

Automatic Musculoskeletal and Neurological Disorder Diagnosis With Relative Joint Displacement From Human Gait

Worasak Rueangsirarak, *Member, IEEE*, Jingtian Zhang, *Student Member, IEEE*,
Nauman Aslam[✉], *Member, IEEE*, Edmond S. L. Ho[✉],
and Hubert P. H. Shum[✉], *Member, IEEE*

Abstract—Musculoskeletal and neurological disorders are common devastating companions of ageing, leading to a reduction in quality of life and increased mortality. Gait analysis is a popular method for diagnosing these disorders. However, manually analyzing the motion data is a labor-intensive task, and the quality of the results depends on the experience of the doctors. In this paper, we propose an automatic framework for classifying musculoskeletal and neurological disorders among older people based on 3D motion data. We also propose two new features to capture the relationship between joints across frames, known as 3D Relative Joint Displacement (3DRJDP) and 6D Symmetric Relative Joint Displacement (6DSymRJDP), such that the relative movement between joints can be analyzed. To optimize the classification performance, we adapt feature selection methods to choose an optimal feature set from the raw feature input. Experimental results show that we achieve a classification accuracy of 84.29% using the proposed relative joint features, outperforming existing features that focus on the movement of individual joints. Considering the limited open motion database for gait analysis focusing on such disorders, we construct a comprehensive, openly accessible 3D full-body motion database from 45 subjects.

Index Terms—Musculoskeletal disorders, neurological disorders, gait analysis, feature selection.

I. INTRODUCTION

GAIT disorders are among the most common causes of balance problems in older people [1] and often lead to injury, disability, loss of independence, which result in poor quality of life. The prevalence of these conditions is expected to rise dramatically as the population ages. At least 30% of people aged 65 and older report difficulty walking three city blocks or climbing one flight of stairs, and approximately 20%

require the use of a mobility aid to move [2]. Musculoskeletal and neurological disorders have proved as some of the major reasons for an abnormal gait. They were detected in approximately 25% of persons between 70 and 74 years of age, and nearly 60% of those between 80 and 84 years of age [3].

Gait analysis is a popular method for diagnosing the musculoskeletal and neurological disorders, but determining abnormal gaits can be challenging. This is because there are no clearly accepted standards to evaluate the gait of older people [4]. While some research investigate on time-distance variables (e.g. walking speed, step length) [5], others carry out detailed biomechanical analysis (e.g. joint rotation, joint position) [6] to investigate musculoskeletal and neurological disorders. Regardless of the methods used, manually analysing gait data is a labour intensive process, and its accuracy depends on the experience of the doctors performing the analysis.

Automatic diagnosis of musculoskeletal and neurological disorders using machine learning technology has shown to be effective in reducing the manpower required for gait analysis and ensuring the reliability of the diagnosis results [7]. Holzeiter and Kohle [8] apply artificial neural networks (ANNs) for the classification of normal and pathological gaits using foot-ground reaction forces. Barton and Lees apply ANNs to differentiate different types of gaits using joint angles [9] to distinguish different gait patterns. Begg *et al.* extract the Minimum Foot Clearance (MFC), the shortest distance from the floor to the toes during the swing phase, as the input of the support vector machine (SVM) classifier [10] for gait balance classification. Khandoker *et al.* [11] combine multiple features for gait analysis. They concatenate a selection of extracted statistical values from the MFC histogram (e.g. Mean, Standard deviation, Maximum and Minimum) as the identification features for the balance impairment classification for older people. Despite many successes, the majority of existing work only evaluates features extracted from individual joints independently, and ignores the relationship between different joints that could be useful in gait analysis. Also, selecting a subset of joints based on expert knowledge, such as [10] and [11], may not be optimal for gait classification from the data point-of-view. Concatenating a large number of features into a single feature vector could have an adversarial effect on classification accuracy due to the noise of some of the features, and could cause low system efficiency.

Manuscript received May 21, 2018; revised October 1, 2018; accepted November 7, 2018. Date of publication November 15, 2018; date of current version December 6, 2018. This work was supported in part by the Erasmus Mundus Action 2 Programme under Grant 2014-0861/001-001, in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/M002632/1, in part by the Royal Society under Grant IES\R2\181024 and Grant IE160609, and in part by Mae Fah Luang University. (Corresponding author: Hubert P. H. Shum.)

W. Rueangsirarak is with Mae Fah Luang University, Chiang Rai 57100, Thailand (e-mail: worasak.rue@mfu.ac.th).

J. Zhang, N. Aslam, E. S. L. Ho, and H. P. H. Shum are with Northumbria University, Newcastle-upon-Tyne NE1 8ST, U.K. (e-mail: jingtian.zhang@northumbria.ac.uk; nauman.aslam@northumbria.ac.uk; e.ho@northumbria.ac.uk; hubert.shum@northumbria.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2018.2880871

In this paper, we utilize features based on the relative movement information for gait analysis. Previous research in analysing the distance between the left and right feet is a solid support for the effectiveness of relative information [12], but they are limited since other joint pairs are not considered. Here, we propose two comprehensive features that capture the relative information between all pairs of joints across frames, which are known as the 3D Relative Joints Displacement (3DRJDP) and the 6D Symmetric Relative Joint Displacement (6DSymRJDP). Compared with existing methods that utilize features extracted from individual joints, our proposed method is capable of evaluating the relationships of joints for all joint pairs, providing a holistic view of all interactions. Such comprehensive information allows us to develop methods that outperform the existing work in gait disorder classification.

We also adapt feature selections algorithms from the field of machine learning to further improve the system performance. Concatenating all possible features for gait analysis has an adverse effect as some features are noisy or even irrelevant. Manually selecting the features based on expert knowledge is sub-optimal from the data point-of-view. Therefore, we evaluate and adopt a number of feature selection algorithms to select an optimal feature set from the input features. In particular, during feature selection, we consider the whole temporal series of the relative features from two joints as a unit. This allows us to create human-understandable results and to ensure a reasonable system training cost. We develop a fully automated framework to evaluate the quality of the features using F-score, Neighborhood Component Analysis (NCA) and ReliefF. The system then iteratively selects the best I features that maximize the classification accuracy.

Finally, in view of the limited open resource for 3D gait analysis, we construct a comprehensive 3D motion database captured from 45 older people, annotated with the subjects' anonymised medical history. Based on the agreed decision from 3 medical doctors, the subjects are diagnosed as healthy, muscle weakness (e.g. muscle strain, underdeveloped muscles), joint problem (e.g. degenerative joint disease, osteoarthritis) and neurological defect (e.g. Parkinson's disease, Alzheimer's diseases). All three classes of disorder result in movement difficulties, which may appear similar at the movement level. However, the underlying causes are very different. The database is open for academic and research uses for free, and is downloadable from our research website.

We present four main contributions in this paper:

- We propose an automatic framework for identifying musculoskeletal and neurological disorders among older people based on 3D motion data.
- We propose two new features called the 3D Relative Joints Displacement (3DRJDP) and the 6D Symmetric Relative Joint Displacement (6DSymRJDP) to capture the relationship of joint pairs across frames.
- We adapt feature selection methods including F-score, Neighborhood Component Analysis and ReliefF, for choosing an optimal feature set from the input features to optimize classification accuracy.

