Neural networks for detection and classification of walking pattern changes due to ageing

R. Begg[1] and J. Kamruzzaman[2]

[1]*Victoria University, City Flinders Campus, Melbourne, Australia*
[2]*Monash University, Gippsland Campus, Churchill, Australia*

Abstract

With age, gait functions reflected in the walking patterns degenerate and threaten the balance control mechanisms of the locomotor system. The aim of this paper is to explore applications of artificial neural networks for automated recognition of gait changes due to ageing from their respective gait-pattern characteristics. The ability of such discrimination has many advantages including the identification of at-risk or faulty gait. Various gait features (e.g., temporal-spatial, foot-ground reaction forces and lower limb joint angular data) were extracted from 12 young and 12 elderly participants during normal walking and these were utilized for training and testing on three neural network algorithms (Standard Backpropagation; Scaled Conjugate Gradient; and Backpropagation with Bayesian Regularization, BR). Receiver operating characteristics plots, sensitivity and specificity results as well as accuracy rates were used to evaluate performance of the three classifiers. Cross-validation test results indicate a maximum generalization performance of 83.3% in the recognition of the young and elderly gait patterns. Out of the three neural network algorithms, BR performed superiorly in the test results with best sensitivity, selectivity and detection rates. With the help of a feature selection technique, the maximum classification accuracy of the BR attained 100%, when trained with a small subset of selected gait features. The results of this study demonstrate the capability of neural networks in the detection of gait changes with ageing and their potentials for future applications as gait diagnostics.

Key words gait analysis, ageing, neural networks, feature selection

Introduction

With age, gait functions reflected in the walking patterns degenerate and threaten balance control mechanisms of the locomotor system. The reported declines in gait measures include: basic spatial-temporal parameters such as stride length, walking speed, stance/swing times[1,2]; joint angular excursions at the hip, knee and ankle joints[3] and kinetic parameters as reflected in the foot-to-ground reaction force-time data such as the vertical and horizontal peak forces[2,4]. Gait analysis has been proposed as a method for identifying individuals with a decline in performance of the locomotor system. Recently, many research projects have been directed at describing age-related changes in gait. The main focus of these research schemes is to devise diagnostic techniques such that these gait features can be used to detect gait degeneration due to ageing and identify the gait changes as a result of the effects of fall-proneness in the elderly individuals. Like other industrialized countries, falls in the older population has been identified as a major public health issue in Australia costing the community about A$2.4 billion per annum[5].

Gait classification using statistical techniques such as Linear Discriminant Analysis (LDA) has limitations, especially when the problem to be studied is linearly non-separable or complex. It has been demonstrated that neural networks techniques have the potential to offer a better alternative in pathological gait pattern classification[6]. The motivation behind this research was, therefore, to apply neural networks (NN) for automatic identification of gait types (young/old) from their gait measures. The success of such discrimination abilities by a neural network could lead to many potential applications, such as for gait diagnostics. As an example, a neural network could be used for early identification of at-risk or faulty gait so that appropriate measures could then be undertaken for gait rehabilitation. In recent years, neural networks have emerged as powerful tools for solving various classifications and modelling problems in biomechanics[7]; for example, to classify normal and pathological gait using force platform measures[8,9] and to simulate various gait types (e.g., leg length discrepancy) from their joint-angle measures[10], with reported excellent

Corresponding author: R. Begg, Centre for Ageing, Rehabilitation, Exercise & Sport, Victoria University City Flinders Campus, PO Box 14428, Melbourne City MC, Victoria 8001, Australia, Tel: +61 3 9919 1116, Fax: +61 3 9919 1110, Email: rezaul.begg@vu.edu.au
Received: 20 April 2005; Accepted: 18 April 2006

success rates. These research outcomes also highlight that gait features inherent in the kinetic ground reaction forces and joint angular excursion measures carry useful information regarding the quality of gait and its functional status such that these characteristics might be used to detect the declines in gait performance due to ageing or pathology.

At present, there is no reported research known to the authors exploring the classification ability of the neural networks in ageing gait. In this paper, we apply three well known neural network algorithms (Standard Backpropagation (BP), Scaled conjugate gradient (SCG) and Bayesian regularization (BR)) for the automated recognition of young/old gait patterns from standard gait measures, and compare their relative suitability as a gait classifier. Performance of the classifiers was evaluated using their accuracy rates in recognizing the young/old gait patterns and areas under the receiver operating characteristics (ROC) curves to examine their strengths and weaknesses in relation to detecting the gait changes with age.

Materials and methods

Participants

Twelve young and twelve older adults participated in the gait data collection. There were equal numbers of male and females within the two age groups. The young adults were recruited from the academic community of Victoria University and the elderly participants were volunteers from local senior citizen clubs. All subjects undertook informed-consent procedures approved by the Victoria University, Melbourne, Human Research Ethics Committee. The subjects had no known injuries or abnormalities that would affect their gait. Means and standard deviations (in brackets) of subject characteristics were as follows; Age (year) - young 28.1(5.6), elderly 68.8(4.6); Height (cm) – young 172 (11.2), elderly 166(10); Body Mass (kg) – young 72.4 (16.4), elderly 67.8(9.1).

