ELSEVIER

Gait and Posture 13 (2001) 102–120

GAIT & POSTURE

www.elsevier.com/locate/gaitpost

Review

A review of analytical techniques for gait data. Part 2: neural network and wavelet methods

Tom Chau [a,b,*]

[a] *Bloorview MacMillan Centre, 350 Rumsey Road, Toronto, Ontario Canada M4G 1R8*

[b] *Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ont., Canada*

Received 7 May 2000; received in revised form 26 June 2000; accepted 16 October 2000

Abstract

Multivariate gait data have traditionally been challenging to analyze. Part 1 of this review explored applications of fuzzy, multivariate statistical and fractal methods to gait data analysis. Part 2 extends this critical review to the applications of artificial neural networks and wavelets to gait data analysis. The review concludes with a practical guide to the selection of alternative gait data analysis methods. Neural networks are found to be the most prevalent non-traditional methodology for gait data analysis in the last 10 years. Interpretation of multiple gait signal interactions and quantitative comparisons of gait waveforms are identified as important data analysis topics in need of further research. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Automatic data processing; Artificial neural networks; Wavelets; Gait analysis

1. Introduction

This paper is the second of a two-part review of emergent data analysis techniques applied to gait data. The analysis of quantitative gait data has traditionally been a difficult problem. Due to a handful of persistent challenges, such as high-dimensionality, temporal dependence and curve correlations [1], numerous alternative data analysis techniques have been recently investigated. In Part 1 of this review [1], fuzzy, multivariate statistical and fractal methods were introduced and critiqued in terms of gait data analysis. While statistical methods are well understood and most widely applied among gait researchers, fuzzy and fractal approaches are not to be dismissed. The present paper expands the review by surveying applications of artificial neural networks and wavelets (Fig. 1) as a means to analyze gait data. Both artificial neural networks and wavelets have played dominant roles in signal and image processing applications in the engineering world (see for example, [2–5]). Their introduction to gait data evaluation, although a contemporary event, has generated enormous interest in the biomechanics community, as evidenced by the volume of publications reviewed herein.

As in Part 1, the review of each method adheres to the following format. A conceptual overview of each methodology will be presented. Partial mathematical formulations will be included only in the simplest cases while references for further exploration will be provided. The discussion then proceeds to summarize recent applications of each method to gait data analysis, closing with a list of perceived advantages and practical issues.

At the conclusion of this review, I will revisit the challenges identified at the beginning of Part 1 [1] and through a series of tables, compare the extent to which each challenge has been overcome by each of the discussed methods. Brief mention will be made of other techniques which may potentially enhance the analysis

Abbreviations: FA, Factor analysis; FD, Fractal dynamics; FC, Fuzzy clustering; MCA, Multiple correspondence analysis; NN, Neural networks; PCA, Principal component analysis; WT, Wavelet transforms.

* Tel.: +1-416-4256220, ext. 3515; fax: +1-416-4251634.

E-mail address: ttkchau@ieee.org (T. Chau).

0966-6362/01/$ - see front matter
PII: S0966-6362(00)00095-3

of gait data. All mentioned and reviewed techniques are organized according to the research or clinical questions that they suitably address. This summary can serve as a quick guide to gait interpretation teams looking to apply alternative data analysis methods.

2. Neural network (NN) applications in gait analysis

2.1. Conceptual overview

The body of literature on artificial neural networks (ANN) is intractably vast. Here, only some very general comments will be made. One specific type of neural network, the multilayer feed-forward neural network [6] does require specific mention. It has been the standard workhorse in a wide range of applications including gait analysis. In the literature, this network has a number of equivalent names, including the multilayer perceptron or back-propagation network. For a general introduction to multilayer feed-forward networks see Hinton [7,8]. Ohno-Machado and Rowland [9] present a brief introduction and review of multilayer neural networks in physical medicine and rehabilitation. The list of different artificial neural networks is ever increasing. Other prominent types include the radial basis function [10], the probabilistic neural network [11] and the self-organizing feature map [12].

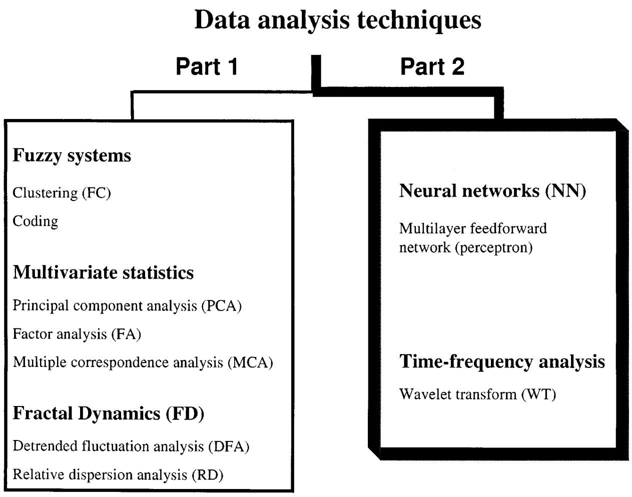

Fig. 1. Data analysis methods in this two-part review. The methods on the right are discussed in this paper, while those on the left were surveyed in Part 1 [1].

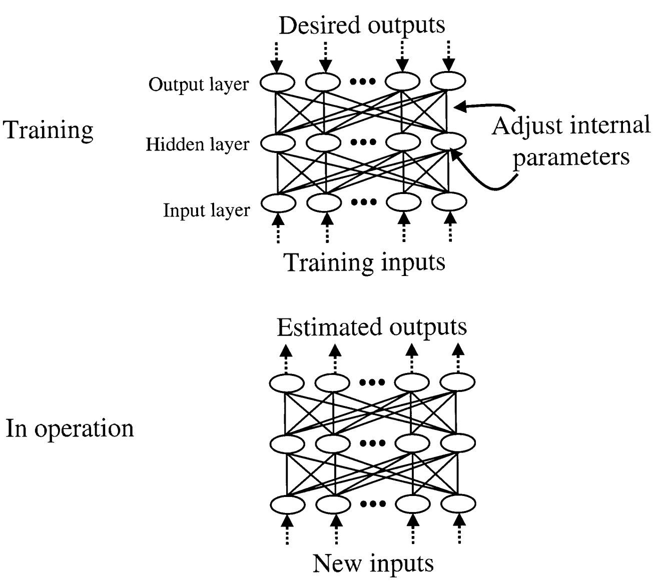

Fig. 2. Typical neural network training and operation.

Artificial neural networks typically have inputs and outputs, with some processing or so-called *hidden* layers in between. In traditional statistical language, the inputs are the independent variables and the outputs are the dependent variables. The following comments focus on multilayer feedforward networks but also apply to many other types of networks.

In general, the ANN can be likened to a flexible mathematical function which has many configurable internal parameters. To accurately represent complicated relationships among gait variables, these internal parameters need to be adjusted through an optimization or so-called learning algorithm. In 'supervised' learning, examples of inputs and corresponding desired outputs are simultaneously presented to the network, which iteratively self-adjusts to accurately represent as many examples as possible. Learning is complete when some criterion such as prediction error falls below a preset threshold.

Once the neural network is trained (i.e. its internal parameters are fine-tuned), it can accept new inputs which it has not previously seen and attempt to predict an accurate output. To produce an output, the trained network simply performs function evaluation. Fig. 2 summarizes this conceptual overview of a neural network. The only assumption in deploying a multilayer feedforward neural network (with one hidden layer) is that there exists a continuous functional relationship between the input and output data. This assumption is so general and unrestrictive that it is seldom mentioned.

2.2. Recent literature

Unlike any previous technology, neural methods endow gait analysis with a highly flexible, inductive, nonlinear modeling ability. This non-linear property has facilitated the study of complicated gait variable relationships which have traditionally been difficult to model analytically. Recent efforts generally fall into three categories of application; (1) classification of human gait; (2) biomechanical modeling; and (3) prediction of gait variables and parameters. This is not an absolute grouping with many works crossing these

Table 1
Summary of neural network classification of gait data

Author	Inputs	Network type	Outputs:categories	Results
Holzreiter and Kolile [13]	128 FFT coefficients	Feed forward (one hidden layer)	(1) Able-bodied gait; (2) Pathological gait	up to 95%
Barton and Lees [14]	1316 maximum pressure values	Feed forward (two hidden layers)	(1) Healthy feet; (2) pes cavus, (3) hallux valgus	77–100%
Barton and Lees [15]	30 FFT coefficients	Feed forward (two hidden layers)	(1) normal walking; (2) 20 mm thick sole; (3) 3.5 kg mass	83.3%
Lafuente et al. [16]	cadence, velocity, five kinetic parameters	Feed forward (one hidden layer)	(1) healthy; (2) ankle arthrosis; (3) knee arthrosis and (4) hip arthrosis	80%

boundaries. Where ambiguity exists, the work has been classified according to the primary efforts.

2.2.1. Classification

There have been several attempts to automatically classify a person's gait or diagnose a walking condition with neural networks. One of the first attempts was undertaken by Holzreiter and Kohle [13] who attempted to categorize gait pathology based on ground reaction forces. They measured two successive ground reaction forces during normal walking from 131 subjects with various lower limb conditions, including calcaneus fracture and limb deficiencies (i.e. prosthesis users). Ninety-four healthy persons complemented the study sample. Holzreiter and Kohle computed fast-Fourier transforms (FFT) of vertical components of the two ground reaction forces. The FFT coefficients served as inputs to a standard network with one hidden layer which, with adequate training, could achieve up to 95% accuracy in discriminating healthy from pathological gait. See Table 1. This early work demonstrated simple two-category gait classification with a fairly large number of input parameters.

With a similar goal of classifying gait pathologies, Barton and Lees [14] extended the classification problem to three output categories. Unlike the three-layer network used by Holzreiter and Kohle, Barton and Lees commissioned a more complex neural network with two hidden layers to categorize maximum pressure prints into one of three categories: healthy feet, pes cavus and hallux valgus. Below-foot pressure patterns were recorded from 18 subjects during normal walking. The patterns were rotated to a common orientation, scaled to a common size and normalized to the interval [0,1]. The network inputs consisted of a massive 1316 measured pressure values, dwarfing the number of inputs used by Holzreiter and Kohle. The reported accuracies ranged from 77 to 100% based on the relatively small volume of test and training samples. Additional work is required to ascertain the practical advantage for such a classification system as the studied foot conditions are routinely identified by observational gait analysis. Furthermore, the use of the second hidden layer was not well motivated. A single hidden layer is known to be theoretically sufficient for learning most functional relationships [10]. This point will be further elaborated upon in Section 2.4.

