SVM MODELS FOR DIAGNOSING BALANCE PROBLEMS USING STATISTICAL FEATURES OF THE MTC SIGNAL

DANIEL T. H. LAI*, REZAUL BEGG† and MARIMUTHU PALANISWAMI*

*Department of Electrical Engineering
The University of Melbourne, Parkville Campus
Melbourne, Victoria 3010, Australia

†Centre for Ageing, Rehabilitation
Exercise and Sport, Victoria University
Melbourne, Victoria 8001, Australia

Received 23 November 2007

Trip-related falls are a major problem in the elderly population and research in the area has received much attention recently. The focus has been on devising ways of identifying individuals at risk of sustaining such falls. The main aim of this work is to explore the effectiveness of models based on Support Vector Machines (SVMs) for the automated recognition of gait patterns that exhibit falling behavior. Minimum toe clearance (MTC) during continuous walking on a treadmill was recorded on 10 healthy elderly and 10 elderly with balance problems and with a history of tripping falls. Statistical features obtained from MTC histograms were used as inputs to the SVM model to classify between the healthy and balance-impaired subjects. The leave-one-out technique was utilized for training the SVM model in order to find the optimal model parameters. Tests were conducted with various kernels (linear, Gaussian and polynomial) and with a change in the regularization parameter, \( C \), in an effort to identify the optimum model for this gait data. The receiver operating characteristic (ROC) plots of sensitivity and specificity were further used to evaluate the diagnostic performance of the model. The maximum accuracy was found to be 90% using a Gaussian kernel with \( \sigma^2 = 10 \) and the maximum ROC area 0.98 (80% sensitivity and 100% specificity), when all statistical features were used by the SVM models to diagnose gait patterns of healthy and balance-impaired individuals. This accuracy was further improved by using a feature selection method in order to reduce the effect of redundant features. It was found that two features (standard deviation and maximum value) were adequate to give an improved accuracy of 95% (90% sensitivity and 100% specificity) using a polynomial kernel of degree 2. These preliminary results are encouraging and could be useful not only for diagnostic applications but also for evaluating improvements in gait function in the clinical/rehabilitation contexts.

Keywords: Trip falls; support vector machines; receiver operating characteristics (ROC); minimum toe clearance (MTC) signal; statistical features.

1. Introduction

Gait analysis involves recording and analysis of human walking patterns. This analysis is frequently undertaken in clinics and rehabilitation centers to gauge the extent
of abnormality in the lower limbs and also to evaluate overall walking capability. Besides disease, gait patterns also have been shown to change with age. In recent years, several studies have been undertaken to investigate relationships between aging and the changes that occur in the various gait patterns. One issue of concern is the decline in gait performance due to aging, which subsequently can lead to tripping and falling seen in the elderly population. Gait deviations in older adults have been reported in many gait measures\textsuperscript{1,2} where a serious consequence of this is falling. Falls in the older population have been identified as a major public health issue in Australia with an estimated 2.4 A$\text{billion per annum}\textsuperscript{3} in injury costs. It is evident that further work needs to be done in order to reduce the number of such incidents by early detection of gait degeneration and studying the effects of fall history on gait patterns in older individuals.

Minimum toe clearance (MTC) during walking (see Fig. 1), which occurs during the mid-swing phase of the gait cycle, is defined as the minimum vertical distance between the lowest point on the shoe and the ground. This has been regarded as an important gait parameter as it can indicate the potential for successful negotiation of the environment during walking. MTC characteristics during walking have been associated with tripping and trip-related falls in the older population.\textsuperscript{4} Trips that cause falls are of more concern as they indicate a failure of the body’s locomotion system to successfully recover from a sudden imbalance. A model based on MTC information could be utilized to diagnose potential falls-prone individuals for gait rehabilitation. Furthermore, the MTC represents a single gait parameter, which is directly obtainable from foot trajectory data as opposed to the more complicated methods required for other variables such as joint angles\textsuperscript{5–7} and hip centers.\textsuperscript{8}

One possible approach to early detection of elderly at risk of tripping is to develop an automated gait detection model using Support Vector Machines (SVM).
The SVM is a supervised learning technique, which can be employed to model the risk of tripping by learning the different classes, i.e. falls-prone and healthy individuals with no-falls history using MTC features. In contrast, most previous works in gait recognition have applied neural networks to classify gait types including normal and pathological gait using force platform measures, simulated gait types (e.g. leg length discrepancy), and joint-angle measures. We have used SVMs because they have been proven to be powerful classifiers with good generalization even when the training data set is not large. In addition, SVMs provide a unique classifier for each set of SVM parameters and can be quickly trained compared with neural networks, which are heavily influenced by the initial conditions of the training algorithm. We have found, for example, that MTC gait kinematics used in conjunction with SVMs can successfully discriminate young and healthy older adults.