Data collection and features

Gait recordings were carried out during normal walking on the laboratory walkway. All subjects completed a minimum of 3 walking trials while their gait characteristics were captured using an AMTI force platform and a PEAK 2D Motion Analysis system (Peak Performance Inc, USA). The walkway used in this experiment was 20m long with the force platform installed in the middle of the walkway. The subjects were given a few practice trials during which a suitable starting position was determined for each subject. This allowed a more natural footfall on the platform during the actual walking trials without targeting the force platform. Altogether 24 gait features were extracted from the gait data describing lower limb joint motion and foot-ground reaction force-time characteristics. Mean values of 3 walking trials were calculated and then used as gait features for training and testing the gait classifiers. The gait features used in this study included stride cycle time-distance data, lower limb joint angles and angular range of motion (ROM), and characteristics of the foot-to-ground reaction forces (GRF):

i) Stride phase gait cycle data – Walking speed, stance, swing and double-stance times and their corresponding normalized data and stride length,
ii) GRF data – Foot-ground reaction forces along vertical and anterior-posterior directions were recorded using one force-sensing platform. Peak forces during key phases of the gait cycle were normalized to body weight – this included vertical peak forces during weight acceptance, mid stance and push-off phases, and horizontal peak forces during braking and propulsive phases, and
iii) Joint angular data - Movement of the lower limb was recorded using the PEAK system and reflective markers attached to lower limb joints and segments (greater trochanter, lateral epicondyle, lateral malleolus, calcaneus, 5th metatarsal head). The angular data included knee and ankle joint angles at key events (heel contact and toe-off), and joint angular excursion or range of motion (ROM) during the stance, swing and stance-to-swing transition phases of the gait cycle. These gait measures have been shown to be useful indicators of ageing in a number of investigations[1, 4, 11-14].

Neural network algorithms

Artificial neural networks can be very useful to realize an input to output mapping when the exact relationship between input and output is unknown or very complex. Because of its ability to learn complex mappings, it has been used as a powerful classifier in many engineering[15, 16], financial[17] and biomedical applications [6-10, 18]. In this study, we applied an artificial neural network for the analysis of human gait and investigated the performance of different neural network learning algorithms in relation to this particular problem.

The multilayer feed forward network (Fig. 1) is one of the most commonly used neural network architectures. It consists of an input layer, an output layer and one or more intermediate layers called the hidden layer(s). All the nodes at each layer are connected to each node at the upper layer by interconnection strengths called weights. All the interconnecting weights between the layers are initialized to small random values at the beginning. During training, input features are presented at the input layer and associated target outputs are presented at the output layer. A training algorithm is used to attain a set of weights that minimizes the difference between the target output and actual output produced by the network.

There are various algorithms proposed in the literature to train a multilayer feed-forward network. There exists a theoretical framework that focuses on estimating the generalization ability of a network as a function of architecture and training set considering the region of weight space consistent with the training set; that is, a particular learning rule might favour some regions over others[19,20]. However, the suitability of a training algorithm in producing good generalization ability, in relation to a

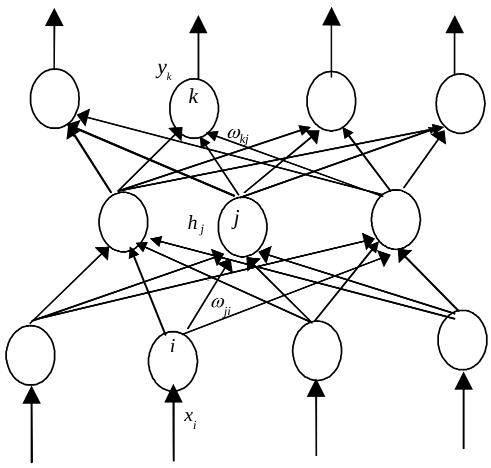


Figure 1. *Architecture of a multilayer feed-forward neural network.*

particular application, is usually determined by experiments. In this study, we experimented using three commonly applied neural network learning algorithms, namely, standard Backpropagation (BP), Scaled Conjugate Gradient Algorithm (SCG) and Backpropagation with Bayesian Regularization (BR) in order to find the best suited algorithm for detecting human gait pattern changes[20]. In the following, we provide a brief description of the three algorithms.

Standard Backpropagation (BP)

The standard Backpropagation[20] algorithm iteratively updates the weights to map a set of input-output pairs $\{(\mathbf{x_1},\mathbf{y_1}), (\mathbf{x_2},\mathbf{y_2}), \ldots, (\mathbf{x_p},\mathbf{y_p})\}$ using gradient descent technique. The input vector $\mathbf{x_p}$, upon multiplied by weight vectors, produces outputs at the hidden layer. Similarly, hidden layer outputs, being multiplied by their respective weights are propagated to the final output layer. In a three-layer network, the activation at each layer is represented by the following equations.

$$\mathbf{y}_\mathrm{p} = f(\mathbf{W}_\mathrm{o}\mathbf{h}_\mathrm{p} + \boldsymbol{\theta}_\mathrm{o}), \tag{1}$$

$$\mathbf{h}_\mathrm{p} = f(\mathbf{W}_\mathrm{h}\mathbf{x}_\mathrm{p} + \boldsymbol{\theta}_\mathrm{h}), \tag{2}$$

where $\mathbf{W_o}$ and $\mathbf{W_h}$ are the output and hidden layer weight matrices, $\mathbf{h}_\mathrm{p}$ is the vector denoting the response of hidden layer for pattern 'p', $\boldsymbol{\theta}_\mathbf{o}$ and $\boldsymbol{\theta}_\mathrm{h}$ are the output and hidden layer bias vectors, respectively and *f(.)* is the sigmoid activation function.

