of standard configuration. Similar to RR data, measurement of gait stride intervals is very tolerant to noise since only peak-to-peak periods of otherwise complex time series are required.

Variability metrics of gait stride intervals is known to be sensitive to changes in neurological functions associated with aging and development of certain neurological diseases [8–11]. Long-range correlation and other measures of stride-interval dynamics could be effective in detecting neurological abnormalities and in quantification of their severity [8–11]. These include Parkinson's (PD) and Huntington's (HD) diseases, amyotrophic lateral sclerosis (ALS), and others.

Remaining challenges in treatment and diagnostics of ALS, PD, HD and other neurological abnormalities maintain significant interest in unobtrusive modalities capable of early diagnostics and robust monitoring of such abnormalities. There-

fore, variability indicators computed from stride-interval time series could provide convenient and robust tool for early diagnostics and monitoring of neurological abnormalities. A generic set of NLD complexity measures and linear indicators used in HRV analysis can be directly applied to gait quantification after recalibration.

However, similar to HRV analysis, accuracy of NLD measures and advanced linear indicators could significantly deteriorate when applied to shorter segments of gait time series. Therefore, such indicators would have serious limitations in detecting early intermittent signatures of neurological abnormalities as well as in monitoring effectiveness of medical treatment and its optimization.

In this work we suggest that multi-complexity measures discovered by boosting-like ensemble learning could be effective for early diagnostics and monitoring of neurological abnormalities even when applied to short segments of gait stride intervals. We illustrate validity of our approach using real gait data of normal subjects and patients with ALS, HD and PD. Similarly, we demonstrate capability of multi-complexity ensemble indicators of detecting slow regime changes using gait time series collected from healthy children and teens of different age groups (from 3 to 14 years old). Finally, we present analysis of long gait time series from 10 healthy adults suggesting possible application of our meta-indicators in gait-based biometrics. All presented results are based on real gait data available at http://www.physionet.org. Current work provides significant extensions and generalizations of our preliminary results previously reported in [20].

2 Variability Analysis of Physiological Time Series: Advantages, Challenges and Multi-Complexity Generalization

Majority of NLD measures and linear indicators used for variability analysis require long time series to achieve desired accuracy and stability. For example, many HRV indicators require long time series for stable calculation [1–3] which could drastically restrict their application scope. Similar restrictions are also relevant for stride-interval analysis. For example, recommendation in one of the recent study [10] is to use segment of at least 600 stride intervals for variability-based diagnostics. Nevertheless, indicators have to be computed on short segments in order to capture early signs of developing and/or intermittent abnormalities or to detect subtle initial effects of treatment procedures. Indeed, indicator computed on a long time series will average out these short-lived effects and will fail to detect them.

The well-known NLD indicators applicable for HRV and gait analysis are based on detrended fluctuation analysis (DFA) [12], multi-scale entropy (MSE) [13], and multi-fractal analysis (MFA) including MFA extension of DFA [14]. The discriminative-ability preservation conclusion extends to advanced linear indicators based on power spectrum analysis of the RR time series [4]. We also successfully used power spectrum measure as one of the base indicators for stride-interval analysis.

DFA was proven to be useful in revealing the extent of long-range correlations in time series including HRV applications [12] and diagnostics of neurological abnormalities [8–11]. First, the investigated time series of length N is integrated. Next, the integrated time series is divided into n boxes. All boxes have the same length. In each box, a least-square line is fitted to the data with y-coordinate denoted by $y_n(k)$ (representing the trend in that box). Finally, the integrated time series, $y(k)$, is detrended as follows:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2}. \tag{1}$$

A linear relationship on the plot of $\log F(n)$ vs. $\log n$ indicates power law (fractal) scaling characterized by a scaling exponent β (slope of the fitted straight line) which is used as HRV indicator.

Multi-scale entropy (MSE) method [13] has been introduced to resolve limitations of traditional single-scale entropy measures. First, a coarse-graining process is applied to the original time series, x_i. Multiple coarse-grained time series are constructed by averaging the data points within non-overlapping windows of increasing duration, τ:

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \tag{2}$$

where τ represents the scale factor and $j = 1,\ldots,N/\tau$. The duration of the coarse-grained time series is N/τ. Next, entropy is calculated for each time series and plotted as a function of the scale factor. Different signatures of this function's curve including originally suggested entropy difference between two scales [13] can serve as HRV and other physiological indicators.

HRV indicators based on frequency-domain analysis are often superior in accuracy and stability to the time-domain linear indicators. One of the widely accepted indicators of this type in HRV analysis is a power spectrum ratio of the low-frequency band (0.04-0.15 Hz) to the high-frequency band (0.4-0.15 Hz) [4]. Due to irregularity of the time grid of the RR time series, it is convenient and more accurate to use a Lomb periodogram for power spectrum calculations instead of Fast Fourier Transform (FFT) [15]. In certain regimes, the accuracy of such power spectrum indicators could be comparable to the best NLD approaches. As discussed in the next section, similar power spectrum measures can also be successfully used as base indicators for stride-interval analysis.

Recently we have illustrated that challenges of variability analysis, when applied to short time series, could be overcome by using a classification framework based on boosting-like ensemble learning techniques that are capable of discovering robust multi-component meta-indicators from a combination of existing variability measures and other incomplete empirical knowledge [6,7]. Unlike most other combination techniques, the use of boosting is capable of discovering an ensemble of complementary models that has both significantly lower bias (higher accuracy) and lower variance (better stability) compared to each individual model. Potentially

more flexible data-driven models (e.g., neural networks) are often unstable due to training data incompleteness, intrinsic non-stationarity, and low signal-to-noise ratio. In addition, such "black-box" systems lack interpretability. In contrast, meta-indicators, discovered by boosting, combine accuracy, stability, and interpretability because they are constructed from the well-understood low-complexity base models.

