Feature Extraction of Autism Gait Data Using Principal Component Analysis and Linear Discriminant Analysis

Suryani Ilias[1,2], Nooritawati Md Tahir[1] and Rozita Jailani[1]
[1]Faculty of Electrical Engineering
Universiti Teknologi MARA (UiTM)
40450 Shah Alam, Selangor, Malaysia
[2]Politeknik Premier Sultan Salahuddin Abdul Aziz Shah (PSA)
Persiaran Usahawan, Seksyen U1, Shah Alam, Selangor, Malaysia
Corresponding Author: nooritawati@ieee.org

Abstract — In this research, the application of machine learning approach specifically support vector machine along with principal component analysis and linear discriminant analysis as feature extractions are evaluated and validated in discriminating gait features between normal subjects and autism children. Gait features of 32 normal and 12 autism children were recorded and analyzed using VICON motion analysis system and a force platform during normal walking. Here, twenty one gait features describing the three types of gait characteristics namely basic, kinetic and kinematic in these children are extracted. Further, with these gait features as input during classification, the ability of SVM as classifier are investigated using three different kernel functions specifically linear, polynomial, and radial basis. Results showed that LDA as feature extraction is the highest accuracy with kinematic parameters as gait features along with polynomial function as kernel for the SVM classifier. This finding proven that LDA is suitable as feature extraction and SVM is indeed apt as gait classifier in classifying the gait pattern autism and normal children.

Keywords: Gait Analysis, Classification, Support Vector Machine, Principal Component Analysis, Linear Discriminant Analysis

I. INTRODUCTION

Gait analysis and classification is vital in addressing the issue of gait patterns, for instance in assisting initial diagnosing [1,2] as well as preliminary clinical decision making [3]. As we know, gait classification is conducted based on a set of defined gait features such as temporal, spatial [4], joint kinematics [5] and kinetics [6] specifically by analyzing and identifying deviations that occurs if compared with the normal gait pattern. Conversely, numerous studies have been conducted for pathological gait too. For instance, Manap H. et al. [12] reported that the most significant gait features to classify gait for normal and PD patients are basic spatiotemporal.

Conversely, research on children's gait analysis is required since early diagnosis is indeed vital to confirm if movement impairment existed [7], abnormal motor development [10] as well as gait deviations in children with autism [4,6,7,8]. In addition, from previous researches, it was found that gait due to autism contributed to altered gait pattern such as velocity [9], cadence and walking speed [6] as well as kinematics such as hip flexor moments and angles [8]. Furthermore, research findings also reported that significant parameter gait pattern in autistic children can be categorized as a sign of this syndrome [7] and may be useful for therapies development [6, 10].

On the other hand gait analysis and classification using machine classifiers offers advantages for more potential applications in biomedical signal and image processing. For instance, support vector machine (SVM) is used in automated identification of gait pathologies as well as for potential interventions [16,27]. As we know, SVM is a generalized linear classifier that is suitable for small sample, non-linear and high-dimension data [21]. SVM with RBF kernel showed promising results for classification between normal and Parkinson disease (PD) gait pattern [12]. Additionally, SVM has demonstrated that it is able to identify normal and pathological gait patterns of young and elderly gait patterns [21]. Besides that, with small number of selected features and appropriate kernel, SVM is able to identify normal and pathological gait patterns too [16]. As reported in [17], with feature selection approach, SVM could classify gait patterns with 91.7% accuracy using basic, kinetic and kinematic gait data [17].

On the other hand, to the best of our knowledge, there has been minimal discussion about application of machine learning in classifying different gait parameters of children with autism. Hence, in this paper, SVM will be evaluated as classifier for discriminating between normal and autism children using gait features from the three categories namely temporal spatial, kinetic and kinematic as input features. In addition, two dimensionality reduction methods are used too specifically Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

978-1-5090-0925-1/16/$31.00 ©2016 IEEE

II. MATERIALS AND METHOD

A. Subjects

In total, 44 subjects participated in this study. As in Table 1, the control group comprised of 32 healthy normal children with mean age of 9.46 ± 1.79 years and height 127.07 ± 12.37 cm, whilst the other group consist of 12 autism children with mean age of 9.85 ± 2.5 years and height 135.22 ± 14.71. All parents or guardians for each subject are informed the related procedures and study protocols as approved by the university Research Ethics Board prior to signing the consent form. Additionally, all subjects are ensured that there are no known injuries or abnormalities that may affect their gait.

