GAIT

BIOMEDICAL ENGINEERING | RESEARCH ARTICLE

Analysis of altered complexity of gait dynamics with aging and Parkinson's disease using ternary Lempel–Ziv complexity

Chandrakar Kamath

BIOMEDICAL ENGINEERING | RESEARCH ARTICLE

Analysis of altered complexity of gait dynamics with aging and Parkinson's disease using ternary Lempel–Ziv complexity

Chandrakar Kamath[1]*

*Corresponding author: Chandrakar Kamath, ShanthaNilaya, 107, First stage, Ananthnagar, Manipal 576104, India
E-mail: kamath.chandrakar@gmail.com

Abstract: Fluctuations in stride interval series show complex dynamical behavior in healthy young adults. Hypothesizing that these stride interval complexity changes would be altered by changes in neurological function associated with aging and certain disease states, we aimed to develop a tool to facilitate clinical judgments to assess the complex dynamical behavior in the stride series in discerning young, elderly, and Parkinson's disease (PD) classes. This novel approach, which employs a new variant of coarse-graining in conjunction with Lempel–Ziv complexity measure, yields useful, reliable, and predictive results. We also show the presence of nonlinear deterministic structures in the stride time series and appropriateness of the application of our nonlinear approach through surrogate data analysis. The findings show that the fluctuations are more complex/random in elderly and PD classes than those in young class. These findings may add to the growing body of literature supporting the clinical utility of this new approach to stride time series.

Subjects: Biomedical Engineering; Computational Neuroscience; Health and Social Care

Keywords: coarse-graining techniques; complex dynamical behavior; Lempel–Ziv complexity; Parkinson's disease; stride interval time series

ABOUT THE AUTHOR

Chandrakar Kamath received his BE degree from Mysore University in the year 1978 and MTech degree from Mangalore University in the year 1989. Till recently he has been working as a professor in the Department of Electronics and Communication Engineering at Manipal Institute of Technology, Manipal University. He has published several papers in International and Indexed Journals. His research interests include nonlinear dynamics, chaotic theory, biosignal processing, in particular nonlinear signal processing.

PUBLIC INTEREST STATEMENT

Despite continuous and extensive preventive efforts, falls remain a major cause of morbidity and mortality among older subjects and neurodegenerative disorder patients. Falls usually lead to fractures, hospitalization, and may be, sometimes even death. This important public health problem which demands right time interventions to reduce the risk of falling has been tried with varying degree of success. The severe effects of falls and these interventions, which often require high cost, necessitate an early diagnosis of subjects who are at high risk of falls. Some measures of gait have been found to be useful in identifying potential fallers. In this work, we propose a simple approach to develop a tool to facilitate clinical judgments to assess the complex dynamical behavior in the stride time series of healthy young controls, healthy elderly subjects, and patients suffering from Parkinson's disease and identify their risk of falls.

1. Introduction

Most of the early childhood of humans is spent on learning how to walk. In adulthood, walking is done almost automatically, but in older adults and also, patients with neurodegenerative diseases more attention is needed to carry out the complex task of walking. Otherwise, they are at the risk of falling (Bridenbaugh & Kressig, 2011; Hausdorff, Rios, & Edelberg, 2001). Despite continuous and extensive preventive efforts, falls remain a major cause of morbidity and mortality among older subjects and neurodegenerative disorder patients. Falls usually lead to fractures, hospitalization, and may be, sometimes even death. The severe effects of falls and these interventions, which often require high cost, necessitate an early diagnosis of subjects who are at high risk of falls. This important public health problem which demands early preventive interventions to reduce the risk of falling has been tried with varying degree of success. Nevertheless, there is still scope for improvement of targeting efficiency so that health care resources are assigned to the most susceptible subjects.

A frailty stride is one of the most prevalent and sensitive risk factors for falls (Bridenbaugh & Kressig, 2011), (Tinetti & Kumar, 2010; *Journal of the American Geriatrics Society*, 2011). Goldberger, Rigney, Mietus, Antman, and Greenwald (1998) suggested that a healthy system has a certain amount of inherent variability and that this variability is not random but contains an order and can be characterized by nonlinear measures. It is this deterministic variability which provides healthy flexibility to the system (Pool, 1989). Further, it is the loss of this flexibility which may be linked different ailments. The fluctuations overlying the cyclic trend in human walking may reveal valuable information about the neuromuscular processes responsible for healthy and pathological locomotor patterns. In this context, gait variability measures have been found to be significant in the evaluation of elderly/disordered gait for predicting falls and/or age-/disease-related changes leading to impairment. To evaluate an abnormal gait it is also equally essential to perform studies on normal gait. In this study, we investigate the differences in the gait of healthy young (normal), healthy elderly (abnormal/disordered), and patients (diseased) with neurodegenerative disease from the point of complexity of the dynamics of the gait time series. For biological systems, complexity refers to the presence of nonrandom fluctuations in the seemingly irregular dynamics of physiological signals (Grassberger, 1991). Such theory has suggested that healthy dynamic stability arises from the combination of specific feedback mechanisms and spontaneous properties of interconnected networks. However, the weak connection between systems or within system is the cause of the disorder/disease which is characterized by an increased complexity/randomness of the time series Pincus, (1995, 2006; Pincus & Goldberger, 1994).

In the literature, different linear and nonlinear theory tools have been used to measure gait parameters in healthy subjects and gait disordered patients. Analysis of linear statistics, however, does not directly address the complexity of the gait time series and thus may potentially miss important inherent information. Gait kinematic parameters have also been extensively treated algorithmically (i.e. smoothing, differentiation, and normalization) to provide a mean picture of the subject's movement, but distorting the temporal structure of variability (Stergiou, Buzzi, Kurz, & Heidel, 2004). On the other hand, measures from nonlinear dynamics estimate how a neuromuscular behavior changes over time and provide information about the structure or organization of the locomotion (Sosnoff, Valantine, & Newell, 2006; Stergiou et al., 2004). Several complexity measures have been proposed in the literature (Eckmann & Ruelle, 1992; Jeong, Kim, & Han, 1998; Tononi, Edelman, & Sporns, 1998). It is very important to note that alteration (a decrease or an increase) in complexity depends upon the metric used to quantify complexity and also, the task constraints within which the system is operating (Goldberger, Peng, & Lipsitz, 2002). Because the clinical patients get tired after a lengthy walk which will interfere with the gait recordings, it is difficult to obtain walking data over long period of time from them. For this reason, the subjects will usually be asked to walk only for 5 or 6 min. As an implication, the measure of complexity used must give reliable results for short signal segments, which is important in most experimental and clinical gait studies (Ferenets et al., 2006). Moreover, the results must be easy to comprehend by most clinicians. These reasons suggest that complexity measures which are relatively easy to compute and comprehend by doctors are needed for complexity analysis of the gait time series. The complexity measure proposed by Lempel and Ziv has

been extensively used to characterize the randomness present in the finite sequences of biomedical signals (Lempel & Ziv, 1976; Hu, Gao, & Principe, 2006; Zhang, Zhu, Thakor, & Wang, 1999; Li et al., 2008). Lempel–Ziv complexity (LZC) is a nonparametric measure of complexity for finite sequences related to the number of distinct substrings and the rate of their occurrence along the sequence and without the drawbacks of the classical nonlinear complexity methods. Hence, LZC can provide a way for estimating a time series structure or organization. LZC analysis has found wide applications in various disciplines (Adiego, Navarro, & De la Fuente, 2004; Louchard & Szpankowski, 1997; Savari, 1997) and also, in biomedical signal analyses, to estimate the complexity of discrete-time signals (Abásolo, Alcaraz, Rieta, & Hornero, 2011; Abásolo, Hornero, Gómez, García, & López, 2006; Aboy, Hornero, Abasolo, & Alvarez, 2006; Ferenets et al., 2006; Hausdorff et al., 2001; Kamath, 2013, 2015; Nagarajan, 2002; Radhakrishnan & Gangadhar, 1998; Torres et al., 2008; Wu & Xu, 1991; Xu, Liu, Ren Liu, & Yang, 1997; Zhang & Roy, 2001a; Zhang, Roy, & Weber, 2001b; Zhang, Zheng, & Jiang, 1998; Zhang et al., 1999; Zhao, Wang, Xu, & Yu, 2011).