A typical boosting algorithm such as AdaBoost [16, 17] for the two-class classification problem starts with equal and normalized weights for all training data. Base classifiers, $h_t(x)$, are trained using a weighted error function and the optimal one is chosen at each iteration t. Here x is an input vector. Data points misclassified by the current interations best model are penalized by the weight-factor adjustment (increase) for the next iteration. Therefore, on each iteration, the algorithm focus is on harder-to-classify samples. The final meta-model, given below, classifies the unknown sample as class +1 when $H(x) > 0$ and as -1 otherwise:

$$H_T(x) = \sum_{t=1}^{T} \alpha_t h_t(x). \tag{3}$$

Here, the constants α_t are the sequence of combination coefficients obtained, and T is the total number of iterations. Regime adjustments together with important regularization procedures also can be introduced to the original boosting algorithm in several ways [18].

A natural choice of base models could be low-complexity base classifiers, where each of the classifiers uses just one complexity measure, β_i, out of several available choices:

$$y = h(\beta_i[p_i], \gamma). \tag{4}$$

Here γ is a threshold level (decision boundary) and p_i is a vector of adjustable parameters of the chosen measure. In our case, β_i may correspond, for example, to either a DFA scaling exponent, a slope of MSE curve, or a power spectrum ratio. Applying boosting steps to a set of such base classifiers (4) with different measures β_i and optimizing over (p_i, γ) on each boosting iteration, we obtain a meta-classifier (3).

Additionally, many different physiological regimes, quantified by individual base classifiers from the ensemble, are implicitly encoded in such multi-complexity meta-indicators. In our recent publications [19, 21] we referred to this utilization of ensemble internal structure as ensemble decomposition learning (EDL). We also outlined possible practical application of EDL concept using one of the single-example learning (SEL) frameworks.

The EDL technique can be summarized as follows. Boosting constructs local experts $h_i(x)$ for different implicit regimes or domains of a whole feature space, which ensures good global performance of the final ensemble. Therefore, partial information of wide variety of dynamical regimes becomes implicitly encoded in the obtained ensemble of classifiers. However, only aggregated output is used for normal-abnormal classification, while the rich internal structure of the ensemble is completely ignored. Extraction of this underutilized knowledge could be formal-

ized in terms of ensemble decomposition learning (EDL) [19]. Formally, one can introduce ensemble decomposition feature vector as follows:

$$D(x) = [\alpha_1 h_1(x), \alpha_2 h_2(x), \ldots, \alpha_T h_T(x)]. \tag{5}$$

Each sample after ensemble classification procedure can be represented by this vector. Although each individual component of this feature vector may not contain explicit and usable information, collectively, these values may provide detailed and informative state representation of the considered system which is not accessible in the aggregated form given by $H(x)$.

Later, we have also pointed out that, besides particular SEL framework, one can effectively utilize this fine-grain knowledge by using EDL metrics in different types of instance-based learning (IBL) and clustering algorithms including graph-based techniques [21]. For example, we have found that the length change of the minimum spanning tree (MST), constructed using ensemble distance metrics, could provide an early indication of the emerging physiological regimes. All provided illustrations were based on real data for several cardiac abnormalities.

MST representation is motivated by the human perception which organizes information with the most economical encoding. A spanning tree is a connected graph containing all vertices of the original graph without loops, i.e., there exists only one path connecting any two pairs of nodes in the graph. If the edges of the graph are weighted, the spanning tree length is defined as the sum of the weights of its edges. MST is a spanning tree with minimal length among all spanning trees connecting the nodes of the graph. MST of the graph can be derived with Prim's or Kruskal's algorithm [22] with subsequent removal of several longest edges to generate clusters.

Advantages of graph-based representation such as MST have been recently demonstrated in financial applications [23–25]. Similarly, MST representation can be used to capture essential dependencies and differences between physiological states quantified by the EDL vector. If the length of time series permits computation of N EDL vectors from N consecutive segments of physiological time series, information from $N(N-1)/2$ numbers of distance matrix d_{ij}, the distances between EDL vectors i and j given by (5), defined as l_1 or l_2 norm in T-dimensional space, will be represented with $(N-1)$ edges of MST which can be used for intuitive cluster visualization and analysis. Besides effective clustering, EDL-based MST representation also offers informative aggregated measure such as normalized tree length that could be a sensitive indicator of emerging new regimes or regime changes:

$$L = \frac{1}{N-1} \sum_{d_{ij} \in T} d_{ij}, \tag{6}$$

where $(N-1)$ is the number of edges present in MST. Indeed, even in the very early stage when a new regime begins manifesting itself only on short intermittent segments, MST length (6) will increase because the distance d_{ij} between EDL vectors of existing and new regimes is significantly higher than the distance between EDL

vectors describing the same regime. This could be used for preventive monitoring of healthy subjects as well as for side effects detection in the beginning of new therapy or drug treatment. The described approach is different from EDL-SEL framework mentioned before. Indeed, no EDL vectors associated with particular abnormalities are required and any emerging regimes different from current ones will be detected.