In this work, we apply SVMs for automated recognition of elderly subjects at risk of falling using statistical features extracted from MTC gait data. Performance of the classifier was evaluated using accuracy rates and measures of receiver operating characteristic (ROC) curves using the leave-one-out procedure. It was hypothesized that an SVM model would be suitable for constructing relationship between MTC gait features and the respective gait types, i.e. healthy older adults and older adults with impaired-balance movement patterns.

2. Experimental Setup

The data for analysis were obtained from the Victoria University Biomechanics database. This particular data set contained 10 healthy older adults (H) and 10 older adults with balance impairments (I). All the participants were >65 years old and were volunteers from the local community and senior citizen clubs. The experimental setup for gait data collection is as reported by Begg et al. In brief, the toe clearance data were collected during steady-state self-selected walking on a treadmill using the PEAK MOTUS 2D (Vicon, USA) motion analysis system. A 50 Hz Panasonic F15 video camera, with a shutter speed of 1/1000s, was positioned 9 m from the treadmill, perpendicular to the plane of foot motion to record unobstructed treadmill walking. Two reflective markers were attached to each subject’s left shoe at the fifth metatarsal head (MH) and the great toe (TM). Each subject completed between 10 and 20 min of normal walking at a self-selected comfortable walking speed. The marker positions and shoe dimensions were used to predict the position of the shoe/foot endpoint, i.e. the position on the shoe travelling closest to the ground at the time when MTC occurs using a 2D geometric model of the foot. The MTC was then calculated by subtracting ground reference from the minimum vertical coordinate during the swing phase (see Fig. 1).

3. SVM Gait Model

In this study, we hypothesize that information contained in a series of MTC samples corresponding to a number of gait cycles can be employed to detect an elderly at
risk of falling. The MTC samples are treated as a time series or discrete signal where time is representative of the gait cycle number, e.g. \( t = 10 \) refers to the 10th gait cycle from the beginning of data collection. Mathematically, we can define the MTC signal as follows:

\[
MTC_n(t) = \{x_1, x_2, \ldots, x_n\},
\]

where \( n \) is the length of the MTC signal or the number of samples and \( x_i \) is the MTC value for the \( i \)th gait cycle. We can extract further information from histogram plots (see next section) of the MTC samples for each subject (see, for example, Fig. 2). The histograms provide an estimate of the probability distribution, which generated the MTC samples under the assumption that (1) is stationary. The objective is to obtain a gait model, which takes an MTC signal as input and determines if the person is at risk of tripping and falling. Since the exact function or model relating MTC variation to tripping falls is unknown, we have elected to use an SVM classifier as a function estimation model. SVMs are very useful to realize an input–output mapping by learning the mapping function from a set of inputs and outputs. The block diagram of the SVM-based diagnostic model is shown in Fig. 3 and can be regarded as consisting of two distinct stages: feature extraction using MTC data and construction of the SVM model to map relationships between MTC features and falls/no-falls category.

Fig. 2. Sample MTC histogram plots for healthy elderly (left) showing a more uniform distribution compared with the MTC distribution of an elderly with a history of tripping falls (right).

Fig. 3. Block diagram of the SVM diagnostic model (H — healthy, I — impaired balance).
3.1. Statistical feature extraction

We extracted nine statistical features from each MTC histogram, namely the values of the first quartile (Q1), second quartile or median (Q2), third quartile (Q3), maximum MTC value (MX), minimum MTC value (MN), mean (M), standard deviation (SD), skewness (S), and kurtosis (K). The last four features are extracted under the assumption that the MTC samples are drawn from some unknown discrete distribution $P(x)$ (not necessarily Gaussian). These features are known as the moments of the distribution defined as

$$
\mu_k(0) = \sum_{i=1}^{n} x_i^k P(x_i).
$$

The first moment $\mu_1$ or $\mu$ is the mean and the second moment $\mu_2$ is the variance of the distribution of which we use its square root value or the standard deviation. The third moment is the skewness of the distribution, which measures the degree of asymmetry in the distribution, while the fourth-order moment is kurtosis, which measures the peakness in the distribution.