LZC is a nonparametric complexity measure in Kolmogorov's sense. It contains the notion of complexity in statistical sense too (Zozor, Ravier, & Buttelli, 2005). LZC is sensitive to temporal amplitude distribution which permits coarse-graining of the original time series. LZC does not require any parameters and it is model-independent: only those differences between activity patterns that make a difference to the underlying system itself are considered, no matter whether the system is dominated by deterministic chaos or stochastic processes (Zhang et al., 2001b). LZC represents a sequence pattern complexity in a dynamic system. Because of these reasons, LZC may be a better choice for the estimation of complexity of biomedical signals, including gait signal.

LZC analysis is based on a coarse-graining of the measurements. Hence, the original time series needs to be transformed into a symbol sequence of finite alphabet before the computation of complexity. One can consider several methods to convert a real-valued time series into a symbolic string. In most of such studies, the signal has been transformed into a two-symbol (binary) sequence by comparison with a specific threshold (Abásolo et al., 2006, 2011, 2011; Aboy et al., 2006; Hu et al., 2006; Nagarajan, 2002; Zhang et al., 1999, 2001b). In few studies, the signal is transformed into a three-symbol (ternary) sequence employing two thresholds determined from the median and the dynamic range of the signal (Abásolo et al., 2006; Zhang et al., 2001b). Very few studies have used an alphabet with more than three symbols (Kamath, 2013, 2015; Sarlabous et al., 2009; Talebinejad, Chan, & Miri, 2011). Sarlabous et al. have found that LZC with two symbols can evaluate only complexity, with a large number of symbols can evaluate only changes in amplitude of the signal, while an intermediate number of symbols can quantify complexity changes as well as amplitude changes of the signal (Sarlabous et al., 2009).

Although LZC has been extensively used in wide variety of applications in different arenas, its application in gait analysis has been very limited. Fatmehsari and Bahrami (2011) used LZC in the nonlinear analysis of gait in patients with Parkinson's disease (PD) and found that gait time series of PD patients had larger LZC values than those of control subjects. In a comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains LZC have been tried as one of the various features by Sejdic, Lowry, Bellanca, Redfern, and Brach (2014). Their study showed that gait accelerometry time series of PD patients have smaller medial–lateral LZC values than those of healthy control subjects.

Fluctuations in the stride interval show complex dynamical behavior in healthy young adults. Aging is associated with a number of neurophysiological changes that may disturb the locomotor system's ability to generate complex dynamical behavior. Though the magnitude of these age-related changes depends to a greater extent on an individual's comorbid medical conditions, these changes to some degree reflect even in normal aging process. Nevertheless, healthy elderly subjects may serve to model subtle changes in neuromuscular control. PD is a neurodegenerative disorder, a disease of the basal ganglia which is characterized by gait impairment. With impairment confined primarily to central nervous system, PD can offer a contrast to aging for the study of the conditions

necessary for the complex dynamical behavior. We hypothesized that these stride time fluctuations would be altered by changes in neurological and physiological functions associated with aging and certain disease states and that these alterations would manifest as changes in fluctuation magnitude of stride-to-stride variability and complexity of fluctuation dynamics of stride time series. To investigate this hypothesis, we compared the complex dynamical behavior of the stride/gait time series of 1) healthy young controls with healthy elderly subjects, 2) healthy young controls with patients suffering from PD, and 3) healthy elderly subjects with Parkinson disease patients. Hausdorff et al. (1996) have shown that the normal human gait time series is extremely inhomogeneous and nonstationary, which fluctuates about the mean in an irregular and complex manner. Hypothesizing that nonlinear deterministic structures are present in the stride time dynamics and that a mere binary coarse-graining is not sufficient to capture the complex dynamical behavior, we employed four variants of coarse-graining transformations in conjunction with LZC measure to assess the complex dynamical behavior in the stride time series. In particular, we examine the impact of four different transformations on the LZC results of young, elderly, and Parkinson's gait time series: two transformations based on two-symbol (binary) sequences and other two based on three-symbol (ternary) sequences. Among the four transformations used for gait time series analysis, one is the widely used two-symbol (binary) conversion and the second is the widely used three-symbol (ternary) conversion. The remaining two, which outperform as we show below, are improvements over the two widely used (Ferrario, Signorini, & Magenes, 2009). We anticipated a more complex dynamical behavior with elderly or diseased population when compared with that of the young.

Linear stochastic processes can generate very complicated looking signals and that not all the structures that we observe in the time series are likely to be due to nonlinear dynamics of the system. To test the presence of nonlinear deterministic structures in the stride time series and thereby ascertain appropriateness of the application of our nonlinear approach, we carried out surrogate data analysis too.

2. Methods and materials

2.1. Database
The database used in this study can be downloaded from the public domain database at PhysioNet internet resource which is a large repository of various physiological signals. The database we used comprises stride time series from 10 healthy young adults, 10 healthy elderly subjects, and 10 PD patients and includes stride time series from 2 groups: Group-I and Group-II (Goldberger et al., 2000; Hausdorff et al., 2000). The protocols used in the two Groups differed a little only in terms of total walking time. Also, the difference in age between classes of the two groups was not statistically significant. Further, the PD group patients were free from other significant medical and neurological diseases. We first test the efficacy and robustness of our method in discerning the young, elderly, and PD classes in Group-I. Next, we validate the proposed method on Group-II. Group-I includes 5 healthy young adults (23–29 years old), 5 elderly subjects (71–77 years old), and 5 PD patients (60–77 years old). It was confirmed that the patients free from other pathologies which might lead to lower extremity weakness only participated. Over the duration of treatment the medication usage was not changed in the case of PD.

The healthy subjects (young and elderly) from the above classes were asked to walk continuously at their normal pace on an obstacle-free level ground along a long hallway for 15 min. The subjects from the PD class, however, were asked to walk at their normal pace up and down a long hallway for 6 min. This is because it is difficult to collect walking data for an extended period of time in clinical patients. The protocols differed slightly because of the use of two different locations. To measure the stride interval, ultrathin force-sensitive resistors were placed inside subject's shoes. These sensors produce a measure proportional to the force applied to the ground during movement. The analog force signal was sampled at 300 Hz with a 12 bit A/D converter, using an ambulatory, ankle-worn microcomputer that also recorded the data. Subsequently, the time between foot-strikes was automatically computed. The method for determining the stride interval is a modification of a previously

validated method that has been shown to agree with force-platform measures, a "gold" standard (Hausdorff et al., 1996).

Group-II is formed by selecting an equal number of stride time series, matched for age, from 5 healthy young adults (21–29 years old), 5 healthy elderly (71–77 years old), and 5 Parkinson disease patients (65–77 years old). Group-II is a subset of the database for neurodegenerative diseases contributed by Hausdorff et al. (2000) and Goldberger et al. (2000) and can be downloaded from the same public domain PhysioNet mentioned above. The subjects from the different classes were asked to walk at their normal pace up and down a 77 m long hallway for 5 min. To measure the gait rhythm and the timing of the gait cycle, force-sensitive insoles were place inside or under subject's shoes. The output from the footswitches which corresponds to force signal is sampled at 300 Hz and digitized using an analog-to-digital converter and then stored in a recorder. The recorded data is then analyzed using a validated software that determined initial and end contact times (and also, stride and swing times) of each stride.

2.2. Pre-processing the gait data
Before the application of the method of analysis it is necessary to pre-process the gait data. To minimize the start-up effects, the samples in the first 20 s of the recordings are removed (Hausdorff et al., 2000). Over the monitoring interval, each time the subject reached the deep curvature or end of the hallway, the subject had to turn around and continue walking. The strides associated with these turning events are to be treated as outliers and should be removed from the rest of the time series. To remove the outliers we employed the three-sigma rule (Hahn & Shapiro, 1994), which states that 99.7% of the normally distributed probability values lie within the range of (mean ± 3.SD), where SD is the standard deviation (SD). This implies that those samples which lie outside the range (median ± 3.SD) are outliers and hence, can be removed. The stride time series before and after median filtering showed a strong correlation which implies the effectiveness of the median filter in preserving the underlying dynamics, while removing the extraneous data points is associated with the changes in gait direction. In the removal process, median value and not mean value of the time series has been used because some outliers possessed large values and will affect the computation of the mean.