3 Diagnostics and Monitoring of Neurological Abnormalities

Several neurologic disorders include abnormalities of movement and gait as prominent symptoms. Huntington's disease is a debilitating and ultimately fatal neurodegenerative disease with an autosomal dominant inheritance pattern. A good family history, symptom history, and physical exam can suggest the diagnosis, which can be then confirmed by molecular tests looking for CAG (codon that codes for the amino acid glutamine) triplet repeats in the Huntington gene [27]. The challenge lies in diagnosing clinical disease onset in an individual who is known to have the genetic abnormality, when classical movement abnormalities may be less prominent [28]. In addition, a key component of active research into potential therapeutic approaches is the ability to monitor symptomatic response in a reliable manner. Currently the United Huntington's Disease Rating Scale is used for this purpose. However, administration of the entire scale is time-consuming. Furthermore, certain items assessed have better discriminatory function than others [26].

Parkinson's disease is another neurologic movement disorder. Diagnosis relies on typical symptoms of motor impairment by physical exam and patient history, after excluding other causes through laboratory and imaging evaluations. A definitive genetic test is not applicable in most cases. Specific medications are available to treat the symptoms of the disease, but making the accurate diagnosis is essential. Patients presenting with gait disturbance as their main initial symptom can have a significant delay in diagnosis (compared to those who present with a tremor), which in turn delays therapy [30]. Automated gait analysis, as opposed to more subjective clinical assessment, has been shown to be of potential use in diagnosing and following Parkinson's disease [29].

ALS is another progressive neurodegenerative disorder, caused by degeneration of motor neurons which leads to muscle weakness and atrophy. The presenting symptoms depend on which muscles are affected first. Electrophysiological testing, which can be invasive and painful, plays a role in early diagnosis of ALS and differentiation from treatable causes of weakness [31]. However, both in clinical practice and research, a less invasive yet reliable method of assessing disease progression and response to treatment is necessary.

Here we demonstrate that single indicators based on NLD complexity measures could partially preserve their discrimination ability (normal vs abnormal) even on short segments of gait time series: down to ≈ 100 stride intervals or even shorter. Nevertheless, the reduced accuracy may not be sufficient for many practical applications. However, combination of complementary complexity measures using

boosting-like algorithms can significantly increase accuracy and stability of indicators operating on short segments of gait time series. Such multi-complexity measures could be effective for early detection and monitoring of wide range of neurological abnormalities.

To illustrate capabilities of our ensemble-based indicator, we use gait data collected from normal subjects and patients with ALS, HD and PD that are available at http://www.physionet.org. This data set includes gait time series from 15 patients with PD, 20 patients with HD, 13 patients with ALS, and 16 healthy subjects. Each time series consists of up to 300 stride intervals. We use segments as short as 128 stride intervals for calculation of DFA (1), MSE (2) and power spectrum measures that are used in base classifiers (4). Due to data limitation, total number of segments is increased in the training phase by overlapping. Also, since low-complexity base classifiers are used, we do not find any significant signs of overfitting on out-of-sample data. In the following, performance metrics are computed on all available data.

We should note that previously reported preliminary results [20] were based on indicator calculation from 128 left-only or right-only stride intervals. While such approach is also valid and could be preferable in some cases (e.g., when one leg has certain problems), here we employ the more natural and practical approach by using segments with both left and right stride intervals. This also means that previous results based on 128 left-only or right-only intervals really used time series segment corresponding to 256 stride intervals. In contrast, the current study is based on the shorter segments consisting of 128 intervals.

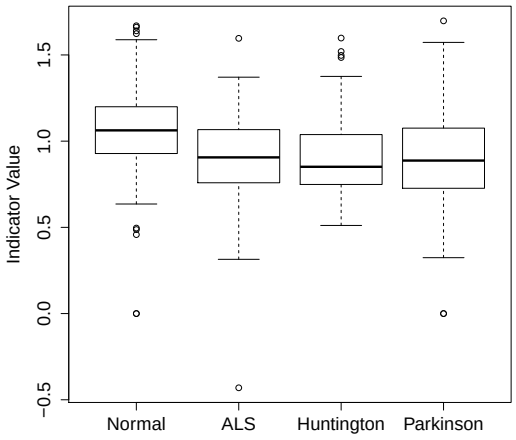

Fig. 1 Single DFA measure computed on each of 128-interval segments of stride data from normal control group and patient groups with ALS, HD and PD.

We apply AdaBoost to discover ensemble meta-indicator given by (3) which consists of complementary base classifiers (4). The classifier from the first boosting iteration is the best single classifier. In our case, it always happens to be DFA-based classifier. To illustrate discrimination ability of the best single complexity measure, we compute DFA exponents for all 128-interval segments of stride time series from different groups of patients and summarize them as the box plot in Fig. 1. Here we use modified box plot from R with all default settings. Thick horizontal segments represent median, while segments below and above represent 25^{th} and 75^{th} percentiles, respectively. Dashed lines extend to max/min points if they are inside the range obtained by adding the difference between 75^{th} and 25^{th} percentiles scaled by 1.5 to 75^{th} percentile and subtracting the same value from 25^{th} percentile. Otherwise, points outside this range are plotted as circles.

It is evident from Fig. 1 that medians of all abnormal groups (ALS, HD, PD) are clearly below the median of the healthy group. This confirms remaining discrimination ability of a single indicator even for short stride-interval segments. However, there is significant overlapping between normal and abnormal groups even for ranges defined by $25^{th} - 75^{th}$ percentiles. Therefore, normal/abnormal discrimination accuracy of a single measure is quite limited.