3.2. SVM fundamentals

The SVM classifier first introduced by Vapnik$^{12,13,17}$ is now a well-known machine learning tool employing kernel methods to implicitly map data to a usually higher-dimensional space. Classification is then performed in this space (usually referred to as feature space) through the construction of a linear separating hyperplane (see Fig. 4). The optimal separating hyperplane is selected from the solution of the

![Fig. 4. The soft-margin SVM classifier with slack variables $\xi$ and support vectors shown.](image-url)
quadratic SVM optimization problem formed from the following data set:

$$\Theta = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\},$$

$$x_i \in \mathbb{R}^k,$$

$$y_i \in \{-1, 1\},$$

(3)

where $x_i$ is a vector of features and the scalar $y_i$ is the corresponding class label. The separating hyperplane has the following universal approximator form:

$$f(x) = \sum_{i=1}^{m} w_i \phi_i(x) + b,$$

(4)

where $w_i$ are the weights of the hyperplane and the scalar $b$ is the hyperplane bias. The function $\phi(x)$ defines a nonlinear mapping from input space to some higher-dimensional feature space, denoted by

$$\phi(x) : x \subset \mathbb{R}^k \rightarrow \mathbb{R}^m, \quad k \ll m.$$

The main idea behind the SVM is to maximize the margin of the hyperplane in order to obtain good classifier performance. The optimal separating hyperplane is found by solving the following optimization problem:

$$\min_{w, \xi} \frac{1}{2} \sum_{i=1}^{m} w_i^2 + C \sum_{i=1}^{n} \xi_i$$

subject to

$$y_i \left( \sum_{i=1}^{m} w_i \phi_i(x) + b \right) \geq 1 + \xi_i,$$

$$\xi_i > 0, \quad \forall i = 1, \ldots, n.$$

(5)

The parameter $C$ controls the tradeoff between generalization capability and the number of training misclassifications while $\xi_i$ is known as a slack variable, i.e. the amount of error we are willing to tolerate in classifying a point $x_i$. Using Lagrangian theory and noting the constraints in (5), the SVM Lagrangian primal problem is

$$L_p(w, \alpha, \xi) = \frac{1}{2} \sum_{i=1}^{m} w_i^2 + C \sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \pi_i \xi_i$$

$$- \sum_{i=1}^{n} \alpha_i \left( y_i \left( \sum_{j=1}^{m} w_j \phi_j(x_i) + b \right) - 1 - \xi_i \right),$$

where $\forall i = 1, \ldots, n$, the Lagrangian multipliers are

$$\alpha_i, \pi_i \geq 0.$$

The dual representation is found by differentiating the primal with respect to the variables and finding the stationary conditions. Using the stationary conditions of
the partial gradients, i.e. setting them to zero, we then eliminate the variables by substitution and solve the Wolfe dual:\(^a\)

\[
\begin{aligned}
\mathcal{L}(\alpha) &= -\sum_{i=1}^{n} \alpha_i + \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \\
\end{aligned}
\]

where

\[
\mathbb{F} = \{ \alpha | 0 \leq \alpha_i \leq C, \alpha^T y = 0 \}.
\]

The trained classifier (machine) then has the following form:

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b \right).
\]

The explicit definition of the nonlinear function \(\phi(.)\) has been circumvented by the use of a kernel function, defined formally as the dot products of the nonlinear functions, i.e.

\[
K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle.
\]

This method is usually attributed to Mercer’s theorem,\(^{18}\) which implicitly computes the inner product of the vectors without the need to explicitly define the mapping of inputs to higher-dimensional space since they could be infinite dimensioned.

4. Experimental Methodology

We first train the SVM using the leave-one-out procedure, which is the recommended cross-validation test for small to medium data sets in order to determine the region of optimal SVM parameters. In this procedure, the data set was divided into 20 subsets consisting of 19 training examples and a single test example. First, a subset was used to train the SVM model, while the remaining data example was used for testing. The process was repeated for the other subsets so that in the end each example had been tested on an SVM model trained using the other data. The accuracy of each test was then combined to obtain an average accuracy reading known as the leave-one-out accuracy. We also report the average sensitivity and specificity of the SVM model, where

\[
\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%,
\]

\[
\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%.
\]

Here TP is the number of true positives, i.e. the elderly who have been correctly identified by the SVM as healthy; TN is the number of true negatives, i.e. the elderly.