Hip–knee joint angle diagrams are characteristic of a subject's gait pattern and therefore could serve as the basis for automated identification of gait patterns. This is the justification of Barton and Lees [15] for exploiting hip–knee joint angles from eight healthy subjects for classification of walking condition. Hip and knee angles were calculated via a set of four reflective markers. Subjects walked on a motorized treadmill at constant speed under three conditions: normal walking, simulated leg length difference and simulated leg weight difference. The angles were preprocessed in three steps, namely, normalization in time, fast-Fourier transform and standardization to the interval [0, 1]. As the angular curve was predominantly of low frequency content, only the low frequency coefficients of the hip and knee angles served as inputs. As in their previous work, Barton and Lees invoked a neural network with two hidden layers without justification for the second hidden layer. The average accuracy of discriminating among the three walking conditions was 83.3%.

Lafuente et al. [16] returned to a standard feedforward network (one hidden layer), asserting its adequacy for four-category gait classification. The sample population consisted of 148 subjects with ankle, knee or hip arthrosis and 88 control subjects without limb pathology. Measurements were made at three different walking speeds either shod or barefoot. The inputs therefore consisted of cadence, velocity and parameterizations of five kinetic magnitudes. Based on these inputs, a trained three-layered neural network distinguished between the four gait categories with an accuracy of 80%, a statistically significant improvement over a traditional bayesian quadratic classifier [17]. These efforts are summarized in Table 1. Collectively, they have established the potential for multicategory classification of complicated pathological gait using standard feedforward neural networks.

2.2.2. Biomechanical modeling

The highly non-linear mapping of neural networks has encouraged investigators to model the elusive relationships between EMG, kinematic and kinetic parameters. The following studies have attempted to use standard feedforward neural networks to capture various aspects of this traditionally abstruse interaction.

Heller et al. [18] assembled a single-hidden layer network in the attempt to reconstruct EMG of the semitendinosus and vastus medialis muscles from kinematic data. Specifically, the kinematic inputs consisted of hip and knee angles, angular velocities, angular accelerations and integrated foot contacts, all measured during normal and fast-paced walking. Data were collected from one healthy subject. Timing and amplitude of the reconstructed signals were accurate and comparable to that predicted by traditional, explicit models of the musculoskeletal system. From this study, the authors highlighted advantages of inductive biomechanical models such as neural networks over deductive, inverse musculoskeletal models.

1. There is no time delay in the generation of EMG.
2. A single neural network can model activations of multiple muscles.

The following disadvantages of inductive biomechanical models were identified.

1. The inductive models are only valid over the range of motion represented in the training examples. Inverse musculoskeletal models are valid over the entire range of movement.
2. Inductive models do not offer biomechanical insight into the locomotion system since the mathematical equations are not based upon the structure of the biomechanical system.

The first disadvantage is not entirely accurate as one of the greatest strengths of neural networks is their ability to generalize to data upon which it has not previously trained [8], granted adequate training examples are available. The second disadvantage is partially true in the sense that the structure of the mathematical equations of an artificial neural network are generic and not specific to human biomechanics. Although often challenging to interpret [19], it is the freely adjustable parameters (weights and biases) of the trained network which do in fact capture the structure of the biomechanical system.

With particular focus on the relationship between muscle activity and lower-limb dynamics, Sepulveda et al. [20] invoked two single-hidden layer neural networks to model correlations between EMG and joint characteristics. Training data were taken from Winter [21] and included EMG from 16 muscles along with moments and angles for the hip, knee and ankle. The network input consisted of 16 normalized EMG values. Training data were sampled from the gait cycle at 20 evenly spaced time intervals. One network was created to model the relationship between EMG and joint moments while another network modeled the interdependence of EMG and joint angles. Sepulveda et al. tested their models with perturbed EMG signals ($\pm 20\%$ random noise and amplitude offset) and observed robust behaviour, with less than 7% deviation in the output angles and moments. The authors also simulated the removal of a particular muscle and found that outputs of the EMG-joint moment model agreed with expectations based on physiological principles. While their model demonstrates modeling of a complex relationship in biomechanics, the authors acknowledge limitations due to a lack of exposure to intersubject variability and pathological gait data and the omission of temporal relationships.

Prentice et al. [22] developed a neural network to model the shaping function of a central pattern generator for human locomotion. Using two sinusoidal inputs at a frequency equal to the stride rate, their network was able to produce a representation of EMG amplitude and timing characteristics of eight lower extremity muscles at various walking speeds over a period of 12 strides. The prediction errors, measured as percentages of the output operating range, were generally below 20%. With very simple networks and a single temporal parameter of stride rate, Prentice et al. demonstrated the potential of simultaneously modeling a cohort of muscle activations for such applications as functional

Table 2
Summary of neural network modeling of gait relationships

Author	Relationship modeled		Network type
	Inputs	Outputs	
Heller et al. [18]	Hip and knee angle, angular velocity and acceleration and integrated foot contact	EMG envelope of semitendinosus and vastus medialis	Feed forward (one hidden layer)
Sepulveda et al. [20]	EMG from 16 muscles (right leg)	Hip, knee and ankle joint angles and moments	Two feed forward networks (one hidden layer each)
Prentice et al. [22]	Two sinusoidal signals at stride rate	EMG envelopes of eight lower extremity muscles	Feed forward (one hidden layer)

Table 3
Summary of neural network prediction of gait parameters[a]

Author	Inputs	Predicted output	Network type	Results
Gioftsos and Grieve [23]	Durations of right, left and double	Walking condition, walking speed	Recurrent network	up to 73% accuracy
Aminian et al. [25,26]	Forward, vertical, lateral and heel accelerations	Incline and speed	Two feed forward networks (one hidden layer)	speed: $\leq 6\%$ (mean error) $r^2_{\text{incline}} = 0.98$
Herren et al. [27]	Ten select parameters of forward, vertical, lateral and heel	Incline and speed	Two feed forward networks (one hidden layer)	$\rho^2_{\text{speed}} = 0.965$ $\rho^2_{\text{incline}} = 0.936$
Savelberg and Herzog [28]	EMG (gastrocnemius)	Tendon force	Feed forward (two hidden layers)	$0.71 \leq r^2 \leq 0.98$

[a] r^2 is the correlation coefficient and ρ^2 is the cross-correlation coefficient.

Table 4
Summary of neural network prediction of gait parameters (continued from Table 3)[a]

Author	Inputs	Predicted output	Network type	Results
Liu et al. [29]	EMG (soleus), ankle joint angles and angular velocities	Tendon force	Feed forward (two hidden layers)	$\rho^2 > 0.9$
Savelberg and de Lange [30]	Insole pressure values (48)	Horizontal foreaft force	Feed forward (two hidden layers)	Typically $\rho^2 > 0.9$
Cottrell et al. [31]	12 joint angles at time $t-\delta$	12 joint angles at time t	Recurrent network (Tau net)	Good qualitative results
Tucker and White [32]	5 EMG (lower limb)	Velocity and cadence condition	Feed forward, self-org. map, fuzzy inference, neuro-fuzzy	94%, 56%, 91.4%, 76%

[a] r^2 is the correlation coefficient and ρ^2 is the cross-correlation coefficient.

electrical stimulation. Table 2 tabulates these modeling attempts using neural networks.

2.2.3. Prediction

The flexible modeling ability of neural networks also facilitates the prediction of gait parameters which are difficult to measure outside of controlled laboratory conditions. As summarized in Table 3 and Table 4, a colourful collection of gait parameters have been predicted using neural networks.

Gioftsos and Grieve [23] investigated the application of recurrent neural networks for the prediction of walking speed and walking condition from temporal measurements. Recurrent networks are feedforward networks in which the outputs feed back to the inputs (see Elman [24]). The networks received three temporal measurements as input in milliseconds, namely, the durations of right single support, left single support and double support. Their study involved 20 subjects with healthy gait, walking at seven different speeds and under three different conditions. In addition to normal walking, each subject's gait was artificially altered by first wearing an ankle weight and subsequently by wearing a knee brace. Gioftsos and Grieve [23] found that they lacked adequate training samples for the recurrent network. In fact, the improvement in prediction accuracies over a simple linear discriminant was not statistically significant.

To assess the energy cost of overground walking, the incline of the terrain and the speed of walking must be known. These parameters are not readily measured outside of a controlled laboratory environment while body accelerations are easily detectable through accelerometry. This is the motivation of Aminian et al. [25,26] and Herren et al. [27] for building neural networks to predict incline and speed from body accelerations.

Aminian et al. [25,26] measured the forward, vertical and lateral accelerations of the trunk and right heel acceleration from a handful of subjects walking at various speeds and inclines. Ten parameterizations of these accelerations were found to be closely correlated with speed and incline, and were thus used as neural network inputs. Aminian et al. constructed two three-layered feedforward neural networks to separately forecast speed and incline of walking from these 10 parameters. Training data were obtained from treadmill walking while testing data consisted of self-paced walking on an outdoor, multiple incline circuit. In general, predicted speeds and inclines agreed very closely with the actual values, both with training and testing data. Aminian et al. showed that neural networks can easily predict difficult-to-measure variables from those that are more easily accessible.

Herren et al. [27] expanded the study to include 20 subjects and also formalized the feature selection process. With an expanded list of 28 parameters derived from the four accelerations, they used correlation coefficients and stepwise regression analysis to select a subset of parameters that are most closely associated with speed and incline. In addition to neural network prediction, Herren et al. also developed a linear model through multiple linear regression to predict speed and incline. While the neural prediction was fairly accurate for both speed and incline, the strictly linear model only predicted speed at a comparable level of accuracy. This suggests that the relationship between bodily accelerations and speed is linear while that with incline is non-linear. The work of Herren et al. specifically illustrated that neural networks cannot be applied blindly. One needs to prudently choose inputs which will be good predictors of the outputs. As well, through the comparison with multiple linear regression, a key feature of neural networks is underlined, namely the ability to model non-linear relationships among variables.