2.3. Measures of fluctuation magnitude of stride-to-stride variability
A linear measure focuses only on the magnitude of variation in a distribution irrespective of the order in which data points accumulate and thus potentially misses important inherent information. Nevertheless, in this study, hypothesizing that the stride interval variability would be altered by changes in neurological function associated with aging and certain disease states, first we investigate the linear statistics of stride time series of young, elderly, and PD subjects. Each gait record, in each group, is divided into segments, with 45 strides per segment. A thumb rule to select segment length is that it must be long enough to reliably estimate the measure of interest, while it must be short enough to accurately capture local activities. For each segment, the variability measures of the stepping patterns, namely the coefficient of variation (CV) and SD of the detrended stride time series ($SD_{detrended}$) are computed and the results of a particular class are averaged.

It is often difficult to use the usual SD to compare measurements from different populations. To get round this problem, two measures are used to assess the magnitude of stride-to-stride variability and gait unsteadiness: (1) CV, the CV of the original stride time series and (2) $SD_{detrended}$, the SD of the $SD_{detrended}$. It is important to note that both of these measures are not sensitive to changes in the ordering of the stride intervals or stride dynamics. That is to say, randomly ordering the time series will not affect these measures.

The CV expresses the SD as a percentage of what is being measured relative to the sample or population mean. CV is a normalized measure of stride-to-stride variability. It is defined as the ratio of the SD σ to mean μ as, CV = σ/μ. It shows the extent of variability in relation to mean of the

population. It provides a measure of relative variability. The only advantage is that it lets you compare the scatter of variables expressed in different units.

The SD of a time series, in general, provides a measure of overall variations in the gait with respect to mean. It is a metric for absolute variability. This measure may be influenced by the trend in the data and may fail to differentiate between a walk with large changes from stride to next and one in which stride changes are small. To minimize effects of local changes in the mean the time series is detrended. The detrended stride time series refers to time series from which the trend is removed. Detrending can be carried out by computing the first difference of the time series or removing the least-squares-fit straight line. In this study, the former method is used for detrending. $SD_{detrended}$ is a measure of variability which minimizes the effects of the local changes in the mean.

2.4. Complexity measure of fluctuation dynamics of stride time series

Fluctuation dynamics is about how the stride interval changes from one stride to the next, independent of the variance. We hypothesized that the stride interval dynamical complexity changes would be altered by changes in neurological function associated with aging and certain disease states. To quantify how the complexity of the fluctuation dynamics changes over time during walk, we employed LZC. The LZC algorithm was proposed by Lempel and Ziv to evaluate the randomness of finite sequences (Lempel & Ziv, 1976). It is rather a simple-to-compute nonparametric measure of complexity suitable for finite length one-dimensional signals related to the number of distinct substrings and the rate of their recurrence along the given sequence. Larger values of LZC imply higher complexity data. Since LZC analyzes finite symbol-sequences, it is essential that the given signal must first be coarse-grained, i.e. the signal to be analyzed is transformed into a sequence whose elements are only a few symbols. Thus, coarse-graining is a crucial aspect in the computation of LZC analysis. It has been found that on one extreme, with two symbols LZC can evaluate only complexity; on the other extreme, with a large number of symbols LZC can evaluate only changes in amplitude of the signal; while an intermediate number of symbols can quantify complexity changes as well as amplitude changes in the signal (Sarlabous et al., 2009). There are several methods to convert a real-valued time series into a symbolic string. The most commonly used computation of LZC is based on binary sequence. We hypothesize that the binary sequence cannot well characterize gait signal and may lose some important information in the signal. As an implication, we examine in this study the impact of four different transformations on the LZC results of young, elderly, and Parkinson's gait time series: (1) conventional two-symbol (binary) sequences, (2) improved two-symbol (binary) sequences (3) conventional three-symbol (ternary) sequences, and (4) improved three-symbol (ternary) sequences.

2.4.1. Conventional and improved binary coarse-graining methods

This is the simplest possible coarse-graining involving the division of data range into two partitions (binary partition with $L = 2$, L being the number of partitions) (Abásolo et al., 2011). Those data which are above a cut-off value (usually, either mean or median) are assigned a symbolic value of '1,' while those below the cut-off value are assigned a symbolic value of '0.' In case the cut-off value is chosen to be equal to the mean or median of the data, x_m, the time series x_i is transformed into the symbolic sequence S_i, where $S_i \in \{0, 1\}$, as given below.

$$S_i = \begin{cases} 1 \ if x_i \geq x_m \\ 0 \ if x_i < x_m \end{cases} \tag{1}$$

The Lempel–Ziv algorithm is then applied on the symbolic sequence, S_i. This binary string is scanned from left to right and a complexity counter $c(N)$ is incremented by one unit every time a new subsequence pattern is encountered in the scanning process, and the immediate next symbol is regarded as the beginning of the next subsequence pattern. The corresponding LZC is called binary LZC (BLZC).

Now, we present improved two-symbol (binary) sequenced coarse-graining method in which time series x_i is transformed into the symbolic sequence S_i based on the following definition.

$$S_{i+1} = \begin{cases} 0 \; if \; x_{i+1} \; \leq \; (1+p).\,x_i \\ 1 \; if x_{i+1} \; > \; (1+p).\,x_i \end{cases} \tag{2}$$

The parameter p represents the minimum quantization level for a symbol change in the coded string. The introduction of this parameter limits the effect of additive noise as well. In this study, p was varied in small steps of 0.0005 from 0 to 1 to arrive at an optimum value for the best separation among the young, elderly, and PD subjects. It is found that $p = 0.035$ yields optimum results.

2.4.2. Conventional and improved ternary coarse-graining methods

As hypothesized above, the conventional binary coarse-graining method may miss significant information in dynamical systems and multi-valued coarse-graining method (with large L) may miss complexity information and hence we employ ternary coarse-graining method ($L = 3$) which is explained below. Let x_i represent the discrete time series with x_{max} and x_{min} as maximum and minimum values, respectively. Then, with $L = 3$ representing the number of partitions for ternary coarse-graining we have,

$$d = (x_{max} - x_{min})/L \tag{3}$$

Let y_j ($j = 1, 2, L$) represent the set of unique symbols with each y_j corresponding to a particular partition. The time series x_i is then transformed into the symbolic sequence S_i where $S_i \in \{y_1, y_2, y_L\}$, as given below.

$$S_{i+1} = \begin{cases} 0 & if \; x_i \; \leq \; x_{min} \; + d \\ 1 & if \; x_{min} + d \; < \; x_i \; \leq \; x_{max} - d \\ 2 & if \; x_i \; > \; x_{max} - d \end{cases} \tag{4}$$

The Lempel–Ziv algorithm is then applied on the symbolic sequence, S_i. This multi-valued string is scanned from left to right and a complexity counter $c(N)$ is incremented by one unit every time a new subsequence pattern is encountered in the scanning process, and the immediate next symbol is regarded as the beginning of the next subsequence pattern. The resulting LZC is designated as multi-valued LZC (MLZC). In the present study, $L = 3$ and the MLZC is called ternary LZC (TLZC).

Now we present improved three-symbol (ternary) sequenced coarse-graining method in which time series x_i is transformed into the symbolic sequence S_i based on the following definition.

$$S_{i+1} = \begin{cases} 0 & if \; x_{i+1} \; < \; (1-p).\,x_i \\ 1 & if \; x_{i+1} \; < \; (1+p).\,x_i \\ 2 & if \, (1-p).\,x_i \; \leq \; x_{i+1} \; \leq \; (1+p).\,x_i \end{cases} \tag{5}$$

As mentioned above, the parameter p represents the minimum quantization level for a symbol change in the coded symbol sequence. The introduction of this parameter restricts the effect of additive noise as well. Like before, p was varied in small steps of 0.0005 from 0 to 1 to arrive at an optimum value for the best separation among the young, elderly, and PD subjects. It is found that the same value of $p = 0.035$ yields optimum results in the ternary case as well.

It is to be noted that irrespective of the method of coarse-graining the LZC algorithm remains the same. Any difference in the coarse-graining method gets reflected during the normalization procedure.

2.4.3. Lempel–Ziv complexity algorithm

Once the symbolic string is ready the LZC can be estimated using the following algorithm (Zhang et al., 1998):

cogent··engineering

(1) Let P denote the original string sequence i.e. $P = \{s_1, s_2, s_3, ...\}$, with s_i defined as in Equation (1). Let S and Q denote two subsequences of P and SQ be concatenation of S and Q. Also, let $SQ\pi$ be a sequence derived from SQ after its last character is deleted (π implying deletion of last character in the sequence) and $\upsilon(SQ\pi)$ denote the vocabulary of all different subsequences of $SQ\pi$.