\(^a\)The Lagrangian dual is a maximization problem and the Wolfe dual is found by taking the negative to give a minimization problem.
D. T. H. Lai, R. Begg and M. Palaniswami

who have been correctly identified by the SVM as a faller; FP is the number of false positives or the number of balance impaired elderly which the SVM has misclassified as healthy and FN where the healthy has been misclassified as fallers. A primary goal when designing our SVM detection system was to reduce the number of false positives since it is dangerous to mislabel the balance impaired elderly as healthy and deny or delay treatment for their gait problems, i.e. want a system with high specificity.

Performance of the classifier model was also represented using ROC curves. ROC plots have been used in many investigations, e.g. Chan et al.\textsuperscript{19} to gauge the predictive ability of a classifier over a wide range of threshold values. The predicted output of the SVM in response to an unknown gait pattern is used to determine the class of that pattern. A threshold value was then applied such that an output below the threshold was assigned into a healthy category whereas a value equal to or above the threshold was assigned into balance-impaired category. ROC curve plots sensitivity against (1 — specificity) as the threshold level of the classifier is varied and is useful for qualitatively examining the effect of threshold variation on the classification performance. In general, the larger the area under the ROC curve, the better the classifier model or the more likely we can pick a threshold, which gives the best classification accuracy.

Experiments were conducted over various $C$ values and also with three kernel types, i.e. linear, polynomial, and Gaussian (or radial basis functions (RBF)), having the following forms, respectively:

\begin{align*}
K(x, y)_{\text{linear}} &= x^T y; \\
K(x, y)_{\text{poly}} &= (x^T y + 1)^d; \\
K(x, y)_{\text{RBF}} &= e^{-\frac{\|x - y\|^2}{2\sigma^2}}.
\end{align*}

It should be noted that the linear kernel is a special case of polynomial kernel with degree, $d = 1$. In this work, the SVM model was trained with examples containing the nine statistical features, which served as inputs features for discriminating the balance impaired from the healthy. The output of the classifier was used to represent the gait types ($+1 = \text{healthy}$, $-1 = \text{balance-impaired}$). All SVM architectures were trained and tested on the Matlab D2CSVM software.\textsuperscript{b}

5. Experimental Results and Discussion

5.1. Full feature set

The leave-one-out accuracy results relating to gait recognition of healthy elderly and elderly with balance problems are presented in Table 1. We have included the results for the three different kernel functions over a selected variation of the regularization

\textsuperscript{b}http://www.ee.unimelb.edu.au/people/dlai/.
Table 1. Accuracy, sensitivity, specificity, and ROC area results over a range of SVM parameters using the leave-one-out procedure.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Parameters</th>
<th>$C$</th>
<th>Acc (%)</th>
<th>Sens (%)</th>
<th>Spec (%)</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td></td>
<td>1</td>
<td>75</td>
<td>60</td>
<td>90</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.94</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\sigma^2 = 1$</td>
<td>1</td>
<td>75</td>
<td>70</td>
<td>80</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 10$</td>
<td>1</td>
<td>75</td>
<td>60</td>
<td>90</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>70</td>
<td>60</td>
<td>80</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>100</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 100$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>75</td>
<td>60</td>
<td>90</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>75</td>
<td>60</td>
<td>90</td>
<td>0.76</td>
</tr>
<tr>
<td>Polynomial</td>
<td>$d = 2$</td>
<td>1</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>$d = 3$</td>
<td>1</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>$d = 4$</td>
<td>1</td>
<td>45</td>
<td>10</td>
<td>80</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>45</td>
<td>10</td>
<td>80</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>45</td>
<td>10</td>
<td>80</td>
<td>0.40</td>
</tr>
</tbody>
</table>