However, boosting-based combination of complementary complexity measures could drastically increase accuracy of meta-classifier given by (3). Box plot of the aggregated ensemble measure (3) computed on 128-interval segments for each group of patients is shown in Fig. 2. Now, not only medians of ALS, PD, and HD groups are well below healthy group, but also $25^{th} - 75^{th}$ percentile ranges of all

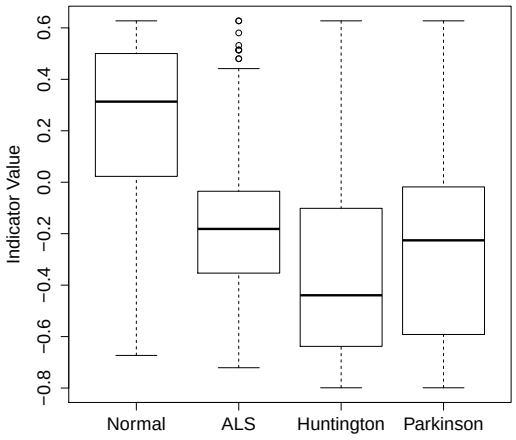

Fig. 2 Aggregated ensemble measure computed on each of 128-interval segments of stride data from normal control group and patient groups with ALS, HD, and PD.

abnormal groups do not overlap with healthy group. Thus, accuracy of the multi-complexity ensemble indicator is significantly increased.

A more formal comparison of the best single measure vs ensemble indicator is presented in Fig. 3. Here we plot detection rates for 3 reasonable false alarm rates: 10%, 20% and 30%. It is clear that boosting-based combination of individual complexity measures can increase detection rate by 40-50%. Such dramatic performance improvement suggests that such meta-indicators based on multiple complexity measures could be a reasonable choice for early detection and monitoring of various neurological abnormalities.

4 Detection of Emerging Physiological States and Regime Changes

Early detection of emerging physiological state or slow regime changes is often more challenging than accurate diagnostics of the developed abnormality. Indeed, many discriminative features used in diagnostics are not yet present when abnormality is in early development stage. Similar challenges are also typical for early detection of changes during personalization of drug treatment or therapy.

High detection rates with acceptable false alarm rates shown in Fig. 3 illustrate ability of the multi-complexity meta-classifier to discriminate between gait time series from normal subjects and subjects with various developed abnormalities. Implicitly, this also suggests potential ability to detect early signs of the developing

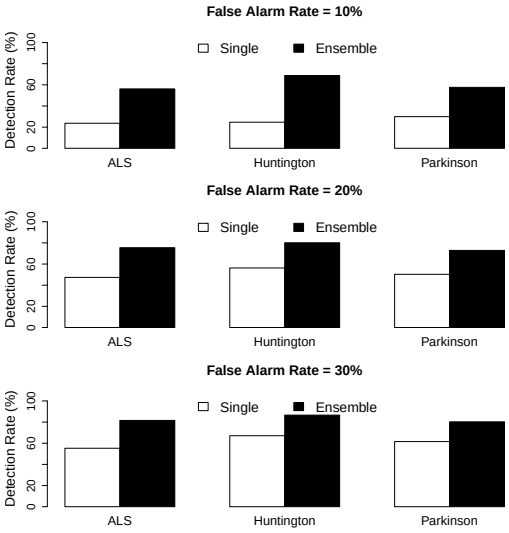

Fig. 3 Abnormality detection rates for a given false alarm rate: The best single measure vs ensemble of multi-complexity measures.

abnormality and other regime changes. However, direct illustration based on data with emerging or intermittent pathologies could be more convincing. While we are not aware of any large open-access databases capturing slow development of neurological abnormalities, other gait databases can be used for illustration of slow physiological regime changes. One of them is gait maturation database first analyzed by Hausdorff et al. [32] and now available at http://www.physionet.org.

Gait maturation database is a collection of gait time series from 50 children of various age groups: from 3 to 14 years old. For each subject, time series is up to 500 stride-intervals long. It is known that in very young children, immature control of posture and gait results in unsteady locomotion [32]. In children ≈ 3 years old, gait appears relatively mature. However, as suggested in [32], the dynamics of walking changes continues beyond this age. This was confirmed by quantitative analysis of 50 children from gait maturity database [32]. Single time- and frequency-domain measures as well as DFA-based measures have been used in that study. It was demonstrated that, while gait in younger age groups resembles that of adults with neurological abnormality, it continuously matures and approaches the dynamical range of healthy young adults as age increases.

In the analysis of gait maturation data, Hausdorff et al. [32] calculated indicators using significantly long segments (at least 256 stride intervals), and there was still wide overlap of indicator values among different age groups. Such overlap could only increase for shorter segments. This overlap is not critical for the main objective of the analysis presented in [32]. Indeed, the authors demonstrated statistically significant trend of gait characteristics approaching those of healthy adults when children age increases. However, for early detection of slow regime change due to developing abnormality or initial treatment effects, insufficient discrimination of single indicators could make them useless in practice.

Thus, gait maturation database offers convenient real-life data to demonstrate advantages of multi-complexity ensemble measures over single indicators in detecting emerging physiological states and slow regime changes. For this purpose, we compare the best single indicator (DFA) and ensemble of multi-complexity measures discovered in the normal-abnormal classification problem in the previous section. It is important to note that none of the gait maturation data were used in the training phase.

We applied these indicators to short (128-interval) segments from different age groups and summarized results as box plots in Figs. 4 and 5.

Fig. 4 demonstrates that single DFA indicator is not capable to detect any clear trend in gait dynamics evolution as children age increases. On the other hand, as evident from Fig. 5, aggregated output, $H(x)$, of the multi-complexity ensemble indicator shows clear trend towards gait dynamics of healthy adults as age increases. Remember that the range of probability-like output, $H(x)$, is $[-1,+1]$, with healthy state corresponding to positive numbers. Therefore, aggregated output of multi-complexity ensemble measures could be a promising metrics for early detection of subtle and/or slow changes in gait dynamics.