(2) At the beginning, the complexity counter $c(N)=1$, $S = s_1$, $Q = s_2$, $SQ = s_1, s_2$, and therefore, $SQ\pi = s_1$.

(3) In general, with $S = s_1, s_2, s_3, ..., s_r$ and $Q = s_{r+1}$, $SQ\pi = s_1, s_2, s_3, ..., s_r$. If Q belongs to $\upsilon(SQ\pi)$ then Q is subsequence of $SQ\pi$ and not a new sequence.

(4) With S intact, change Q to s_{r+1}, s_{r+2} and check if Q belongs to $\upsilon(SQ\pi)$ or not.

(5) Keep repeating previous steps until Q does not belong to $\upsilon(SQ\pi)$. Now $Q = s_{r+1}, s_{r+2}, ..., s_{r+i}$ is not a subsequence of $SQ\pi = s_1, s_2, ..., s_{r+i-1}$. So increase $c(N)$ by 1.

(6) Thereafter, S is renewed to $S = s_1, s_2, ..., s_{r+i}$ and Q to $Q = s_{s+i+1}$.

(7) Repeat the previous steps until Q is the last character. At this point in time, the number of subsequences in P is $c(N)$, which corresponds to measure of complexity.

To arrive at a measure of complexity independent of sequence length, $c(N)$ must be normalized. If the length of the sequence is n and the number of different symbols is α, it has been shown that the upper bound of $c(N)$ is (Zhang et al., 1998)

$$c(N) < N/((1 - \epsilon N)\log_\alpha(N)) \tag{6}$$

where ε_N is a small quantity and $\varepsilon_N \to 0$ ($N\to\infty$). In general, $N/\log_\alpha(N)$ is the upper limit of $c(N)$, i.e.

$$\lim N \to \infty c(N) = b(N) = N/\log_\alpha(N) \tag{7}$$

For a binary conversion $\alpha = 2$, $b(N) = N/\log_2(N)$ and the resulting LZC is BLZC. For a ternary conversion $\alpha = L = 3$, $b(N) = N/\log_3(N)$ and the resulting LZC gives TLZC.

$c(N)$ can be normalized by $b(N)$ as

$$C(N) = c(N)/b(N) \tag{8}$$

$C(N)$, the normalized LZC, reflects the arising rate of new patterns along with the sequence and thus captures the temporal structure of the sequence. A larger value of LZC means that the chance of generating a new pattern is greater, so the sequence is more complex, and vice versa (Savari, 1997). In this work, we evaluate the evolution of new patterns in gait time series in young, elderly, and PD subjects.

2.5. Statistical and receiver operating characteristic analyses

Kruskal–Wallis tests are used to evaluate the statistical significances among LZCs of the stride time series of young, elderly, and PD classes. If statistical significances are found then the statistical difference of LZCs between different classes are evaluated using Mann–Whitney rank sum tests. These nonparametric tests are used because they make no assumption about the underlying distribution of the data. A p-value ≤ 0.05 is considered statistically significant. If significant differences between classes are found, then the ability of the nonlinear analysis method to discriminate gait of young, elderly, and PD classes is evaluated using receiver operating characteristic (ROC) plots in terms of area under ROC curve (AUC) (Zweig & Campbell, 1993). ROC curves are obtained by plotting sensitivity values (which represent that proportion of states identified as class-1) along the y axis against the corresponding (1-specificity) values (which represent the proportion of the correctly identified class-2) for all the available cut-off points along the x-axis. Accuracy is a related parameter that quantifies the total number of states (both class-1 and class-2) precisely classified. The AUC

measures this discrimination, that is, the ability of the test to correctly classify stride of class-2 and class-1 subjects and is regarded as an index of diagnostic accuracy. The optimum threshold is the cut-off point in which the highest accuracy (minimal false negative and false positive results) is obtained. This can be determined from the ROC curve as the closet value to the left top point (corresponding to 100% sensitivity and 100% specificity). An AUC value of 0.5 indicates that the test results are better than those obtained by chance, where as a value of 1.0 indicates a perfectly sensitive and specific test.

2.6. Surrogate data analysis

If the dynamics that generated the time series is not known or if the time series is noisy, in that case it is essential to investigate whether the amount of nonlinear deterministic dependencies is worth analyzing further or to treat the time series as stochastic. Hence, one of the first steps before applying the nonlinear technique to the data is to investigate if the application of such technique is justified. The main reason behind this rationale is that linear stochastic processes can generate very complicated looking signals and that not all the structures that we observe in the data are likely to be due to nonlinear dynamics of the system. The method of surrogate data test, introduced by Theiler, Eubank, Longtin, Galdrikian, and Farmer (1992), has been a popular validating test to address this issue. This test facilitates to find out if the regularity of the data is most likely due to nonlinear deterministic structure or due to variations in system parameters or due to random inputs to the system.

This section presents a brief sketch of the idea in that connection. The starting point is to create an ensemble of random nondeterministic surrogate data-sets that have the same mean, variance, and power spectrum as the original time series. The measured topological properties of the surrogate data-sets are compared with those of the original time series. If, in case, the surrogate data-sets and original data yield the same values for the topological properties (within the SD of the surrogate data-sets) then the null hypothesis that the original data are random noise cannot be ruled out. On the other hand, if the data under test are generated by a nonlinear process, the value for the topological property would be different from that of the surrogate data, and the null hypothesis that a linear method characterizes the data can be rejected.

The method of computing surrogate data-sets with the same mean, variance, and power spectrum as the original time series, but otherwise random is as follows: First find the Fourier transform of the original time series, then randomize the phases, and find the inverse Fourier transform. The resulting time series is that of the surrogate data. More details can be found in (Theiler et al., 1992). However, Rapp et al. have shown that inappropriately constructed random phase surrogates can lead to false-positive rejections of the surrogate null hypothesis (Rapp, Cellucci, Watanabe, Albano, & Schmah, 2001). They found that numerical errors in the computation of Fourier transform were the cause for this problem and that Welch windowing the data can eliminate false-positive rejections of the surrogate null hypothesis. Hence, in this study, we made sure that Welch window was introduced before the computation of the Fourier transform of the stride interval segment whose surrogate needs to be found.

3. Results and discussion

After pre-processing the gait data as explained in Section 2.2, it is necessary to normalize the data by subtracting from each sample the mean of the time series and dividing the result by the SD of the time series. Normalization removes most of the very large within and across-subject variability in the signal under consideration. Normalization also tends to produce values that are more exchangeable across different laboratories and research studies.

Unlike a linear measure which focuses on the magnitude of variation in a distribution irrespective of the order in which data points accumulate, a nonlinear measure is explicitly concerned with the temporal evolution of structure of the data variability and hence, may unravel more meaningful information.

3.1. Linear statistics

In this study, first we investigate the linear statistics of stride time series of young, elderly, and PD classes from Group-I. Each gait record in each class is divided into segments, with 45 strides per segment. A thumb rule to select segment length is that it must be long enough to reliably estimate the measure of interest, while it must be short enough to accurately capture local activities. For each segment, the variability measures are computed and the results of a particular class are averaged. Table 1 shows the linear measures of fluctuation magnitude of stride-to-stride variability characteristic of young, elderly, and PD gait time series of Group-I. All the variability measures are expressed as mean ± SD. The table shows differences among the three examined classes, possibly indicating differences in aging process and neuropathology. It is found that the average stride time is longer in the PD class compared to those of young and elderly classes. The two measures of fluctuation magnitude, CV and $SD_{detrend}$, are also considerably increased in the PD class compared to those of the young and elderly classes. The CV and $SD_{detrend}$ of the young class are the lowest while those of PD are the highest. The elderly class showed intermediate values for both the parameters. These results indicate that the magnitude of stride-to-stride variability in elderly is increased by the aging process and in PD subjects is significantly increased by the neurodegenerative disease. These significantly increased values suggest higher susceptibility for falls. These measures are based on the notion that a stable walking pattern has less variability and higher variability, as in elderly and PD subjects, is associated with higher vulnerability for falls. These results are in agreement with previous studies (Gabell & Nayak, 1984). Kruskal–Wallis tests are performed to evaluate the statistical differences between the different measures of the three classes. The test detected significant class differences for all the measures. In the case of stride time CV, $p = 0$ and chi-square = 128.42. In the case of stride time $SD_{detrended}$, $p = 0$ and chi-square = 137.86. It is to be noted that these statistical measures do not provide insight into how the motor system responds to disturbances that are present in the walking patterns over time. This implies that measures which can encapsulate time-dependent changes in the walking patterns, like LZC may serve as better metric for assessing balance.