Backpropagation minimizes the sum of squared error defined as

$$\mathrm{E} = \frac{1}{2}\sum_{\mathrm{p}}(\mathbf{t}_\mathrm{p} - \mathbf{y}_\mathrm{p})^\mathrm{T}(\mathbf{t}_\mathrm{p} - \mathbf{y}_\mathrm{p}) \tag{3}$$

where $\mathbf{t}_\mathrm{p}$ is the target output vector for pattern 'p' and T denotes transpose of a matrix. Denoting the fan-in weights to a single neuron by a weight vector **w**, its update in the *t*-th epoch is governed by the following equations.

$$\Delta\mathbf{w}_t = -\eta\nabla\mathrm{E}(\mathbf{w})\big|_{\mathbf{w}=\mathbf{w}_t} + \alpha\Delta\mathbf{w}_{t-1} \tag{4}$$

The parameter η, called learning rate, controls the step size in each iteration process and α, called momentum factor, reduces the oscillation in the learning process as well as provides faster training. Learning speed in Backpropagation largely depends on the suitable choice of these parameters. In training Backpropagation for larger-scale problems, the error surface may contain long ravines with sharp curvature and gently sloping floor causing slow convergence.

Scaled conjugate gradient (SCG)

The scaled conjugate gradient technique is designed to achieve faster convergence in training in multilayer feed-forward network. In conjugate gradient methods, a search is performed along conjugate directions[21]. The new search direction is determined by combining the new steepest descent direction with the previous search direction so the current and previous search directions are conjugate. This technique is based on the assumption that the error in the neighborhood of a given point is locally quadratic. The weight changes in successive steps are given by the following equations.

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha_t\mathbf{d}_t \tag{5}$$

$$\mathbf{d}_t = -\mathbf{g}_t + \beta_t\mathbf{d}_{t-1} \tag{6}$$

with

$$\mathbf{g}_t \equiv \nabla E(\mathbf{w})\big|_{\mathbf{w}=\mathbf{w}_t} \tag{7}$$

$$\beta_t = \frac{\mathbf{g}_t^\mathrm{T}\mathbf{g}_t}{\mathbf{g}_{t-1}^\mathrm{T}\mathbf{g}_{t-1}} \quad \text{or} \quad \beta_t = \frac{\Delta\mathbf{g}_{t-1}^\mathrm{T}\mathbf{g}_t}{\mathbf{g}_{t-1}^\mathrm{T}\mathbf{d}_{t-1}} \quad \text{or} \quad \beta_t = \frac{\Delta\mathbf{g}_{t-1}^\mathrm{T}\mathbf{g}_t}{\mathbf{g}_{t-1}^\mathrm{T}\mathbf{g}_{t-1}} \tag{8}$$

where $\mathbf{d}_t$ and $\mathbf{d}_{t-1}$ are the conjugate directions in successive iterations, and $\mathbf{g}_t$ and $\mathbf{g}_{t-1}$ are the corresponding gradient directions. The step size is governed by the coefficient α_t and the search direction is determined by β_t. In scaled conjugate gradient the step size α_t is calculated by the following equations.

$$\alpha_t = -\frac{\mathbf{d}_t^\mathrm{T}\mathbf{g}_t}{\delta_t} \tag{9}$$

$$\delta_t = \mathbf{d}_t^\mathrm{T}\mathbf{H}_t\mathbf{d}_t + \lambda_t\|\mathbf{d}_t\|^2 \tag{10}$$

where λ_t is the scaling co-efficient and $\mathbf{H}_t$ is the Hessian matrix at iteration *t*. The term λ is a multiplicative factor introduced to make the Hessian matrix positive definite. With sufficiently large λ, the modified Hessian is guaranteed to be positive (δ > 0). If the error function is not quadratic or δ <0, λ can be increased to make δ >0. In case of δ <0, Moller[22] suggested the appropriate scale coefficient $\bar{\lambda}_t$ to be

$$\bar{\lambda}_t = 2\left(\lambda_t - \frac{\delta_t}{\|\mathbf{d}_t\|^2}\right) \tag{11}$$

Rescaled value $\bar{\delta}_t$ of δ_t is then expressed as

$$\bar{\delta}_t = \delta_t + (\bar{\lambda}_t - \lambda_t)\|\mathbf{d}_t\|^2 \tag{12}$$

The scaled coefficient $\bar{\delta}_t$ needs adjustment to validate the local quadratic approximation. A detailed description of scaled conjugate gradient algorithm together with how scaled coefficients can be adjusted have been given by Moller[22] and Bishop[23].