Table 1: Characteristics for each group

Parameter	*Normal (32 subjects)*		*Autism (12 subjects)*	
	Mean	SD	Mean	SD
Age(years)	9.46	1.79	9.85	2.55
Height(cm)	127.07	12.37	135.22	14.71
Weight(kg)	28.91	10.85	36.29	14.07
Gender	14 male & 18 female		8 male & 4 female	

B. Gait data acquisition

Firstly, motion capture and force plate data were collected using eight camera, 100Hz Vicon MX motion capture system and two AMTI force plates. All marker trajectories were filtered using Vicon Workstation's Woltring filter. The overview of experimental procedure using Vicon Motion Capture System is as shown in Figure 1. During experimental, subjects were required to walk along the walkway for each individual trial until they understood the instructions by walking along the walkway. Next, each participant performed 10 trials for them to be familiar with the walkway and experimental setting using their normal gait. The selected gait features were calculated using the mean of three trials. Twenty one gait features describing the basic, kinetic and kinematic aspects of gait data are as listed in Table 2. Ground reaction forces (GRF) lateral F_x, horizontal F_y and vertical F_z were recorded using two force sensing platforms (AMTI, USA) However, in this study only horizontal and vertical GRF are considered as the kinetic features to the classifier [22]. Movement of the walking was recorded using a motion analysis system via 16 reflective markers adhered to the subjects' skin at lower limb joints and segments (hip, knee, and ankle). An infrared camera will be traced this marker for kinematic analysis purpose. Three-dimensional image of lower limb tracks by motion capture cameras shown in Figure 2.

Furthermore, in this study, all gait features are normalized using similar normalization as mentioned by A. Carriero et al. [23] prior to classification. In order to eliminate the influence of subjects' body weight to GRF reading which can be different for each person regardless their health status, GRF values or kinetics data are normalized to body weight[12].

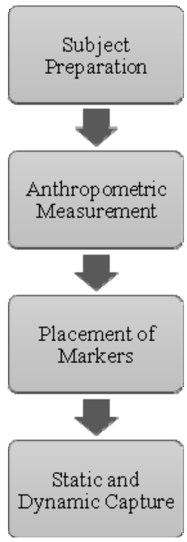

Figure 1: Overview of experimental procedure using Vicon Motion Capture System

Table 2: Gait Parameters Extracted During Experimental Analysis

Basic Temporal Spatial Features (4)
- Stride time s)
- Cadence (steps/min)
- Step length (m)
- Walking speed(m/s)

Kinetic Features (5)
- Maximum vertical loading response (F_{z1})
- Maximum vertical mid stance (F_{z2})
- Maximum vertical terminal stance (F_{z3})
- Maximum horizontal mid stance (F_{y1})
- Maximum horizontal terminal stance (F_{y2})

Kinematic Features (12)
- Hip angle at heel strike
- Hip angle at toe off
- Maximum hip flexion
- Maximum hip extension
- Knee angle at heel strike
- Knee angle at toe off
- Maximum knee flexion
- Maximum knee extension
- Ankle angle at heel strike
- Ankle angle at toe off
- Maximum ankle plantarflexion
- Maximum ankle dorsiflexion

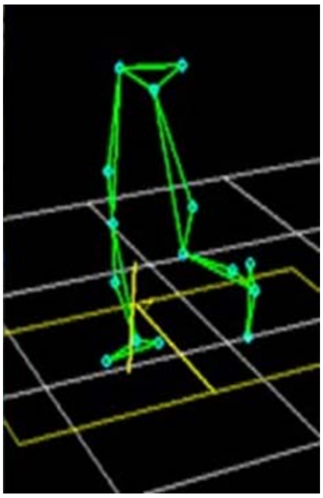

Figure 2: Three-dimensional image of lower limb tracks by motion capture cameras

C. Feature Extraction

As stated earlier, Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA) are two techniques applied as feature extraction. Then new data after reduction of dimension and selection the important features will be used for the training and testing of SVM classifier. PCA purposes is to find the directions that maximize the variance in a dataset since it ignores class labels [20,24]. It can be described as an unsupervised algorithm as contrast to LDA that is supervised. LDA computes the linear discriminants direction between multiple classes then represent the axes that maximize the separation [19].