- We construct an openly accessible, comprehensive 3D gait database with the anonymised medical history of the subjects. The subjects are diagnosed as healthy, muscle weaknesses, joint problems and neurological defects by 3 medical doctors.

The rest of the paper is organized as follows. In Section II, we introduce some related features and approaches for gait analysis. Section III gives an overview of the proposed system. In Section IV, we give the information about our data collection process for creating the motion database. Section V provides the details of our proposed features on relative joint relationships. Section VI introduces different feature selection algorithms. Section VII explains different classification kernels that are commonly used for motion classification. Section VIII presents the experimental results. Section IX concludes the paper and discusses possible future directions.

II. RELATED WORKS

Extensive research efforts have been made towards gait disorder analysis. In this section, we first discuss the different features adopted in gait analysis. Then, we present a summary of machine learning methods for improving gait analysis, including feature selection algorithms and gait pattern classification algorithms.

A. Features for Gait Analysis

Gait features are important for an objective gait assessment and analysis. The core of many contemporary features for gait analysis is the measurement of joint kinematics and kinetics such as the Conventional Gait Model and the Cleveland Clinic Model [13]. Among many gait features, symmetry is an important gait characteristic and is defined as a perfect agreement between the actions of the two lower limbs [14]. To calculate symmetry, mobility parameters (e.g. single joint rotation) and spatiotemporal parameters (e.g. step length) can be used [15]. Since it is difficult to diagnose the class of disorder solely based on asymmetric gait, balance and walking stability are also used. Multiple balance and stability measures including RMS acceleration, jerk (the derivative of acceleration), sway (a measure on how much a person leans his/her body), step and stride regularity and variability are proposed [16]. Mobility parameters such as cadence and step length are also important indicators to quantify gait [17]. Notice that these kinds of balance and walking stability parameters are hand-crafted and require expert knowledge. In our research, we propose a new, generic feature based on relative joint information, which is an important addition to the currently developed parameters in identifying gait abnormalities.

B. Gait Feature Selection Methods

Simply concatenating all the gait features typically results in high-dimensional data, and some dimensions may not be relevant for the problem. It is therefore advantageous to identify the important features for gait analysis, thereby removing features that convey little or redundant discriminatory information. Some techniques [18], [19] involve choosing a subset of original features that can represent the original

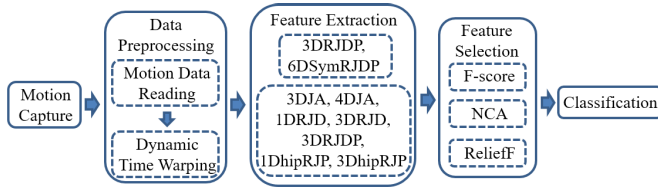


Fig. 1. The overview of our automatic method for gait disorder diagnosis.

data under certain criteria. They mainly consider conventional dimensionality reduction or statistical tools, such as principal component analysis (PCA) [20], [21] and F-score [22]. Robnik-Å ikonja *et al.* propose ReliefF, which weights different features by maximizing the distance between the data that belongs to different classes [23]. Yang *et al.* [24] propose Neighborhood Component Analysis (NCA) to learn a feature weighting vector by maximizing the expected leave-one-out classification accuracy using a regularization term. In our research, we adapt F-score, ReliefF and NCA for feature selection, and evaluate their respective performance.

C. Automatic Diagnosis Methods

Many machine learning algorithms have been adapted to automatically diagnose gait disorder. In [25], non-hierarchical cluster analysis is used to categorize four subgroups based on the temporal-spatial and kinematic parameters of walking. Similarly, hierarchical cluster analysis is conducted in [26], identifying three groups of subjects with homogeneous levels of dysfunction. Artificial neural networks (ANN) are used in [27] to classify post-stroke subject's gait into three categories based on the types of foot positions on the ground at first contact. In [28] and [29], the ability of ANN and Support Vector Machine (SVM) in distinguishing gait patterns for Parkinson Disease is discussed. Gait features from wavelet analysis and kinematic parameters are extracted, which are passed to the SVM for classification [11]. We also utilize SVM in our method as it has great usability in clinical routines without necessitating complex apparatus.

III. SYSTEM OVERVIEW

An overview of the proposed system is presented in Fig. 1. We capture the 3D motion of walking from the subjects and collect their anonymised medical histories. Then, we pre-process the captured data using inverse kinematics and dynamic time warping. We extract different types of features from the captured data, including 3DRJDP and 6DSymRJDP we proposed, as well as other common joint-based features. We adopt 3 feature selection methods and evaluate their effectiveness. Finally, the selected feature set is passed into a gait classifier.

IV. DATA COLLECTION

In this section, we first introduce the information about the invited subjects. Then, we give details on our process to capture motion data.

A. Subjects

The data was collected from a total of 45 subjects in a voluntary manner, whose protocol was agreed by the Faculty

TABLE I
THE CAUSES OF DISORDER AND SUBJECTS' STATISTICS

Muscle Weakness	Joint Problem	Neurological Defect	Healthy
18	4	13	10

	Min.	Max.	Average
Age	61	91	70.26
Weight	33	67	54.06
Height	138	198	154.26
MMSE-Thai*	15	30	24.16
FES-I**	16	26	18.96

* Mini-Mental State Examination ** Thai Fall Efficacy Scale-International

of Associated Medical Sciences Ethics Committee at Chiang Mai University. The experiment was conducted in Chiang Mai, Thailand. Applicants were Thai older people living in dwelling communities and nursing homes in Chiang Mai.

3 medical doctors from the Faculty of Associated Medical Sciences, Chiang Mai University attended the assessment. They diagnosed the subjects and agreed on the respective causes of the gait disorder, which were classified into healthy, muscle weakness, joint problem and neurological defect. This classification scheme was suggested as it was very useful as an initial diagnosis. With our system, patients can be efficiently and accurately directed to the relevant departments for further evaluations, in order to understand the specific causes of the disorder. In developing countries where there are limited budget and manpower for health-care, such an initial diagnosis scheme can effectively screen patients and relieve the stress from the front line. We discuss the possibility of employing depth camera based motion sensing system for markerless motion capturing or even everyday movement tracking in Section IX.

The voluntary applicants were screened and approved by medical experts using standard clinical test. Applicants were only included if they satisfy the following requirement: (1) They could walk without any assistance from physio-therapist or gait aids for at least 10 meters. (2) They had no cognitive impairment as tested by the Mini-Mental State Examination (MMSE-Thai 2002), i.e., $MMSE-Thai \leq 14$ for older people who were uneducated, $MMSE-Thai \leq 17$ for older people who graduated in primary school level, and $MMSE-Thai \leq 22$ for older people who graduated in high-school level or above [30]. (3) They had no fear of falling as tested by Thai Fall Efficacy Scale-International (FES-I), i.e., $Thai\ FES-I < 23$ [31]. (4) They had no other medical history that affected walking than those we considered. (5) They had no pain while walking.