3.2. Coarse-grained LZC transformations

Now, we investigate the complexity of particularly short stride-to-stride sequences in the order of 45 consecutive stride intervals by means of the four different coarse-grained LZC transformations. For this, each gait record is divided into segments each with 45 samples, in the case of young, elderly, and PD gait records of Group-I. The four LZC transformations: BLZC, IBLZC TLZC, and ITLZC are applied to each of these segments from different classes to examine the impact of these transformations and to decide whether a particular segment belongs to young, elderly, or PD class. The performance of these transformations on the LZC results of young, elderly, and Parkinson's gait time series are shown in Table 2. The p-values from Kruskal–Wallis tests for BLZC ($p = 0.9378$) and TLZC ($p = 0.1012$) are insignificant. This clearly implies that conventional BLZC and TLZC are not capable of separating properly young, elderly, and PD gait time series from each other. On the other hand, the improved versions IBLZC ($p = 1.11 \times 10^{-16}$) and ITLZC ($p = 0$) show their high discriminating capability indicating that they are a better sensitive measure of age- and disease-related changes in gait.

Table 1. Gait rhythm variability parameters of young, elderly, and Parkinson disease classes of Group-I

Variability parameter	Class			Kruskal–Wallis test results
	Young1	**Elderly1**	**PD1**	
Stride time T_{stride}, ms	1104.00 ± 36.96	1017.00 ± 31.77	1124.00 ± 69.0	$p = 0$
				Chi-sq = 97.13
Stride time CV, (%)	1.58 ± 0.16	1.65 ± 0.32	4.22 ± 1.60	$p = 0$
				Chi-sq = 128.42
Stride time $SD_{detrended}$, ms	20.38 ± 2.23	22.07 ± 3.76	65.67 ± 27.93	$p = 0$
				Chi-sq = 137.86

Note: All values are expressed as mean ± SD.

cogent ·· engineering

| Table 2. Kruskal–Wallis test results of performance of four different transformations on the LZC results of young, elderly, and Parkinson's gait time series of Group-I: (1) BLZC, (2) IBLZC, (3) TLZC, and (4) ITLZC				
Transformed LZC	**Class**			**Kruskal–Wallis test results**
	Young1	**Elderly1**	**PD1**	
BLZC	1.241 ± 0.083	1.241 ± 0.082	1.241 ± 0.092	$p = 0.9378$
				Chi-sq = 0.13
IBLZC	0.470 ± 0.049	0.620 ± 0.083	1.029 ± 0.095	$p = 1.11 \times 10^{-16}$
				Chi-sq = 73.47
TLZC	0.939 ± 0.052	0.865 ± 0.046	0.861 ± 0.058	$p = 0.1012$
				Chi-sq = 4.58
ITLZC	0.438 ± 0.066	0.562 ± 0.081	1.018 ± 0.058	$p = 0$
				Chi-sq = 83.45

For this reason, the further discussion will be confined to mostly IBLZC and ITLZC. The distributions of IBLZC and ITLZC values for young, elderly, and PD classes from Group-I are illustrated in Figure 1(a) and (b) using box and whisker plots. It is seen that in either case young class has lower complexity/randomness, PD class the higher, and the elderly class an intermediate value. This shows that the complexity/randomness nature is low in the stride series of the young class, increases with aging as seen in the stride series of the elderly class, and increases significantly because of the progression of the disease as seen in the stride series of the PD class. The elderly subjects chosen in this study were free from any overt neurological disease. Thus, the increased complexity may be due to alterations in gait dynamics caused by subtle changes in neurophysiological control. It is to be understood that low complexity in young class implies a more organized class than other classes. One can attribute a new interpretation to this behavior. Since LZC, in general, represents rate of appearance of new patterns in a time series, a lower value of LZC signifies smaller chance of occurrence of the new patterns and thus a less complex dynamical behavior. This means that the young class exhibits a less complex dynamical behavior for walking continuously at their normal pace on an obstacle-free level ground. This complex dynamical behavior increases with aging as seen in the elderly class stride time series and increases significantly because of the progression of the disease as seen in the PD class stride time series. LZC accounts for only those activities affecting the patterns of the underlying system. It does not actually matter whether the system is deterministic or stochastic. It is found from Figure 1 that the boxes of the three classes show a better separation between healthy (young and elderly) and diseased (PD), in the case of ITLZC.

3.3. Surrogate analyses and their implications

To test the presence of deterministic structures in the gait time series and thereby ascertain appropriateness of the application of our nonlinear approach, we carried out surrogate data analysis using IBLZC and ITLZC on both the Groups I and II (Table 3). Fifteen surrogate series for each of the original gait series are constructed as explained in the Section 2.6. The mean of surrogates IBLZC and ITLZC values for the 15 surrogate series are computed and compared with those of the original series. Tables 4 and 5 show results of surrogate data analyses for young, elderly, and PD classes from both the Groups I and II. The values are expressed as mean ± SD. The statistical significance of the differences between the nonlinear measures of the original and surrogate series, during walking, in the three classes of the two groups investigated using Mann–Whitney rank sum tests are also specified in Tables 4 and 5. Comparison between the respective IBLZC and ITLZC of the original and surrogate gait series for young and elderly, in both the Groups I and II, reveals highly significant differences (*p*-value < 0.0001) implying that the relevant patterns in the original time series of young and elderly cannot be considered present by chance. This indicates that the fluctuations observed in the original gait time series of young and elderly classes are not randomly derived, but they are deterministic in nature. This justifies the application of our nonlinear approach. Interestingly, comparison between

Figure 1. The distributions of (a) IBLZC and (b) ITLZC values for young, elderly, and PD classes from Group-I using box and whisker plots.
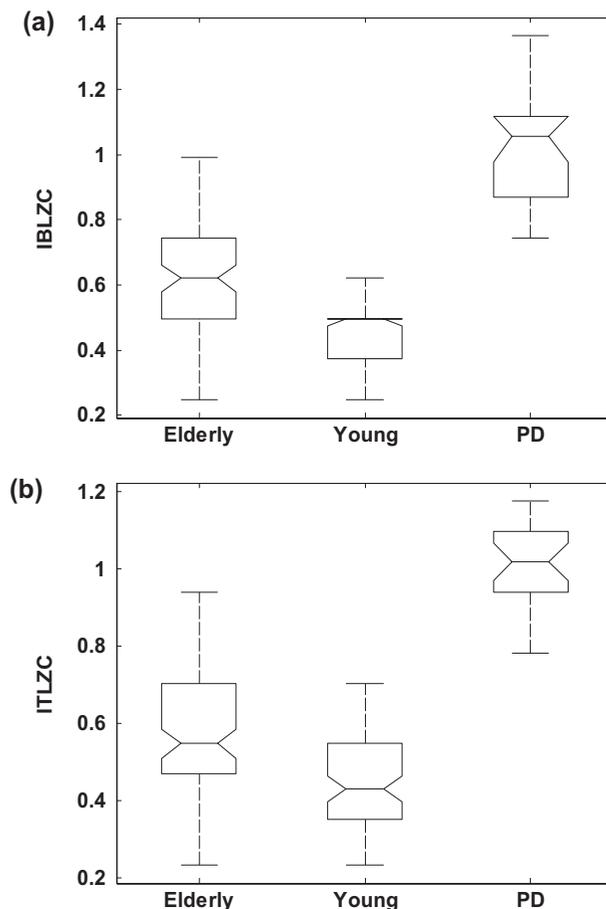
the respective IBLZC and ITLZC of the original and surrogate gait series for PD class, in both the groups, reveals no significant difference (p-value > 0.05). This means that the deterministic nature in the PD patients might have been either considerably decreased or lost because of the disease. One can use p-value as a marker of the order of deterministic nature. Comparison of the respective IBLZC and ITLZC of the original and surrogate gait series for young, elderly, and PD classes shows that the p-value is the least for the young class, intermediate for elderly class, and the highest for PD class. This shows that the deterministic nature is prominent in the young class, decreases with aging as noticed in the elderly class, and further diminishes or may break down because of the progression of the disease as observed in the PD class.