D. Support Vector Machine (SVM)

SVM main concepts are to find an optimal separating hyperplane (OSH) between the two data sets by mapping the predictors onto a new, higher-dimensional space in which they can be separated linearly differ between other classifiers using a flat plane to separate classes by hyperplanes. Support vectors are lists of the predictor values taken from cases which the greatest impact on the location and the closest to the decision boundary that separating the classes. The detail theory of SVM is as described in [12, 16].

E. Cross validation

In this study, 10-fold cross-validation technique is applied during classification due to small sample size and to maximize the classification performance [6,25]. As mentioned earlier, gait features of all 44 subjects are divided into ten subsets and further used as training and testing dataset. For each classification stage, five subsets are chosen as the training set and the remaining subsets as the testing set. This process is repeated ten times to evaluate the generalization ability of the classifier. Finally, all ten results from the folds are averaged to obtain single performance estimation.

F. Performance Measure

Performance of machine classifier will be verified using three statistical indices specifically accuracy, sensitivity and specificity as outline below:

$$\text{Accuracy} = \frac{TAR + TRR}{TAR + FRR + TRR + FAR} \times 100\% \quad (1)$$

$$\text{Sensitivity} = \frac{TAR}{TAR + FRR} \times 100\% \quad (2)$$

$$\text{Specificity} = \frac{TRR}{TRR + FAR} \times 100\% \quad (3)$$

where:

TAR= True Acceptance Rate
FRR = False Rejection Rate
TRR=True Acceptance Rate
FAR= False Acceptance Rate

The representation of TAR, FRR, TRR and FAR can be briefly described as:
TAR: Classifier identifies Autism as Autism
FRR: Classifier identifies Normal as Autism
TRR: Classifier identifies Normal as Normal
FAR: Classifier identifies Autism as Normal

III. RESULTS AND DISCUSSION

In this section, feature extraction via PCA and LDA along with the capability of SVM to classify the gait pattern of autism children versus normal children will be discussed. Recall that all three kernels are utilized as tabulated in Table 3. Firstly, LDA surpass PCA as feature extraction except for temporal spatial gait features with SVM using polynomial as kernel. In addition, all highest accuracy rate attained using LDA as feature extraction are based on kinematic features and again with SVM using polynomial kernel the classification rate attained is the highest specifically 100%. Recall that kinematic features comprised of hip, knee and ankle features. Additionally, amongst the three gait features, the next highest accuracy rate using LDA as feature extraction is based on temporal spatial that representing the stride time, cadence, step length as well as walking speed. The accuracy rate attained is 87% with radial basis function (RBF) as kernel for SVM.

Table 3: Comparison of the classification results of LDA and PCA-Based SVM using different kernel functions

Gait features		*Linear*			*Polynomial*			*RBF*		
		Acc	*Sens*	*Spec*	*Acc*	*Sens*	*Spec*	*Acc*	*Sens*	*Spec*
Basic Temporal Spatial	Original	67	69.2	60	76.5	85	55	76.5	80.33	60
	PCA	71.5	74.2	60	**77**	85	55	76.5	80.33	60
	LDA	**84.3**	90.8	65	72.5	79.2	55	**87**	88.33	80
Kinetic	Original	72.2	81.7	55	74.3	79.2	60	**78.2**	92.5	40
	PCA	57.3	60.8	50	70	74.2	55	69	70.83	60
	LDA	**79.5**	79.2	85	**82**	87.5	70	73.3	75.00	75
Kinematic	Original	91	86.7	100	93.8	90.8	100	89.7	88.3	95
	PCA	86	85	90	93.5	90.83	100	89.5	89.17	90
	LDA	**96**	94.2	100	**100**	100	100	**96**	94.17	100