parameter $C$ and kernel parameters, e.g. width of the Gaussian kernel, $\sigma^2$ and order of polynomial, $d$. Note that in all cases, the full feature set containing nine statistical features has been used. High values of $C$ mean that we are penalizing the training errors more which tends to result in biasing the classifier to the data (overfitting). Conversely, lower values of $C$ result in underfitting of the model, which may be desirable for better generalization for highly nonseparable data. In view of this, we have elected to use the range $1 \leq C \leq 100$, which was empirically found to give acceptable results. It can be seen from Table 1 that the Gaussian kernel with $\sigma^2 = 10$ gives the highest accuracy of 90% when used with $C = 100$. Furthermore, this SVM model attains 100% specificity, i.e. it correctly identifies all the elderly people with balance problems and gives an ROC area of 0.98. However, the performance of the Gaussian kernel varies significantly (by up to 15%) when different values of $C$ and $\sigma^2$ are used. On average, these accuracies were within the range 70–85% across the different kernel widths. Furthermore, larger kernel widths, i.e. $\sigma^2 = 100$, gave poorer results with it predicting every test example wrongly for the case $C = 1$. Figure 5 depicts the performance for various SVM models based on the Gaussian kernel with $\sigma^2 = 10$ within the range of $C = 1,10,100$. The increasing ROC areas demonstrate the sensitivity of the Gaussian kernel to variation of the $C$ parameter.
This is undesirable since one will be tempted to select a large $C$ in order to obtain an accurate SVM model thereby inadvertently overfitting the model to the training data.

In contrast, we observe instead that the linear and polynomial kernel tend to give more stable accuracies and specificities (85% accuracy and ROC areas of 0.92 on average) over the range of $C$. The insensitivity to the $C$ parameter is demonstrated further in Fig. 5 where the polynomial kernel of degree 3 was found to be highly insensitive to $C$ parameter variation. The linear kernel initially shows large variation in accuracy for small values of $C$ but the accuracy saturates to 85% for $C > 50$. These results indicate that the family of polynomial kernels have achieved maximum discrimination using the full nine statistical features within the range of $C$.

We also found that the SVM was more successful in classifying fallers as indicated by the higher specificity values. This trend was observed to persist across the types of kernels and $C$ parameters. One possible reason for this is that the MTC data from a particular healthy elderly may closely resemble that of a faller indicating possible risk of tripping and falling in the future. However, further investigation using a larger sample size may be required to verify this.

### 5.2. Feature selection

It has been shown that the full set of features may not be required for achieving maximum classification accuracy. In fact, in many reported applications, the presence of redundant features in the training set causes a degradation in model performance. Several techniques exist in order to select a good set of features, such as hill climbing, principal component analysis (PCA), Sheffe’s post hoc test, fuzzy mutual information, and so on. We first examine the linear separability of each individual feature by computing their ROC values using the feature
Table 2. ROC values for single features where the thresholding is computed across the values of each feature.

<table>
<thead>
<tr>
<th>Feature No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature description</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>M</td>
<td>SD</td>
<td>MN</td>
<td>MX</td>
<td>S</td>
<td>K</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.79</td>
<td>0.80</td>
<td>0.82</td>
<td>0.80</td>
<td>0.92</td>
<td>0.59</td>
<td>0.73</td>
<td>0.56</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Higher values show better linear separability of the two classes using the respective feature alone.

values themselves as thresholds (see Table 2). It was found that among the nine features, the standard deviation (SD) had the highest linear separability (0.92) while distribution skewness (S) had the lowest (0.56).

We next used the hill climbing procedure described in Ref. 15 to obtain the feature set, which gave the highest leave-one-out accuracy. In this method, we began with the most separable feature and trained an SVM model to classify the balance impaired elderly. This feature was then combined with the remaining features one at a time and the two features that gave the highest accuracy was retained. The combinatorial procedure of training and testing was then repeated, this time starting with the two best features and selecting in turn the remaining features. The procedure continues as such until all nine statistical features are used to represent the training set.

For $C = 100$, it was found on average that 2–3 features were adequate to give significantly improved accuracies across the linear and polynomial kernels. We found that when using a linear kernel, we could achieve a maximum accuracy of 90% using just Feature 5 (SD) and Feature 7 (MX). It was also found that if training vectors with Feature 5 (SD) and Feature 6 (MN) were used, the polynomial kernel ($d = 2$) gave an improved accuracy of 95%. The same feature set interestingly gave 90% accuracy when used with a higher polynomial kernel ($d = 4$). Comparatively, this was a significant improvement over the 45% accuracy achieved previously with this kernel when using all nine statistical features (see Table 1). However, no significant improvement was observed for a Gaussian kernel when using feature selection with accuracies staying in the range of 75–80% except for the case where 90% was achieved using the full feature set and a width of $\sigma^2 = 10$. It was also observed that features such as Feature 8 (S) and Feature 9 (K) often reduced the classifier accuracy during the hill climbing procedure. This can be explained by noting that these two features had the lowest separability, i.e. ROC areas of 0.56 and 0.61, respectively, which forces the classifier to find a more curved decision surface thereby reducing the generalization capability of the SVM. Note that an ROC area of 0.5 means that a linear hyperplane would make a classification mistake half the time and is therefore not very useful.