We performed a randomly sampled, population-based study. In particular, we randomly selected 45 older people with ages ranging between 61 and 91 from the approved list of the applicants, which resulted in 5 male subjects and 40 female subjects. The gender bias in the database reflects that of the voluntary applicants. The details information of the subjects including the causes of the disorder, age, weight, height, MMSE-Thai and FES-I are summarized in Table I.

B. Data Acquisition

Direct interviews with the subjects were conducted using a structured questionnaire [32] before their walking sessions.

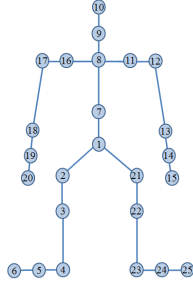


Fig. 2. The skeleton structure.

The collected data covered known diseases, medication history and Activities of Daily Living (ADLs).

A clinical and functional assessment with motion capture was carried out. The motion data was collected using the Motion Analysis® optical motion capture system [33] with fourteen Raptor-E optoelectronic cameras sampling at 100 Hz. The subjects agreed to wear a motion capture suit, attached with a set of reflective markers on their body based on Helen Hayes marker set structure [34]. The output of the motion capture system is a set of 3D marker's positions in the temporal domain.

Adapting the protocol from [35], all subjects were asked to walk naturally along a 10 meter walkway with their normal gait speed. The length is bounded by the capturing volume of the motion capture system. Four trials of walking were performed and a rest period of two minutes was given after each trial. The first trial was for practising and the last was for cooling down. We only consider the second and third trials as they better represent the subjects' normal walking motions.

V. SKELETON-BASED FEATURE EXTRACTION

In this section, we explain how we process the data and extract the features from a skeleton-based motion format.

A. Data Preprocessing

The data from the motion capture system is expressed in a 3D marker position format. Such a format is inefficient for motion classification as it depends heavily on the body size and phenotype. Following existing method on motion analysis such as [36], we convert the marker position based format into a skeletal format, such that we can focus on the movement of joints instead of body surfaces.

The conversion to a skeletal format is facilitated by inverse kinematics [37], a process that calculates the angle of a skeletal joint from multiple markers on the body surface. Motion retargeting is performed to obtain a skeleton with predefined dimensions. In our system, all these functionalities are provided by the software Autodesk MotionBuilder.

The skeletal definition we adapted consists of 25 joints corresponding to real-human joint positions, as shown in Fig. 2. Fig. 3 shows sampled frame of the skeleton in a gait cycle. We store the skeletal joint rotation data in the BioVision (BVH) format, which is a popular format readable by a large number of software libraries [38]. Notice that the end effectors do not contain any rotation information as they



Fig. 3. Sampled keyframes in one walking cycle.

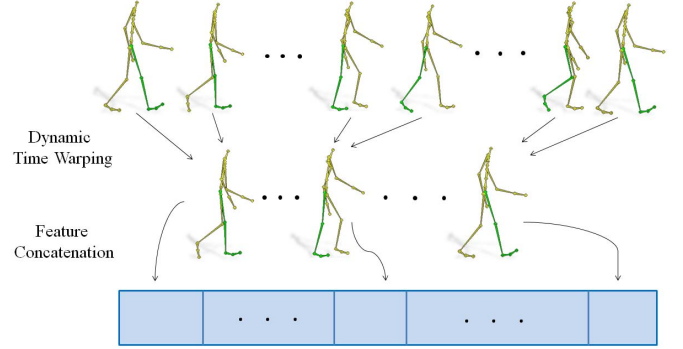


Fig. 4. The extraction of the feature vector.

do not have children joints, and they are excluded in the feature extraction process.

We normalize the motion data of the walking cycles in the spatial domain by aligning the first frame to 0 degree, such that the gait classification process is not affected by the initial walking direction. Furthermore, to normalize the temporal variation, we perform dynamic time wrapping (DTW) [39] to wrap all captured motion into the mean duration of the walking cycles, such that the classifier is not affected by the duration of the motion. These processes are supported by the software MATLAB® version 2016b [40].

B. Feature Extraction

In the two sections below, we explain how the proposed features based on relative joint information, known as 3DRJDP and 6DSymRJDP, are extracted. We also describe other traditional features that are considered in this paper, for which the features are extracted from individual joints.

The considered features below are defined for each frame of the motion. In order to evaluate the continuous sequence of motion features over time, we concatenate the frame-based features over time into a long feature vector. This process is visualized in Fig. 4.

In the following explanation, i and j represents joints. (x_i, y_i, z_i) represents the 3D position of joint i , and (x_j, y_j, z_j) represents that of joint j . ϕ_i , θ_i , ω_i are the roll, pitch and yaw angles of joint i respectively. $(x_{hips}, y_{hips}, z_{hips})$ represents the 3D position of the hips joint. n is the total number of joints.

C. The Proposed Features

3D Relative Joint Displacement (3DRJDP) is defined as the displacement between all the possible joint pairs in the skeletal hierarchy, excluding the pairs connecting to the same joint. We extract relative joint displacement as:

$$D_{3DRJDP}(i, j) = \{(x_i - x_j), (y_i - y_j), (z_i - z_j)\}. \quad (1)$$

3DRJDP within one frame consists of $n(n-1)$ items:

$$\{D_{3DRJDP}(1, 2), D_{3DRJDP}(1, 3), D_{3DRJDP}(1, 4), \dots, D_{3DRJDP}(2, 1), D_{3DRJDP}(2, 3), \dots, D_{3DRJDP}(n, n-1)\}. \quad (2)$$

6D Symmetric Relative Joint Displacement (6DSymRJDP) is defined as the pair-wise displacement between all the possible joint pairs, excluding the pairs connecting to the same joint. Also, because of the symmetric nature of the feature, we also exclude the pairs with joint identifier $i > j$:

$$D_{6DSymRJDP}(i, j) = \{(x_i - x_j), (y_i - y_j), (z_i - z_j), (x_j - x_i), (y_j - y_i), (z_j - z_i)\}. \quad (3)$$

6DSymRJDP within one frame consists of $\frac{n(n-1)}{2}$ items:

$$\{D_{6DSymRJDP}(1, 2), D_{6DSymRJDP}(1, 3), D_{6DSymRJDP}(1, 4) \dots D_{6DSymRJDP}(2, 3) \dots D_{6DSymRJDP}(n-1, n)\}. \quad (4)$$

D. Existing Features Considered in This Work

3D Joint Angle (3DJA) is defined as the concatenation of Euler joint rotation angle for each joint:

$$D_{3DJA}(i) = \{\varphi_i, \theta_i, \omega_i\}. \quad (5)$$

3DJA within one frame is therefore:

$$\{D_{3DJA}(1), D_{3DJA}(2), \dots, D_{3DJA}(n)\}. \quad (6)$$

4D Joint Angle (4DJA) is extracted by converting the Euler 3DJA rotation into the quaternion representation. The quaternion representation avoids problems such as ambiguity and gimbal locks in the Euler angles system:

$$D_{4DJA}(i) = \{w_i, a_i, b_i, c_i\}, \quad (7)$$

where the quaternion parameters are calculated as in (8), shown at the bottom of this page. 4DJA within one frame is therefore:

$$\{D_{4DJA}(1), D_{4DJA}(2), \dots, D_{4DJA}(n)\}. \quad (9)$$

3D Relative Joint Distance (3DRJD) is defined as the concatenation of the Euclidean distance in each dimension between the all possible joints pairs, except the self-connecting pairs and the neighboring joint pairs as the distances of these pairs are constant:

$$D_{3DRJD}(i, j) = \left\{ \sqrt{(x_i - x_j)^2}, \sqrt{(y_i - y_j)^2}, \sqrt{(z_i - z_j)^2} \right\}. \quad (10)$$