The aforementioned approaches focussed on predicting static parameters. In contrast, Savelberg and Herzog [28] were interested in estimating a dynamic variable over a period of time. Their specific goal was to predict dynamic tendon forces using EMG signals of the gastrocnemius muscles of a cat. Tendon forces and EMG signals were recorded from three cats while walking at four different speeds. Savelberg and Herzog employed a four-layer (two hidden unit layers) feedforward neural network. The input to the network consisted of rectified and averaged EMG values from the current and previous 29 time steps. The corresponding desired output was the tendon force at the current time. After training the neural network, Savelberg and Herzog investigated intrasession, intrasubject and intersubject generalization ability. While the neural network could accurately predict tendon force from EMG in all three scenarios, they noted that to achieve good prediction, it was necessary to first determine the variables which impact the EMG–force relationship. For example, for intersubject tests, the body mass of the cat was required to account for differences in muscle force magnitudes between animals. This study underscored the importance of prudently choosing neural network inputs. It also exemplified the feasibility of predicting time-varying gait signals and the modeling of highly non-linear relationships, such as that between EMG, force and speed of walking.

The artificial neural network prediction of time-varying tendon force was further investigated by Liu et al. [29]. In addition to EMG, they incorporated five ankle joint angles and angular velocities as neural network inputs for the prediction of soleus tendon force. As in Savelberg and Herzog, a two-hidden layer neural network was constructed and experimental data consisted of force, EMG and kinematics from three cats walking at four speeds. They tested force prediction under the same three levels of generalization and found that the addition of kinematics only selectively improved prediction. Based on the good intersubject predictions, Liu et al. concluded that the EMG–force relationship was similar among cats walking at the same speed. While their work further portrays the ability to predict forces from EMG without an explicit muscle model, the authors acknowledge a principal drawback of neural networks, namely, that insights into the modeled relationship, i.e. between EMG and tendon force, are not automatically provided by the neural network.

It is difficult to evaluate activities of daily living with laboratory-bound force plates. Numerous trials are usually required to obtain a few representative force measurements. These disadvantages, among others, motivated Savelberg and de Lange [30] to invoke a neural network to predict horizontal fore-aft forces from insole pressure patterns. They believed that if horizontal forces could be predicted accurately from insole pressures (which are related to the vertical forces through the sensor area), one would have a more portable alternative than force plates for obtaining vertical and horizontal force components. Savelberg and de Lange measured spatially averaged insole pressure patterns from eight areas under the foot along with the ground reaction force. Data were collected from four subjects walking at various speeds. A four-layer feedforward neural network (two hidden unit layers) was employed. The 48 inputs consisted of eight pressure values at the current and previous five time steps. The horizontal ground reaction force constituted the single network output. With their very small sample size, Savelberg and de Lange noted good within-subject prediction but poor intersubject prediction. They measured the quality of the prediction by the coefficient of cross-correlation between actual and predicted force curves. Their work substantiates the existence of a relationship between insole pressure patterns and horizontal force and demonstrates the potential of neural network prediction in this context. However, to assess the practical accuracy of this mapping, one would require, as the authors admit, a much larger experimental sample. Training a neural network with an inadequate number of samples leads to poor generalization ability. In these last two works, no justification is given for the commissioning of a second hidden layer.

Gait signals do not fluctuate at a fixed rate and are therefore difficult to predict analytically. This phenomenon was briefly examined by Cottrell et al. [31] using an adaptive recurrent network. They trained the network to predict 12 joint angles representing level walking at a self-selected, constant pace. The outputs of the network were the angles at time t while the inputs were the angles at some earlier time $t - \delta$ where δ was an adjustable delay. Once the network was trained on signals at the self-selected pace, it was able to predict joint angles at different walking speeds. Cottrell et al. achieved real-time adaptive prediction by automatically modulating the network's time constant and delay parameter δ. Although significantly more complex than standard feedforward networks, Cottrell et al. demonstrated that subtle rate variations in gait signals can be automatically incorporated into the predicted output of a recurrent network. This ability could be exploited in the detection of subtle pathological anomalies in gait.

Recently, the flavours of neural networks in gait analysis have begun to diversify. In addition to the standard multilayer feedforward neural network, Tucker and White [32] have applied three other neurocomputational approaches for kinematic prediction based on EMG. These newcomers to gait analysis included a self-organizing map [12], a powerful topographical clustering algorithm, a fuzzy inference system [33,34], a system which reasons approximately with rules and fuzzy sets, and a neuro-fuzzy hybrid system [35], one which combines neural networks with fuzzy processing. They collected EMG data from five lower extremity muscles from 16 subjects walking at nine combinations of velocities and cadences. The normalized EMG linear envelopes were fed into the network in their entirety. This input is similar to the 30 point temporal signal used by Savelberg and Herzog. The output of the network was a single number coded to represent the nine possible combinations of velocity and cadence. Tucker and White found that the best prediction occurred with the feedforward neural network with the fuzzy inference system close behind. Their work has opened the door to future investigations of different neural network algorithms for gait data analysis and further highlights the ability to study the entire gait waveform. Recent work in prediction is recapped in Table 3 and Table 4.

2.3. Benefits to gait analysis

With the potential to address most of the traditional data analysis challenges, artificial neural networks have much to offer to gait analysis.

The three-layered feedforward network is a universal function approximator [36,37]. This means that given enough processing units, commonly known as hidden units, a three-layer network can approximate any continuous function, regardless of its complexity. In the context of gait analysis, this property allows one to model any relationship among gait variables, provided adequate data are available and requisite network complexity is computationally feasible. As evidenced in the neural modeling of muscle activity and kinematic interactions [20], this property facilitates the study of traditionally analytically unmanageable relationships.

Neural networks can handle vast amounts of gait data, demonstrated most notably by the large studies conducted by Holzreiter and Kohle [13] and Lafuente et al. [16]. High dimensional data can also be manipulated, as seen in Barton and Lees [14], where an astounding 1316 pressure values were input into a neural network. Robustness to errors in the data due to miscalculations or instrumentation fault is yet another beneficial property. Two final benefits to gait analysis include the inherent non-linear mapping ability and the demonstrated aptitude at capturing temporal dependence [28].

Neural networks capture patterns in the data within their internal parameters, which are known as weights and biases [8]. Past work has not attempted to interpret the weights and biases of a trained network, an effort which could potentially yield new insight into underlying gait patterns. Alternatives for interpreting weights and biases include neural rule extraction [38,39], Hinton diagrams [40,41] and bond diagrams [42].

2.4. Application issues

Despite the numerous potential and proven benefits of neural networks to gait data analysis, one must be cognizant of the key issues regarding their application.

In general, a neural network cannot directly process raw gait data. Proper pre-processing of input and post-processing of output variables are essential for good generalization performance of the neural network [10]. Ranges of variables are usually harmonized and normalized to avoid saturation [43] of the processing units and to generally facilitate optimization. In the reviewed gait analyses, pre-processing encompassed a host of techniques such as scaling, normalization, fast Fourier transforms, rectification and averaging. The choices of appropriate pre- and post-processing of gait data, while significantly affecting performance, are not always obvious, requiring a combination of experience and trial-and-error.

As part of pre-processing, neural network application is oftentimes preceded by a judicious selection of input variables. Typically there are insufficient samples to warrant the use of all available variables, i.e. there will not be enough training data. Alternately, the use of all available variables results in a prohibitively large network that will be difficult to train with the available computing resources. The work of Aminian et al. [25], and Savelberg and Herzog [28] underline the importance of selecting the appropriate input variables.

Selection involves discarding irrelevant variables and retaining only those that are potentially good predictors of the desired output variables. Although formal feature selection methods from statistical pattern recognition exist (see Siedlecki and Sklansky [44]), they have not been widely applied in gait analysis. Possible reasons are the complexity and heuristic nature of the algorithms. Most often it appears that the selection of independent gait variables relies heavily on clinical experience and is generally not a well documented subject. The ultimate performance of the neural network is highly sensitive to the choice of input gait variables.

The network architecture must be chosen prudently as network performance is highly architecture-sensitive [45]. Although no concrete rules exist, some general prescriptions ([10], pp. 372–380; [46], pp. 176–180) and ([43], pp. 145–147) provide guidance. The network architecture is intimately related to the ability to generalize to new data. The use of two hidden layer architectures, such as that employed by Barton and Lees[15] and Savelberg and Herzog [28] have not been adequately justified. As discussed earlier, networks with one hidden layer are universal approximators and the added computation associated with a second hidden layer is only warranted for approximating discontinuous functions [37].

Since multilayer feedforward networks are trained by an optimization algorithm, one must contend with the prevalent problems of non-convergence and local minima traps [43]. With multilayer perceptrons, the most popular training algorithm is back-propagation [40] and its variants. Other alternatives such as the Levenberg–Marquardt algorithm (see Bishop [10]) provide faster training and better probability of convergence under certain conditions. Alternatively, one may employ other types of neural networks such as the probabilistic neural network [11], which do not require optimization during training.

In the reviewed works, neural network classification accuracies, although significantly better than statistical alternatives in some instances [16], are not yet adequate for use as an automatic clinical diagnostic tool. Further improvements and innovations are necessary to improve reliability of classifying pathological gait.

Almost all the reported studies have exclusively exploited standard feedforward neural networks. Other proven alternatives such as the radial basis function [10] which also boasts the universal approximation property [47], have not yet been investigated. In other disciplines, these networks have exhibited certain performance advantages over the standard backpropagation network (see for example Shaffer et al. [48], Blue et al. [49]).

3. Gait analysis with wavelet transforms (WT)

3.1. Conceptual overview

Traditional spectral analysis methods such as the Fourier transform (FT) tell us which frequency components are contained in a signal. However, they do not tell us at what time those frequency components are present in the signal. This information is important in analyzing non-stationary signals, where the frequency content changes over time. Examples of non-stationary signals include the transient EMG [50], the EMG associated with 50–80% maximum voluntary contraction [51], the EMG associated with local muscle fatigue during a sustained contraction [52] and velocity and acceleration signals with sharp high-frequency transients [53]. The wavelet transform overcomes this deficiency in the FT by providing both frequency and time information simultaneously. Like the FT, the wavelet transform comes in both continuous and discrete flavours. Gait data analysis has only focussed on the discrete transform, as will the ensuing discussion.