5 Biometrics Based on Gait Time Series Analysis

In the previous section, we have demonstrated that multi-complexity ensemble measures, obtained as binary classifiers for normal-abnormal diagnostics problem, could have much wider application scope. Indeed, the aggregated ensemble output could be a sensitive and robust indicator of physiological regime changes. In addition, as shown previously in the context of cardiac abnormalities, multi-complexity ensembles could be also used for multi-class discrimination and rare-state quantification using EDL representation [19,21]. Similar approach could be also effective for physiological state discrimination based on gait time series. Gait-based biometrics is an example of such practical problem where effective discrimination among multiple physiological states or classes is crucial.

Recently, gait recognition has become a topic of interest within the computer vision applications, due to its growing importance as a biometric modality [36–38]. An important motivation for gait recognition research has been provided by psychophysical experiments with Moving Light Displays (MLDs) pioneered by Johansson [33]. Johansson's experiments demonstrated the ability of humans to recognize the type of movement of a person solely from observing the 2D motion pattern generated by light bulbs attached to the person. Similar experiments indicated that even the identity of a familiar person, as well as the gender of the person, might be recognizable from MLDs [34]. These experiments provided insight into motion perception in the human visual system and suggested that motion patterns generated by the human gait encode information that could be unique to the moving person.

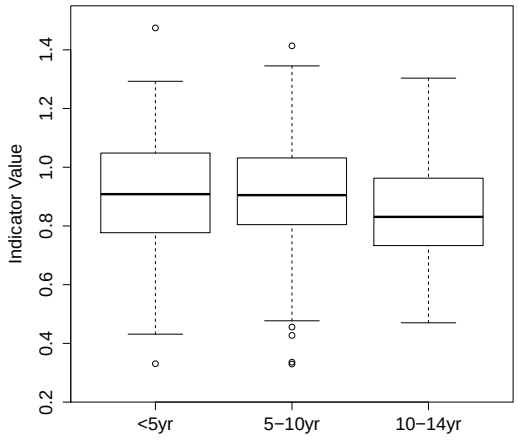

Fig. 4 Single DFA measure computed on each of 128-interval segments of stride data from 3 different age groups of healthy subjects.

The fact that each person seems to have a distinctive (idiosyncratic) way of walking is hardly surprising from a biomechanics standpoint [35]. Human ambulation consists of synchronized integrated movements of hundreds of muscles and joints in the body. Although these movements follow the same basic pattern for all humans, they seem to vary from one individual to another in certain details such as their relative timing and magnitudes. Much research in biomechanics and clinical gait analysis is devoted to the study of the inter-person and intra-person variability of gait, mainly to determine normal vs. pathological ranges of variation. There is an increased interest in gait as a biometric, mainly due to its non-intrusive and non-concealable nature as well as possibility of remote biometrics. Considerable research efforts are being devoted in the computer vision community to characterize and extract gait dynamics automatically from video [36–38]. Biometric systems for human identification at a distance are in increasing demand in various real-life applications. Many biometric modalities such as face recognition, iris, fingerprints, palm prints, and hand geometry have been systematically studied and employed in many practical systems. However, these approaches suffer from two main disadvantages: (1) high failure rate when only low resolution images and pictures taken at a distance are available, and (2) necessity of subject cooperation for accurate results. For these reasons, innovative biometric recognition methods for human identification at a distance have significant potential and appeal for surveillance and forensic applications [38].

The most common categories of gait recognition, currently used and discussed in the literature, are appearance-based and model-based approaches [36, 37]. Among the two, the appearance-based approaches suffer from changes in the appearance

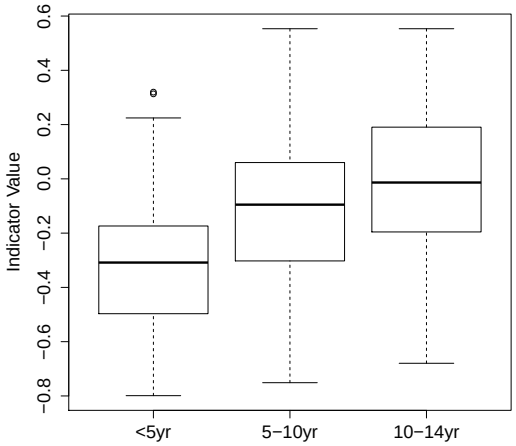

Fig. 5 Aggregated ensemble measure computed on each of 128-interval segments of stride data from 3 different age groups of healthy subjects.

owing to the change of the viewing or walking directions [36, 37]. Model-based approaches extract the motion of the human body by means of fitting their models to the input images. Model-based approaches are view- and scale-invariant as well as take into account the kinematic characteristics of walking manner [36, 37]. In general, a gait is considered as being composed of a sequence of kinematic characteristics of human motion and most systems in existence recognize it by the similarity of these characteristics.

However, for accurate recognition, both approaches require significant amount of details to be extracted from sequential video images. Unfortunately, in many practical cases the quality of available video may be poor due to insufficient lighting, covering of motion details by cloth, and other reasons. In such circumstances, one of the gait characteristics that could still be extracted is time period between consecutive steps. This is similar to RR intervals that could be quite accurately measured even from very noisy ECG time series and still used for cardiac diagnostics unlike ECG waveforms. Therefore, multi-complexity ensemble measures for gait time series analysis could be potentially employed as complementary approach in biometric applications, especially in cases where quality of the video prohibits usage of more traditional techniques.