Bayesian Reguralization (BR)

In classification problems, the main target is to build a network that, once trained, is capable of recognizing not only the training data but also the test data, i.e., generalizes well on the unseen data. In order to achieve better generalization in multilayer feed-forward network training, Mackay[24] proposed a method to constrain the size of network parameters by regularization. Regularization technique forces the network to settle to a set of weights and biases having smaller values. This causes the network response to be smoother and less likely to overfit[22] and capture noise. In the regularization technique, the cost function F is defined as

$$F = \gamma E_D + (1-\gamma) E_W \tag{13}$$

where E_D equals E defined in Eq. (3), $E_w = \|\mathbf{w}\|^2/2$, and is the sum of squares of the network weights, and γ (<1.0), called regularization parameter, determines the emphasis of the training on regularization. A large γ will drive the error E_D to small value whereas a small γ will put excessive emphasis on weight size reduction at the expense of higher error. Mackey[24] proposed Bayesian framework to automatically determine the optimum regularization parameter. The Bayesian framework considers a probability distribution over the weight space assuming some initial prior distribution. Let D={$\mathbf{x}_m$, $\mathbf{t}_m$} be the data set of the input-target pair, m being a label running over the pair and M be a particular NN model. The posterior probability distribution for the weight $p(\mathbf{w}|D,\gamma,M)$ is given according to the Bayesian rule.

$$p(\mathbf{w}\,|\,\mathrm{D},\gamma,M) = \frac{p(\mathrm{D}\,|\,\mathbf{w},\gamma,M)\,p(\mathbf{w}\,|\,\gamma,\mathrm{M})}{p(\mathrm{D}\,|\,\gamma,\mathrm{M})} \tag{14}$$

where $p(\mathbf{w}|\gamma,M)$ is the prior distribution, $p(D|\mathbf{w},\gamma,M)$ is the likelihood function and $p(D|\gamma,M)$ is a normalization factor, which guarantees that the total probability is 1. In the Bayesian framework, the optimal weight should maximize the posterior probability $p(\mathbf{w}|D,\gamma,M)$, which is equivalent to minimizing the function in Eq.(13). The performance ratio parameter is optimized by applying the Bayes' rule

$$p(\gamma\,|\,\mathrm{D},M) = \frac{p(\mathrm{D}\,|\,\gamma,M)\,p(\gamma\,|\,\mathrm{M})}{p(\mathrm{D}\,|\,\mathrm{M})} \tag{15}$$

Assuming a uniform prior distribution $p(\gamma|M)$ for the regularization parameter γ, maximizing the posterior probability is achieved by maximizing the likelihood function $p(D|\gamma,M)$. Since all probabilities have a Gaussian form it can be expressed as

$$p(D\,|\,\gamma,M) = (\pi/\gamma)^{-N/2}[\pi/(1-\gamma)]^{-L/2} Z_F(\gamma) \tag{16}$$

where L is the total number of parameters in the NN. Supposing that F has a single minimum as a function of $\mathbf{w}$ at $\mathbf{w}^*$ and has the shape of a quadratic function in a small area surrounding that point, Z_F is approximated as[24],

$$Z_F \approx (2\pi)^{L/2} \det{}^{-1/2} H^* \exp(-F(\mathbf{w}^*)) \tag{17}$$

where $H=\gamma\nabla^2E_D + (1-\gamma)\nabla^2E_W$ is the Hessian matrix of the objective function. Using Eq. (17) into Eq. (16), the optimum value of γ at the minimum point can be determined. To approximate the Hessian matrix, Foresee and Hagan[25] proposed to apply Gauss-Newton approximation, which can be conveniently implemented if the Lebenberg-Marquart optimization algorithm[26] is used to locate the minimum point. A detailed discussion of Bayesian Regularization techniques is presented by Mackey[24] and Bishop[23].

In this study, gait classification between young and older adults was performed using BP, SCG and BR algorithms to find out which algorithm would be most effective in gait data training for recognizing ageing effects. Each NN model had an input layer consisting of 24 input neurons corresponding to the input characteristics (features) of the gait patterns, one hidden layer and an output layer unit representing the gait types (+1 =young, -1=elderly). All 24 features were normalized using their equivalent z-scores to have unity variance before applying them for developing the NN models and for subsequent testing of the models to assess their generalization ability.

Classifier performance testing

Cross-validation (k-fold) tests were performed to evaluate performance of the classifiers. In this experiment, all subjects' data were divided into six segments (6-fold) of equal length; five segments' data were used to train the classifier whereas the remaining data segment was used to test the accuracy of prediction. This was then repeated until all subjects' data appeared in the test sample. Experiments were conducted with varying number of hidden units, however, 3 hidden units produced the best results. In designing a neural network, an important issue is to select the appropriate number of hidden units and layers as the network size influences the generalization capability of the network. Using Kolmogorov theorem, it has been shown that a neural network with a single hidden layer can learn any non-linearly separable problem provided sufficient number of units are used in that hidden layer[27]. However, there are no widely accepted rules or analytical methods for determining the optimal number of hidden units. A network with too few hidden units will be unable to learn the input-output mapping well, whereas too many hidden units will generalize poorly on any unseen data. Most applications use a single hidden layer and the usual practice is to train networks with different number of hidden units and select the one that yields the best results. Similar practice has been adopted in the current application which produced best results with three units in the hidden layer. All the results presented here are therefore based on three hidden units. Neural network's performance also depends on the selection of initial weights – therefore, all tests were run 20 times starting with different initial random weights. Average and maximum classification accuracies of 20 trials were used to report the overall performance of the machine classifier. A statistical classifier (Linear Discriminant Analysis, LDA) was also built to classify the same data and its performance was compared with that of neural networks.