This effect can be better observed by graphically plotting the SVM decision surfaces for two features. Figure 6 depicts four decision surfaces plotted using different kernels and feature sets. The top row shows the decision surfaces for the linear kernel and polynomial kernel with leave-one-out accuracies of 90% as before. It can be
seen that the decision surfaces are simple, which allows for better generalization as depicted by the accuracies. The bottom row depicts more complex decision surfaces, which would potentially misclassify unseen data. This is indeed the case for the two models where features with lower separability have been intentionally included. It was found that the Gaussian kernel achieved only 80% using Features 3 and 7 while the polynomial kernel with Features 1 and 9 achieved only 70% accuracy. In summary, features with poor separability can degrade the generalization performance of the classifier by forcing it to find a more curved decision surface, which indirectly overfits the SVM model to the data.

5.3. Further discussion

In this study, the MTC variable was used as a single gait variable together with the SVM to classify the elderly population at risk of falling from the healthy elderly. Statistical information was then extracted from the MTC histograms and used as
input features to the classifier. This is in contrast to previous studies, which have favored the use of kinetic and kinematic contemporary variables such as measurements of joint angles and walking speed. The MTC variable is a critical measurement in the gait cycle because low MTC values have been linked to the potential of tripping when walking. In fact, the MTC has also been treated as a constraint in a gait model\textsuperscript{16} for walking where tripping was said to occur if the MTC value becomes negative.

The presence of higher-order moments such as skewness and kurtosis for MTC data demonstrates that the MTC data are not normally distributed. It was found that the MTC of all 20 subjects were positively skewed and the non-normal characteristic was further confirmed by the large ROC values for their standard deviation, i.e. large variations in standard deviations. This trend has also been recently found in studies on MTC data for young and elderly subjects.\textsuperscript{4} It was used to conclude that statistical features such as standard deviation and mean would not be effective features. However, in this work, we have shown that the variations in standard deviation could be effectively used by the SVM to classify the healthy from the balance impaired elderly. The success of the SVM is mainly due to the highly nonlinear separation boundaries induced by the kernels, which work better if there is a greater feature variation between the two classes. Hence, even though direct comparison of the standard deviation may not be encouraging for discriminating gait as suggested by Begg et al.,\textsuperscript{4} we show here that it can be successfully used with a nonlinear SVM classifier to detect elderly fallers. It should be noted that while the SVM provides high detection accuracies, its performance is largely affected by parameter selection. Proper selection of parameters such as the leave-one-out cross-validation method used is required to ensure optimal SVM performance.

The proposed model in this work could be applied to detection of other disorders by statistically modeling a time series of measurements, for example, by calculating statistical quantities from measurements of ST-T deflections from a series of QRS complexes and using an SVM to detect possible cardiovascular problems. In gait studies, this method could be applied in conjunction with several scoring methods such as posterior probabilities or mutual information to obtain a risk of falling. The benefit of such a future study is a model, which can be used as a falls risk prevention diagnostic tool and also as a measure of recovery for any gait correction procedures performed on the elderly individual. This will greatly assist the elderly group through a reduction in injury costs and easy monitoring of gait rehabilitation.

6. Conclusion

We have addressed the problem of modeling the relationship between gait patterns and elderly fallers and nonfallers using a single gait variable known as minimum foot clearance (MTC). By treating a series of MTC measurements as a discrete signal, we have extracted nine statistical quantities and used them as inputs for training a SVM classifier. It was found that an SVM trained using a Gaussian kernel
using all features gave a high accuracy of 90%. Further investigations showed that the selection of gait features to be incorporated could influence the performance of the classifier. The accuracy was further improved using a hill climbing feature selection method where it was found that two features alone were adequate to give 95% accuracy using a polynomial kernel. The results were encouraging because it demonstrated that a nonlinear model could capture the complex relationship using characteristics from a single gait variable. In this study, we could only use 20 subjects’ gait features for developing the classification model, due to the time-intensive data collection procedure as reported in Sec. 2. However, we plan to conduct further tests with a larger sample size to investigate and validate our proposed SVM classification model.

Acknowledgment

The authors are grateful to Simon Taylor for recording the gait data for the Victoria University Biomechanics database.

References