3DRJD within one frame is therefore:

$$\{D_{3DRJD}(1, 2), D_{3DRJD}(1, 3), D_{3DRJD}(1, 4) \dots D_{3DRJD}(n-1, n)\}. \quad (11)$$

1D Relative Joint Distance (1DRJD) is the 1D version of 3DRJD by combining the 3D distance into a 1D distance:

$$D_{1DRJD}(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \quad (12)$$

1DRJD within one frame is therefore:

$$\{D_{1DRJD}(1, 2), D_{1DRJD}(1, 3), D_{1DRJD}(1, 4) \dots D_{1DRJD}(n-1, n)\}. \quad (13)$$

3D Hips Relative Joint Position (3DhipRJP) is defined as the concatenation of the Euclidean distance of each dimension of all joints with respect to the hip position, except hips joint itself and the neighboring joints of the hips as they have constant distances with the hips:

$$D_{3DhipRJP}(i) = \left\{ \sqrt{(x_{\text{hips}} - x_i)^2}, \sqrt{(y_{\text{hips}} - y_i)^2}, \sqrt{(z_{\text{hips}} - z_i)^2} \right\}. \quad (14)$$

3DhipRJP within one frame is therefore:

$$\{D_{3DhipRJP}(1), D_{3DhipRJP}(2), \dots, D_{3DhipRJP}(n)\}. \quad (15)$$

1D Hips Relative Joint Position (1DhipRJP) is the 1D version of 3DhipRJP by combining the 3D distance into a 1D distance:

$$D_{1DhipRJP}(i) = \sqrt{(x_{\text{hips}} - x_i)^2 + (y_{\text{hips}} - y_i)^2 + (z_{\text{hips}} - z_i)^2}. \quad (16)$$

1DhipRJP within one frame is therefore:

$$\{D_{1DhipRJP}(1), D_{1DhipRJP}(2), \dots, D_{1DhipRJP}(n)\}. \quad (17)$$

VI. FEATURE SELECTION ALGORITHMS

To extract the useful information from the raw data in a lower dimensional format, we apply feature selection algorithms before doing classification. These algorithms are independent of the input feature types. Given a type of input feature (e.g. 3DJA, 3DRJD, 3DRJDP, 6DSymRJDP), these algorithms obtain a subset of the features to be used in the classification algorithm in the next stage. Using 3DJA as an example, one possible solution from these algorithms could be considering only the lower body joint angles but discarding the upper body

$$\begin{bmatrix} w_i \\ a_i \\ b_i \\ c_i \end{bmatrix} = \begin{bmatrix} \cos(\frac{\varphi_i}{2}) \cos(\frac{\theta_i}{2}) \cos(\frac{\omega_i}{2}) + \sin(\frac{\varphi_i}{2}) \sin(\frac{\theta_i}{2}) \sin(\frac{\omega_i}{2}) \\ \sin(\frac{\varphi_i}{2}) \cos(\frac{\theta_i}{2}) \cos(\frac{\omega_i}{2}) - \cos(\frac{\varphi_i}{2}) \sin(\frac{\theta_i}{2}) \sin(\frac{\omega_i}{2}) \\ \cos(\frac{\varphi_i}{2}) \sin(\frac{\theta_i}{2}) \cos(\frac{\omega_i}{2}) + \sin(\frac{\varphi_i}{2}) \cos(\frac{\theta_i}{2}) \sin(\frac{\omega_i}{2}) \\ \cos(\frac{\varphi_i}{2}) \cos(\frac{\theta_i}{2}) \sin(\frac{\omega_i}{2}) - \sin(\frac{\varphi_i}{2}) \sin(\frac{\theta_i}{2}) \cos(\frac{\omega_i}{2}) \end{bmatrix}. \quad (8)$$

ones. The underlying motivation is that some dimensions of the features are more relevant to the classification problem, while some may either be irrelevant or noisy. By selecting only the feature subset that is helpful for classification, the size of the input data can be reduced and the classification accuracy can be improved.

In this research, we employ three algorithms to measure the discriminativeness of each feature and select the optimal subset, including (1) F-score [22], (2) Neighborhood Component Analysis (NCA) [24], and (3) ReliefF [23]. These methods have shown great successes in other problems, and we evaluate their performances in human motion analysis.

A. F-Score

Here, we explain on our implementation of F-score, which measures the discrimination power within a set of training data with predefined criterion functions to characterize the intrinsic properties of the training data.

Given the training vectors $p_k, k = 1, \dots, m$, and the number of members in each of the C different gait classes (e.g. healthy, muscle weakness) as n_c , the F-score of the l^{th} feature is defined as:

$$w_l = \frac{\sum_{c=1}^C (\bar{p}_l^{(c)} - \bar{p}_l)^2}{\sum_{c=1}^C \frac{\sum_{k=1}^{n_c} (p_{k,l}^{(c)} - \bar{p}_l^{(c)})^2}{n_c - 1}} \quad (18)$$

where C is the total number of gait classes, \bar{p}_l and $\bar{p}_l^{(c)}$ are the average values of the whole dataset and the gait class c respectively, $p_{k,l}^{(c)}$ is the l^{th} feature of the k^{th} instance in the gait class c .

The numerator in (18) represents the discrimination among all category sets, and the denominator represents the discrimination within each of the sets. Since more discriminative features are represented by larger F-score values, we use the scores to rank the importance of the joints features and select the best I features that maximize the classification accuracy.

B. Neighborhood Component Analysis (NCA)

Neighborhood Component Analysis (NCA) is a dimensional reduction method that improves the predicting performance of the K-Nearest Neighbor (KNN) classifier being used in the feature selection process.

In [41], NCA has been proposed as the method to learn a Mahalanobis distance measurement for maximizing a stochastic variant of the leave-one-out cross-validation within KNN in the training dataset. This Mahalanobis distance can be calculated using inverse square roots and represented as symmetric positive semi-definite matrices. Here, the probability of a training vector, p_i , selecting another training vector, p_j , as its reference point is defined as:

$$P_{ij} = \begin{cases} \frac{K(D_w(p_i, p_j))}{\sum_{k \neq i} D_w(p_i, p_k)} & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \quad (19)$$

where $K(\Delta) = \exp(-\Delta/\sigma)$ is a kernel function, the kernel width σ is an input parameter that influences the probability of each point being selected as the reference point, w is a

weighting vector, and D_w is the weighted distance between two samples.