A discrete wavelet transform (DWT) operates on a discrete signal, i.e. one that has values at discrete instants in time. The length of a discrete signal is the number of time instants over which the signal is measured. The two essential elements of a wavelet transform are the wavelet and scaling signals.

The wavelet basis signal is a finite energy signal with compact support. This means that the signal only exists over a finite period of time, unlike Fourier basis functions which have infinite extent. The set of basis functions are obtained by translating and contracting or dilating a prototype wavelet. In the simplest case, these basis functions are orthonormal.

The following is the filter bank interpretation of the discrete wavelet transform. Suppose we have a signal $x(n)$ of length N and maximum frequency f_{max}. The transform decomposes the signal $x(n)$ into two subsignals, namely the fluctuation and trend signals, each of length $N/2$. The fluctuation signal d_1 contains the upper half of frequency components in the original signal. It is obtained by passing the original signal through a high pass or bandpass filter with passband, $[f_{max}/2, f_{max}]$. This is equivalent to the convolution of the wavelet basis signal with the original signal.

Similarly, the low frequency trend signal a_1 is obtained by filtering the original signal through a low pass filter with passband $[0, f_{max}/2]$. This is equivalent to the convolution of the scaling signal with the original signal. The low and bandpass filters used in a wavelet transform are quadrature mirror filters.

The two subsignals constitute the first level of decomposition as shown in Fig. 3. The subsignals are downsampled by a factor of 2, meaning that every other sample is discarded. This is justified by the Nyquist sampling rule. Since the bandwidth of each subsignal is half that of the original signal, the subsignals actually contain more samples than necessary to capture the constituent frequencies in the subsignals.

The decomposition is repeated only on the first trend signal to produce a second trend and fluctuation signal. Again the signal bandwidth and signal length are

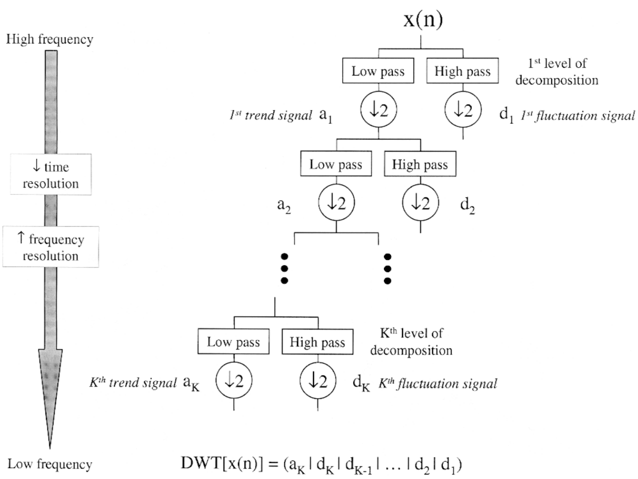

Fig. 3. Action of the discrete wavelet transform.

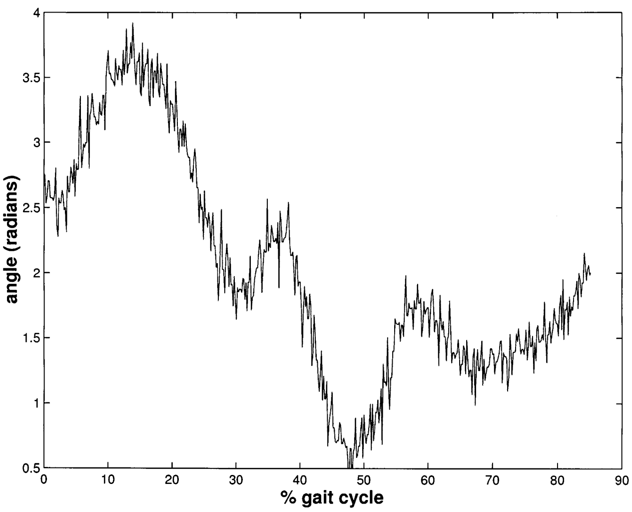

Fig. 4. Example of a noisy kinematic signal.

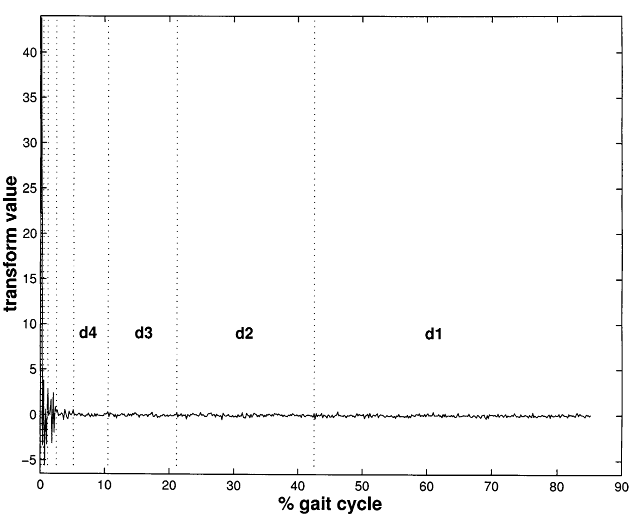

Fig. 5. Discrete wavelet transform of a noisy kinematic signal.

halved. This process is repeated on subsequent trend signals until a subsignal of length 1 is obtained. For a signal of length N, there will be $K = \log N / \log 2$ levels of decomposition.

Note that by bisecting the bandwidth of each subsignal, the frequency resolution is doubled, i.e. we are focusing on a finer band of frequencies. Likewise, subsampling by a factor of 2 reduces the number of time samples and hence decreases time resolution. In other words, we have fewer time samples to represent the signal over its duration. This tradeoff between time and frequency resolution is the hallmark of the wavelet transform. Low frequency components are more difficult to resolve in the frequency domain, and thus finer frequency resolution is desirable. On the other hand, high frequency components are more difficult to resolve in the time domain, demanding better time resolution. The wavelet transform meets these requirements, providing enhanced frequency resolution at low frequencies and better time resolution at high frequencies.

Finally, the discrete wavelet transform of the signal $x(n)$ is obtained by concatenating the last trend signal with all the fluctuation signals. Specifically,

$$\mathrm{DWT}\,[x(n)] = (a_K|d_K|d_{K-1}|\cdots|d_2|d_1). \tag{1}$$

The original signal can be reconstructed from the transform values through multiresolution analysis (MRA). Essentially, averaged and detail signals are constructed and summed together to reproduce the original signal. Averaged signals are expansions in terms of the scaling signals with values of the trend signals as coefficients. Similarly, detail signals are expansions in terms of the wavelet signals with values of the fluctuation signals serving as coefficients.

Fig. 4 and Fig. 5 portray a noisy kinematic signal and its discrete wavelet transform. Since the original signal contained 512 values, there are nine levels of decomposition. The first few fluctuation signals (d_1, d_2,...) are labelled and separated by dotted lines for illustrative purposes.

Another way to view this discrete wavelet transform is to plot the transform values by decomposition level, over time, as in Fig. 6. Only the coefficients of the fluctuation signal are plotted. The height of the spikes is proportional to the magnitude of the transform value while the direction, upward or downward, is indicative of its sign, either positive or negative. This plot gives a rough idea of the frequency content in the signal over time. Generally, the more the signal is decomposed, the lower the frequency of the subsignal. The energy of the subsignal at a given level is given by the sum of squares of the transform values. Hence, for the noisy kinematic signal example, the large spikes at deeper levels of decomposition (Fig. 6) suggest that most of the signal energy resides in lower frequencies at distinct instances in time. On the other hand, the high frequency noise

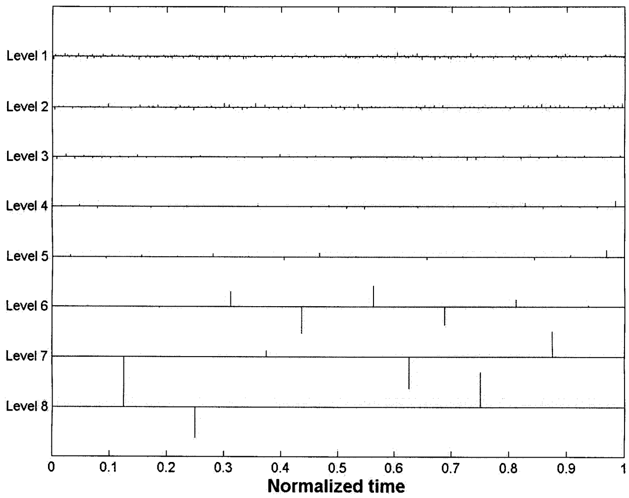

Fig. 6. Plot of wavelet transform values at each decomposition level. Only coefficients of the fluctuation signal are plotted.

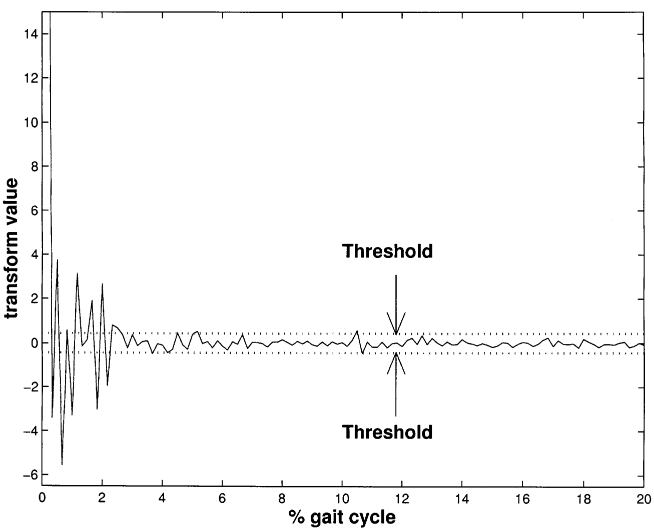

Fig. 7. Thresholding the wavelet transform to remove high frequency noise.

components (smaller spikes in levels 1 and 2) occur throughout the signal's duration.