Table 3. Comparison of IBLZC with its surrogate IBLZC using Mann–Whitney rank sum tests for classes of Group-I and Group-II			
Class	**IBLZC**	**IBLZC_surrogate**	**p-value**
Young1	0.470 ± 0.049	0.854 ± 0.081	5.16×10^{-21}
Elderly1	0.620 ± 0.083	0.951 ± 0.048	4.18×10^{-17}
PD1	1.029 ± 0.095	1.098 ± 0.090	0.803
Young2	0.707 ± 0.094	0.879 ± 0.092	7.90×10^{-04}
Elderly2	0.620 ± 0.091	0.859 ± 0.079	6.85×10^{-04}
PD2	0.888 ± 0.059	1.079 ± 0.058	0.153

Note: All values are expressed as mean ± SD.

cogent ·· engineering

Table 4. Comparison of ITLZC with its surrogate ITLZC using Mann–Whitney rank sum tests for classes of Group-I and Group-II

Class	ITLZC	ITLZC_surrogate	*p*-value
Young1	0.438 ± 0.066	0.830 ± 0.031	3.79×10^{-28}
Elderly1	0.562 ± 0.081	0.847 ± 0.051	1.98×10^{-16}
PD1	1.018 ± 0.058	1.001 ± 0.030	0.361
Young2	0.704 ± 0.115	0.870 ± 0.058	8.94×10^{-04}
Elderly2	0.570 ± .104	0.861 ± 0.087	2.90×10^{-06}
PD2	0.954 ± 0.089	0.990 ± 0.094	0.770

Note: All values are expressed as mean ± SD.

Table 5. Statistical significance of IBLZC and ITLZC between young, elderly and PD classes of Group-I and Group-II using Mann–Whitney rank sum test

Parameter	Class-1	Class-2	*p*-value
IBLZC	Young1	Elderly1	5.79×10^{-05}
	Young1	PD1	2.20×10^{-14}
	Elderly1	PD1	3.28×10^{-12}
ITLZC	Young1	Elderly1	5.92×10^{-07}
	Young1	PD1	9.54×10^{-15}
	Elderly1	PD1	3.87×10^{-13}
IBLZC	Young2	Elderly2	0.02
	Young2	PD2	4.29×10^{-06}
	Elderly2	PD2	3.49×10^{-05}
ITLZC	Young2	Elderly2	0.005
	Young2	PD2	5.32×10^{-08}
	Elderly2	PD2	8.84×10^{-07}

3.4. Statistical significance of IBLZC and ITLZC measures

Next, we evaluated the statistical significance of IBLZC and ITLZC between different classes using Mann–Whitney rank sum test. The results for Group-I and Group-II are depicted in Table 5. The *p*-values of the tests show that IBLZC and ITLZC have diagnostic potential to separate one class from other classes. It is observed that LZC values in both the Groups exhibit a significant difference between healthy young and healthy elderly classes. However, it is found that between healthy young/ elderly and PD classes the statistical difference is considerably significant in both the Groups. This is because the alterations in gait dynamics in the elderly may be due to due to subtle changes in the neuromuscular control, while in the PD subjects this clearly reflects neuropathology.

Since significant differences between classes are found, the ability of the nonlinear analysis methods (IBLZC and ITLZC) to discriminate gait of young, elderly, and PD classes is evaluated using ROC plots. The corresponding ROC plots are shown in Figure 2(a) for the case of IBLZC and in Figure 2(b) for the case of ITLZC. The results of evaluation of diagnostic quality of the IBLZC and ITLZC in separating young, elderly, and Parkinson's gait time series from Group-I and Group-II are summarized in Tables 6 and 7. It is found that the IBLZC and ITLZC perform extremely well in their diagnostic ability. In discerning young and elderly of Group-I, IBLZC (ITLZC) yielded the following diagnostic parameters: AROC = 0.6731 (0.7166), accuracy = 65.7% (67.4%), specificity = 54.9% (61.5%), sensitivity = 77.4% (73.8%), and precision = 61.3% (63.9%). In separating elderly and PD of Group-I, IBLZC (ITLZC) yielded the following diagnostic parameters: AROC = 0.9444 (0.9662), accuracy = 84.6% (92.3%), specificity = 81.3% (93.4%), sensitivity = 96.2% (88.5%), and precision = 59.5% (79.3%). In

Figure 2. The ROC plots (a) for the case of IBLZC and (b) for the case of ITLZC for young, elderly, and PD classes from Group-I. The diagonal line (dotted line) from (0, 0) to (1, 1) represents ROC curve that cannot discriminate between the two compared states.
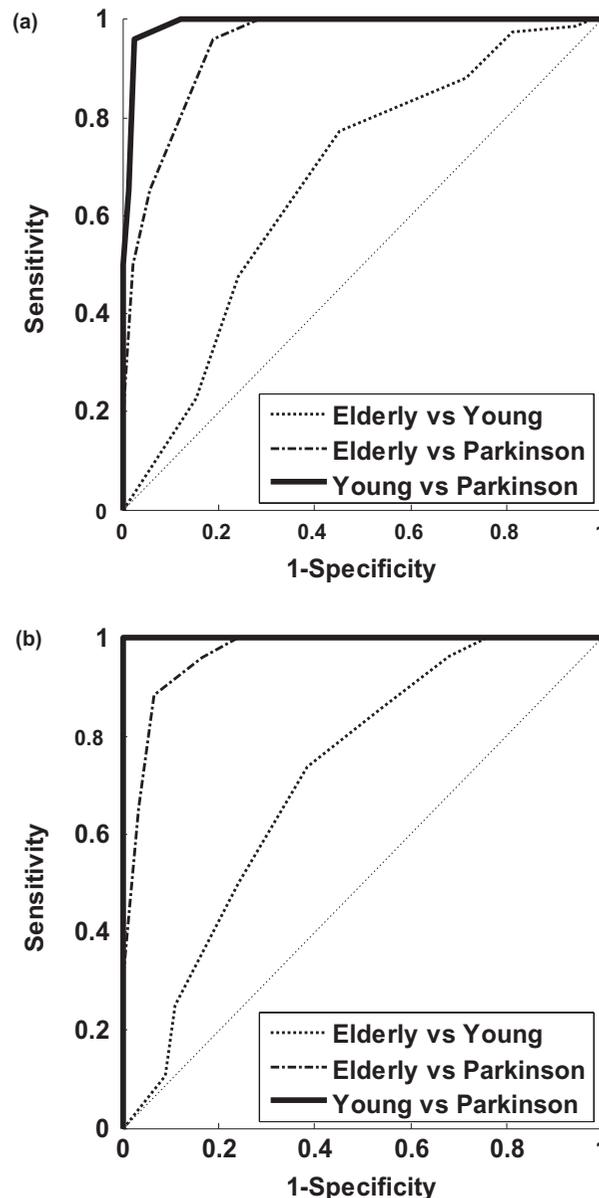
discerning young and PD of Group-I, IBLZC (ITLZC) yielded the following diagnostic parameters: AROC = 0.9908 (1.0000), accuracy = 97.3% (100%), specificity = 97.6% (100%), sensitivity = 96.2% (100%), and precision = 92.6% (100%). Comparison shows that ITLZC performs better than IBLZC. In discerning young and elderly of Group-II, IBLZC (ITLZC) yielded the following diagnostic parameters: AROC = 0.6309 (0.6448), accuracy = 65.5% (67.3%), specificity = 82.1% (60.7%), sensitivity = 48.1% (74.1%), and precision = 72.2% (64.5%). In separating elderly and PD of Group-II, IBLZC (ITLZC) yielded the following diagnostic parameters: AROC = 0.8264 (0.8900), accuracy = 71.7% (81.1%), specificity = 67.9% (94.1%), sensitivity = 76.0% (96.4%), and precision = 67.9% (79.3%). In discerning young and PD of Group-II, IBLZC (ITLZC) yielded the following diagnostic parameters: AROC = 0.8681 (0.9378), accuracy = 82.7% (83.7%), specificity = 70.4% (81.5%), sensitivity = 96.0% (84.0%), and precision = 75.0% (80.8%). This shows the robustness of our approach in discerning healthy young, healthy elderly, and Parkinson patients and also justifies our hypotheses. A possible explanation could be that more information from the gait time series is retained during coarse-graining process