The objective function of KNN with the approximate leave-one-out classification accuracy can be written as:

$$\xi(w) = \frac{1}{Q} \sum_i \sum_j y_{ij} P_{ij}, \quad (20)$$

where $y_{ij} = 1$ if and only if $y_i = y_j$ and $y_{ij} = 0$ otherwise. Q is the total number of samples.

We follow [24] to adopt the NCA strategy into a feature selection task by including a weighting score, which results in a nearest neighbor-based feature weighting methods for reducing the high dimensionality of the input vector. Such a method modifies the original KNN and improves the classification performance in the leave-one-out cross-validation method. Here, Eq. 20 is modified for approximating the leave-one-out classification accuracy in KNN as:

$$\xi(w) = \sum_i \sum_j y_{ij} P_{ij} - \lambda \sum_{l=1}^d w_l^2, \quad (21)$$

where $\lambda > 0$ is a regularization parameter and can be tuned via cross-validation, which replaces the functionality of the coefficient $\frac{1}{Q}$ in Eq. 20, w_l is the scores that defines the ranking of the features, and it is obtained by gradient decent method. Essentially, by using the feature weighting method within the context of KNN, we identify the extent of redundancy in the features and select the optimal feature subset for better classification.

C. ReliefF

The original RELIEF algorithm [42] is used to evaluate the features of the data samples, resulting in a value that distinguishes a sample from its neighbors. The quality of the features is represented by its weight, which is estimated from two neighbor samples, including one nearest value from the same class (i.e. nearest hit) and another nearest value from a different class (i.e. nearest miss). The weighting score vector of the feature l on the sample r , $w_l(r)$, is estimated using the normalization of all Q training samples with the following equation:

$$w_l(r) = w_l(r) - \frac{\text{diff}(l, r, h)}{Q} + \frac{\text{diff}(l, r, m)}{Q}, \quad (22)$$

where r is the feature value of the considering sample, h is the feature value of the nearest hit sample, and m is the feature value of the nearest miss sample, $\text{diff}(l, r, h)$ calculates the distance between samples r and h using the l^{th} feature, $\text{diff}(l, r, m)$ calculates that between samples r and m . However, RELIEF is limited to only two-class problems and is sensitive to noisy data samples.

Therefore, ReliefF [43] is proposed to improve the reliability in estimating the weight and has been extended to handle multi-class data sets while retaining the same computational complexity. Here, the K-nearest neighbor (KNN) concept is adopted to find the k nearest hit observations and k nearest miss observations, instead of using one nearest neighbor in each class as in Eq. 22, which improves the reliability of

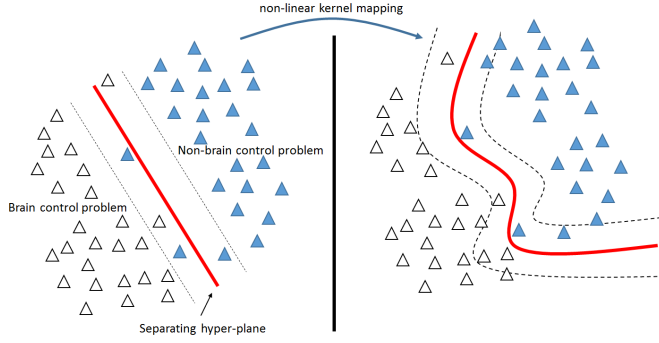


Fig. 5. Support vector machine classifiers.

weight approximation. A multiple-class problem is formulated as finding one nearest miss, $M(\Phi)$, for each different class Φ and averaging their contribution for updating the estimated w_l :

$$w_l(r) = w_l(r) - \frac{\text{diff}(l, r, h)}{Q} + \sum_{\Phi \neq \text{class}(r)} \frac{\frac{P(\Phi)}{1 - P(\text{class}(r))} \times \text{diff}(l, r, M(\Phi))}{Q}, \quad (23)$$

The final weighting score of the feature, w_l is the summation of difference among the sample and its neighbors from all Q samples. It estimates the ability of features to separate different classes.

VII. MOTION CLASSIFICATION

In supervised learning-based classification problems, the Support Vector Machine (SVM) is a powerful classifier that is commonly used to classify different categories of the data. It maps the data to a higher dimensional space and separates the data using hyperplanes. Such hyperplanes are optimal boundaries that categorise new sets of data. The optimised hyperplanes would maximise the margin among classes in the training data.

Given an instance-label pair (p_i, q_i) , $i = 1, \dots, l$ where p_i is a sample vector and $q_i \in \{1, -1\}^l$ represent one of the two categories in the training data, the original SVM [44], [45] obtains the solution as an optimization problem as follow:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & q_i (w^T \phi(p_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \quad (24)$$

where the feature vectors p_i are mapped into the higher dimensional space using the function ϕ , $C > 0$ is the penalty parameter for the error. w is known as the weight vector, b is the bias and ξ is the maximum margin. In Eq. 24, the idea is to apply a linear hyperplane to identify the margin of each data class as shown in Fig. 5 left.

However, in a multiple classes problem, the linear SVM with one hyperplane cannot separate multiple data classes. As a solution, the *one-against-one* approach is proposed [46], [47], in which for a M classes classification problem, $M(M-1)/2$ classifiers are constructed and each of

them trains data from two classes. Given training data from the i^{th} and the j^{th} classes, the solution can be found by the following objective function:

$$\begin{aligned} \min_{w^{ij}, b^{ij}, \xi^{ij}} \quad & \frac{1}{2} (w^{ij})^T w^{ij} + C \sum_t (\xi_t^{ij}) \\ \text{subject to} \quad & (w^{ij})^T \phi(p_t) + b^{ij} \geq 1 - \xi_t^{ij}, \\ & \text{if } p_t \text{ in the } i^{\text{th}} \text{ class,} \\ & (w^{ij})^T \phi(p_t) + b^{ij} \leq -1 + \xi_t^{ij}, \\ & \text{if } p_t \text{ in the } j^{\text{th}} \text{ class,} \\ & \xi_t^{ij} \geq 0. \end{aligned} \quad (25)$$

where w^{ij} is the weighting vector, b^{ij} is the bias and ξ^{ij} is the maximum margin. A voting strategy is implemented to find the class label that fits the best to the testing data. In particular, the class label that has a maximum number of votes from all binary classifiers is considered to be the best fit class label.

In a complex data space, it may be difficult to apply linear hyperplanes to separate the classes. Therefore, kernel functions, $K(p_i, p_j) \equiv \phi(p_i)^T \phi(p_j)$, are proposed to construct higher dimensional hyperplanes, as shown in Fig. 5. In this study, we evaluate three popular kernel functions including linear, polynomial, and radial basis function (RBF) as defined below:

- Linear:

$$K(p_i, p_j) = p_i^T p_j \quad (26)$$

- Polynomial:

$$K(p_i, p_j) = (\gamma p_i^T p_j + r)^d, \quad \gamma > 0 \quad (27)$$

- Radial basis function (RBF):

$$K(p_i, p_j) = \exp(\gamma \|p_i - p_j\|^2), \quad \gamma > 0 \quad (28)$$

where, γ , r , and d are kernel parameters that need to be optimized during the construction of the model. In this research, we implemented a multi-class SVM using the library LIBSVM [47]. We conducted a grid-search to tune the kernel parameters.