Classification outcomes were also represented using receiver operating characteristics (ROC) curves. ROC plots have been used in many investigations[28] to gauge the predictive ability of a classifier over a wide range of threshold values. The predicted output of the neural network in response to an unknown gait pattern resulted in an output in the range between -1.0 and 1.0. A threshold value was then applied such that an output above the threshold was assigned into a young category whereas a value equal to or below the threshold was assigned into an elderly category. In regard to ROC plots, the following measures[29] were used to evaluate the overall performance:

True positive (TP): A neural network identifies an older gait that was labeled as older
True negative (TN): A neural network identifies a young gait that was labeled as young
False positive (FP): A neural network identifies an older gait that was labeled as young
False negative (FN): A neural network identifies a young gait that was labeled as older

TP rate (*Sensitivity*) is defined as a measure of the ability of the classifier to identify an older gait, whereas TN rates or *Specificity* is a measure of the classifier to detect young gait characteristics. *Selectivity* measures classifier's ability to reject false detection of older gaits and *Detection_rate* is defined as an average of sensitivity & specificity. These parameters can be calculated as:

$$Sensitivity = \frac{TP}{TP+FN} x100\%$$

$$Specificity = \frac{TN}{TN+FP} x100\%$$

$$Selectivity = \frac{TP}{TP+FP} x100\%$$

$$Detection\_Rate = \frac{Sensitivity + Specificity}{2} x100\%$$

These measures were calculated for various threshold values. In addition, sensitivities at two selected specificities[28] of 0.9 and 0.75 were calculated to compare the classifiers' performance. ROC curve plots sensitivity against (1 – specificity) as the threshold level of the classifier is varied. ROC curves were plotted for all three classifiers to examine qualitatively the effect of threshold variation on the classification performance. Furthermore, ROC areas were calculated numerically to compare the three classifiers quantitatively. All neural network architectures were developed, trained and tested using routines written in Matlab toolbox 6.12 (The MathWorks, Natick, MA).

Results

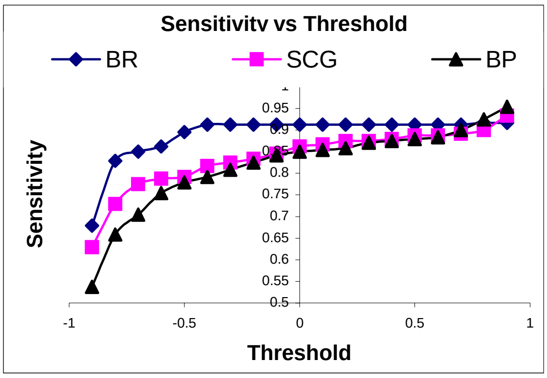
The neural networks were trained using three different algorithms (Standard BP, SCG and BR) and each algorithm was tested 20 times. The average and maximum accuracy rates are shown in Table 1. Although there were some minor differences among the three classifiers with regard to average classification accuracy, the maximum accuracy rate was the same for all the classifiers (83.3%). These classifiers results were better when compared with the accuracy results obtained with LDA-based statistical classifier (75%).

Table 1. *Correct classification rates by neural networks to differentiate young/elderly gait patterns, BP – Backpropagation, LDA – Linear Discriminant Analysis. Average accuracy results are presented from 20 trials. The standard deviation over accuracy among the trials is shown within the bracket.*

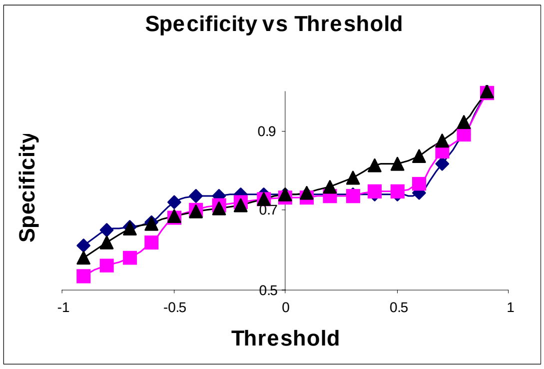
Algorithms	Average Accuracy (%)	Maximum Accuracy (%)
Standard BP	79.4 (4.16)	83.3
Scaled Conjugate Gradient	79.6 (3.55)	83.3
Baysian Reguralization	82.7 (1.52)	83.3
LDA	75.0	75.0

Figures 2(a)-(d) plot sensitivity, specificity, selectivity and average detection rates for the 3 classifiers as a function of threshold variation (-1 to +1). BR showed greater sensitivity and detection rates over a wide range of thresholds, and also consistently higher sensitivities at two selected specificities of 0.9 and 0.75 (see Table 2). ROC plots of the three classifiers (Figure 3) display graphical representation of the relationship between sensitivity and (1 – specificity) for a range of thresholds and confirm higher sensitivity results by the BR classifier particularly for higher specificities. ROC area was higher for the BR (ROC_{area}=0.9) compared to the other two classifiers (0.82 for BP and 0.84 for SCG), however all three were considerably higher than that of the LDA (0.77). Area under the ROC curve represents performance of the classifier over a range of thresholds. In general, the larger the area, the better the classification performance. Out of the three neural network classifiers, BR displayed improved classification performance when trained and tested with all the 24 gait features.