Table 6. Descriptive results of ROC analysis using IBLZC for discriminating young, elderly, and PD classes of Group-I and Group-II					
Comparison between	**AUC**	**Average sensitivity (%)**	**Average specificity (%)**	**Average precision (%)**	**Average accuracy (%)**
Elderly1 and young1	0.6731	77.4	54.9	61.3	65.7
Elderly1 and PD1	0.9444	96.2	81.3	59.5	84.6
Young1 and PD1	0.9908	96.2	97.6	92.6	97.3
Elderly2 and young2	0.6309	48.1	82.1	72.2	65.5
Elderly2 and PD2	0.8264	76.0	67.9	67.9	71.7
Young2 and PD2	0.8681	96.0	70.4	75.0	82.7

Table 7. Descriptive results of ROC analysis using ITLZC for discriminating young, elderly, and PD classes of Group-I and Group-II					
Comparison between	**AUC**	**Average sensitivity (%)**	**Average specificity (%)**	**Average precision (%)**	**Average accuracy (%)**
Elderly1 and young1	0.7166	73.8	61.5	63.9	67.4
Elderly1 and PD1	0.9662	88.5	93.4	79.3	92.3
Young1 and PD1	1.0000	100.0	100.0	100.0	100.0
Elderly2 and young2	0.6448	74.1	60.7	64.5	67.3
Elderly2 and PD2	0.8900	96.4	94.1	75.0	81.1
Young2 and PD2	0.9378	84.0	81.5	80.8	83.7

when a three-symbol conversion is employed. Obviously, from these results it is evident that between the IBLZC and ITLZC, the latter showed the better performance.

3.5. Findings of this study and their implications

The chief findings of this study are as follows: (1) The coarse-graining procedure has a strong influence on the estimation of LZC. (2) The coarse-graining with three-symbols shows better diagnostic performance than that with two-symbols. (3) In the case of gait time series, the conventional BLZC and TLZC did not show significant results. (4) The improved versions, IBLZC and ITLZC, both showed higher significant results. (5) Between IBLZC and ITLZC, ITLZC exhibited better performance in discriminating young, elderly, and PD groups. (6) The deterministic nature is prominent in the young class, decreases with aging as seen in the elderly class, and further diminishes or may break down because of the progression of the disease as seen in the PD class. (7) Thus, aging and disease both, alter stride interval dynamics, which manifests as increased fluctuation magnitude and altered fluctuation dynamics. (8) It is found that the young class exhibits a less complex dynamical behavior for walking continuously at their normal pace on an obstacle-free level ground. This complex dynamical behavior increases with aging as observed in the healthy elderly class stride time series and increases significantly because of the progression of the disease as seen in the PD class stride time series.

This clearly shows that the alterations in gait dynamics in the elderly are due to subtle changes in the neuromuscular control making them prone to falls (for no apparent reason). On the other hand, the prominent alterations in gait dynamics of the PD subjects reflect neuropathology which increases their risks of falling. These findings could be of importance for clinical diagnostics, in algorithms for gait fall risk stratification, and for therapeutic and fall-preventive tools of next generation.

3.6. Applications of this approach

This tool can facilitate judgments to assess the complex dynamical behavior in the stride time series in discerning young, elderly, and PD classes in clinical settings. This study has application in clinical geriatrics and future gerontological research. We believe that this study has implications in sedentarism, perhaps somewhat like advances aging and disease.

3.7. Limitations

The present study focused on only level walking at their normal pace on an obstacle-free path. Therefore, further study is necessary to evaluate changes in gait pattern among different classes using different experimental designs, including gait adaptation strategy by gait condition. Another limitation of this study is the small sample size. Factors like high variance, age differences, and differing male-to-female ratios between groups will have an impact on the results when statistical analyses are carried out on small sample sizes. However, it has been shown that the effect of gender on usual gait patterns is considerably small (Gabell & Nayak, 1984). Though the effect of age on gait is complex, the effect of neurodegenerative disorder considerably predominates over the aging effects. This implies that the discrimination using this nonlinear method stands irrespective of the above limitations.

4. Conclusion

Though previous studies, in general, claim that a two-symbol sequence conversion to study dynamic complexity of a signal is suffice, our results show that a three-symbol sequence might give a more detailed insight into the differences among young, elderly, and PD classes stride time series. Thus, while using LZC in analysis there is a need to investigate the method of coarse-graining which is appropriate for the time series used. We found that the stride time series in young exhibits a less complex dynamical behavior for walking continuously at their normal pace on an obstacle-free level ground. This behavior may change with physiologic aging and as found in elderly subjects dynamical complexity increases resulting in local instability making them more disposed to falls. In the pathological case, dynamical complexity is significantly increased resulting in increased risk of falls.

Author details
Chandrakar Kamath[1]
E-mail: kamath.chandrakar@gmail.com
[1] ShanthaNilaya, 107, First stage, Ananthnagar, Manipal 576104, India.

Citation information
Cite this article as: Analysis of altered complexity of gait dynamics with aging and Parkinson's disease using ternary Lempel–Ziv complexity, Chandrakar Kamath, *Cogent Engineering* (2016), 3: 1177924.

Cover image
Source: Author.

References
Abásolo, D., Hornero, R., Gómez, C., García, M., & López, M. (2006). Analysis of EEG background activity in Alzheimer's disease patients with Lempel–Ziv complexity and central tendency measure. *Medical Engineering & Physics, 28*, 315–322.

Abásolo, D., Alcaraz, R, Rieta, J. J., & Hornero, R. (2011). Lempel–Ziv complexity analysis for the evaluation of atrial fibrillation organization. In *8th IASTED International Conference on Biomedical Engineering*. Innsbruck.

Aboy, M., Hornero, R., Abasolo, D. D., & Alvarez, D. (2006). Interpretation of the Lempel–Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering, 53*, 2282–2288. http://dx.doi.org/10.1109/TBME.2006.883696

Adiego, J., Navarro, G., & De la Fuente, P. (2004). Lempel–Ziv compression of structured text. In *IEEE Proceedings of the Data Compression Conference* (pp. 112–121). Snowbird, UT.

Bridenbaugh, S. A., & Kressig, R. W. (2011). Laboratory review: The role of gait analysis in seniors' mobility and fall prevention. *Gerontology, 57*, 256–264. http://dx.doi.org/10.1159/000322194

Eckmann, J. P., & Ruelle, D. (1992). Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Physica D: Nonlinear Phenomena, 56*, 185–187. http://dx.doi.org/10.1016/0167-2789(92)90023-G

Fatmehsari, Y. R., & Bahrami, F. (2011). Lempel–Ziv complexity criteria for nonlinear analysis of gait in patients with Parkinsons disease. *18th Iranian Conference of Biomedical Engineering ICBME-2011*. (pp. 137–141). Tehran: IEEE.

Ferenets, R., Lipping, T., Anier, A., Jantti, V., Melto, S., & Hovilehto, S. (2006). Comparison of entropy and complexity measures for the assessment of depth of sedation. *IEEE Transactions on Biomedical Engineering, 53*, 1067–1077.
http://dx.doi.org/10.1109/TBME.2006.873543

Ferrario, M., Signorini, M. G., & Magenes, G. (2009). Complexity analysis of fetal heart rate variability: Early identification of severe intrauterine growth-restricted fetuses. *Medical & Biological Engineering & Computing, 47*, 911–919.

Gabell, A., & Nayak, U. S. L. (1984). The effect of age on variability in gait. *Journal of Gerontology, 39*, 662–666.
http://dx.doi.org/10.1093/geronj/39.6.662

Goldberger, A. L., Rigney, D. R., Mietus, J., Antman, E. M., & Greenwald, S. (1998). Nonlinear dynamics in sudden cardiac death syndrome: heart rate oscillations and bifurcations. *Experientia, 41*, 983–987.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., & Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation, 101*, e215–e220.
http://dx.doi.org/10.1161/01.CIR.101.23.e215

Goldberger, A. L., Peng, C. K., & Lipsitz, L. A. (2002). What is physiologic complexity and how does it change with aging and disease? *Neurobiology of Aging, 23*, 23–26.
http://dx.doi.org/10.1016/S0197-4580(01)00266-4

Grassberger, P. (1991). Information and complexity measures in dynamical systems: Information dynamics. In H. Atmanspacher, H. Scheingraber(Eds). New York, NY: Plenum. 15–34.
http://dx.doi.org/10.1007/978-1-4899-2305-9

Hahn, G. J., & Shapiro, S. S. (1994). *Statistical models in engineering*. Hoboken, NJ: Wiley.