VIII. EXPERIMENTAL RESULTS

Using the database constructed in Section IV, we evaluate the performance of our proposed method. Our *primary experiment setup* utilizes only the three unhealthy classes for classification, while our *secondary setup* utilizes all four classes including the healthy class. The former is considered practically important to the service providers as the majority of patients only access health-care services only when they are unhealthy. The latter includes a healthy control group to verify the performance of the system.

A. Evaluation on Different Kinematics Features

Table II summarises the numerical statistics of all features extracted from the participants' motion. For example, for 6DSymRJDP, the average value of the features is 17.2 ± 6.81 millimetre in the muscle weakness category. Notice that 3DRJDP and 6DSymRJDP have the same statistics because of

TABLE II

THE NUMERICAL STATISTICS OF THE EXTRACTED FEATURES. P-VALUE AND BASELINE CLASSIFICATION REFERRED TO THE PRIMARY SETUP

Feature Types	Muscle Weakness	Joint Problem	Neurological Defect	Healthy	p-value	Baseline Classification
3DJA	-0.2 ± 3.08	0.2 ± 4.93	0.89 ± 3.47	0.36 ± 3.86	0.871	57.14
4DJA	-1.2 ± 0.54	-0.3 ± 0.84	-0.4 ± 0.93	-0.4 ± 0.34	0.043	51.43
1DRJD	60.6 ± 1.63	62.5 ± 2.63	64.2 ± 2.58	62.5 ± 2.94	0.578	55.86
3DRJD	16.7 ± 6.73	12.6 ± 2.61	13.8 ± 3.21	15.7 ± 5.68	0.001	61.43
1DhipRJP	50.9 ± 0.76	51.7 ± 0.98	51.6 ± 0.98	49.2 ± 0.67	0.168	50.82
3DhipRJP	-3.9 ± 1.98	-6.1 ± 3.75	-5.2 ± 3.67	-5.4 ± 2.11	0.389	57.14
3DRJDP	17.2 ± 6.81	9.7 ± 2.65	12.8 ± 4.21	11.7 ± 3.78	0.001	63.29
6DSymRJDP	17.2 ± 6.81	9.7 ± 2.65	12.8 ± 4.21	11.7 ± 3.78	0.001	67.14

the similar ways the features are defined, but the former has more items in the feature vector, which affects the performance of classification.

We analyse the values of the features using ANOVA to identify the significance of variance using the primary setup. We define $p\text{-value} \leq 0.05$ as an indication of significant variance, showing that the features can be used to differentiate different types of gait disorders. The results are shown in the fifth column of Table II. It can be observed that features such as 3DJA and 4DJA have high p-values, meaning that there is not sufficient evidence to reject the null hypothesis. The three better features for gait classifications are 6DSymRJDP, 3DRJDP, and 3DRJD.

Next, we test the performance of the features by using the whole feature vectors for SVM classification. For each type of feature, we experiment different SVM kernels including linear, polynomial and radial basis function (RBF), and select the best classification result. The accuracy is obtained using the leave-one-out cross-validation strategy. As shown in the last column of Table II, both of our proposed features, 3DRJDP and 6DSymRJDP, outperform the other features. In particular, 6DSymRJDP performs the best by delivering an accuracy of 67.14%. This demonstrates that the relationship between joint pairs carries more information for gait classification comparing to the absolute values of individual joint features. Comparing with 3DRJDP, 6DSymRJDP has only a half number of items in the feature vector by pairing up logically relevant ones, making it easier for the SVM systems to model the data and interpret the diagnostic results.

B. Evaluation on Different Feature Selection Methods

We compare different features under four different feature selection strategies: the baseline method without any feature selection, F-score, NCA and ReliefF. The selected features are concatenated as the feature vector for representing each motion based on the selected key frames. They are fed as an input vector into the SVM machine learning mechanism. For each type of feature, we experiment with different numbers of features and obtain the optimal value. We also experiment with different kernels including linear, polynomial and RBF, and use the value from the best one.

As shown in Table III, generally, the results with feature selection outperform the baseline ones. Also, the best results appear in the use of the F-score feature selection method. Among eight different features, 6DSymRJDP and 3DRJDP perform better than the other features under most of feature selection methods. The highest accuracy 84.29% (primary setup) and 79.17% (secondary setup) appears when

TABLE III

CLASSIFICATION ACCURACY OF EACH FEATURE USING DIFFERENT FEATURE SELECTION METHODS FOR (UPPER) THE PRIMARY SETUP (LOWER) THE SECONDARY SETUP

Features	Baseline	F-score	NCA	ReliefF
3DJA	57.14	83.17	75.71	78.57
4DJA	51.43	51.43	52.86	72.86
1DRJD	55.86	78.57	71.43	80.00
3DRJD	61.43	81.43	71.43	65.71
1DhipRJP	50.82	52.86	51.43	48.29
3DhipRJP	57.14	64.29	60.00	62.86
3DRJDP	63.29	82.43	72.86	76.86
6DSymRJDP	67.14	84.29	74.82	80.00

Features	Baseline	F-score	NCA	ReliefF
3DJA	49.71	78.61	72.81	74.56
4DJA	46.32	47.82	45.67	68.43
1DRJD	47.78	74.54	68.31	71.34
3DRJD	55.16	75.43	72.56	58.43
1DhipRJP	41.56	44.65	47.32	51.71
3DhipRJP	48.91	59.71	53.45	57.43
3DRJDP	57.76	77.33	65.89	71.47
6DSymRJDP	60.17	79.17	69.78	72.43

we use 6DSymRJDP under F-score feature selection method. With NCA, the best accuracy achieved is 75.71% (primary setup) and 72.81% (secondary setup) with the 3DJA feature type, while our proposed feature 6DSymRJDP achieves comparable values. With ReliefF, by using the default value of $k = 10$ into KNN [43], the best classification performance is 80.00% of accuracy on both 1DRJD and 6DSymRJDP (primary setup), and 74.56% on ReliefF (secondary setup).

F-score utilizes the whole training set in evaluating the suitability of a feature in classification, while NCA and ReliefF consider different sub-parts of the training set under the KNN algorithm. Since many databases in this field are small, F-score has an advantage of utilizing the full data for a more robust performance. This explains why it outperforms other feature selection methods in our experiments.

C. Kernel and Classifier Analysis

We evaluate the kernel selection process using F-score as the feature selection algorithm, as it performs the best in the previous experiment. We evaluate the features that perform well with F-score ($>80\%$), including 3DJA, 3DRJD, and our proposed features 3DRJDP and 6DSymRJDP.

Table IV shows the experiment of all kernels. The polynomial kernel generates the best results with 6DSymRJDP using F-score, with the classification accuracy of 84.29% (primary setup) and 79.17% (secondary setup). In general, the polynomial kernel performs better than the linear and RBF kernels, showing that the polynomial kernel models the motion features better.