Conversely, besides accuracy, sensitivity and specificity rate are also calculated in evaluating the classifier's performance. Recall that sensitivity is the proportion of true positive that is correctly recognized by SVM classifier. For this research, autism is the true positive. Therefore, sensitivity resembled the ability of classifier to identify autism gait pattern as autism children whilst specificity is the true negative that correctly classify normal gait pattern as normal subjects. Here, for sensitivity, LDA as feature extraction with kinematic as the feature inputs outshine other kernel function of SVM with 100% successive accuracy rate. Specificity rate is similar too. Hence, these results proven that SVM with LDA as feature extraction are able to discriminate gait pattern between normal children and autism subjects.

IV. Conclusion

In conclusion, a gait classification method based on SVM classifier along with PCA and LDA as feature extraction are investigated. Results attained proven that SVM can effectively classify gait parameters between normal children and autism children. Next, both PCA and LDA are suitable as feature extraction in this study and LDA outperformed PCA in extracting significant gait features. Amongst the three gait features namely spatial temporal, kinematic and kinetic, kinematic contributed the highest accuracy rate in recognizing autism gait pattern versus normal gait as well as specificity and sensitivity rate. Future research will focus on larger data sample with specific age range. Findings from this research could assist both medical practitioner and physiotherapy department in planning suitable gait analysis program for autistic children.

Acknowledgment

This research is funded by Ministry of Higher Education (MOHE) Malaysia under the Niche Research Grant Scheme (NRGS) Project No: 600-RMI/NRGS 5/3 (8/2013). The authors wished to thank Human Motion Gait Analysis (HMGA) Laboratory, IRMI Premier Laboratory, Research Management and Innovation (IRMI), Universiti Teknologi MARA (UiTM), Malaysia for the instrumentations and experimental facilities provided, Faculty of Electrical Engineering UiTM Shah Alam for all the support given during this research. The first author also thanked Polytechnic Education Department Malaysia as well as Premier Polytechnic Sultan Salahuddin Abdul Aziz Shah (PSA) for the scholarship awarded.

References

[1] A. Moreno, I. Quiñones, G. Rodríguez, L. Núñez, and A. I. Pérez, "Development of the Spatio-Temporal Gait Parameters of Mexican Children between 6 and 13 years old Data Base to be included in Motion Analysis Softwares," no. 2, 2009

[2] M. J. Weiss, M. F. Moran, M. E. Parker, and J. T. Foley, "Gait analysis of teenagers and young adults diagnosed with autism and severe verbal communication disorders." Front. Integr. Neurosci., vol. 7, no. May, p. 33, 2013.

[3] V. L. Chester and A. T. Wrigley, "The identification of age-related differences in kinetic gait parameters using principal component analysis." Clin. Biomech. (Bristol, Avon), vol. 23, no. 2, pp. 212–20, Feb. 2008

[4] V. L. Chester and M. Calhoun, "Gait Symmetry in Children with Autism," Autism Research and Treatment, vol. 2012. pp. 1–5, 2012.

[5] M. Shetreat-Klein, S. Shinnar, and I. Rapin, "Abnormalities of joint mobility and gait in children with autism spectrum disorders." Brain Dev., vol. 36, no. 2, pp. 91–6, Feb. 2014.

[6] M. Calhoun, M. Longworth, and V. L. Chester, "Gait Patterns in Children with Autism." Clin. Biomech. (Bristol, Avon), vol. 26, no. 2, pp. 200–6, Feb. 2011.

[7] A. N. Bhat, R. J. Landa, and J. C. Galloway, "Current perspectives on motor functioning in infants, children, and adults with autism spectrum disorders.," Phys. Ther., vol. 91, no. 7, pp. 1116–29, Jul. 2011.