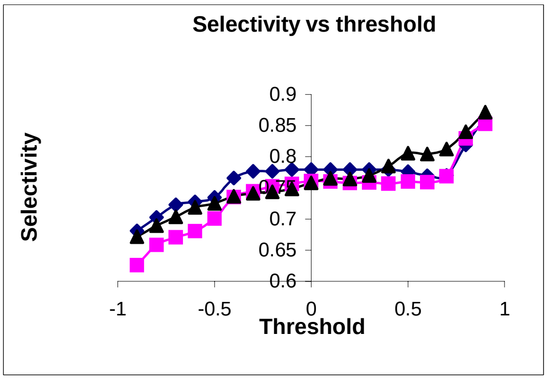
A forward feature selection algorithm revealed that the performance of the BR classifier was dependent on the number of features used to train and test the classifier (see Figure 4). In this feature selection method, a feature was



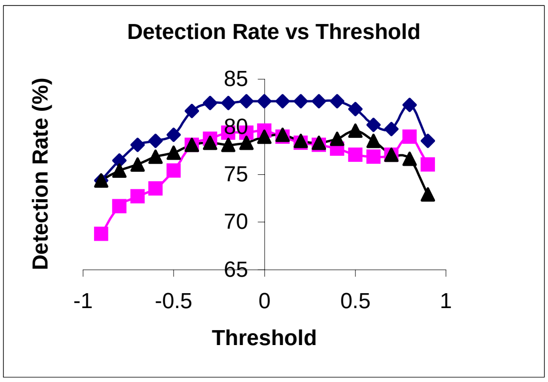
(a)



(b)



(c)



(d)

Figure 2. *Performance measures: a) Sensitivity, b) Specificity, c) Selectivity and d) Detection rate as a function of thresholds (-1 to +1). BP - Standard Backpropagation, SCG - Scaled Conjugate Gradient, BR - Bayesian Regularization*

Table 2. *Comparison of performance using ROC area and sensitivities at selected specificities.*

Classifier	BP	SCG	BR	LDA
ROC Area	0.82	0.84	0.90	0.77
Sensitivity at specificity of 0.9	0.60	0.67	0.73	0.31
& 0.75	0.75	0.79	0.86	0.67

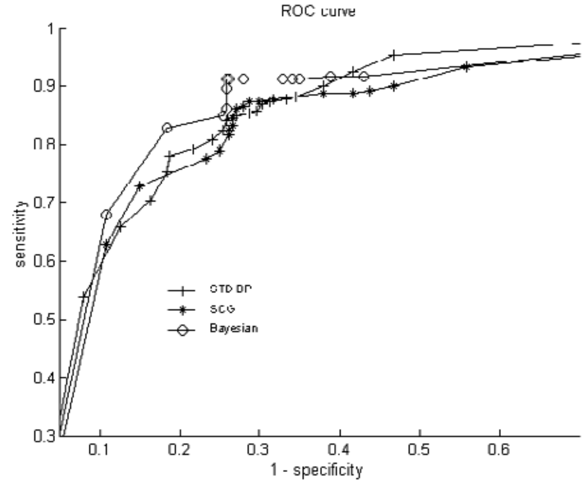


Figure 3. *ROC curves for the three classifiers (Std BP - Standard Backpropagation, SCG - Scaled conjugate, Baysian Regularization) over a range of threshold selections. Sensitivity =True positive rates, Specificity = True negative rates.*

added one at a time that provided the maximum classification accuracy[28]. It is interesting to note that the %accuracy rate was in fact higher with a few selected features compared to that obtained when trained with all 24 features. With 3 selected features the performance curve peaked showing 100% generalization accuracy. The three features that were selected were *knee range of motion* during the swing phase, *maximum horizontal (anterior-posterior) push-off force* and *step length*. However, some features were found to negatively affect the classifier in that their addition to the classifier resulted in a decrement of detection accuracy (see Figure 4). A comparison of ROC plots for the BR classifier using all the 24 features and using only the 3 selected features can be seen in Figure 5 which suggests consistently better performance (ROC_{area}=0.9 versus ROC_{area}=0.999) with a few selected features across all the thresholds.

Discussion

The major aim of this research was to test whether an artificial neural network could be applied to detect gait changes due to ageing using standard gait features that are recorded during gait analysis. The results of the cross-validation test and from the three neural network classifiers suggest that an artificial neural network is able to differentiate young/old gait with an accuracy of 83.3%

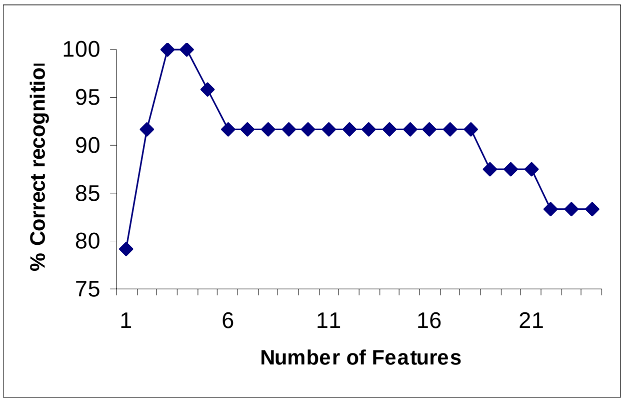


Figure 4. *Graph illustrating the reliance of accuracy rates in gait classification on the number of features selected by the feature selection algorithm. The first 3 features selected for maximum accuracy were: knee range of motion during the swing phase, maximum horizontal (anterior-posterior) push-off force and step length.*