Hausdorff, J. M., Purdon, P. L., Peng, C. K., Ladin, Z., Wei, J. Y., & Goldberger, A. L. (1996). Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *Journal of Applied Physiology, 80*, 1448–1457.

Hausdorff, J. M., Cudkowicz, M. E., Firtion, R., Wei, J. Y., & Goldberger, A. L. (1998). Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. *Movement Disorders, 13*, 428–437.
http://dx.doi.org/10.1002/mds.870130310

Hausdorff, J. M., Lertratanakul, A., Cudkowicz, M. E., Peterson, A. L., Kaliton, D., & Goldberger, A. L. (2000). Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of Applied Physiology, 88*, 2045–2053.

Hausdorff, J. M., Rios, D. A., & Edelberg, H. K. (2001). Gait variability and fall risk in community-living older adults: A 1-year prospective study. *Archives of Physical Medicine and Rehabilitation, 82*, 1050–1056.
http://dx.doi.org/10.1053/apmr.2001.24893

Hu, J., Gao, J. B., & Principe, J. C. (2006). Analysis of biomedical signals by the Lempel–Ziv complexity: The effect of finite data size. *IEEE Transactions Biomedical Engineering, 53*, 2606–2609.

Jeong, J., Kim, S. J., & Han, S. H. (1998). Non-linear dynamical analysis of the EEG in Alzheimer's disease with optimal embedding dimension. *Electroencephalography and Clinical Neurophysiology, 106*, 220–228.
http://dx.doi.org/10.1016/S0013-4694(97)00079-5

Kamath, C. (2013). A classification system to discriminate epileptic patients using multi-valued coarse-graining Lempel–Ziv complexity. *International Journal of Biomedical Engineering and Technology, 11*, 96–106.

http://dx.doi.org/10.1504/IJBET.2013.053717

Kamath, C. (2015). A risk stratification system to discriminate congestive heart failure patients using multivalued coarse-graining Lempel–Ziv complexity. *Journal of Engineering Science and Technology, 10*, 12–24.

Lempel, A., Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory, 22*, 75–81.
http://dx.doi.org/10.1109/TIT.1976.1055501

Li, Y., Tong, S., Liu, D., Yi, G., Xiuyuan, W., Jijun, W., ... Yisheng, Z. (2008). Abnormal EEG complexity in patients with schizophrenia and depression. *Clinical Neurophysiology, 119*, 1232–1241.

Louchard, G., & Szpankowski, W. (1997). On the average redundancy rate of the Lempel–Ziv code. *IEEE Transactions on Information Theory, 43*, 2–8.
http://dx.doi.org/10.1109/18.567640

Nagarajan, R. (2002). Quantifying physiological data with Lempel–Ziv complexity-certain issues. *IEEE Transactions on Biomedical Engineering, 49*, 1371–1373.
http://dx.doi.org/10.1109/TBME.2002.804582

Panel on Prevention of Falls in Older Persons AGS, British Geriatrics S. (2011). Summary of the updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *Journal of the American Geriatrics Society, 59*, 148–157.

Pincus, S. M. (1995). Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 5*, 110–117.
http://dx.doi.org/10.1063/1.166092

Pincus, S. M. (2006). Approximate entropy as a measure of irregularity for psychiatric serial metrics. *Bipolar Disorders, 8*, 430–440.
http://dx.doi.org/10.1111/bdi.2006.8.issue-5p1

Pincus, S. M., & Goldberger, A. L. (1994). Physiological time-series analysis: What does regularity quantify? *American Journal of Physiology-Heart and Circulatory Physiology, 266*, H1643–H1656.

Pool, R. (1989). Is it healthy to be chaotic? *Science, 243*, 604–607.
http://dx.doi.org/10.1126/science.2916117

Radhakrishnan, N., & Gangadhar, B. N. (1998). Estimating regularity in epileptic seizure time-series data. *IEEE Engineering in Medicine and Biology Magazine, 17*, 89–94.
http://dx.doi.org/10.1109/51.677174

Rapp, P. E., Cellucci, C. J., Watanabe, T. A. A., Albano, A. M., & Schmah, T. I. (2001). Surrogate data pathologies and the false-positive rejection of the null hypothesis. *International Journal of Bifurcation and Chaos, 11*, 983–997.
http://dx.doi.org/10.1142/S021812740100250X

Sarlabous, L., Torres, A., Fiz, J. A., Gea, J., Galdiz, B., Jané, R. (2009). Multistate Lempel–Ziv index interpretation as a measure of Amplitude and complexity changes. In *31th Annual Conference IEEE-EMBS-2009*. Minneapolis, MN.

Savari, S. (1997). Redundancy of the Lempel–Ziv incremental parsing rule. *IEEE Transactions on Information Theory, 43*, 9–21.
http://dx.doi.org/10.1109/18.567642

Sejdic, E., Lowry, K. A., Bellanca, J., Redfern, M. S., & Brach, J. S. (2014). A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22*, 603–612.
http://dx.doi.org/10.1109/TNSRE.2013.2265887

Sosnoff, J. J., Valantine, A. D., & Newell, K. M. (2006). Independence between the amount and structure of variability at low force levels. *Neuroscience Letters, 392*, 165–169.
http://dx.doi.org/10.1016/j.neulet.2005.09.010

Stergiou, N., Buzzi, U. H., Kurz, M. J., & Heidel, J. (2004).

Innovative analyses of human movement. Champaign: Human Kinetics.

Talebinejad, M., Chan, A. D. C., & Miri, A. (2011). A Lempel–Ziv complexity measure for muscle fatigue estimation. *Journal of Electromyography and Kinesiology, 21,* 236–241. http://dx.doi.org/10.1016/j.jelekin.2010.12.003

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: The method of surrogate data. *Physica D: Nonlinear Phenomena, 58,* 77–94. http://dx.doi.org/10.1016/0167-2789(92)90102-S

Tinetti, M. E., & Kumar, C. (2010). The patient who falls: Its always a trade-off. *JAMA, 303,* 258–266. http://dx.doi.org/10.1001/jama.2009.2024

Tononi, G., Edelman, G. M., & Sporns, O. (1998). Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Sciences, 2,* 474–484. http://dx.doi.org/10.1016/S1364-6613(98)01259-5

Torres, A., Fiz, J. A., Jané, R., Laciar, E., Galdiz, J. B., Gea, J., & Morera, J. (2008). Rényi entropy and Lempel–Ziv Complexity of mechanomyographic recordings of diaphragm muscle as indexes of respiratory effort. In *Proceedings of 30th Annual International Conference on IEEE-EMBS* (pp. 2112–2115). Vancouver.

Wu, X., & Xu, J. (1991). Complexity and brain functions. *Acta Biochimica et Biophysica Sinica, 7,* 103–106.

Xu, J., Liu, Z. R., Ren Liu, R., & Yang, Q. F. (1997). Information transmission in human cerebral cortex. *Physica D: Nonlinear Phenomena, 106,* 363–374. http://dx.doi.org/10.1016/S0167-2789(97)00042-0

Zhang, X. S., & Roy, R. J. (2001a). Derived fuzzy knowledge model for estimating the depth of anesthesia. *IEEE Transactions on Biomedical Engineering, 48,* 312–323. http://dx.doi.org/10.1109/10.914794

Zhang, J. W., Zheng, C. X., Jiang, D. Z. (1998). EEG complexity measurement of focal ischemic cerebral injury. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1998, 20,* 2027–2029.

Zhang, X. S., Zhu, Y. S., Thakor, N. V., & Wang, Z. Z. (1999). Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering, 46,* 548–555. http://dx.doi.org/10.1109/10.759055

Zhang, X., Roy, R. J., & Weber, E. (2001b). EEG complexity as a measure of depth of anesthesia for patients. *IEEE Transactions on Biomedical Engineering, 48,* 1424–1433. http://dx.doi.org/10.1109/10.966601

Zhao, H., Wang, G., Xu, C., & Yu, F. (2011). Voice activity detection method based on multi-valued coarse–graining Lempel–Ziv complexity. *Computer Science and Information Systems, 8,* 869–888.

Zozor, S., Ravier, P., & Buttelli, O. (2005). On Lempel–Ziv complexity for multidimensional data analysis. *Physica A: Statistical Mechanics and its Applications, 345,* 285–302. http://dx.doi.org/10.1016/S0378-4371(04)00994-X

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic ROC plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39,* 561–577.

Cogent Engineering (ISSN: 2331-1916) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com