TABLE IV

THE PERFORMANCE OF SVM KERNELS WITH F-SCORE FOR (UPPER) THE PRIMARY SETUP (LOWER) THE SECONDARY SETUP

Features	Linear	Polynomial	RBF
3DJA	74.29	83.17	51.43
3DRJD	81.43	80.00	55.71
3DRJDP	76.57	82.43	59.17
6DSymRJDP	81.43	84.29	62.86

Features	Linear	Polynomial	RBF
3DJA	70.15	78.61	50.65
3DRJD	77.64	75.43	46.34
3DRJDP	74.14	77.33	54.23
6DSymRJDP	78.45	79.17	56.46

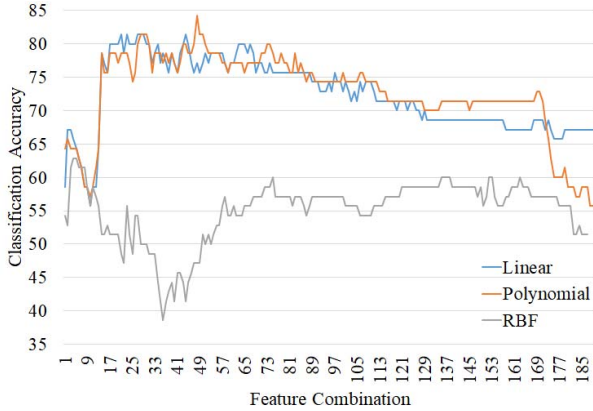


Fig. 6. Feature selections according to different kernels for 6DSymRJDP with F-score.

TABLE V

THE PERFORMANCE OF DIFFERENT CLASSIFIERS WITH F-SCORE FOR (UPPER) THE PRIMARY SETUP (LOWER) THE SECONDARY SETUP

Features	SVM	Neural Network	Decision Tree
3DJA	83.17	51.32	79.19
3DRJD	81.43	54.89	77.52
3DRJDP	82.43	53.13	79.63
6DSymRJDP	84.29	56.87	81.24

Features	SVM	Neural Network	Decision Tree
3DJA	78.61	45.71	71.15
3DRJD	75.43	48.95	73.34
3DRJDP	77.33	50.67	72.65
6DSymRJDP	79.17	52.57	74.33

Fig. 6 shows the classification accuracy against the number of features selected using F-score with 6DSymRJDP using the primary setup. It can be observed that selecting more features may not necessary improve the classification accuracy. The classification accuracy starts to drop when the number of features exceeds the modelling power of the classification system. The best result is obtained with the polynomial kernels with 48 features selected, indicated by the orange line.

To evaluate different classifiers, we compare the performance of SVM (using the polynomial kernel), neural network and binary decision tree. The neural network is implemented using a 4-layer structure. The neuron number of the input (i.e. first) layer and the output (i.e. last) layer is the feature dimensions and the number of classes respectively. The second layer has 50% neurons of the first one, and the third layer has 50% neurons of the second one, thereby implementing a typical triangle structure. The sigmoid function is used as the activation function. The binary decision tree

does not require any particular parameter setup. Table V shows the results and SVM achieves the best accuracy of 84.29% (primary setup) and 79.17% (secondary setup). Neural networks typically require a larger amount of training data, and therefore perform sub-optimally on smaller clinical databases. Decision trees have weaker generalization power comparing to polynomial SVM, as they only utilize planer decision boundaries.

IX. CONCLUSION AND DISCUSSION

In this paper, we propose an automatic gait analysis framework for musculoskeletal and neurological disorder diagnosis. We capture the gait motion from 45 Thai people with ages ranging between 61 and 91 according to four disorder categories: healthy, muscle weakness, joint problem and neurological defect. After that, we map all the gait motion into fixed length sequence with dynamic time warping and represent the warped gait sequences with different gait features. Then, we experiment with several combinations of the optimal joint set, feature extraction methods, and classifiers for the disorder diagnosis. The results show that using our proposed feature 6D Symmetric Relative Joint Displacement (6DSymRJDP) can achieve better results (84.29%) than using other gait features.

Since it is inconvenient and uncomfortable for the older people to wear suits and markers for motion capture, one future direction is to explore RGBD motion sensing hardware such as the Microsoft Kinect, which does not require capture suits nor calibrations. A benefit is that such a type of motion capture equipment can be deployed in care homes for everyday motion tracking and pre-diagnosis, thereby helping to identify disorders in the early stage.

Also, while our system allows the machine to automatically extract features for disorders classification, it may be enhanced by introducing prior medical knowledge to the existing features. We envisage that a combination of human knowledge and machine understanding would give a better description on these problems.

While there can be overlapping between different classes of disorder, when selecting subjects for motion capture, we found that there were not enough subjects suffering from more than one disorder to form a representative class. Therefore, we did not include those subjects. One of our future directions is to gather enough data of subjects suffering from multiple types of disorder and design a corresponding classification system. With enough data, one possible idea is to implement one binary classifier for each disorder to tell if the patient is suffering from such a disorder or not. Multiple binary classifiers are then combined to form the full system.

The gender bias in our database is unfortunately unpreventable due to the local culture. In Thailand where the data is collected, females have much stronger local social networks comparing to males. These networks are more open to voluntary works including ours - working with technicians to capture motion data. In our current database, we do not have sufficient data to evaluate if the gender bias has affected the experimental accuracy. One future direction is to analyze if (1) there is a correlation between genders and the types of gait disorder, and (2) the gender ratio would affect the performance of the proposed framework.