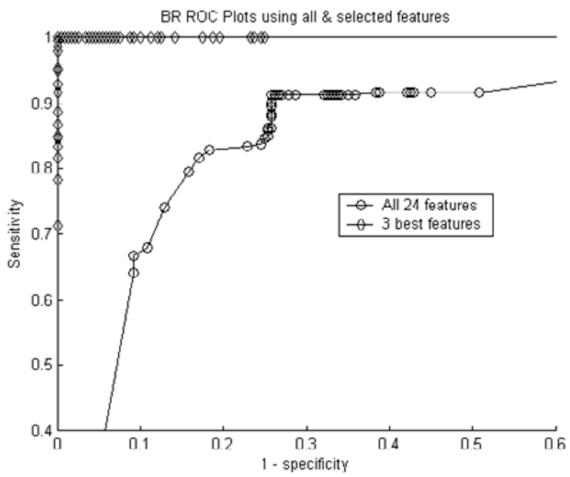


Figure 5. *ROC plots of BR classifier using all 24 features and 3 key features selected by the feature selection algorithm.*

across all the subjects. The results also suggest that although there were differences in accuracy rates in individual trials, presumably due to variation of initial weights of the networks[30], the maximum accuracy was the same with all three BP algorithms. Scaled Conjugate Gradient was found to be the fastest learner of the three. While previous research has supported neural network's ability to differentiate between normal and pathological gait from respective force platform recordings[8,9], this research suggests that a neural network can also be quite effective in detecting gait changes due to ageing. Such automatic gait classification capability has many potential benefits including, monitoring gait deterioration due to ageing, identifying at-risk or faulty gait and also in the evaluation of the effectiveness of intervention outcomes.

Receiver operating characteristics and other associated measures (e.g. selectivity and average detection rate) on the test data suggest that out of the three classifiers, a BR trained classifier would be most effective for recognizing gait changes as a result of ageing. This is particularly evident in ROC, sensitivity, selectivity and detection rate plots of the classifiers (see Fig. 2 & 3). It also appears that BR would be particularly more sensitive for recognizing elderly gait characteristics as well as in its capacity to reject false detection of elderly gait characteristics. This is due to higher sensitivity, selectivity and detection rate results displayed by the BR classifier across a wide range of threshold values (Fig. 2). Furthermore, BR's performance remained invariant over a wide threshold range (–0.4 to 0.4) whereas performance for the SCG and BP classifiers was variable with threshold changes.

Among other factors, complexity of the model is one that can affect the generalization performance of a classifier; therefore in order to achieve the best generalization it is important to optimize the complexity of the classifier[19]. Models, which are either too simple or too complex, will exhibit poor generalization performance. One approach to control the complexity of a neural network classifier is through the use of regularization technique that penalizes a highly complex model and searches for a balance between the training error and the classifier complexity. A classifier trained by Bayesian Regularization estimates the degree of regularization and hence controls the classifier complexity by optimizing the regularization parameter. The regularization technique is formulated to produce smooth network mappings by favouring small values for network weights[19]. The BP and SCG trained classifier do not incorporate such constraints in their learning rules. As a result, without any constraints in weight values, a network may settle to a set of weights with huge variation in their values. In such cases, a slight variation in gait pattern may result in a completely different classification decision by a BP and SCG trained classifier. A BR trained classifier achieves a better performance in identifying gait patterns through the use of regularization technique in the learning rule.

In this study, feature selection significantly influenced the classifier's ability in distinguishing the young and old gait patterns. As demonstrated in the ROC and %accuracy plots (Fig. 4 & 5), a small subset selected from the original features could provide improved performance compared to the entire set of input features. This also highlights the importance of selecting relevant features before applying the features to the machine classifier for the classification task. Such dependence of classification performance on features has also been highlighted in other applications, such as classification of hand movements[31].

An artificial neural network classifier, particularly with BR learning algorithm, appears to have high potential for applications in automated gait recognition from its features. In this application involving age-related gait classification, the neural network's performance was considerably higher than that offered by an LDA classifier. Such observation is also in line with previous finding involving pathological (ankle arthrodesis) gait classification[6]. Future research in this area may include: applying neural network algorithms for recognition of gait changes due to falling behavior and various pathologies. Other feature selection algorithms, such as backward elimination techniques[28] and genetic algorithms[31] may be applied for separating the relevant and irrelevant features in order to improve the effectiveness of gait detection and classification tasks.

Conclusion

In this study we explored gait pattern recognition and classification abilities using three neural network algorithms: standard Backpropagation, scaled conjugate gradient and Backpropagation with Bayesian regularization. Gait features for training and testing the neural networks were extracted from gait recordings in young and elderly subjects using standard techniques and instrumentation. All three neural network algorithms provided very good classification results, however the Bayesian regularization proved to be superior in performance measured in ROC areas/ plots and sensitivity and specificity measures. The classification performance was significantly enhanced when a subset of selected features was used as classifier inputs. These results demonstrate neural networks' potential in mapping the relationship between ageing effects and gait characteristics and provide support for its future applicability as gait diagnostics in ageing and pathological populations.