Besides remote gait-based biometrics from video sequences, gait recognition from accelerometer data, that are currently available in all standard smartphones and other wearable devices, becomes comparably important [39]. Indeed, remote acquisition of accelerometer time series could be used to identify the person currently carrying the phone and in related applications. Until now, the most common approaches to analyze such time series included direct feature extraction using wavelet transform or similar techniques with subsequent application of standard classification algorithms [39]. However, when quality of time series is poor, it would be beneficial to extract the most robust feature such as stride interval time series with subsequent application of suitable techniques for further analysis. Again, multi-complexity measures discussed in this paper could be applicable to this set of gait-based biometric problems.

Success of multi-complexity measures in normal-abnormal classification does not warranty its effectiveness in gait-based biometrics. Indeed, stride-interval dynamics is quite different in healthy subjects and patients with neurological abnormalities. This explains success in binary "normal-abnormal" classification. However, differences between same-class (e.g., healthy) subjects could be much more subtle. Also, here we are dealing with multi-class classification problem that is often significantly more challenging than binary (2-class) classification. Nevertheless, our preliminary results presented below indicate that multi-complexity ensemble measures could differentiate between different individuals using their stride interval time series. This suggests applicability of our approach to gait-based biometrics problems discussed above. Our results are based on the analysis of gait time series from 10 healthy subjects available at http://www.physionet.org. This database consists of 3.5×104 stride intervals.

Multi-class classification algorithms are generally based on the reductions to the binary case. A common choice, applicable to both single classifiers and boosting

algorithms, is "one-against-all" reduction where separate classifiers for each of the considered classes are built [40]. After test sample is presented to such set of classifiers, the sample is considered belonging to the class for which probability-like output of the corresponding classifier is the largest among all available classifiers.

First, we demonstrate possibility of discrimination among subjects based on ensembles of multi-complexity measures that are built using "one-against-all" approach. We apply our framework to build 10 ensemble classifiers given by (3) corresponding to each of the 10 healthy subjects. In the test phase, we present data from each of the subjects to this set of 10 classifiers. If multi-complexity measures are capable to discriminate among different subjects, the maximum output (up to +1) should be obtained from the classifier corresponding to the subject whose data are currently presented. In other words, the output differences between the classifier corresponding to the subject of the presented data and all other classifiers (total, 9 numbers) should be positive with values up to +2. This is exactly what we observe in Fig. 6 where results of our multi-class classification test are presented. Indeed, all data points, except one, are positive numbers. Therefore, these preliminary results suggest effectiveness of multi-complex measures in the context of "one-against-all" multi-class classification that can be used in gait-based biometrics.

While "one-against-all" multi-class classification is valid approach to gait-based biometrics, it has important limitations. Indeed, it is applicable only to limited number of classes (individuals) since the data from each class (individual) should be available in training phase to build classifiers for each class. Therefore, other indi-

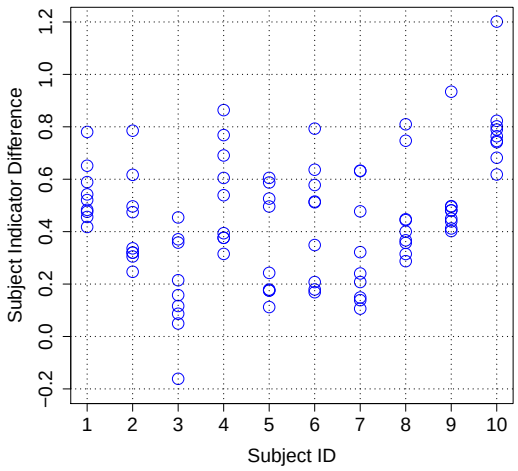

Fig. 6 Difference between median of the aggregated outputs of the current subject classifier and medians of outputs from other subjects classifiers computed on 128 stride interval segments. All classifiers are computed using current subject data. Calculations repeated for each 10 healthy subjects.

viduals that are not present in training data cannot be identified. However, in many important practical cases one would be interested to measure the similarity between two gait time series to conclude whether these time series belong to the same individual or not. Such similarity metrics should be robust enough to work for individuals that are not included in the training phase. This is possible if, instead of aggregated outputs of boosting-based classifiers, one uses EDL vectors (5) of the ensemble classifiers.

Here we describe and test one of the possible EDL-based frameworks for gait-based biometrics. Previously, in the context of cardiac abnormalities, we have shown that normalized MST tree length (6) computed from distance matrix of consecutive EDL vectors could be used as sensitive indicator of emerging patterns and slow regime changes [21]. The same approach can be used in EDL-based biometrics with slight modification. Here, instead of using distance metrics of EDL vectors from the same individual, we can create cross-subject distance (proximity) matrix where distances between all EDL vectors of one subject (individual) and all EDL vectors of another subject (individual) are computed. Then we still apply MST technique for low-dimensional representation of such matrix and compute MST tree length as an aggregated measure of EDL-based distance or proximity measure between two individuals.

Operationally, such EDL-based biometrics framework can be implemented as follows. Given gait time series from N subjects (individuals), we build N multi-complexity ensembles as in "one-against-all" multi-class classification approach.

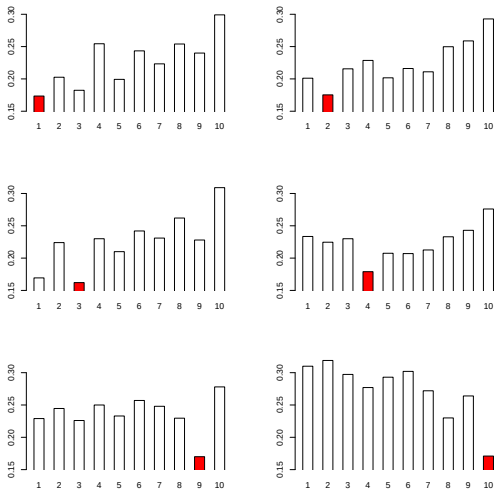

Fig. 7 Normalized MST length computed from the cross-subject proximity matrix. Proximity matrix is defined as l_2 distance between all pairs of EDL feature vectors computed on 128 stride interval segments. Red column indicates current reference subject whose proximity to other subjects including himself is measured by MST length.