REFERENCES

- [1] B.-C. Lee, A. Fung, and T. Thrasher, "The effects of coding schemes on vibrotactile biofeedback for dynamic balance training in Parkinson's disease and healthy elderly individuals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 153–160, Jan. 2018.
- [2] Centers for Disease Control and Prevention, "Prevalence and most common causes of disability among adults—United States, 2005," *Morbidity Mortality Weekly Rep.*, vol. 58, no. 16, pp. 421–426, 2009.
- [3] J. Verghese, A. LeValley, C. B. Hall, M. J. Katz, A. F. Ambrose, and R. B. Lipton, "Epidemiology of gait disorders in community-residing older adults," *J. Amer. Geriatrics Soc.*, vol. 54, no. 2, pp. 255–261, 2006.
- [4] N. B. Alexander, "Gait disorders in older adults," *J. Amer. Geriatrics Soc.*, vol. 44, no. 4, pp. 434–451, 1996.
- [5] K. M. Ostrosky, J. M. VanSwearingen, R. G. Burdett, and Z. Gee, "A comparison of gait characteristics in young and old subjects," *Phys. Therapy*, vol. 74, no. 7, pp. 637–644, 1994.
- [6] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. IEEE 5th Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 155–162.
- [7] T. D. Pham, "Texture classification and visualization of time series of gait dynamics in patients with neuro-degenerative diseases," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 188–196, Jan. 2018.
- [8] S. H. Holzreiter and M. E. Köhle, "Assessment of gait patterns using neural networks," *J. Biomech.*, vol. 26, no. 6, pp. 645–651, 1993.
- [9] J. G. Barton and A. Lees, "An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams," *Gait Posture*, vol. 5, no. 1, pp. 28–33, 1997.
- [10] R. K. Beggs, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 828–838, May 2005.
- [11] A. H. Khandoker, D. T. H. Lai, R. K. Beggs, and M. Palaniswami, "Wavelet-based feature extraction for support vector machines for screening balance impairments in the elderly," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 587–597, Dec. 2007.
- [12] T. N. Hung and Y. S. Suh, "Inertial sensor-based two feet motion tracking for gait analysis," *Sensors*, vol. 13, no. 5, pp. 5614–5629, 2013.
- [13] R. Baker, "Gait analysis methods in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 3, no. 4, p. 4, 2006.
- [14] T. Liikavainio et al., "Loading and gait symmetry during level and stair walking in asymptomatic subjects with knee osteoarthritis: Importance of quadriceps femoris in reducing impact force during heel strike?" *Knee*, vol. 14, no. 3, pp. 231–238, 2007.
- [15] K. K. Patterson, W. H. Gage, D. Brooks, S. E. Black, and W. E. McIlroy, "Evaluation of gait symmetry after stroke: A comparison of current methods and recommendations for standardization," *Gait Posture*, vol. 31, no. 2, pp. 241–246, 2010.
- [16] R. P. Hubble, G. A. Naughton, P. A. Silburn, and M. H. Cole, "Wearable sensor use for assessing standing balance and walking stability in people with parkinson's disease: A systematic review," *PLoS ONE*, vol. 10, no. 4, p. e0123705, 2015.
- [17] I. Papavasiliou, W. Zhang, X. Wang, J. Bi, L. Zhang, and S. Han, "Classification of neurological gait disorders using multi-task feature learning," in *Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Technol. (CHASE)*, Jul. 2017, pp. 195–204.
- [18] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
- [19] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intell.*, vol. 97, pp. 245–271, Dec. 1997.
- [20] B. Bhanu and J. Han, "Human Recognition on Combining Kinematic and Stationary Features," in *Proc. 4th Int. Conf. Audio-Video-Based Biometric Person Authentication (AVBPA)*, Berlin, Germany: Springer-Verlag, 2003, pp. 600–608. <http://dl.acm.org/citation.cfm?id=1762222.1762301>
- [21] M. S. Nixon, T. Tan, and R. Chellappa, *Human Identification Based on Gait*, vol. 4. Springer, 2010.
- [22] Y.-W. Chen and C.-J. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction: Foundations and Applications*, I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds. Berlin, Germany: Springer, 2006, pp. 315–324. [Online]. Available: https://doi.org/10.1007/978-3-540-35488-8_13
- [23] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.
- [24] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161–168, 2012.
- [25] S. Mulroy, J. Gronley, W. Weiss, C. Newsam, and J. Perry, "Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke," *Gait Posture*, vol. 18, no. 1, pp. 114–125, 2003.
- [26] S. Kinsella and K. Moran, "Gait pattern categorization of stroke participants with equinus deformity of the foot," *Gait Posture*, vol. 27, no. 1, pp. 144–151, 2008.
- [27] K. Kaczmarczyk, A. Wit, M. Krawczyk, and J. Zaborski, "Gait classification in post-stroke patients using artificial neural networks," *Gait Posture*, vol. 30, no. 2, pp. 207–210, 2009.
- [28] N. M. Tahir and H. H. Manap, "Parkinson disease gait classification based on machine learning approach," *J. Appl. Sci.*, vol. 12, no. 2, pp. 180–185, 2012.
- [29] S.-H. Lee and J. S. Lim, "Parkinson's disease classification using gait characteristics and wavelet-based feature extraction," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7338–7344, 2012.
- [30] O. Kuha, B. Bunmeepitip, C. Vanichvarotm, and N. Thamanavat, "Comparative study of mini-mental state examination thai 2002 (MMSE-thai 2002) and thai mini-mental state examination (TMSE) in elderly screening test for cognitive impairment," *J. Gerontol. Geriatric Med.*, vol. 10, no. 1, pp. 19–24, 2009.
- [31] L. Thiamwong, "Psychometric testing of the falls efficacy scale-international," *Songklanagarind Med. J.*, vol. 29, no. 6, pp. 277–287, 2012.
- [32] W. Rueangsirarak, A. S. Atkins, B. Sharp, N. Chakpitak, K. Meksamoot, and P. Pothongsunun, "Clustering the clusters—Knowledge enhancing tool for diagnosing elderly falling risk," *Int. J. Healthcare Technol. Manage.*, vol. 14, nos. 1–2, pp. 39–60, 2013.
- [33] M. Analysis. (Aug. 2017). *Leading the Industry in Optical Motion Capture Solutions*. [Online]. Available: <http://www.motionanalysis.com>
- [34] T. D. Collins, S. N. Ghoussayni, D. J. Ewins, and J. A. Kent, "A six degrees-of-freedom marker set for gait analysis: Repeatability and comparison with a modified Helen Hayes set," *Gait Posture*, vol. 30, no. 2, pp. 173–180, Aug. 2009.
- [35] K. J. Mickle, B. J. Munro, S. R. Lord, H. B. Menz, and J. R. Steele, "Gait, balance and plantar pressures in older people with toe deformities," *Gait Posture*, vol. 34, no. 3, pp. 347–351, 2011.
- [36] F. Han, B. Reily, W. Hoff, and H. Zhang, "Space-time representation of people based on 3D skeletal data: A review," *Comput. Vis. Image Understand.*, vol. 158, pp. 85–105, May 2017.
- [37] J. Lee and S. Y. Shin, "A hierarchical approach to interactive motion editing for human-like figures," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Techn.*, Reading, MA, USA: Addison-Wesley, 1999, pp. 39–48.
- [38] M. Gleicher. (Aug. 2017). *Biovision BVH*. [Online]. Available: <https://research.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff/BVH.html>
- [39] D. Ellis. (Aug. 2003). Dynamic time warp (DTW) in matlab. [Online]. Available: <http://www.ee.columbia.edu/ln/rosa/matlab/dtw/>
- [40] T. M. Inc. (Aug. 2016). *R2016b at a Glance*. [Online]. Available: <https://uk.mathworks.com/products/newproducts/release2016b.html>
- [41] J. Goldberger, G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," in *Advances in Neural Information Processing Systems*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA, USA: MIT Press, 2005, pp. 513–520. [Online]. Available: <http://papers.nips.cc/paper/2566-neighbourhood-components-analysis.pdf>
- [42] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. 9th Int. Workshop Mach. Learn.*, San Francisco, CA, USA: Morgan Kaufmann, 1992, pp. 249–256. [Online]. Available: <http://dl.acm.org/citation.cfm?id=141975.142034>
- [43] I. Kononenko, M. Robnik-Šikonja, M. Robnik, and U. Pompe, "ReliefF for estimation and discretization of attributes in classification, regression, and ILP problems," 1996.
- [44] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory (COLT)*, New York, NY, USA, 1992, pp. 144–152, doi: 10.1145/130385.130401.
- [45] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1023/A:1022627411411.
- [46] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: A stepwise procedure for building and training a neural network," in *Neurocomputing*, F. F. Soulié and J. Héroult, Eds. Berlin, Germany: Springer, 1990, pp. 41–50.
- [47] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>