A simple way to filter noise from the signal is to threshold the transform values. Hence, only transform values above a specified noise threshold are retained, others are set to zero. The inverse wavelet transform is then applied to the thresholded transform values. Continuing with the above example, the thresholded transform and the resultant reconstructed signal are shown in Fig. 7 and Fig. 8. This smoothing method is known as hard thresholding of the wavelet transform. Extending this simple scheme, soft thresholding [54] or wavelet shrinkage [55] also pushes all coefficients outside the thresholds towards zero by a preset amount. For a recent review of techniques for smoothing differentiated biomechanical signals via hard thresholding, refer to Wachowiak et al. [53].

There are numerous different types of discrete wavelet transforms, such as the Haar transform, Daubechies family of transforms and Coiflets transform[56]. The differences among the transforms lies in the definitions of the wavelet and scaling basis functions, i.e. the filter banks. The number following the wavelet transform name is related to the number of coefficients in the quadrature mirror filter. Some prescriptions on the choice of wavelet basis is given in Walker [57].

This highly simplified overview has merely brushed the surface of the filter bank interpretation of the wavelet transform. The technique has deep theoretical foundations in multiresolution signal decomposition [58], real analysis [59] and filter bank design [60]. For a review of wavelet applications in the biomedical arena, see Unser and Aldroubi [61]. Introductions to fundamental principles can be found in Rioul and Vetterli [62], Walker [57] and Graps [63].

3.2. Recent work

There have been two types of wavelet applications in the analysis of movement data, signal smoothing and signal discrimination. Each is reviewed below and summarized in Table 5.

Kinesiological data exhibit sharp spikes corresponding to impacts such as heel strikes. Butterworth and spline filters, while generally successful at smoothing biomechanical signals, undesirably attenuate impact signals and amplify edge effects. This is the motivation cited by Wachowiak et al. [64] in their investigation of wavelet-based smoothing of displacement data. Using an experimental rig, they measured the displacement of a falling cylinder capped with half a rubber ball. The set-up was intended to simulate impact signals. Wa-

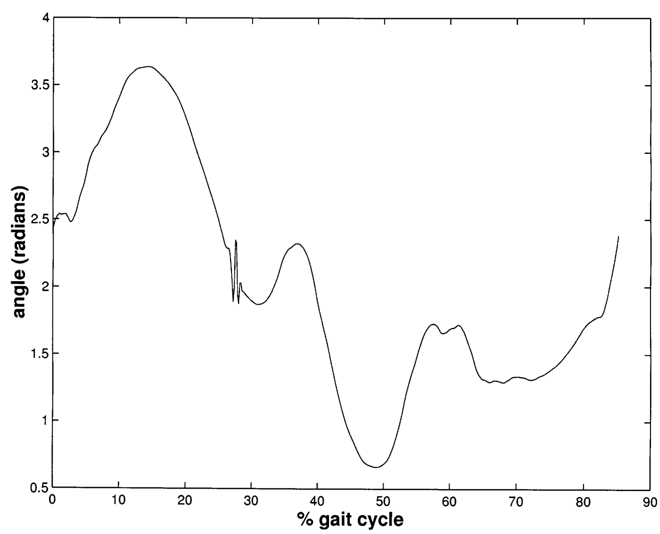

Fig. 8. Reconstructed, smooth signal after wavelet thresholding.

Table 5
Summary of wavelet analysis of gait signals

Author	Data	Transform	Demonstrated capability
Wachowiak et al. [64]	Displacement at heel strike (simulated)	Haar and Daubechies 4	Signal smoothing while retaining impact signals
Ismail and Asfour [65]	Angular displacement	Biorthogonal, Coiflet and Daubechies	Signal smoothing while resolving δ-function peaks
Tamura et al. [66]	Acceleration signals	Daubechies 3	Signal discrimination via energy of selected decomposition levels
Marghitu and Nalluri [67]	Joint angles	Coiflet	Signal discrimination via wavelet energy distributions

chowiak et al. applied the Haar and fourth order Daubechies wavelet to the displacement data. Through hard-thresholding, they found that the Haar wavelet produced a signal with negligible boundary effects and retained transient accelerations. However, the transform did not yield a smooth signal, having retained some unwanted noise components. Wachowiak et al. reported less satisfactory results with the Daubechies wavelet but concluded that the technique showed promise, on the basis of removing initial noise and mitigating boundary effects. Their results are somewhat expected as the Haar transform is better suited to filtering discontinuities [57]. They concluded that the challenge in applying wavelets to kinesiological data amounted to the determination of an appropriate wavelet basis function.

To further justify the use of wavelets for smoothing kinematic data, Ismail and Asfour [65] identified a number of limitations encountered with common Butterworth filters. These included the problems of highly sensitive derivatives in the filtered signal, large root mean square errors (RMSE) and a lack of agreed upon cut-off frequency and filter order. In their experiments, Ismail and Asfour [65] applied Biorthogonal, Coiflet and Daubechies wavelets of different orders to simulated angular displacement data. Wavelet coefficients were thresholded by wavelet shrinkage. For comparison, the same data were also filtered with Butterworth filters. Their results demonstrated that the wavelet transforms were remarkably superior to Butterworth filters at simultaneously reducing noise content in all frequency ranges while minimizing loss of signal energy. In particular, the fourth order Daubechies wavelet at only the second level of decomposition provided the best results, in terms of maximizing the retained energy while minimizing the RMSE. In general, the Butterworth filters achieved overall smoothing but could not resolve the delta function peaks, introduced end-point errors and thus produced very large RMSE. Similar to the conclusion of Wachowiak et al., Ismail and Asfour remark that the central challenge lies in the choice of an appropriate wavelet function. In addition, they also highlight the danger of over-smoothing the data via wavelets.

The discrete wavelet transform decomposes a continuous signal into coefficients which carry spectral and temporal information about the original signal. Tamura et al. [66] noted that these informative coefficients could therefore serve as discriminatory features for signal classification. To demonstrate the point, Tamura et al. exploited wavelets in the classification of acceleration signals. The magnitude of acceleration close to the body's center of gravity was recorded from 20 subjects during normal walking, walking up and walking down a stairway. The third-order Daubechies wavelet transform was applied to the acceleration signal with decomposition at nine levels. While it was difficult to distinguish the walking conditions in either time or frequency domain alone, the time–frequency representation of wavelets offered new insight. Specifically, Tamura et al. found that coefficients at the fourth and fifth decomposition levels were suitable for accurate classification. The sum of squared wavelet coefficients at these levels served as the discriminating features. No formal accuracies were reported but the visible discriminability in sum of squares graphs alone, shows promise in this technique. Since the sum of the squared coefficients is related to signal energy, Tamura et al. hypothesized that the discrimination of the different types of walking is related to differences in energy consumption.

Adopting this energy interpretation of squared wavelet coefficients, Marghitu and Nalluri [67] applied the wavelet transform in the detection of differences between normal greyhound gait and that affected by tibial nerve paralysis. Coxofemoral, femorotibial and tarsal joint angles were computed for six quadrupedal subjects. The six-coefficient Coiflet wavelet was applied to the kinematic time series, yielding eight wavelet decomposition levels. The percentage energy contribution of frequency components at each level was defined as the sum of squared coefficients in each level divided by the total energy in the original signal. Through signal reconstruction with selected wavelet coefficients, Marghitu and Nalluri found that low frequency components among subjects were very similar while differences manifested themselves at high frequencies. To contrast normal and affected gait, energy distributions at each decomposition level were compared within subjects. Marghitu and Nalluri identified significant differences only in the signal energy of the tarsal joint angle at decomposition levels 3 and 4. Their work demonstrated the novel use of wavelet coefficients to detect and quantify subtle abnormalities in gait, which would have otherwise been overlooked by conventional graphical comparisons.

3.3. Benefits and application issues

The greatest advantage of wavelet transforms over conventional Fourier transforms, is the provision of localized time and frequency information about a gait signal. Dynamic changes in walking, such as modified acceleration, can be easily detected [66] as shifts in local signal energy. Thus, wavelet transforms hold promise for revealing obscure pre- and post-intervention changes in gait. Further, using wavelets, gait signals can be locally de-noised with minimal signal energy loss [65]. This ability permits for example, the retention of impact signals while discarding confounding contaminations at similar frequencies occurring at different times in the gait cycle, a task beyond the capabilities of conventional filters.

The wavelet packet transform [68], a natural generalization of the wavelet transform, has not been applied in gait analysis. By decomposing both the trend and fluctuation signals, the time–frequency plane can be arbitrarily tiled to suit the specific signal under analysis. The wavelet packet transform has been successful in deciphering EMG signals [50,51].

In the application of wavelet transforms to gait signals, the selection of appropriate wavelet and scaling basis functions is a central, open question. The simplest Daubechies transform (i.e. with four wavelet numbers) has been the most successful and popular choice in the few gait studies reviewed. In general, the Daubechies and other smooth bases are well-suited to smoothly varying time series while the Haar transform is recommended for time series with discontinuous jumps [57]. Further, complex-valued wavelet bases are better adapted for capturing oscillatory behaviour whereas real-valued wavelet basis can more adeptly isolate peaks or discontinuities in the signal [69]. Horgan [70] give additional practical tips on selection of appropriate basis functions from the perspective of optimal signal smoothing.

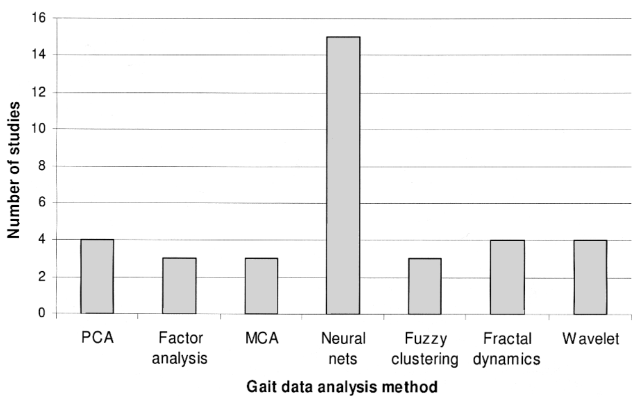

Fig. 9. Number of studies exploiting each gait data analysis method over the last 10 years.