Then, we combine N EDL vectors of these classifiers to obtain one multi-feature EDL vector. Such EDL vector can be applied not only to N subjects used in training but to quantify proximity of any two gait time series. Indeed, even for relatively small number of individuals used in training, their EDL vectors provide rich multi-feature (multi-regime) representations that can be generically applied to quantify any gait time series.

For demonstration of the effectiveness of such EDL-based approach, we apply it to the same gait database of 10 healthy individuals as in "one-against-all" multi-class classification discussed earlier. We combine 10 EDL vectors into a single EDL vector that should capture large number of different micro-regimes and their combinations. In testing phase, we compute consecutive EDL vectors from gait time series of a subject that needs to be identified. Then, we compare these consecutive EDL vectors with EDL vectors of reference subjects by construction MST tree from cross-subject proximity matrix and computing MST tree length (6). The lowest MST tree length among reference subjects will identify the tested subject. Please note that the set of reference subjects can be increased without retraining for obtaining new EDL vector.

In our series of tests we use gait data from each of 10 subjects for EDL-based identification. In all 10 tests, the minimum MST tree length correctly identifies tested individual. Details of 6 out 10 tests are presented in Fig. 7. We see that the lowest MST tree length point to the correct reference subject, i.e., if data from the n-th subject is presented, the lowest MST tree length will be obtained from cross-subject proximity matrix with the n-th reference individual. Thus, our results provide preliminary indication that generic high-dimensional feature vector can be constructed using training data from reasonable number of individuals and used in the wide range gait-recognition applications without any additional retraining.

6 Conlusions

We have demonstrated that boosting-based combination of multi-complexity measures could significantly improve quantitative analysis of short gait time series and could be applied for early detection of neurological abnormalities and their monitoring. Our conclusions are illustrated on real gait data from healthy subjects and patients with several neurological abnormalities. We have also demonstrated that multi-complexity ensemble measures can be effective in early detection of slow physiological regime changes and in gait-based biometric applications. More detailed analysis of our ensemble-based multi-complexity indicators on larger gait data sets is warranted.

References

1. Seely, A.J.-E., Macklem P.T.: Complex systems and the technology of variability analysis. Critical Care 8, 367-384 (2004)
2. Glass, L., Milton. J.: Dynamical Disease: Mathematical Analysis of Human Illness. AIP Press, New York (1995)
3. Voss, A., Schulz, S., Schroederet, R., Baumert, M., Caminal, P.: Methods derived from nonlinear dynamics for analysing heart rate variability. Philosophical Transactions of the Royal Society A 367, 277-296 (2008)
4. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93, 1043-1065 (1996)
5. Kantz, H., Schreiber, T.: Nonlinear Time Series Analysis. Cambridge University Press, Cambridge (1997)
6. Gavrishchaka, V.V., Senyukova, O.: Robust algorithmic detection of the developed cardiac pathologies and emerging or transient abnormalities from short periods of RR data. In: International Symposium on Computational Models for Life Sciences, Toyama City, October 2011. AIP Conference Proceedings, vol. 1371, pp. 215-224. AIP Publishing LLC, New York (2011)
7. Gavrishchaka, V.V., Senyukova, O.V.: Robust Algorithmic Detection of Cardiac Pathologies from Short Periods of RR Data. In: Pham, T.D., Jaim, L.C. (eds.) Knowledge-Based Systems in Biomedicine and Computational Life Science. Studies in Computational Intelligence, vol. 450, pp. 137-153. Springer, Heidelberg (2013)
8. Hausdorff, J.M., Mitchell, S.L., Firtion, R., Peng, C.K., Cudkowicz, M.E., Wei, J.Y., Goldberger, A.L.: Altered fractal dynamics of gait: reduced stride interval correlations with aging and Huntington's disease. J. Appl. Physiol. 82(1), 262-269 (1997)
9. Hausdorff, J.M., Lertratanakul, A., Cudkowicz, M. E., Peterson, A.L., Kaliton, D., Goldberger, A.L.: Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. J. Appl. Physiol. 88, 2045-2053 (2000)
10. Damouras, S., Chang, M.D., Sejdic, E., Chau, T.: An empirical examination of detrended fluctuation analysis for gait data. Gait & Posture 31, 336-340 (2010)
11. Ota, L., Uchitomi, H., Suzuki, K., Hove, M.J., Orimo, S., Miyake, Y.: Relationship between Fractal Property of Gait Cycle and Severity of Parkinson's Disease. In: IEEE/SICE International Symposium on System Integration (SII), 236-239 Dec. 2011 Kyoto
12. Peng, C.-K., Havlin, S., Stanley, E.H., Goldberger, A.L.: Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5, 82-87 (1995)
13. Costa, M., Goldberger, A.L., Peng, C.-K.: Multiscale entropy analysis of physiologic time series. Physical Review Letters E 71, 021906 (2005)
14. Makowiec, D., Dudkowska, A., Zwierz, M., Galaska, R., Rynkiewicz, A.: Scale invariant properties in heart rate signals. Acta Physica Polonica B37(5), 1627-1639 (2006)
15. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge (1992)
16. Schapire, R.E.: The Design and Analysis of Efficient Learning Algorithms. MIT Press, Cambridge, MA (1992)
17. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
18. Jin, R. Liu, Y., Si, L., Carbonell, J., Hauptmann, A.G.: A new boosting algorithm using input-dependent regularizer. In: 20th International Conference on Machine Learning, AAAI Press, Palo Alto (2003)
19. Senyukova, O. Gavrishchaka, V.: Ensemble Decomposition Learning for Optimal Utilization of Implicitly Encoded Knowledge in Biomedical Applications. In: IASTED International Conference on Computational Intelligence and Bioinformatics, pp. 69-73. ACTA Press, Calgary (2011)
20. Gavrishchaka, V., Davis, K., Senyukova, O.: Multi-complexity measures for early detection and monitoring of neurological abnormalities from gait time series. In: 2013 International