References

1. Hageman, P.A. and Blanke, D.J., *Comparison of gait of young women and elderly women*, Phys Ther. Vol. 66: 1382-7, 1986.
2. Winter, D.A., *The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological*, University of Waterloo Press, Waterloo, 1991.
3. Öberg, T., Karsznia, A. and Öberg, K., *Joint angle parameters in gait: Reference data for normal subjects,* 10-79 years of age, J Rehab Res Dev. Vol. 31: 199-213, 1994.
4. Judge, J. O., Davis, R. B., and Õunpuu, S., *Step length reductions in advanced age: The role of ankle and hip kinetics*, J Gerontol Med Sci. vol. 51: 303-12,1996.
5. Fildes, B. *Injuries among older people: Falls at home and pedestrian accidents,* Dove Publications, Melbourne, 1994.
6. Wu, W.-L., Su. F.-C., Chou, C. K., *Potential of the back propagation neural networks in the assessment of gait patterns in ankle arthrodesis,* Clin Biomech. Vol. 15: 143-5, 2000.
7. Chau, T., *A review of analytical techniques for gait data. Part 2: neural network and wavelet methods*, Gait Posture, vol. 13: 102-20, 2001.
8. Schöllhorn, W. I. *Applications of artificial neural nets in clinical biomechanics,* Clin Biomech. Vol. 19: 876-98, 2004.
9. Holzreiter, S. H. and Kohle, M. E., *Assessment of gait pattern using neural networks,* J Biomech, vol. 26: 645-51, 1993.
10. Barton, J. G., Lees, A., *An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams,* Gait Posture, vol. 5: 28-33, 1997.
11. Begg, R., Kamruzzaman, J., *A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data,* J Biomech. Vol. 8: 401-408, 2005.
12. Begg, R. K., Sparrow, W. A., Lythgo, N. D., *Time-domain analysis of foot-ground reaction forces in negotiating obstacles,* Gait Posture. Vol. 7: 99-109, 1998.
13. Nigg, B. M., Fisher, V., Ronsky, J. L., *Gait characteristics as a function of age and gender,* Gait Posture. Vol. 2: 213-20, 1994.
14. Whittle, M., *Gait Analysis: An Introduction.* Butterworth Heinemann, Oxford, 1991.
15. Fogel, D. B. and Robinson, C. J., *Computational Intelligence.* IEEE, USA, 2003.
16. Reich, Y. and Barai, S. V., *Evaluating machine learning models for engineering problems,* Artif Intell Eng. Vol. 13: 257-72, 1999.
17. White, H. and Racine, J., *Statistical inference, the bootstrap, and neural-network modelling with application to foreign exchange rates,* IEEE Trans Neural Net. Vol. 12: 657-73, 2001.
18. Sepulveda, F., Wells, D., Vaughan, C. L., *A neural network representation of electromyography and joint dynamics in human gait,* J Biomech. vol. 26: 101-9, 1993.
19. Hetrz, J., Krogh, A. and Palmer, R. G., *Introduction to the Theory of Neural Computation,* Addison-Wesley, 1991.
20. Begg, R. K., Kamruzzaman J., Sarker, R., *Movement pattern recognition using neural networks,* Chapter 10, In (Begg, Kamruzzaman & Sarker eds.) Neural Networks in Health Care: Potential and Challenges, IGI Publishing: Hershey, PA, 217-237, 2006.
21. Hagan, M. T., Demuth, H. B., Beale, M. H., *Neural Network Design.* PWS Publishing, Boston, 1996.
22. Moller, A. F., *A scaled conjugate gradient algorithm for fast supervised learning,* Neural Net. Vol. 6: 525-33, 1993.
23. Bishop, C. M., *Neural Network for Pattern Recognition.* Oxford University Press, New York, 1995.
24. Mackay, D. J. C., *A practical Bayesian framework for backpropagation networks,* Neural Comp. Vol. 4: 415-47, 1992.
25. Foresee, F. D., and Hagan, M. T., *Gauss-Newton approximation to Bayesian learning*. Proc. IEEE ICNN, vol. 3: 1930-5, 1997.
26. More, J. J., *The Levenberg-Marquart algorithm: implementation and theory, in: G.A. Watson, ed., Numerical analysis. Lecture Notes in Mathematics 630.* Springer-Verlag: London, 1977.
27. Cybenko, G., *Approximation by superpositions of a sigmoidal function,* Mathematical Control Signal Systems, Vol. 2: 303-314, 1989.
28. Chan, K., Lee, T. W., Sample, P. A., Goldbaum, M. H., Weinreb, R. N., and Sejnowski, T. J., *Comparison of Machine Learning and Traditional Classifiers in Glaucoma Diagnosis,* IEEE Trans Biomed Eng. Vol. 49: 963-74, 2002.
29. Pang, C. C. C., Upton, A. R. M., Shine, G., Kamath, M. V., *A comparison of algorithms for detection of spikes in the electroencephalogram,* IEEE Trans Biomed Eng. Vol. 50: 521-6, 2003.
30. Kecman, V., *Learning and soft computing: support vector machines, neural networks and fuzzy logic models.* IEEE, New Jersey, 2002.
31. Yom-Tov, E., and Inbar, G. F., *Feature selection for the classification of movements from single movement-related potentials,* IEEE Trans Neural Sys. Vol. 10: 170-7, 2002.