4. Comparisons and summary

A number of technical challenges in quantitative analysis of gait data were identified in Part 1 of this review [1]. This section summarizes the extent to which individual methods in this two-part review have met the identified challenges. The present limitations of the methods are also highlighted. For detailed elaborations on principal components, factor analysis, multiple correspondence analysis, fuzzy clustering and fractal dynamics, please refer to the first paper [1] in this two-part review.

4.1. Prevalence of methods

In contrasting the various methods, it is interesting to first observe the dominance of neural network-based analyses of gait data over the past 10 years (Fig. 9), based on the reviewed literature. This prevalence speaks to the relative ease with which neural networks are applied, their widespread availability in commercial and educational software and their generic applicability to a broad spectrum of analysis problems. Fig. 9 also indicates that most of the analyses are not of an exploratory nature (since neural networks are not good exploratory tools) but driven by well-defined clinical and research questions. This lop-sided concentration suggests that more effort is required in non-neural network applications to more accurately gauge their practical benefits and limitations in gait data analysis. Further, objective comparisons of multiple methods in addressing a common clinical or research question have not yet been conducted. Such studies would be invaluable in defining the future directions of quantitative methods for analyzing gait data. Incidentally, PCA is the only technique whose history in gait data analysis dates back more than a decade. Its ongoing utility was reaffirmed by the recent work of Deluzio et al. [71,72].

4.2. Issues addressed

The following tables compare the extent to which the reviewed methods have addressed each of the identified

Table 6
Addressing the issue of high-dimensionality

Method	Maximum number of dimensions		Interpretability of results	
	In gait studies	Potential	Visualization	Mathematical result
PCA	101	Theoretically unlimited	Good only if two-dimensional projection	Interpretable linear equations [1]
FA	16	Theoretically unlimited	Good only if two-dimensional factor plane	Interpretable linear equations [1]
MCA	21	Theoretically unlimited	Good only if two-dimensional factor plane	Matrix of profile points not interpretable without plotting
NN	1316	Theoretically unlimited; practically limited by available computational power	Can only view outputs for given inputs	Non-linear equation is not interpretable

Table 7
Addressing the issue of high-dimensionality (continued from Table 6)

Method	Maximum number of dimensions		Interpretability of results	
	In gait studies	Potential	Visualization	Mathematical result
FC	4	Theoretically unlimited	Can easily visualize two or three-dimensional clusters	Interpretable cluster centers; multidimensional membership functions not easily interpretable
FD	1	Two and three dimensional analysis demonstrated in physiology	Always use bivariate plots to view dependence	Fractal dimension easily interpretable
WT	1	Two-dimensional transform demonstrated in image processing	Can view energy profiles, plots of coefficients	Coefficients easily interpretable via plots of coefficients

Table 8
Addressing the issue of temporal dependence (‘P’ indicates potential to analyze data type, but not yet demonstrated in gait studies)

Method	Admissible data types				Temporal dependence	
	Kinematic	Kinetic	EMG	Other	Incorporated	Extracted information
PCA	Yes	P	Yes	P	Yes [71]	Significant[c] portion of gait cycle
FA	Yes	P	Yes	P	No	NA
MCA	Yes	Yes	P	Yes	Partial (via fuzzy coding)	Significant[c] time windows
NN	Yes	Yes	Yes	P	Yes	Difficult to extract explicit time relationships
FC	Yes	P	P	Yes	No	NA
FD	Yes	P	P	P[a]	Yes	Nature of fluctuations over long times
WT	Yes	P	Yes	P[b]	Yes	Frequency content in local time interval

[a] Must be a time-varying quantity.
[b] Significance of a variable in PCA is defined in terms of the contributed variance.
[c] Significance of a profile point in MCA is defined in terms of contributed inertia.

challenges. In Table 6 and Table 7 the demonstrated and potential ability to manipulate multivariate data is summarized. Note that methods which can theoretically analyze any number of variables are limited practically by visualization and computational constraints. The interpretability of the analysis is closely tied to its usefulness with multivariate data. Here, interpretation is measured in terms of appeal to human visual perception and comprehensibility of numerical results. Immediately, one sees that the neural network offers the greatest ability to handle high-dimensional data but sacrifices interpretability [19]. The statistical methods PCA, FA and MCA yield the most readable results if the data can be adequately represented in two-dimensional displays. PCA is likely the recommended choice as a first step in data reduction and often proceeds FA, MCA and neural network modeling.

Another comparison concerns admissible types of data and is related to the issue of incorporating temporal dependence. Table 8 outlines the different gait data which have been analyzed by the various methods. A ‘P’ designates the potential to analyze a type of data, although not yet demonstrated in published gait studies. 'Other' data refers to discrete entities such as gender, constants for a given trial, such as anthropometric data and other continuous, time-evolving quantities such as metabolic measurements. The last column indicates the method's ability to handle explicit time dependence. Although all methods have the potential to analyze the various types of gait data, the versatility of neural networks has been most widely demonstrated. With the exception of fuzzy clustering, all methods can incorporate time dependence to some extent. However, the temporal information extracted by each method is different. For general exploration of temporal dependencies and the detection of subtle differences over time, a combination of time domain methods (such as PCA and MCA) along with time–frequency analysis (wavelet transform) is recommended. If one is only interested in modeling the temporal evolution of gait signals without a need for explicit interpretation, then a neural network is the tool of choice.

In quantitative gait analysis, one of the common end goals is to detect changes due to interventions such as therapy, orthoses or surgery. A suitable analysis method for this purpose is one which can quantitatively and directly compare entire gait waveforms and pinpoint differences. As shown in Table 9 few methods can analyze the entire gait waveform adequately. Furthermore, neural networks cannot easily isolate specific differences between curves. Conceivably, one could assign network outputs to measure different portions of a curve, but even then the localization would be crude. Note that MCA can provide qualitative but not quantitative comparisons between curves. PCA and the wavelet transform can identify local differences in variability and energy/frequency, respectively. However, these are only limited aspects of the gait waveform.

Table 9
Addressing the issue of measuring differences between curves

Method	Differences between curves	
	Ability to quantify differences	Ability to localize differences
PCA	Yes, as differences in principal component scores	Yes [71]; locate differences at specific instances of gait cycle
FA	No	No
MCA	No	No
NN	Yes, as differences in neural network outputs	Difficult to automatically pinpoint differences
FC	No	No
FD	Yes, as differences in fractal dimension	No, fractal dimension is a global measure
WT	Yes, as differences in wavelet coefficients or total energy	Yes, can detect local differences in energy distributions

Further work is required to supplement these tools with an objective, standard measure of differences between complete gait curves. To this end, canonical correlation analysis for functions [73] and regression analysis for functional responses [74] are two alternatives worthy of deeper exploration.

Another identified challenge is the variability in gait data. The multivariate statistical methods, PCA, FA and MCA, explicitly deal with variance in the data. PCA finds projections to maximally capture total variance. FA finds factors to maximally reveal covariance structure while MCA finds projections that maximally preserve inertia (total variance). Since variance is effectively 'information' for these methods, large variability in the data may lead to the identification of false patterns. This same sensitivity to variability, however, could also act as a useful tool for identifying extraordinary fluctuations in specific measurements early in the analysis.

Variability is indirectly accounted for in the feedforward neural network during training. Statistically, network training can be likened to learning the distribution of the expected value of the outputs conditioned on the inputs [75]. Through this 'averaging' process, the feedforward neural network can be robust to large variability in the data . However, this robustness is contingent upon the use of abundant training data.

Fractal analysis has specifically focussed on long-range intrasubject variability, without assuming that the variability is random. In particular, fractal analysis can detect subtle changes in stride interval variability [76]. Fuzzy clustering can identify grouping tendencies amid natural intersubject variability [77]. With the reviewed methods, there is indeed potential for assessing and understanding gait data variability. This potential nonetheless hinges heavily upon the depth of the user's domain knowledge. Although able to deal with some aspects of the data's variability, the suppression of rampant variability is not the central focus of the reviewed methods. Care should be taken, as in Hausdorff et al. [78] and West and Griffin [79], to verify that observed patterns have not arisen from spurious or random effects.

The ability to detect and model non-linear relationships in gait data is important for gaining deeper insights into pathologies and physiological mechanisms. With the exclusion of PCA and FA, the methods do not make assumptions about the linearity of the data's relationships. Again, the type of relationship detected or modeled is different for each method. The neural network seeks a relationship between a general set of inputs and outputs, and in this sense, is the most generic tool among the methods. The downside is the lack of readability of the modeled relationship. MCA could reveal interesting non-linear relationships among

Table 10
Addressing the issue of detecting non-linear relationships in gait data

Method	Non-linear relationships	
	Ability to detect?	Extracted information
PCA	No	Not applicable
FA	No	Not applicable
MCA	Yes	Relationship between column and row profile points (gait variable values or fuzzy sets)
NN	Yes	Functional relationship between inputs and outputs (dependent and independent gait variables)
FC	Yes	Grouping inclinations (inherent similarities) among subjects or observations
FD	Yes	Relationship between fluctuations in measurements separated by long times
WT	Yes	Relationships between time and frequency characteristics of gait signals

Table 11
Limitations of alternative gait data analysis methods

Method	Limitations
PCA	Only detects linear relationships in data; Heavy reliance on subjective interpretation of components
FA	Only detects linear relationships in data; Heavy reliance on creative labeling of factors
MCA	Factor plane becomes cluttered and difficult to interpret with moderate data volumes; Sensitive to coding of gait signals; Reliance on subjective identification of associations among profile points
NN	Captured relationships are generally difficult to interpret
FC	Number of groups needs to be specified a *priori* Fuzziness parameter needs to be chosen arbitrarily
FD	Utility demonstrated only for stride interval time series Has only been applied to univariate signals
WT	Little guidelines on selection of wavelet basis for gait signals Has only been applied to univariate signals

gait variables insofar as those patterns can be aptly interpreted by the user. The wavelet transform, fractal analysis and fuzzy clustering offer valuable insight into specific types of possibly non-linear relationships. If modeling non-linearity is a primary objective, the neural network is the clear preference. For exploring potential non-linearities, the remaining methods all have relative merits, especially where clinical interests match the type of extractable information, see Table 10.