[8] S. Vernazza-Martin, N. Martin, a Vernazza, a Lepellec-Muller, M. Rufo, J. Massion, and C. Assaiante, “Goal directed locomotion and balance control in autistic children.,” J. Autism Dev. Disord., vol. 35, no. 1, pp. 91–102, Feb. 2005.
[9] M. Nobile, P. Perego, L. Piccinini, E. Mani, A. Rossi, M. Bellina, and M. Molteni, “Further evidence of complex motor dysfunction in drug naive children with autism using automatic motion analysis of gait.,” Autism, vol. 15, no. 3, pp. 263–83, May 2011
[10] Mari, M., Castiello, U., Marks, D., Marraffa, C., & Prior, M. (2003). The reach-to-grasp movement in children with autism spectrum disorder. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358(1430), 393–403.
[11] H. HazfizaManap, N. Md Tahir, and R. Abdullah, “Anomaly gait classification of Parkinson disease based on ANN,” 2011 IEEE Int. Conf. Syst. Eng. Technol., pp. 5–9, Jun. 2011
[12] Manap, H. H., Tahir, N., & Yassin, A. I. M. (2011). Anomalous Gait Detection based on Support Vector Machine, (Iccaie), 623–626.
[13] Katarzyna Kaczmarczyk1, Andrzej Wit1 and 3 and Jacek Zaborski4 Maciej Krawczyk1, “Artificial Neural Networks (ANN) Applied for Gait Classification and Physiotherapy Monitoring in Post Stroke Patients,” 2010.
[14] S. N. Oğulata, C. Şahin, and R. Erol, “Neural Network-Based Computer-Aided Diagnosis in Classification of Primary Generalized Epilepsy by EEG Signals,” J. Med. Syst., vol. 33, no. 2, pp. 107–112, 2009
[15] A. Bouzalmat, J. Kharroubi, and A. Zarghili, “Comparative Study of PCA, ICA, LDA using SVM Classifier,” J. Emerg. Technol. Web Intell., vol. 6, no. 1, pp. 64–68, 2014.
[16] R. K. Begg, M. Palaniswami, S. Member, and B. Owen, “Support Vector Machines for Automated Gait Classification,” vol. 52, no. 5, pp. 828–838, 2005.
[17] R. Begg and J. Kamruzzaman, “A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data,” J. Biomech., vol. 38, no. 3, pp. 401–408, 2005.
[18] K. J. Deluzio and J. L. Astephen, “Biomechanical features of gait waveform data associated with knee osteoarthritis. An application of principal component analysis,” Gait Posture, vol. 25, pp. 86–93, 2007.
[19] K. Ueki, T. Hayashida, and T. Kobayashi, “Two-dimensional Heteroscedastic Linear Discriminant Analysis for Age-group Classification,” 18th Int. Conf. Pattern Recognit., vol. 2, pp. 585–588, 2006.
[20] S. Fernandes and J. Bala, “Performance Analysis of PCA-based and LDA-based Algorithms for Face Recognition,” Int. J. Signal Process. Syst., vol. 1, no. 1, pp. 1–6, 2013.
[21] J. Wu and J. Wang, “PCA-based SVM for automatic recognition of gait patterns,” *J. Appl. Biomech.*, vol. 24, no. 1, pp. 83–7, 2008.
[22] H. hazfizaMd Tahir, Nooritawati,Manap, “PD Gait Classification based on Machine Learning Approach.pdf.” 2012.
[23] A. Carriero, A. Zavatsky, J. Stebbins, T. Theologis, and S. J. Shefelbine, “Determination of gait patterns in children with spastic diplegic cerebral palsy using principal components.,” Gait Posture, vol. 29, pp. 71–75, 2009.
[24] A. M. Martinez and A. C. Kak, “PCA versus LDA,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 228–233, 2001.
[25] A. Subasi and M. I. Gursoy, “EEG signal classification using PCA, ICA, LDA and support vector machines,” Expert Syst. Appl., vol. 37, no. 12, pp. 8659–8666, 2010.
[26] Courant, Richard, and David Hilbert. "Methods of Mathematical Physics (Interscience, New York, 1953)." Vol. I 63 (1953).
[27] a. M. S. Muniz, H. Liu, K. E. Lyons, R. Pahwa, W. Liu, F. F. Nobre, and J. Nadal, “Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait,” J. Biomech., vol. 43, no. 4, pp. 720–726, 2010.