Symposium on Computational Models for Life Sciences (CMLS-13), Sydney, Australia, November 2013. AIP Conference Proceedings, vol. 1559, pp. 47-56. AIP Publishing LLC, New York (2013)
21. Senyukova, O., Gavrishchaka, V., Koepke, M.: Universal Multi-Complexity Measures for Physiological State Quantification in Intelligent Diagnostics and Monitoring Systems. In: Proceedings of The First International Aizu Conference on Biomedical Informatics and Technology (ACBIT 2013), Aizu-Wakamatsu, Septmber 2013. Communications in Computer and Information Science, vol. 404, pp. 69-83. Springer, Heidelberg (2013)
22. Theodoridis, S. and Koutroumbas, K.: Pattern Recognition, Academic Press (1999).
23. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., Kanto, A.: Dynamics of market correlations: Taxonomy and portfolio analysis. Phys Rev E 68, 056110 (2003)
24. Tumminello, M., Lillo, F., Mantegna, R.N.: Correlation, hierarchies, and networks in financial markets. Journal of Economic Behavior & Organization 75, 40-58 (2010)
25. Mantegna, R.N.: Hierarchical structure in financial markets. Eur. Phys. J. B 11, 193-197 (1999)
26. Vaccarino, A.L., Anderson, K., Borowsky, B., Duff, K., Giuliano, J., Guttman, M., Ho, A.K., Orth, M., Paulsen, J.S., Sills, T., van Kammen, D.P., Evans, K.R., PREDICT-HD and REGISTRY Investigators Coordinators: An item response analysis of the motor and behavioral subscales of the unified Huntington's disease rating scale in Huntington disease gene expansion carriers. Movement Disorders 26(5), 877-884 (2011)
27. Novak, M.J.U., Tabrizi, S.J.: Huntington's Disease: Clinical Presentation and Treatment. In: Brotchie, J., Bezard, E. and Jenner, P. (eds.) International Review of Neurobiology, pp. 297-323. Academic Press (2011)
28. Biglan, K.M., Zhang, Y. Long, J.D., Geschwind, M., Kang, G.A., Killoran, A., Lu, W., McCusker, E., Mills, J.A., Raymond, L.A., Testa, C., Wojcieszek, J., Paulsen, J.S., PREDICT-HD Investigators of the Huntington Study Group: Refining the diagnosis of Huntington disease. Frontiers in Aging Neuroscience, 5(12) (2013)
29. Klucken, J., Barth, J., Kugler P., Schlachetzki, J., Henze, T., Marxreiter, F., Kohl, Z., Steidl, R., Hornegger, J., Eskofier, B., Winkler, J.: Unbiased and Mobile Gait Analysis Detects Motor Impairment in Parkinson's Disease. PLoS ONE 8(2) (2013)
30. Breen, D.P., Evans, J.R. Farrell, K., Brayne, C., Barker, R.A.: Determinants of Delayed Diagnosis in Parkinson's Disease. Journal of Neurology 260(8), 1978-1981 (2013)
31. Bouche, P., Le Forestier, N., Maisonobe, T., Fournier, E., Willer, J.C.: Electrophysiological diagnosis of motor neuron disease and pure motor neuropathy. Journal of Neurology 246(7), 520-525 (1999)
32. Hausdorff, J.M., Zemany, L., Peng, C., Goldberger, A.L.: Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children. J Appl Physiol 86(3), 1040-1047 (1999)
33. Johansson, G.: Visual Perception of Biological Motion and a Model for its Analysis. Perception and Psychophysics 14(2), 201-211 (1973)
34. Barclay, C., Cutting, J., Kozlowski, L.: Temporal and Spatial Factors in Gait Perception that Influence Gender Recognition. Perception and Psychophysics 23(2), 145-152 (1978)
35. Winter, D.: The Biomechanics and Motor Control of Human Gait. Univesity of Waterloo Press, Waterloo, ON, Canada (1987)
36. Rani, M. P., Arumugam, G.: An Efficient Gait Recognition System For Human Identification Using Modified ICA. International Journal of Computer Science & Information Technology (IJCSIT) 2(1), 55-67(2010)
37. BenAbdelkader, C., Cutler, R.G., Davis, L.S.: Gait Recognition Using Image Self-Similarity. EURASIP Journal on Applied Signal Processing 4, 572-585 (2004)
38. Bouchrika, I., Goffredo, M., Carter, J., Nixon, M.: On Using Gait in Forensic Biometrics. J Forensic Sci 56(4), 882-889 (2011)
39. Boyle, M., Klausner, A., Starobinski, D., Trachtenberg, A., Wu, H.: Gait-Based Smartphone User Identification. Proceedings of the 9th International Conference on Mobile systems, Applications, and Services (MOBISYS'11), Bethesda, Maryland, USA, pp. 395-396 (2011)
40. Schapire, R.E., Freund,Y.: Boosting: Foundations and Algorithms. MIT Press (2012)