4.3. Limitations

A review of the limitations of each method can also help to determine appropriate avenues of application. Table 11 highlights the critical limitations already mentioned in the foregoing discussions. For the wavelet transform and fractal analysis, the limited scope of application to date is itself an obstacle in objectively gauging the usefulness of these methods in gait data analysis.

4.4. Other emergent data analysis methods

In addition to the emergent techniques in gait data analysis reviewed herein, the potential of several other prominent data analysis methods from statistics, engineering and computer science remain untapped.

Projection pursuit is an exploratory multivariate statistical method [80] that seeks low-dimensional data projections most different from multivariate normality [81]. This method is more general than PCA as it can uncover multiple levels of structure.

Decision trees are fundamental techniques in machine learning where a set of easily readable rules are automatically induced from a data set (see for example Quinlan [82], Quinlan [83] or Bohren et al. [84]). Decision trees have been used to develop locomotion control rules for functional electrical stimulation [85] and could be exploited in the interpretation of multivariate gait data.

Pattern discovery refers to a host of algorithms that automatically search for non-random relationships in the data and present those patterns in a comprehensible format (see for example Chau and Wong [86] or Wong and Yang [87]). These methods could be applied to the automatic detection and deciphering of high-dimensional inter-relationships among gait variables.

Non-linear projection describes a family of methods, including learning vector quantization [12] and Sammon mapping [88], that seek data projections which are not linearly related to the original data. They can thus uncover more general relationships in data.

In addition, the previously mentioned radial basis functions [10], a versatile alternative to feedforward neural networks, have not been commissioned in gait data analysis.

4.5. Choosing an analysis methodology

To assist readers in the selection of suitable analysis methodologies, the various methods are organized into specific research or clinical analysis needs in Table 12. The leftmost column groups together tasks that involve fundamental processing of gait signals. The middle column comprises techniques which seek unknown information from the gait data and is therefore labelled 'exploratory'. The final column lists methods which fulfill specific goals, such as classifying subjects or predicting certain parameters. The superscript 'a' indicates methods which have not yet been applied to gait data analysis, but which have demonstrated potential for analyzing continuous signals. Bearing in mind the above summaries of issues addressed, Table 12 can serve as a quick selection guide for candidate analysis options.

Note that a single analysis method, such as the discrete wavelet transform, may serve multiple purposes. These groupings serve only as a guide and are not necessarily mutually exclusive. For example, detecting differences in gait waveforms, a task classified as 'processing', may also have exploratory objectives. Furthermore, a given analysis may employ multiple methods in a hybrid scheme, such as fuzzy coding followed by MCA [89,90].

5. Conclusion

Quantitative analysis and clinical interpretation of gait data can be formidable tasks. Traditional methods such as signal parameterizations and qualitative inspection of bivariate plots have fallen short in meeting persistent challenges of quantitative and objective analysis. Only in recent years have gait analysis studies begun to take advantage of non-traditional analysis methods. These methods present a variety of fresh perspectives on gait data but are not without individual shortcomings. In particular, neural networks have proven to be a dominant and promising alternative tool in numerous studies, but lack overall readability. Multivariate statistical methods such as PCA and FA, although limited by linearity assumptions, have demonstrated a practical ability to provide powerful interpretations into healthy and pathological human gait. Other methods such as fractal analysis, wavelet transform and fuzzy clustering still require additional studies to clarify their potential in enhancing gait data analysis. However, it is already clear that these methods can offer additional insights into gait data and human locomotion that are simply not achievable with traditional analyses. The most notable example is the fractal dynamics of gait. At the present time, no emergent technique in isolation can sufficiently meet all the challenges of quantitative gait analysis. Much work remains to be done in elevating the level of analysis to widespread clinical acceptance and applicability. From this review, the most marked on-going needs are in visualization and interpretation of multiple gait signal interactions and the quantitative comparisons of gait waveforms. As in many other disciplines, hybrid approaches to data analysis (for example Miyoshi et al. [91], Rao [92], Tzafestas et al. [93] and Al-Fahoum and Howitt [94]) exploiting the collective strengths of multiple techniques, may define the future trend for quantitative analysis of gait data.

Acknowledgements

The author acknowledges the very constructive suggestions and comments of the editor and anonymous referees on earlier versions of the manuscript. The author is also indebted to Stephen Naumann, Kent Campbell, Kim Parker and Alan Morris for feedback on initial drafts. This work was supported by the

Table 12
Data analysis methods arranged by analysis needs

Processing	Exploratory	Task-driven
Reduction of gait data	**Finding long-range correlations**	**Modeling relationships among gait variables**
PCA, FA, Non-linear projection[a], projection pursuit[a]	Fractal dynamics	Feedforward neural networks, radial basis functions[a], recurrent neural networks
Noise removal; smoothing	**Automatic grouping of subjects**	**Prediction of gait parameters**
Discrete wavelet transform	Fuzzy clustering, Self-organizing map	Feedforward neural network, recurrent neural network
Measuring differences in entire gait waveforms	**Search for patterns**	**Classification of pathological conditions**
PCA, Fractal dynamics	PCA, FA, MCA, Pattern discovery[a], decision trees[a]	Feedforward neural network, decision trees[a]
Feature extraction	**Studying relationship between EMG and kinematics**	**Confirming a suspected pattern in data**
Discrete wavelet transform, PCA	Feedforward neural network, decision trees[a]	MCA, FA, Decision trees[a]
	Studying muscle coordination patterns	
	Factor analysis	

[a] Methods which have not been applied to gait data analysis but have demonstrated potential in analyzing continuous signals.

Natural Sciences and Engineering Research Council of Canada, the Hospital for Sick Children Foundation and the Bloorview Children's Hospital Foundation.

References

[1] Chau T. A review of analytical techniques for gait data. part 1: Fuzzy, statistical and fractal methods, Gait and Posture 2001.
[2] de Ridder D, Duin RPW, Verbeek PW, van Vliet U. The applicability of neural networks to non-linear image processing. Pattern Analysis and Applications 1999;2(2):111–28.
[3] Lu XQ, Mo JY, Yang C, Kang JW, Gao JZ. Orthogonal wavelets analysis of electroanalytical signals. Analytical Letters 2000;33(6):1167–86.
[4] Zaknich A. Introduction to the modified probabilistic neural network for general signal processing. IEEE Transactions on Signal Processing 1998;46(7):1980–90.
[5] Lina JM. Image processing with complex daubechies wavelets. Journal of Mathematical Imaging and Vision 1997;7(3):211–23.
[6] Rumelhart DE, McClelland JL. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1&2. Cambridge, MA: MIT Press, 1986.
[7] Hinton GE. Connectionist learning procedures. Artificial Intelligence 1989;40:185–234.
[8] Hinton GE. How neural networks learn from experience. Scientific American 1992;267:144–51.|iles
[9] Ohno-Machado L, Rowland T. Neural network applications in physical medicine and rehabilitation. American Journal of Physical Medicine and Rehabilitation 1999;78(4):392–8.
[10] Bishop CM. Neural Networks for Pattern Recognition. Oxford University Press, 1995.
[11] Specht DF. Probabilistic neural networks. Neural Networks 1990;3(1):109–18.
[12] Kohonen T. Self-organizing Maps. Berlin: Springer-Verlag, 1995.
[13] Holzreiter SH, Kohle ME. Assessment of gait patterns using neural networks. Journal of Biomechanics 1993;26(6):645–51.
[14] Barton JG, Lees A. Development of a connectionist expert system to identify foot problems based on under-foot pressure patterns. Clinical Biomechanics 1995;10(7):385–91.
[15] Barton JG, Lees A. An application of neural networks for distinguishing gait patterns on the basis of hip–knee joint angle diagrams. Gait and Posture 1997;5:28–33.
[16] Lafuente R, Belda JM, Sánchez-Lacuesta J, Soler C, Prat J. Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis. Clinical Biomechanics 1997;13(3):216–20.
[17] Duda RO, Hart PH. Pattern Classification and Scene Analysis. New York: John Wiley & Sons, 1973.
[18] Heller BW, Veltink PH, Rijkhoff NJM, Rutten WLC, Andrews BJ. Reconstructing muscle activation during normal walking: a comparison of symbolic and connectionist machine learning techniques. Biological Cybernetics 1993;69:327–35.
[19] Narazaki H, Yamamoto M, Watanabe T. Reorganizing knowledge in neural networks: an explanation mechanism for neural networks in data classification problems. IEEE Transactions on Systems, Man and Cybernetics 1996;26(1):107–17 Part B.
[20] Sepulveda F, Wells DM, Vaughan CL. A neural network representation of electromyography and joint dynamics in human gait. Journal of Biomechanics 1993;26(2):101–9.
[21] Winter DA. The Biomechanics and Motor Control of Human Gait. Ontario: University of Waterloo Press, 1987.
[22] Prentice SD, Patla AE, Stacey DA. Simple artificial neural network models can generate basic muscle activity patterns for human locomotion at different speeds. Experimental Brain Research 1998;123:474–80.
[23] Gioftsos G, Grieve DW. The use of neural networks to recognize patterns of human movement: gait patterns. Clinical Biomechanics 1995;10(4):179–83.
[24] Elman JL. Finding structure in time. Cognitive Science 1990;14:179–211.
[25] Aminian K, Robert P, Jéquier E. Incline, speed and distance assessment during unconstrained walking. Medicine and Science in Sports and Exercise 1995;27(2):226–34.
[26] Aminian K, Robert P, Jéquier E, Schutz Y. Estimation of speed and incline of walking using neural network. IEEE Transactions on Instrumentation and Measurement 1995;44(3):743–6.
[27] Herren R, Sparti A, Aminian K, Schutz Y. The prediction of speed and incline in outdoor running in humans using accelerometry. Medicine and Science in Sports and Exercise 1999;31(7):1053–9.
[28] Savelberg H, Herzog W. Prediction of dynamic tendon forces from electromyographic signals: an artificial neural network approach. Journal of Neuroscience Methods 1997;78:65–74.
[29] Liu MM, Herzog W, Savelberg HHCM. Dynamic muscle force predictions from emg: an artificial neural network approach. Journal of Electromyography and Kinesiology 1999;39:391–400.
[30] Savelberg HHCM, de Lange ALH. Assessment of the horizontal, fore-aft component of the ground reaction force from insole pressure patterns by using artificial neural networks. Clinical Biomechanics 1999;14(8):585–92.
[31] Cottrell GW, Nguyen M, Tsung F. Tau net: the way to do is to be. In: Proceedings of the 15th Annual Conference of the Cognitive Science Society. New Jersey: Lawrence Erlbaum Associates, 1993:365–70.
[32] Tucker CA, White SC. Neurocomputational approaches to pattern recognition and time-series analysis of electromyographic data obtained during treadmill walking. In: International Society of Biomechanics XVIIth Congress. Calgary: ISB, 1999:214.
[33] Kosko B. Neural Networks and Fuzzy Systems. Englewood Cliffs, NJ: Prentice Hall, 1992.
[34] Yager RR, Filev DP. Essentials of Fuzzy Modeling and Control. New York: John Wiley & Sons, 1994.
[35] Jang JSR. Anfis — adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics 1993;23(3):665–85.
[36] Lapedes A, Farber R. How neural networks work. In: Anderson DZ, editor. Neural Information Processing Systems. New York: American Institute of Physics, 1988:442–56.
[37] Cybenko G. Approximation by superpositions of a sigmoidal functionmathematics of control. Signals and Systems 1989;2:303–14.
[38] Andrews R, Cable R, Diederich J, Geva S, Golea M, Hayward R, HoStuart C, Tickle A. Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowledge-based Systems 1995;8(6):373–89.
[39] Bologna G, Pellegrini C. Three medical examples in neural network rule extraction. Physica Medica 1997;13(Suppl. 1):183–7.
[40] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533–6.
[41] Bremner FJ, Gotts SJ, Denham DL. Hinton diagrams — viewing connection strengths in neural networks. Behavior Research Method Instruments and Computers 1994;26(2):215–8.
[42] Wejchert J, Tesauro G. Visualizing processes in neural networks. IBM Journal of Research and Development 1991;35:244–53.
[43] Hertz JA, Krogh AS, Palmer RG. Introduction to the Theory of Neural Computation. Addison-Wesley Publishing Company, 1991.
[44] Siedlecki W, Sklansky J. On automatic feature selection. International Journal of Pattern Recognition and Artificial Intelligence 1988;2(2):197–220.
[45] Cybenko G. Complexity theory of neural networks and classification problems. In: Almeida LB, Wellekens CJ, editors. Neural Networks: EURASIP Workshop 1990 Proceedings. New York: Springer-Verlag, 1990:26–43.

[46] Haykin S. Neural Networks: A Comprehensive Foundation. New York: Maxwell Macmillan Canada, 1994.
[47] Park J, Sandberg IW. Universal approximation using radial basis function networks. Neural Computation 1991;3:246–57.
[48] Shaffer R, Rose-Pehrsson S, McGill R. A comparison study of chemical sensor array pattern recognition algorithms. Analytica Chimica Acta 1999;384(3):305–17.
[49] Blue JL, Candela GT, Grothier PJ, Chellappa R, Wilson CL. Evaluation of pattern classifiers for fingerpring and OCR applications. Pattern Recognition 1994;27(4):485–501.
[50] Englehart K, Hudgins B, Parker PA, Stevenson M. Classification of the myoelectric signal using time–frequency based representations. Medical Engineering and Physics 1999;21:431–8.
[51] Karlsson S, Yu J, Akay M. Enhancement of spectral analysis of myoelectric signals during static contractions using wavelet methods. IEEE Transactions on Biomedical Engineering 1999;46(6):670–84.
[52] Basmajian JV, Luca CJD. Muscles Alive: Their Functions Revealed by Electromyography, 5th edition. Baltimore: Williams and Wilkins, 1985.
[53] Wachowiak MP, Rash GS, Quesada PM, Desoky AH. Wavelet-based noise removal for biomechanical signals: a comparative study. IEEE Transactions on Biomedical Engineering 2000;47(3):360–8.
[54] Donoho DL. De-noising by soft-thresholding. IEEE Transactions on Information Theory 1995;41(3):613–27.
[55] Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association 1995;90:1200–24.
[56] Daubechies I. Ten Lectures on Wavelets. Philadelphia: PA: SIAM, 1992.
[57] Walker JS. A Primer on Wavelets and Their Scientific Applications. New York: Chapman and Hall/CRC, 1999.
[58] Mallat SG. A theory of multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989;11(7):674–93.
[59] Cohen A, Kovacevic J. Wavelets: the mathematical background. Proceedings of the IEEE 1996;84(4):514–22.
[60] Vetterli M, Herley C. Wavelets and filter banks: theory and design. IEEE Transactions on Signal Processing 1992;40(9):2207–32.
[61] Unser M, Aldroubi A. A review of wavelets in biomedical applications. Proceedings of the IEEE 1996;84(4):626–38.
[62] Rioul O, Vetterli M. Wavelets and signal processing, IEEE Signal Processing Magazine 1991;8(4):14–38.
[63] Graps A. An introduction to wavelets. IEEE Computational Science and Engineering 1995;2(2):50–61.
[64] Wachowiak M, Rash G, Desoky A, Quesada P. Wavelet transforms for smoothing kinesiological data. In: Proceedings of the 21st Annual Meeting of the American Society of Biomechanics. South Carolina: ASB: Clemson Univeristy, 1997.
[65] Ismail AR, Asfour SS. Discrete wavelet transform: a tool in smoothing kinematic data. Journal of Biomechanics 1999;32:317–21.
[66] Tamura T, Sekine M, Ogawa M, Togawa T, Fukui Y. Classification of acceleration waveforms during walking by wavelet transform. Methods of Information in Medicine 1997;36:356–9.
[67] Marghitu DB, Nalluri P. An analysis of greyhound gait using wavelets. Journal of Electromyography and Kinesiology 1997;7(3):203–12.
[68] Wickerhauser MV. Best-adapted wavelet packet bases. In: Daubechies I, editor. Different Perspectives on Wavelets. Providence, RI: American Mathematical Society, 1993:155–72.
[69] Torrence C, Compo GP. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 1998;79(1):61–78.
[70] Horgan GW. Using wavelets for data smoothing: a simulation study. Journal of Applied Statistics 1999;26(8):923–32.
[71] Deluzio KJ, Wyss UP, Zee B, Costigan PA, Sorbie C. Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns. Human Movement Science 1997;16:201–17.
[72] Deluzio KJ, Wyss UP, Costigan PA, Sorbie C, Zee B. Gait assessment in unicompartmental knee arthroplasty patients: Principal component modeling of gait waveforms and clinical status. Human Movement Science 1999;18:701–11.
[73] Leurgrans SE, Moyeed RA, Silverman BW. Canonical correlation-analysis when the data are curves. Journal of the Royal Statistical Society Series B-Methodological 1993;55(3):725–40.
[74] Faraway JJ. Regression analysis for a functional response. Technometrics 1997;39(3):254–61.
[75] White H. Learning in artificial neural networks: a statistical perspective. Neural Computation 1989;1:425–64.
[76] Hausdorff JM, Mitchell SL, Firton R, Peng CK, Cudkowicz ME, Wei JY, Goldberger AL. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and huntington's disease. Journal of Applied Physiology 1997;82(1):262–9.
[77] O'Malley MJ, Abel MF, Damiano DL, Vaughan CL. Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. IEEE Transactions on Rehabilitation Engineering December 1997;5(4):300–9.
[78] Hausdorff JM, Purdon PL, Peng CK, Ladin Z, Wei JY, Goldberger AL. Fractal dynamics of human gait: stability of long-range correlation in stride interval fluctuations. Journal of Applied Physiology 1996;80:1448–57.
[79] West BJ, Griffin L. Allometric control, inverse power laws and human gait. Chaos Solitons and Fractals 1999;10(9):1519–27.
[80] Friedman JH, Tukey JW. A projection pursuit algorithm for exploratory data analysis. IEEE Transactions on Computers 1974;29(9):881–9.
[81] Huber PJ. Projection pursuit. Annals of Statistics 1985;13:435–525.
[82] Quinlan JR. Discovering rules by induction from large classes of examples. In: Michie D, editor. Expert Systems in the Microelectronic Age. Edinburgh: Edinburgh University Press, 1979:168–201.
[83] Quinlan JR. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
[84] Bohren BF, Hadzikadic M, Hanley EN. Extracting knowledge from large medical databases: an automated approach. Computers and Biomedical Research 1995;28:191–210.
[85] Jonić S, Janković T, Gajić V, Popović D. Three machine learning techniques for automatic determination of rules to control locomotion. IEEE Transactions on Biomedical Engineering 1999;46(3):300–10.
[86] Chau T, Wong AKC. Pattern discovery by residual analysis and recursive partitioning. IEEE Transactions on Data and Knowledge Engineering 1999;11(6):833–52.
[87] Wong AKC, Yang W. High-order pattern discovery from discrete valued data. IEEE Transactions on Data and Knowledge Engineering 1997;9(6):877–93.
[88] Sammon J. A non-linear mapping for data structure analysis. IEEE Transactions on Computers 1969;18:401–9.
[89] Loslever P. Error and data coding in the multi-dimensional analysis of human movement signals. Proceedings of the Institution of Mechancial Engineers. Journal of Engineering in Medicine 1993;207:103–10 Part H.
[90] Loslever P, Laassel E, Angue JC. Combined statistical study of joint angles and ground reaction forces using component and multiple correspondence-analysis. IEEE Transactions on Biomedical Engineering 1994;41(12):1160–7.
[91] Miyoshi T, Ichihashi H, Nagasaka K. Fuzzy projection pursuit density estimation by eigenvalue method. International Journal of Approximate Reasoning 1999;20(3):237–48.

[92] Rao DH. Fuzzy neural networks. JETE Journal of Research 1998;44(4–5):227–36.
[93] Tzafestas ES, Nikolaidou A, Tzafestas SG. Performance evaluation and dynamic node generation for principal component analysis neural networks. Mathematics and Computers in Simulation 2000;51(3–4):145–56.
[94] Al-Fahoum AS, Howitt I. Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias. Medical and Biological Engineering and Computing 1999;37(5):566–73.